Spatiotemporal dynamics of epidemiology diseases: mobility based risk and short-term prediction modeling of COVID-19

Melissa Silva[1], Cláudia M. Viana[1], Iuria Betco[1], Paulo Nogueira[2,3,4], Rita Roquette[5] and Jorge Rocha[1]*

[1]Associated Laboratory TERRA, Institute of Geography and Spatial Planning, University of Lisbon, Lisbon, Portugal, [2]Associated Laboratory TERRA, Nursing Research, Innovation and Development Centre of Lisbon (CIDNUR), Nursing School of Lisbon, Lisbon, Portugal, [3]Instituto de Saúde Ambiental (ISAMB), Faculdade de Medicina, Universidade de Lisboa, Lisbon, Portugal, [4]Escola Nacional de Saúde Pública, ENSP, Centro de Investigação em Saúde Pública, CISP, Comprehensive Health Research Center, CHRC, Universidade NOVA de Lisboa, Lisbon, Portugal, [5]NOVA IMS Information Management School, NOVA University of Lisbon, Lisbon, Portugal

Nowadays, epidemiological modeling is applied to a wide range of diseases, communicable and non-communicable, namely AIDS, Ebola, influenza, Dengue, Malaria, Zika. More recently, in the context of the last pandemic declared by the World Health Organization (WHO), several studies applied these models to SARS-CoV-2. Despite the increasing number of researches using spatial analysis, some constraints persist that prevent more complex modeling such as capturing local epidemiological dynamics or capturing the real patterns and dynamics. For example, the unavailability of: (i) epidemiological information such as the frequency with which it is made available; (ii) sociodemographic and environmental factors (e.g., population density and population mobility) at a finer scale which influence the evolution patterns of infectious diseases; or (iii) the number of cases information that is also very dependent on the degree of testing performed, often with severe territorial disparities and influenced by context factors. Moreover, the delay in case reporting and the lack of quality control in epidemiological information is responsible for biases in the data that lead to many results obtained being subject to the ecological fallacy, making it difficult to identify causal relationships. Other important methodological limitations are the control of spatiotemporal dependence, management of non-linearity, ergodicy, among others, which can impute inconsistencies to the results. In addition to these issues, social contact, is still difficult to quantify in order to be incorporated into modeling processes. This study aims to explore a modeling framework that can overcome some of these modeling methodological limitations to allow more accurate modeling of epidemiological diseases. Based on Geographic Information Systems (GIS) and spatial analysis, our model is developed to identify group of municipalities where population density (vulnerability) has a stronger relationship with incidence (hazard) and commuting movements (exposure). Specifically, our framework shows how to operate a model over data with no clear trend or seasonal pattern which is suitable for a short-term predicting (i.e., forecasting) of cases based on few determinants. Our tested models provide a good alternative for when explanatory data is few and the time component is not available, once they have shown a good fit and good short-term forecast ability.

KEYWORDS

mobility, risk, non-linearity, ergodic, forecasting, simple exponential smoothing

1 Introduction

The periodic re-emergence of infectious diseases such as tuberculosis, diphtheria, malaria and cholera (1), as well as the emergence of new pathogens, such as Severe Acute Respiratory Syndrome (SARS) or Human immunodeficiency virus (HIV) (2), reinforce the need for epidemiological science endowed with increasingly complex models (3). The growing technological sophistication has contributed to the development of models that allow the incorporation of health, economic, social and behavioral contexts, among others, in the disease modeling process (4). Among these advances is the growing importance of the spatial dimension in epidemiological modeling to capture the spatiotemporal dynamics of infectious processes (5). Epidemiological modeling allows the temporal and dynamic study of the evolution of diseases and the creation of scenarios related to their transmission, constituting knowledge for the control of the same. Nowadays, epidemiological modeling is applied to a wide range of diseases, communicable and non-communicable, namely AIDS, Ebola, influenza, cancer, Dengue, Malaria, Zika (6). More recently, in the context of the last pandemic declared by the World Health Organization (WHO), several studies applied these models to SARS-CoV-2 (7–9).

SARS-CoV-2 was the most recent world's leading public health problem declared as a pandemic on March 12, 2020, by the WHO. From the pandemic's start until January 22, 2022, there have been 364,191,494 confirmed cases and 5,631,457 deaths worldwide in 117 countries (10–12). Despite the increasing number of researches using spatial analysis applied to SARS-CoV-2, some constraints persist that prevent more complex modeling such as capturing local epidemiological dynamics or capturing the real patterns and dynamics (13). For example the unavailability of: (i) epidemiological information such as the frequency with which it is made available; (ii) sociodemographic and environmental factors (e.g., population density and population mobility) at a finer scale which influence the evolution patterns of infectious diseases (14, 15); or (iii) the number of cases information that is also very dependent on the degree of testing performed, often with severe territorial disparities and influenced by context factors (16). Moreover, the delay in case reporting and the lack of quality control in epidemiological information is responsible for biases in the data that lead to many results obtained being subject to the ecological fallacy (15), making it difficult to identify causal relationships. Other important methodological limitations are mentioned in (17), such as the control of spatiotemporal dependence, management of non-linearity, ergodicy, among others, which can impute inconsistencies to the results. In addition to these issues, social contact, the real responsible for the transmission of SARS-CoV-2, is still difficult to quantify in order to be incorporated into modeling processes.

Recent reports on communicable diseases demonstrate that inequalities constitute conditions for the transmission of infectious diseases which supports the notion that pandemics are social rather than health care problems (18, 19). For example, human concentration in cities and urban areas is a factor that leads to the existence of a suitable environment for the transmission of infectious agents. The behavioral of individuals is critical for the emergence, dispersion and containment of diseases (20). The patterns of population movement and the way in which they occur in the territory (within the parish, between municipalities, by personal car or public transport) are characteristics that potentially can influence the spread of infectious diseases (21). Population movements are essential for social contact to occur, so mobility, in the context of infectious diseases, becomes a determining factor (22). Particularly, urbanization rate can act as a determinant for the occurrence of infectious outbreaks (23) and may play different roles in the context of disease transmission (24), nevertheless is the population density that has been identified as a significant factor for the proliferation of infectious diseases (25).

Information on population mobility patterns during the pandemic can constitute a proxy for social contact, however, the availability of this information in the current patterns of periodicity, aggregation and scales of analysis has some shortcomings that are not allow their use for large-scale predictive epidemiological modeling (13, 26). In this context, network diffusion model becomes particularly relevant because social contacts are transmission channels for the dissemination of any phenomenon (27). With great suitability for transmitting ideas on social networks (28), the network diffusion model is also representative of the transmission of diseases through social contact in physical proximity being an important model in infectious contexts (29, 30).

Accordingly, this study aims to explore a modeling framework that can overcome some of the above-methodological limitations to allow more accurate modeling of epidemiological diseases. Based on Geographic Information Systems (GIS) and spatial analysis, our model is developed to identify group of municipalities where population density (vulnerability) has a stronger relationship with incidence (hazard) and commuting movements (exposure). Specifically, our framework shows how to operate a model over data with no clear trend or seasonal pattern (31) which is suitable for a short-term predicting (i.e., forecasting) of cases base on few determinants. This study introduces a novel modeling framework designed to overcome these limitations. Utilizing Geographic Information Systems (GIS) and spatial analysis, our model identifies clusters of municipalities where population density, disease incidence, and commuting patterns are closely interlinked. Unlike existing models that require extensive data and temporal trends, our approach is optimized for short-term forecasting based on a limited set of determinants. Preliminary tests indicate a strong model fit and reliable short-term predictions. This framework offers a robust alternative for scenarios lacking comprehensive data and time components, thereby contributing to more accurate and adaptable epidemiological modeling. The model has significant implications for public health policy, particularly in addressing local epidemiological dynamics and territorial disparities.

2 Materials and methods

2.1 Study area

Portugal is located at the southwestern tip of Europe. It has a land border with Spain to the North and East and a maritime border to the South and West with the Atlantic Ocean (Figure 1). The Portuguese territory includes two autonomous regions: the archipelagos of the Madeira and Azores (32) (which were not considered in the study). Portugal has a total area of 92,225 km2, of which 89,102 km2 correspond to mainland Portugal (33).

FIGURE 1
Context of Mainland Portugal in Europe.
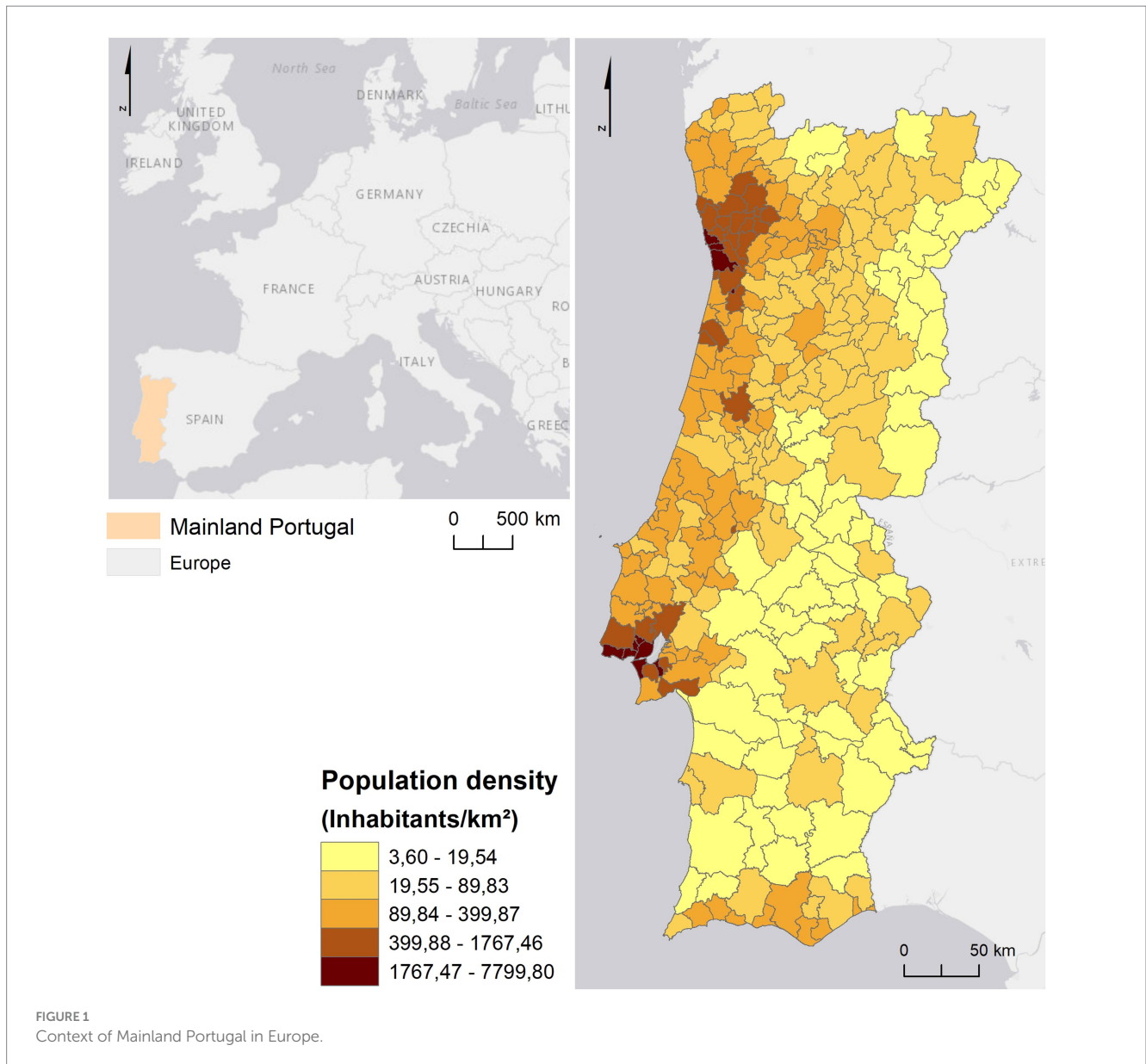
The resident population in mainland Portugal was 9,860,175 inhabitants in 2021 (34). The territory is characterized by a coastal/inland dichotomy regarding population distribution (35), since, in general, the coast has a higher population density than the interior, as shown in Figure 1. However, the population distribution along the Portuguese coast also presents a heterogeneous pattern, since it is evident the contrast between the metropolitan areas of Lisbon and Porto in relation to the remaining territory of mainland Portugal, in which most municipalities have more than 399.88 individuals per km2.

Metropolitan Areas (MAs) polarize employment and potentiate commuting since these areas are dominated by a central municipality. In the Lisbon Metropolitan Area (MAL), the difference between central and peripheral municipalities is evident since the neighboring municipalities of Lisbon have higher percentage of inhabitants working outside the municipality. This also stands for Porto in the Metropolitan Area of Porto (MAP), but in a lower extend (Figure 2A). Also, in MAL it can be seen a higher average commuting time

(between 26 and 33 min) compared to the peripheral municipalities and the central municipality (Lisbon; between 21 and 25 min). This is because the municipality of Lisbon is the one with the highest job offer. In the municipalities of the MAP, the average commuting time is also higher than the other municipalities (between 21 and 25 min; Figure 2B).

In general, the average commuting time decreases as the distance to the central municipalities of each of the MA increases (Figure 3).

2.2 Model framework

The model framework was developed to identify group of municipalities where population density has a stronger relationship with incidence and commuting movements as well as to estimate the probability of infection and predict short-term evolution of the pandemic at national and regional level. The framework includes five main stages: (1) collection and pre-processing of spatial data, (2)
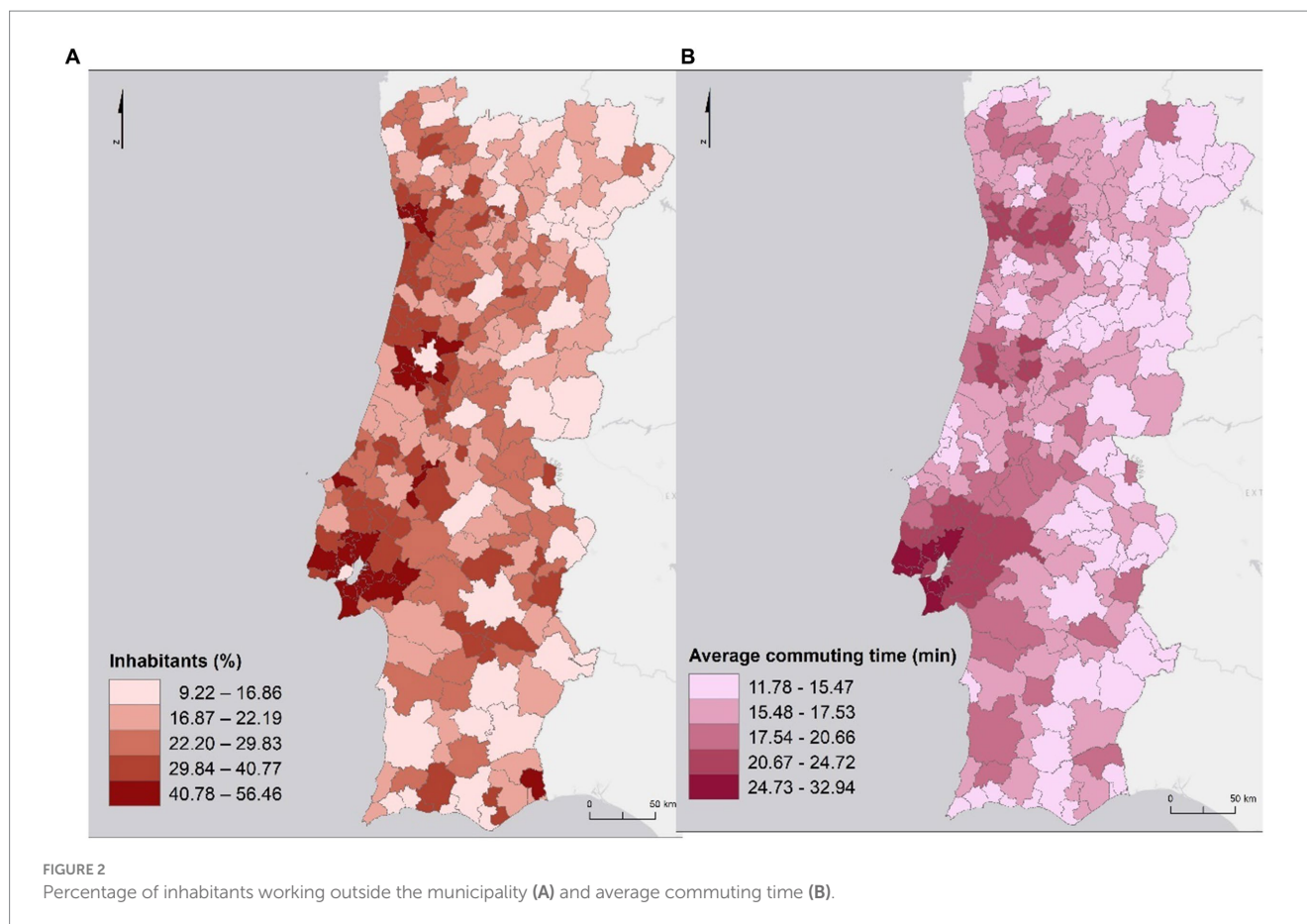
probability density modeling, (3) risk modeling, and (4) global and regional forecasting model. The workflow of the process is shown in Figure 4.

2.2.1 Data

The number of confirmed SARS-CoV-2 infected cases by municipality was collected from July 2020 to July 2022. These data come from the dssg-pt repository, available at https://github.com/dssg-pt/covid19pt-data (36), and sourced from the General Directorate of Health (Direcção Geral de Saúde – DGS) dashboard and the ESRI (Environmental Systems Research Institute) database. We also used mobility data in this study to represent the origin and destination matrix of movements commuting (home-work/school) between municipalities. Due to the importance of movement flows for the transmission of diseases, during the occurrence of epidemic outbreaks, restrictions on movement are among the earliest measures applied, in order to reduce contact between susceptible and infected individuals (37). This data was collected from INE Census 2011, since it was the most recent data made available. It represents the number of individuals who commute daily for work or study between their place of residence (origin) and their workplace or educational institution (destination), categorized by transport mode and sector of activity. It was gathered through an exhaustive inquiry to all the population living permanently in Portugal.

To examine distinct mobility patterns within the country, we combined the data to create different layers: (1) the total number of people traveling between municipalities using any transport mode;

(2) the number of people traveling between municipalities using collective transport modes such as train, metro, bus, and boat. Additionally, we gathered statistical data on major sectors of activity based. The sectors of activity are divided into 21 categories. Based on the restrictions imposed by the Portuguese government, we identified the sectors considered essential and not suspended during lockdowns: (i) Human Health and Social Support sector, (ii) the Construction sector, (iii) the Public Administration and Defense sector and (iv) the Students sector.

Moreover, data regarding population density was retrieved for mainland Portugal, on a municipality level, for the year 2020. Data is freely available from the Statistics Portugal (Instituto Nacional de Estatística—INE) portal (38).[1]

2.2.2 Probability density model

We first perform the inferential statistical model to verify which fitted curve best expressed the distribution of the number of SARS-CoV-2 infected cases occurrence as a function of mobility (total commuting movements) for each municipality (39). Therefore, 60 statistics were tested according to the Kolmogorov Smirnov validation method. Our results show that the Burr distribution presents the best fit (Table 1). According to this result, we model the number of SARS-CoV-2 cases by using the density probability equation given by
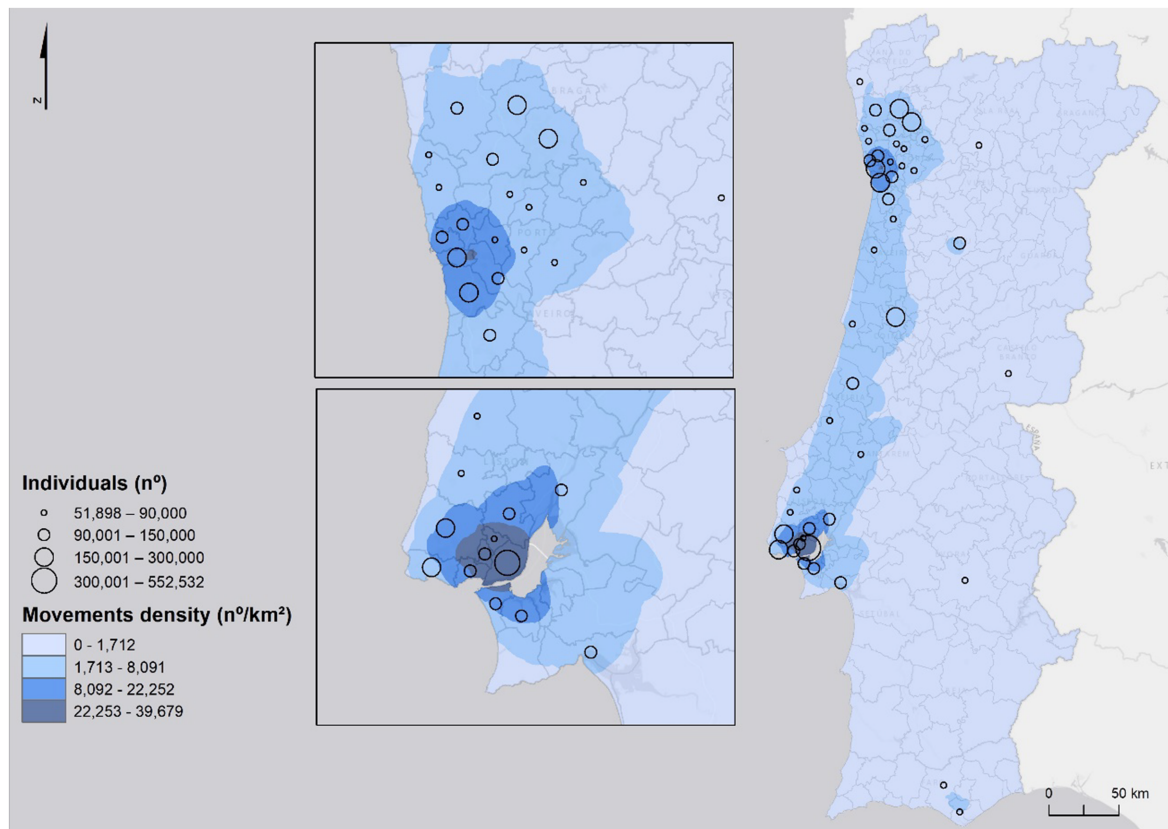

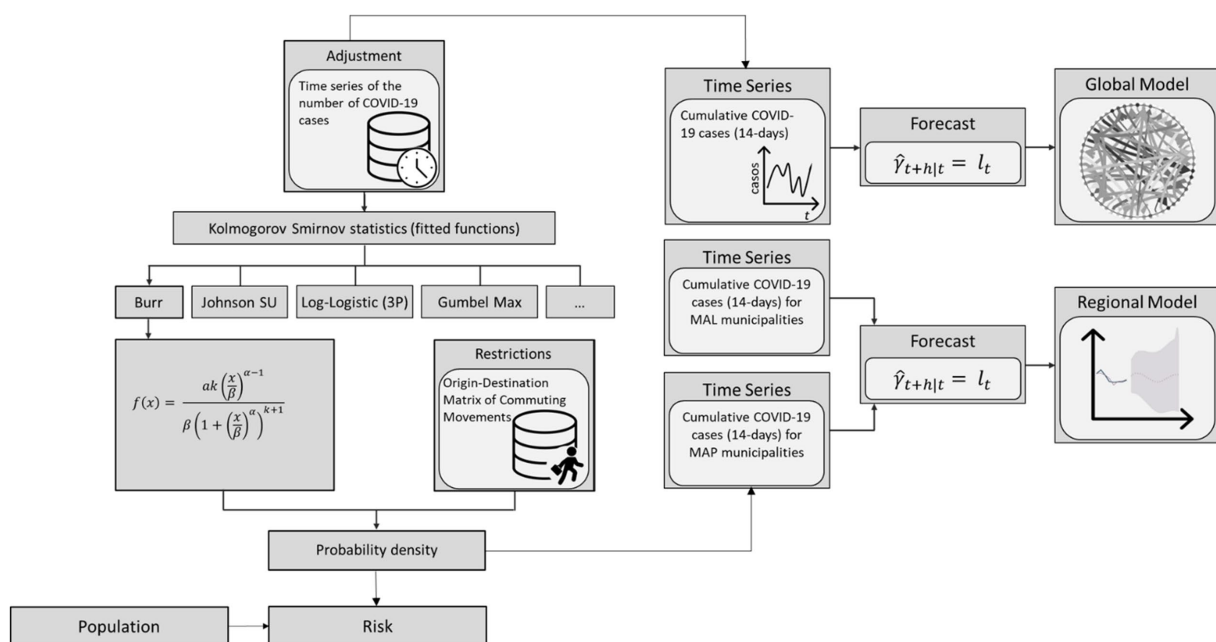1 https://www.ine.pt/

FIGURE 3
Home work/school commuting.

FIGURE 4
Methodological workflow scheme.

TABLE 1 The 10 best-fitting equations and their rank.

Distribution	Kolmogorov Smirnov	Ranking
Burr	0.04183	1
Johnson SU	0.04508	2
Log-Logistic (3P)	0.04696	3
Gumbel Max	0.04785	4
Pearson 5 (3P)	0.05058	5
Pearson 6 (4P)	0.05059	6
Inv. Gaussian	0.05110	7
Pearson 6	0.05213	8
Lognormal (3P)	0.05265	9
Log-Logistic	0.05330	10

the Burr distribution considering only commuting movements made by Human Health and Social Support sector, Construction sector, Public Administration and Defense sector, and Students sector.

2.2.3 Risk model

The concept of risk in epidemiology designates the probability that an individual in a defined population will develop a disease or other adverse health problem (40). In the context of exposure, vulnerability and risk classifications in the face of an epidemiological phenomenon, it is common to relate the incidence rate in a given period with the exposed population. In this way, a relative risk is obtained that estimates the probability of developing the disease in an exposed population group in relation to another that has not been exposed to the disease. Portugal followed the municipal risk classification according to the European Centre for Disease Prevention and Control (ECDC) categories. Based on the 14-day cumulative incidence values per 100,000 inhabitants, the municipalities are classified according to the risk of infection (< 120 Reduced; 120–239.9 Moderate; 240–479.9 High; 480–959.9 Very high; > 960 Extremely High). Although the concept of risk in epidemiology and public health is conceptually more restricted compared to the concept of risk associated with risk analysis and management (41), which presupposes the combination of different dimensions of risk (exposure, hazard, vulnerability), the classification presented by the DGS is not capable of translating any risk component to SARS-CoV-2 infection. Indeed, the indicator provided by the DGS, which can be considered as an incidence rate weighted by the population, only monitors the epidemiological situation based on transmission, so it should not be considered the only criterion for defining health measures. In addition, the use of the number of inhabitants is highly penalizing for municipalities with a low population. Therefore, we developed a risk classification index used for the definition of containment measures with a view to mitigating the incidence. This index considers how the characteristics of the territories can be aggravating or mitigating transmission, adapting the rigidity of the restrictions in relation to the real danger. In this municipal index of predisposition to infection the most relevant determinants guarantee the worsening or reduction of the index based on the relationship between the determinants and the incidence.

Thus, it is possible to analyze the pandemic waves and cycles for each one of the municipalities and to look for synchronization among them (42). Shaldehi and Dastjerdi (43) calls the attention to the

intrinsically ergodicity of synchronized systems. Nonetheless, one not found studies focusing on the pandemic ergodic behavior, and dealing with the commuting movements as a determinant. Ergodicity is a primary notion of physical models that accounts for some degree of randomness in the system. In probability theory, an ergodic system is a stochastic procedure where the average over all its potential states, is the same as its average over time. Therefore, if the expected undulating behavior of the pandemic maintains the same pattern at country our municipality level, recurrently over time and in specifically regarding the peaks, the required circumstances for the pandemic periodicity can be described through the ergodicity concept. However, geographical neighboring does not ensure commuting movements between municipalities (44, 45).

Because the sample function of the number of cases is assumed to follow a Burr distribution, where the shape factor k represents the totality of all determining factors (including the mobility factor); and the approximation of one sample wave of the SARS-CoV-2 pandemic, to the Burr distribution is clearly observable from our calculations, we used a mobility factor for the purpose of integrating commuting movements in risk analysis. This mobility factor (M) is assumed to have an exponential propensity toward k. The rational is that the infection evolution degree is connected to the Burr distribution shape factor and M need to be linked to k for adjusting to the prediction model. This mathematical model undertakes an exponential relationship between k and M to highpoint its significance, $M = re^{\beta k}$ where β is the scale factor.

Since all states are recurrent and aperiodic (46) M can be characterized as a matrix, representing the commuting movements between all the municipalities. The ones with high SARS-CoV-2 incidence are committed by a supplementary factor $(1 + q)$, where q stands for the fuzzy weight of incidence of each municipality within Portugal mainland context; for municipalities with no cases $(q = 0)$. The commuting between distinct municipalities can then be recognized as $M_{ij}^1, M_{ij}^2, \ldots, M_{ij}^n$, supposing a normalized element in a fuzzy system $0 \le M_{ij}^n \le 1$, where i is the population moving to the municipality and j the population moving from the municipality.

The net value of M for the municipality (i) can be computed by means of the arithmetic average of the product of $(1 + q)$ and the interaction M_{ij}^n:

$$M_i^n = \left(\frac{1}{n-1} \right) \sum_{j=1}^{n} (1 + q) M_{ij}^n$$

2.2.4 Global and regional forecast model

As the spread of infectious diseases is a mixed process of random and individual nature (47), the description of transmission between individuals is better adjusted using probabilistic models (48). The behavior of an infectious disease in a population reflects an average of all infected cases. Although the impact of each infected individual on the outbreak is unpredictable, the overall population dynamics typically align with mathematical predictions. This is because random variations among individuals tend to balance out as the number of infected people grows, following the law of large numbers (49). Directed and weighted human movement networks, integrating human flow intensity with neighborhood diversity metrics can help pinpointing super-spreader and super-susceptible locations (50). As the number of individuals in the network increases, the law of large

numbers helps to ensure that these identified locations reliably reflect the true dynamics of disease spread within the population. Hence, even if the distribution of secondary cases is highly uneven, the epidemic will generally progress smoothly as long as the expected number of new cases per observation period is sufficiently large. Conversely, if the infection rate is low, the epidemic may exhibit more complex and fluctuating dynamics (49).

Stochastic approaches are particularly suitable when the population under study is small. These models consider the extinction of a disease to be possible in a finite time, unlike deterministic models that tend toward a situation of endemic equilibrium (51). Therefore, we applied the Simple Exponential Smoothing (SES) to estimate the prediction values for a future period. Particularly, we chose this model as it was deemed most appropriate for data that does not follow a clear trend or seasonality pattern. This predictive model consists of an algorithm that comprises a stochastic data generation process, which can be used to produce the entire probability distribution for a future time interval $n + h$. Thus, a point forecast is obtained via the mean (or median) of the probability distribution. This model also allows the calculation of forecast intervals with a given confidence level. The predictions are weighted averages of past observations. The term exponential smoothing assumes the fact that the weights decrease exponentially as the observations are older (52), i.e., the more recent the observation the more importance is given to the value. Thus, by applying the weighted average, all predictions for the future are equivalent to the simple average of the observed series (53–56):

$$\hat{y}_{T+h|T} = \frac{1}{T}\sum_{t=1}^{T} y_t,$$

for $h = 1, 2, \ldots$. Thus, the averaging method assumes that all observations are of equal importance and assign equal weights when generating future predictions. The predictions are calculated based on weighted averages, in which the weights decrease exponentially as observations come from the oldest records. The smallest weights are associated with the oldest observations:

$$\hat{y}_{T+1|T} = \alpha y_T + \alpha(1-\alpha)y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \ldots,$$

where $0 \le \alpha \le 1$ is the smoothing parameter; the "one step forward" prediction for time $T + 1$ is the weighted average of all observations in the series $y_1, \ldots, y_T$; and the speed at which the weights decrease is controlled by the parameter α.

The application of any exponential smoothing method requires that the smoothing parameters and initial values be set. In particular for SES, for which the values of α and l_0 are selected. All predictions can be computed from the data since we know these values. For the following methods, it is usual to have more than one smoothing parameter and more than one initial component to be chosen (56).

In some cases, smoothing parameters must be chosen subjectively - for example, based on the value of previously tested smoothing parameters. Nevertheless, a more trusted and objective way to obtain unknown parameter values is to estimate them from observed data. Thus, the unknown parameters and initial values can be estimated by minimizing the sum of squared errors (SEE). The residuals are specified by $e_t = y_t - \hat{y}_{t|t-1}$ for $t = 1, \ldots, T$. Therefore, the values for the unknown parameters and the initial values that minimize are found (56):

$$SEE = \sum_{t=1}^{T}\left(y_t - \hat{y}_{t|t-1}\right)^2 = \sum_{t=1}^{T} e_t^2$$

The GM was first evaluated through the coefficient of determination $\left(r^2\right)$ that evaluates the dispersion of the data points around the fitted regression line. The r^2 is applied in computing the explaining capability of the independent variables in the model. This varies between 0 and 1, ant the higher the values, the smaller the difference between the observed data and the predicted values and, therefore, the better the model. Complementarily, Root Mean Squared Error (RMSE) was used to calculated the root mean squared of the errors between observed (actual) values and predictions (hypotheses). Both RMSE and r^2 allows to quantifies the level of agreement (i.e., fitness) between the model and the dataset. However, the RMSE gives the effectiveness of the model in predict the response variable in absolute values while r^2 shows how good are the explaining capabilities of the predictor variables regarding the changes in the response variable.

A feature of RMSE is that errors (actual–predictions) are squared before being averaged. Therefore, different weights will be assigned to the sum and, as the error values of the instances increase, the RMSE increases accordingly. If there is an outlier in the dataset, its weight will be high within RMSE calculation and, as a result, it will affect the metric. This can be problematic when comparing RMSE results calculated on test samples of different sizes, which is often the case in real case modeling, i.e., this metric alone may not describe the mean error. To avoid this problem we applied the Akaike Information Criterion (AIC) to evaluate the goodness-of-fit of the model (57). This method is based on the concept of entropy, i.e., information theory, and measure the loss of information when a given model is used to define reality. It can be described as the exchange between bias and variance in the construction of the model, or simplifying, between the accuracy and complexity of the model. The bias is asymptotically given by k, where k is the number of parameters to be estimated in the model (if the estimated model is a linear regression, k is the number of regressors, including the intercept), and defined its information criterion as:

$$AIC = -2\ln\left[f\left(x|\varnothing\right)\right] + 2k,$$

where x is the observed data, $\varnothing$ is the maximum likelihood estimation, i.e., $f\left(x|\varnothing\right)$ is the probability of the observed data due to the number of parameters k or the likelihood of the parameters given the dataset, and $\ln$ is the natural logarithm.

Overall, the SES model was applied to all municipalities in Continental Portugal, obtaining a Global Model (GM), and also for the municipalities of MAL and MAP, making a prediction for the future of 1 year, obtaining a Regional Model (RM). However, the RM was evaluated using the $F - statistic$ that for an independent variable with k groups evaluates whether the group means are meaningfully different. The $F - statistic$ is computed as:

$$F = \frac{MSR}{MSE},$$

where MSR is the regression mean square and MSE the mean square error. To considerer the $F-$value one need to look at $p-$value. If the latter is low (i.e., lower than the $\alpha-$level we can reject the null hypothesis and accept the model as valid) (58). When the F-value is below α, then the null hypothesis is true, i.e., when it presents very high values, then the variation between the group averages is greater than expected and the variation within the group is greater than a random variation. The high values of F may mean that the null hypothesis is not true (the data are not sampled from populations with the same mean), or that the random sample has large values in some groups and small values in other groups, which seems to corroborate the results of the spatiotemporal analysis where only oscillating patterns were found.

3 Results

3.1 Probability density and risk model

When modeling the density probability of SARS-CoV-2 infected cases as a function of mobility for certain economic activities groups a similar spatial distribution pattern is found for Human Health and Social Support, Construction and Public Administration, and Defense, and a different pattern for Students (Figure 5). If only workers of Human Health and Social Support move (Figure 5A), the density probability of case is higher (above 0.001851) in municipalities on the North and Central coast, in the Algarve, and also in some municipalities in the Central interior. If only individuals working in the Construction sector commute (Figure 5B), the municipalities with a higher density probability of SARS-CoV-2 infected individuals are mostly in the Central region, northwestern Alentejo region, and Algarve. The coastal strip from Mira to Caldas da Rainha presents lower values. As for AMP and AML, there was only one municipality with high values in each, namely Espinho and Alcochete, respectively. If only individuals employed in the Public Administration and Defense made commuting movements, then the density probability of SARS-CoV-2 incidence would be higher in some municipalities in the North (e.g., Monção, Bragança, Amares); Center (e.g., Águeda, Tondela, Anadia; Peniche); Alentejo (e.g., Portalegre; Beja) and Algarve (e.g., Lagos and Tavira) regions. In AMP only the municipality of Trofa presents a higher probability and in AML all municipalities present a very low probability (less than 0.000210).

If only students are commuting (Figure 5C), then the spatial pattern of spread of the density probability of SARS-CoV-2 infection cases is different from the other sectors under analysis, with a generally higher probability in inland municipalities. In this sense, municipalities with a higher density probability are mostly from the eastern region of Terras de Trás os Montes (e.g., Vimioso and Miranda do Douro), Douro (e.g., Torre de Moncorvo, Vila Nova de Foz Côa), Beiras and Serra da Estrela (e.g., Figueira de Castelo Rodrigo, Trancoso) and Alentejo Central (e.g., Alandroal, Redondo), and the Western region of Baixo Alentejo (e.g., Aljustrel, Castro Verde) and Algarve (e.g., Aljezur and Vila do Bispo). If only students were commuting (Figure 5D), the spread patterns could be quite different from the others since there are many educational institutions in the inland municipalities to which commuting takes place.

The classification of municipalities based on the risk interpretation shows that the most important group for all sectors were both

metropolitan areas and surrounding municipalities (Figure 6). The second group, the density highlights most of the remaining territory within the functional dependencies of metropolitan areas (not so much in the case of MAL), where spatial contiguity and adjacency play an important role, and relevant municipalities in the national urban hierarchy, such as district capitals (Coimbra and Aveiro), with a high economic dynamic (Leiria and Caldas da Rainha) and some important cities in the Algarve region (Faro and Portimão). The remaining group, concerns the majority of the municipalities in the interior, especially the ones near the border with small size and/or low population density. The risk information shows that higher density territorial units present more probability of facing intense and prolonged outbreaks, as opposed to less dense areas with rural characteristics.

3.2 Forecast model

In the GM scheme, it can be seen that the municipalities most related to each other are Viseu (1823) and Lisbon (1106) and, on the opposite, Castanheira de Pêra (1007), Tomar (1418), and Vila de Rei (0510; Figure 7). The coefficient of determination of the global predictive model reflects a good fit since it varies between 0.88 and 1 for all municipalities. As for the RMSE, it can be seen that most of the occurrences have an RMSE close to 0, which suggests a good predictive ability. Regarding the AIC of the predictive model, this was the one with the lowest county-by-county values compared to others that were run, with most observations being between 56.38 and 225.5. The fit for each municipality follows a normal distribution.

Figure 8 presents the results of the RMSE prediction performed for the municipalities of MAL and MAP, from which two municipalities were selected for each of the MAs. Predicted results refer to outcomes that are computed based on the model and then compared to actual events. Simulated results represent the expected future behavior of the system based on current knowledge and assumptions. These results are used to test hypotheses or explore scenarios that are difficult to study directly.

All predictions take the same value as in the last level component. The selected municipalities presented values of r2 equal to 1, i.e., they present a complete adjustment to the variation in the data. According to the prediction made for the municipalities of MAP, the behavior of the prediction curve for the municipality of Santo Tirso suggests that the number of cases of SARS-CoV-2 infection may range from 0 to 1,000 cases accumulated at 14 days per 10 thousand inhabitants. Reaching the maximum prediction value in October 2021. For this, the F value is high, around $4.89 (\alpha = 1.4673)$. However, this prediction presents a high p-value. The municipality of Arouca also presents an oscillation movement of the predictive values, as far as the prediction is concerned, between 0 and 800 accumulated cases at 14 days per 10 thousand inhabitants, and it is expected to reach the maximum value in October 2021, and the minimum value in February 2022, rising again in July 2022. This municipality presented an F value of 1.72.

Compliant with the prediction made for AML municipalities, it is estimated that, in Mafra, the number of cases of SARS-CoV-2 infection will vary between 0 and 3 thousand cases accumulated at 14 days per 10 thousand inhabitants. It reaches its maximum value at the end of September 2021, the minimum in January 2022, rising again to approximately 2,500 cases in April 2022. This prediction has an F value close to 1 (0.67). The municipality of Alcochete also
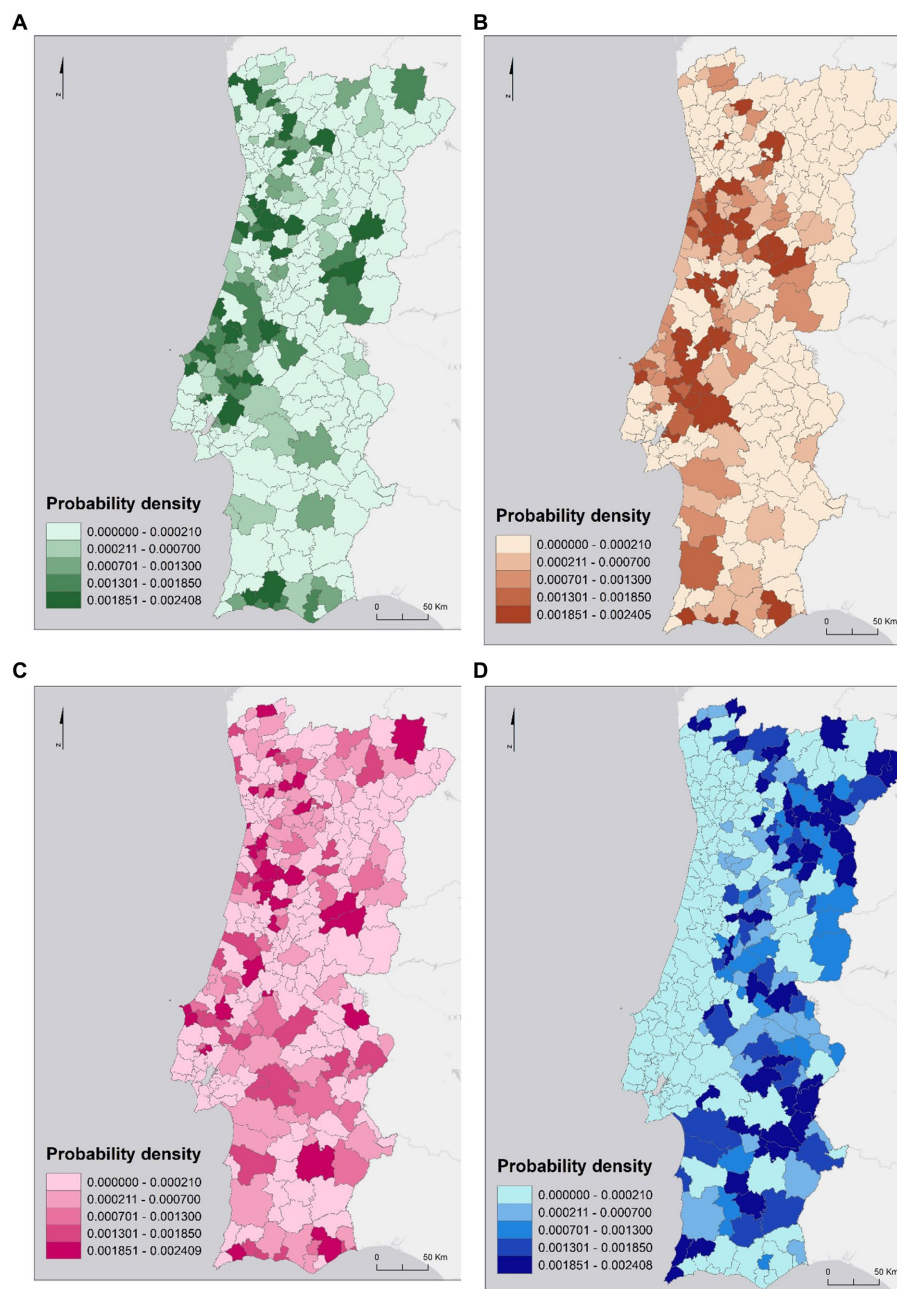
FIGURE 5
Density probability of the number of SARS-CoV-2 infected cases by commuting movements in the human health and social support sector **(A)**, the construction sector **(B)**, the public administration and defense sector **(C)** and the students sector **(D)**.

presents an oscillation of the predictive values, suggesting that the values vary from 0 to approximately 900, reaching the maximum in late September 2021, the minimum during January and early February, rising again to approximately 800 in May 2022. The *F*-value for the prediction for this municipality is 2.41. In general, according to the predictive model applied, the number of accumulated 14-day cases of SARS-CoV-2 infection is expected to oscillate in all municipalities, reaching higher relative maximum values in Mafra (3,000 accumulated 14-day cases of SARS-CoV-2 infection per 10 thousand inhabitants) and lower values in Arouca (800 accumulated 14-day cases of SARS-CoV-2 infection per 10 thousand inhabitants). All estimated values remain in the same range of values of the observed values.

According to the law of large numbers, as the number of individuals in the network increases, the aggregated data ensures that these identified locations reliably reflect the true dynamics of disease spread, thereby aligning both the predicted results and the simulated results with actual epidemic patterns.

4 Discussion

Restrictions and limitations on population displacement are one of the main non-therapeutic intervention measures adopted in epidemic contexts (59, 60). During the evaluation of the SARS-CoV-2
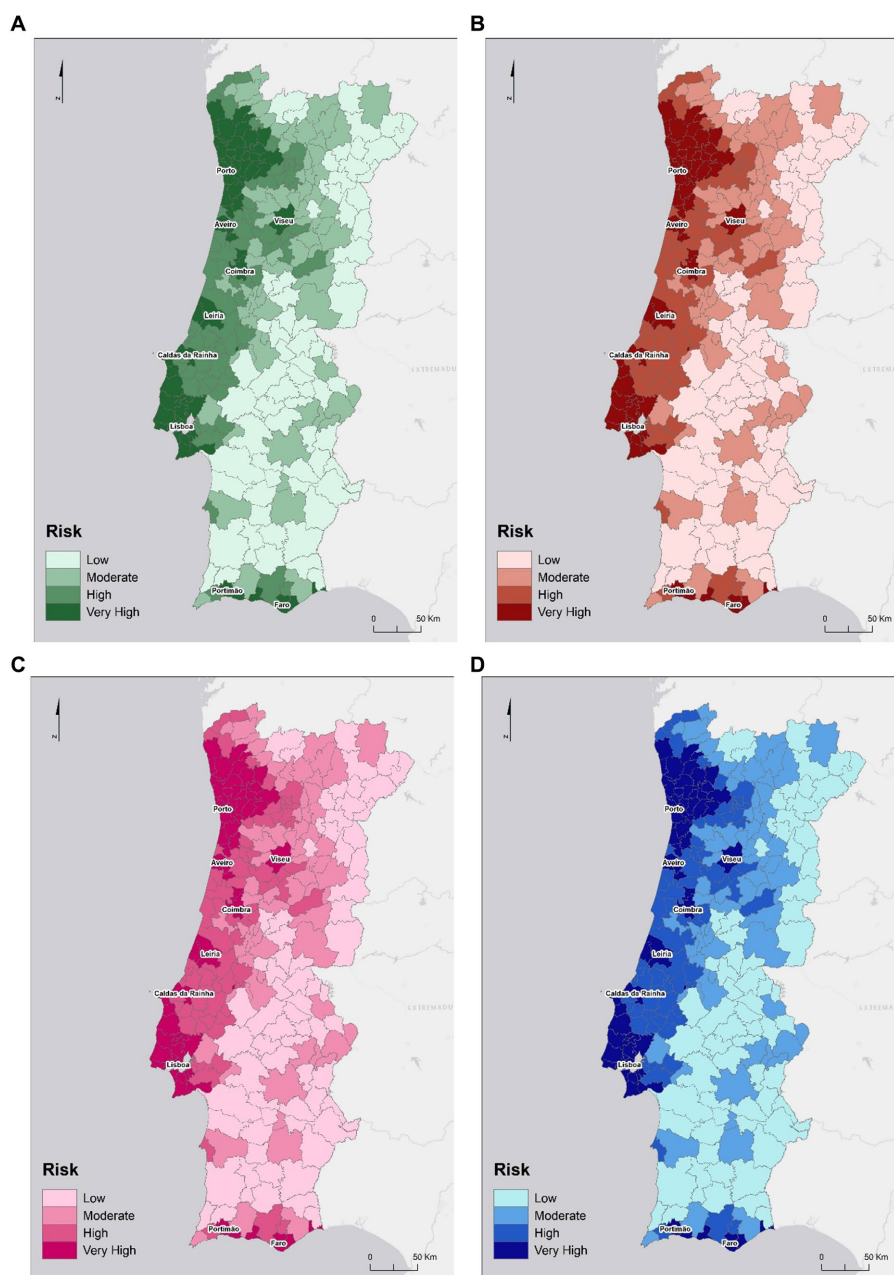
FIGURE 6
Risk of SARS-CoV-2 infections considering existing cases, population density and commuting movements in the human health and social support sector **(A)**, the construction sector **(B)**, the public administration and defense sector **(C)** and the students sector **(D)**.

pandemic several barriers to the free movement of the population were adopted, both internally and internationally. The objective was to promote social distancing and isolation (61), due to the existence of evidence that social distance resulting from the application of mobility restrictions reduces the virus transmission (62). Jia et al. (63) were among the first to associate population movements with the spatiotemporal distribution in China. The authors demonstrated how data on population mobility flows explain the spatiotemporal evolution of incidence. In fact, mobility is an important factor at multi-scale, i.e., from local to global, with connectivity between locations constituting an explanatory factor for the territorial variation of infections (64, 65). Thus, population displacement has been

identified as a determinant for the spread of SARS-CoV-2 since the beginning of the pandemic.

Our results reinforce the conclusions of several studies (66, 67) that stressed the existence of a relationship between mobility and the number of infections, showing the relevance of considering mobility indicators in epidemiological modeling of SARS-CoV-2, although with temporal and spatial variations. The value of the mobility factor can be allocated after considering various interactions between municipalities. Once a municipality assumes an ergodic behavior the majority of its population becomes infected. The possible inexistence of preventing or mitigation measures can turn the scenario even worse. Moreover, due to its high transferring rate SARS-CoV-2 carries

FIGURE 7
Global model scheme.
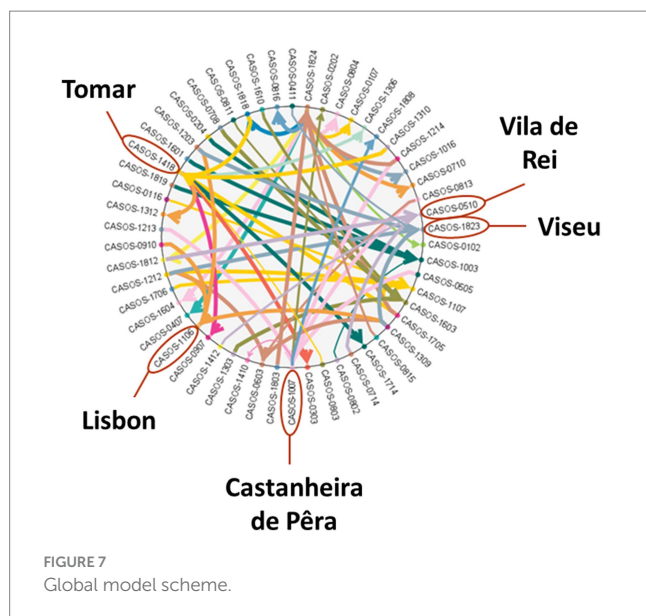
a high probability for random (perhaps chaotic) outbreaks and faster spatial diffusion.

In the specific case of mobility, the use of "real-time" data, for example based on the geolocation of mobile phones, would be very useful in the event of an epidemic, as it makes it possible to assess, at any time, the degree of mobility of the population and its geographic patterns (68, 69). Another relevant aspect of mobility is the movement, and presence, of a large volume of individuals associated with certain events (super-spreading events) that increase the possibilities of contagion to an outbreak scale (70). However, in this case the incorporation into the modeling processes is more difficult because it underlies a dimension for which it is difficult to obtain reliable information.

Numerous simulation models argue the basis of pandemics spreading behavior as single peaked, i.e., without considering the probability of an episodic recurrence. Increased commuting movements' increases the prospective of occurring random outbreaks and, thus, the spreading of the virus. Once the municipalities ease movement restrictions, social distancing or lockdown measures, the odds of a new outbreak in the same administrative region escalates. International mobility and the increasing flows of interaction associated with globalization, by reducing epidemiological differences between regions (71), constitute a difficulty in the modeling of infectious diseases that has to consider how the flows of population movements can be important covariates, for the prediction of future epidemic evolution. Thus, epidemiological models must have a spatial dimension to predict the spatiotemporal spread of pathogens.

The metapopulation version of dynamical deterministic/stochastic mathematical models (72) and the Simple Exponential Smoothing (SES) model offer complementary approaches to understanding and predicting epidemic dynamics. Metapopulation models, which account for multiple interacting subpopulations, provide a detailed representation of disease spread, incorporating spatial heterogeneity and local variations in transmission dynamics. These models can handle complex scenarios, such as varying infection rates and mobility patterns, and can simulate both deterministic and stochastic effects, making them powerful tools for exploring different intervention

strategies and understanding epidemic behavior at a coarse level (73). On the other hand, the Simple Exponential Smoothing model is a time series forecasting method that excels in capturing short-term trends and smoothing out random fluctuations in data. It is particularly useful for producing quick, computationally efficient forecasts based on recent data, without the need for detailed knowledge of the underlying transmission mechanisms. SES is ideal for situations where rapid predictions are needed, and it works well with stable data that follows an exponential trend. By leveraging both approaches, researchers and public health officials can gain a comprehensive understanding of epidemic dynamics, combining detailed spatial and stochastic insights with efficient short-term predictions.

In an epidemic context, it is essential to implement spatiotemporal surveillance systems that prioritize interventions in areas at high risk of infection (67). Likewise, the territory can be classified by its predisposition to the transmission of the virus, according to the known relationship with the determinants of infection. The idea of classifying the territory to the danger of contagion results from the need to delineate spatial-based Public Health measures. Nonetheless, these measures should consider the characteristics of the territories on which they affect because what works well in one set of circumstances may be inappropriate for another, therefore, universal panaceas are not recommended (66), i.e., one size does not fit all. Our results regarding the risk index shows that higher density territorial units present more probability of facing intense and prolonged outbreaks, as opposed to less dense areas with rural characteristics. This finding is in line with the epidemiological theories presented before (74). Living in higher density places implies a greater probability of social contact and generally corresponds to areas with an aging population and where income disparities are higher (75). In metropolitan contexts, the structured connections in the transport networks, in economic terms and in terms of commuting, are pointed out as explaining the central-periphery patterns of infections (76), namely when these movements are carried out via public transport, for example. Greater exposure to contagion (25). Similar, Mourao and Bento (77) identified the relationship between the contiguity between territorial units and the progression of pandemic spread in Portugal, highlighting the importance of mobility between municipalities for contagion, in line with the conclusions of Silva et al. (45) and Oliveira et al. (44).

Networking models prove the validity of traditional spatial diffusion models, but also helps in the interpretation of complex patterns, based on vast ties of social relationships, of which spatial diffusion models are not representative. In a process of infectious disease spreading having knowledge of the networks of social relationships of individuals is crucial to identify suspected cases and determine the full extent of the transmission chains (27). In the context of research into the relationship between SARS-CoV-2 and its determining factors, the adoption of global models continues to be dominant (78). At this stage, there is the advantage of using spatial regression models by incorporating the importance of distance for the dependent variable, similar to others applied to the context of SARS-CoV-2 (79–81), determining the spatial variability of the relationship between incidence and factors previously identified as most relevant and significant for disease patterns in mainland Portugal. This is particularly evident when mixed diffusion processes whose intensity and importance vary in space and time are manifested. HIV-AIDS is an example of a pandemic phenomenon whose spread is shown to
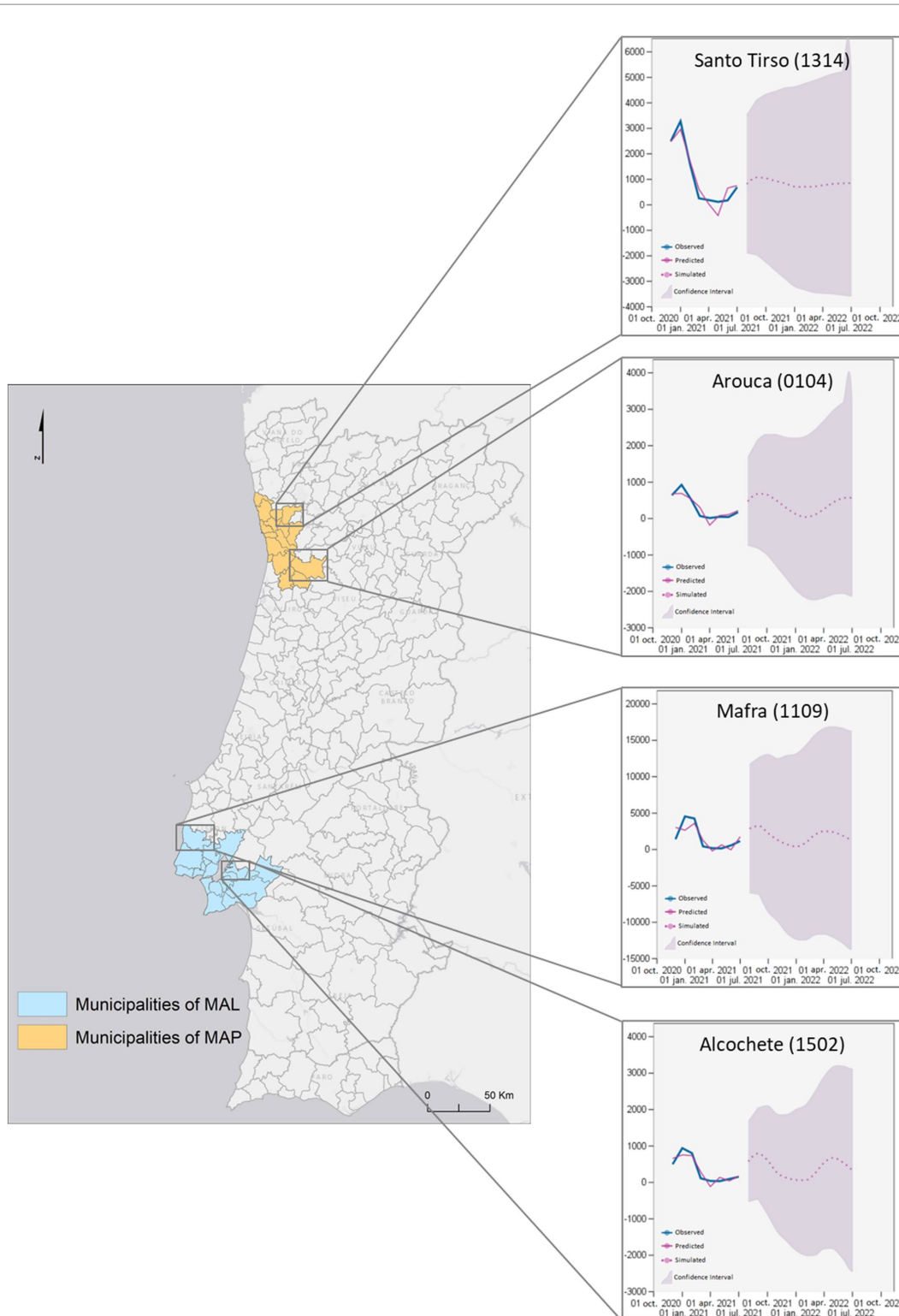
FIGURE 8
Regional model forecast for two municipalities of MAP and two of MAL.

result from the combination of expansion (contagion and hierarchical) and relocation processes (82). The occurrence of these processes derives from different forms of transmission, but whose diffusion increases with more exposure, i.e., the denser the social networks of the individuals (29).

Regression models are also commonly used to measure the relationship between the prevalence of a disease in a given context and possible risk factors - a set of independent explanatory variables (5). Based on the quantification of this relationship, it is also possible to predict future developments both in time and in space. Regressions

can assume different formulations depending on the type of relationship expressed between the epidemic process and the explanatory factors (83). Nevertheless, the behavior of a pandemic is somehow chaotic, due to the difficulty in extracting trends that can be applied to other contexts and scales of analysis. The different waves of the disease are conditioned by the occurrence of those that precede them (84), but as there is no symmetry between them, nor a fixed period, i.e., their future occurrence entails greater forecast risks. The statistical complexity of the phenomenon is compounded by the fact that the permanence of the disease in a territory has never shown to be static. For instance, predictions are only suitable for time series without the trend and seasonality components (53–56). SES model gives more importance to recent occurrences than the latter ones, transposing to the GIS fourth dimension, i.e., time, the spatial first law of geography, that everything is related but near things are probably more related (85).

Brinks and Ibert (86) indicated the need to integrate the spatial component in conceptualizing the risks associated with the pandemic, as Mourao and Bento (77) and Marques (66) reinforced the importance of adopting public health measures aimed at the specificities of the territories. This is because the unequal incidence of SARS-CoV-2 results from different levels of vulnerability of municipalities to infection, resulting from areas such as external exposure, population of the territory, sociodemography, interaction through mobility and exposure through occupation (31, 44, 66, 87), then it is essential that containment policies consider these differences. Nevertheless, the selection of disease transmission models for a given approach under study must be the result of a trade-off between complexity and precision. An adequate model for the study of a given pathology must result from what is intended to be achieved, in a compromise between simplicity, omitting details and demonstrating global qualitative patterns, and the detail, for specific situations and quantitative predictions (88).

Overall, SES requires relatively fewer data points compared to more complex models. It can generate reliable short-term forecasts even with limited or infrequent epidemiological data, making it suitable when data is not consistently available. For example, despite it does not explicitly incorporate sociodemographic or environmental factors, it can effectively capture and smooth trends in case numbers, reflecting the aggregate impact of these factors without needing detailed, granular data. It inherently smooths out random fluctuations and noise, which helps in managing non-linearity and temporal dependencies without the need for complex adjustments. Therefore, provides a straightforward approach to capturing the overall trend, which can be particularly useful when more sophisticated methods are hindered by data limitations.

The proposed method focuses on observed trends and recent data, which can help mitigate the impact of inconsistencies in testing rates and reporting quality (e.g., can smooth out the noise caused by reporting delays, giving a better picture of the global trend). By giving more weight to recent observations, it makes it easier to detect changes in the epidemic's trajectory despite reporting inconsistencies and can adapt to changes in testing policies and practices more dynamically than more complex models that might require a stable testing regime. Finally, while SES does not directly address ergodicity, its simplicity and focus on recent trends can help reduce some biases caused by data quality issues and reporting delays, offering more reliable short-term forecasts compared to more complex but data-sensitive models.

5 Conclusion

Density functions for data fitting must be selected with extreme precision, particularly in situations such as pandemic outbreaks. Simple linear or logistic regressions are not suitable for this type of data due to autocorrelation and the influence of an excessive number of zeros. Our tested models provide a good alternative for when explanatory data is few and the time component is not available, once they have shown a good fit and good short-term forecast ability. The effective use of time is of utmost importance for the study of phenomena such as the SARS-CoV-2 pandemic since it enables the interpretation of the event over time for the area under study. The patterns of virus spread would be similar if only people from the Human Health and Social Support, Construction and Public Administration, and Defense sectors commuted. However, if only students were commuting, the pattern of distribution would be quite different and more concentrated in the interior, since there are many educational institutions in the country's inland municipalities.

We can state that human mobility and transportation patterns are associated with the dissemination of diseases in the population. The patterns of virus dissemination vary between the continental territories and especially, in metropolitan areas and between their municipalities, indicating different mobility dynamics, which should be considered in health strategies and public policies. Over time the tendency is that the number of cases will continue to oscillate, one of the reasons being the application and readjustment of the restriction measures adopted by decision-makers. In this context, mathematical modeling of infectious diseases can provide important insight into the stage of an epidemic and its evolution. Together, these models provide a robust toolkit. For instance, metapopulation models can offer in-depth insights and scenario analysis, while the SES model can deliver quick, reliable forecasts. By leveraging both approaches, researchers and public health officials can gain a comprehensive understanding of epidemic dynamics, combining detailed spatial and stochastic insights with efficient short-term predictions.

Despite the results we have achieved, some limitations were identified, namely the lack of data that could be integrated into the predictive model as explanatory variables (e.g., mobility) since they are not available for time intervals. Another limitation was the fact that data on cases of SARS-CoV-2 infection is available at the municipal level and aggregated in periods of 14 days, which does not reflect the daily reality of the number of cases, making a more detailed analysis impossible.

Further research can be used to study SARS-CoV-2 in a GIS environment, e.g., daily mobility data from social networks or mobile networks, to enable the integration in the predictive model. In addition, it would also be interesting to apply a multi-agent system to model the possible future behavior of the virus dispersion. Through this methodology, the agents reproduce an autonomous behavior (through decision making) and interact with each other through the definition of some functionalities (coordination, cooperation, competition, and negotiation), enabling machine learning. In this way it would be possible to insert limiting factors (such as restrictions on mobility) and predict the repercussions these constraints would have.

In short, we can argue that SES is not a replacement for more detailed and comprehensive models, it offers a practical and robust alternative for short-term forecasting, especially in contexts where data is incomplete, inconsistent, or delayed. Its simplicity and adaptability make it a valuable tool in the context of epidemiological

modeling, providing quick and reliable insights that can inform public health decisions almost in real-time.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: https://github.com/dssg-pt/covid19pt-data.

Author contributions

MS: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Writing – original draft. CV: Data curation, Formal analysis, Methodology, Validation, Writing – original draft. IB: Data curation, Formal analysis, Writing – original draft. PN: Methodology, Supervision, Validation, Writing – review & editing. RR: Supervision, Validation, Writing – review & editing. JR: Conceptualization, Funding acquisition, Supervision, Validation, Writing – review & editing.

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Santana P. Introdução à GEOGRAFIA DA SAÚDE Território, Saúde e Bem-Estar. Coimbra: Imprensa da Universidade de Coimbra (2014).

2. Morens DM, Fauci AS. Emerging pandemic diseases: how we got to COVID-19. *Cell*. (2020) 182:1077–92. doi: 10.1016/j.cell.2020.08.021

3. Earickson R. Medical geography In: R Kitchin, Thrift NBT-IE of HG, editors. International encyclopedia of human geography. Oxford: Elsevier (2009). 9–20.

4. Garner MG, Hamilton SA. Les principes de la modélisation épidémiologique. *Rev sci tech Off int Epiz*. (2011) 30:407–16. doi: 10.20506/rst.30.2.2045

5. Pfeiffer DU, Robinson TP, Stevenson M, Stevens KB, Rogers DJ, Clements ACA. Spatial analysis in epidemiology. Oxford, UK: Oxford University Press (2008). 142 p.

6. Almeida R, Brito da Cruz AMC, Martins N, Monteiro MTT. An epidemiological MSEIR model described by the Caputo fractional derivative. Int. *J Dyn Control*. (2019) 7:776–84. doi: 10.1007/s40435-018-0492-1

7. Ajbar A, Alqahtani RT, Boumaza M. Dynamics of an SIR-based COVID-19 model with linear incidence rate, nonlinear removal rate, and public awareness. *Front Phys*. (2021) 9:215. doi: 10.3389/fphy.2021.634251

8. Ram V, Schaposnik LP. A modified age-structured SIR model for COVID-19 type viruses. *Sci Rep*. (2021) 11:15194. doi: 10.1038/s41598-021-94609-3

9. Rihan FA, Alsakaji HJ, Rajivganthi C. Stochastic SIRC epidemic model with time-delay for COVID-19. *Adv Differ Equations*. (2020) 2020:502. doi: 10.1186/s13662-020-02964-8

10. Jebril N. World Health Organization declared a pandemic public health menace: a systematic review of the coronavirus disease 2019 COVID-19. *SSRN Electron J*. (2020). doi: 10.2139/ssrn.3566298

11. Kim J, Zhang J, Cha Y, Kolitz S, Funt J, Chong RE, et al. Coronavirus disease - 2019 (COVID-19). *ChemRxiv*. (2020) 2019. doi: 10.26434/chemrxiv.12037416.v1

12. WHO. WHO coronavirus (COVID-19) dashboard. (2021). Available at: https://data.who.int/dashboards/covid19/cases (Accessed February 02, 2022).

13. Ribeiro AI, Santos CJ. The importance of spatial analysis of COVID-19 pandemic for health geography: challenges and perspectives. *Finisterra*. (2020) 115:37–42. doi: 10.18055/finis20318

14. Dalziel BD, Kissler S, Gog JR, Viboud C, Bjørnstad ON, Metcalf CJE, et al. Urbanization and humidity shape the intensity of influenza epidemics in U.S. cities. *Science*. (2018) 362:75–9. doi: 10.1126/science.aat6030

15. Fatima M, O'Keefe KJ, Wei W, Arshad S, Gruebner O. Geospatial analysis of COVID-19: a scoping review. *Int J Environ Res Public Health*. (2021) 18:2336. doi: 10.3390/ijerph18052336

16. Cordes J, Castro MC. Spatial analysis of COVID-19 clusters and contextual factors in new York City. *Spat Spatiotemporal Epidemiol*. (2020) 34:100355. doi: 10.1016/j.sste.2020.100355

17. Barceló MA, Saez M. Methodological limitations in studies assessing the effects of environmental and socioeconomic variables on the spread of COVID-19: a systematic review. *Environ Sci Eur*. (2021) 33:108. doi: 10.1186/s12302-021-00550-7

18. Quinn SC, Kumar S. Health inequalities and infectious disease epidemics: a challenge for Global Health security. *Biosecurity Bioterrorism Biodefense Strateg Pract Sci*. (2014) 12:263–73. doi: 10.1089/bsp.2014.0032

19. Singu S, Acharya A, Challagundla K, Byrareddy SN. Impact of social determinants of health on the emerging COVID-19 pandemic in the United States. *Front Public Health*. (2020) 8:406. doi: 10.3389/fpubh.2020.00406

20. Bedson J, Skrip LA, Pedi D, Abramowitz S, Carter S, Jalloh MF, et al. A review and agenda for integrated disease models including social and behavioural factors. *Nat Hum Behav*. (2021) 5:834–46. doi: 10.1038/s41562-021-01136-2

21. Shi P, Keskinocak P, Swann JL, Lee BY. The impact of mass gatherings and holiday traveling on the course of an influenza pandemic: a computational model. *BMC Public Health*. (2010) 10:778. doi: 10.1186/1471-2458-10-778

22. Cliff AD, Ord JK, Haggett P, Versey GR. Spatial diffusion: An historical geography of epidemics in an island community, vol. 14 CUP Archive. Cambridge, United Kingdom: Cambridge University Press. (1981). p. 248.

23. Reyes R, Ahn R, Thurber K, Burke TF. Urbanization and infectious diseases: general principles, historical perspectives, and contemporary challenges BT - challenges in infectious diseases In: IW Fong, editor. New York. New York, NY: Springer (2013). 123–46.

24. Santiago-Alarcon D, MacGregor-Fors I. Cities and pandemics: urban areas are ground zero for the transmission of emerging human infectious diseases. *J Urban Ecol*. (2020) 6:1–3. doi: 10.1093/jue/juaa012

25. Nathan M. The city and the virus. *Urban Stud*. (2021) 60:1346. doi: 10.1177/00420980211058383

26. Tamagusko T, Ferreira A. Data-Driven Approach to Understand the Mobility Patterns of the Portuguese Population during the COVID-19 Pandemic. *Sustainability*. (2020) 12:9775. doi: 10.3390/su12229775

27. Centola D. How behavior spreads New Jersey (USA): Princeton University Press (2018). 312.

28. Jackson MO, Yariv L. Diffusion on social networks. US: Economie Publique. Citeseer (2005).

29. Schneider JA, Cornwell B, Ostrow D, Michaels S, Schumm P, Laumann EO, et al. Network mixing and network influences Most linked to HIV infection and risk behavior in the HIV epidemic among black men who have sex with men. *Am J Public Health*. (2012) 103:e28–36. doi: 10.2105/AJPH.2012.301003

30. Latkin C, Mandell W, Vlahov D, Oziemkowska M, Celentano D. People and places: behavioral settings and personal network characteristics as correlates of needle sharing. JAIDS. *J Acquir Immune Defic Syndr*. (1996) 13:273–280.

31. Barbosa B, Silva M, Capinha C, Garcia RAC, Rocha J. Spatial correlates of COVID-19 first wave across continental Portugal. *Geospat Health*. (2022) 17:1073. doi: 10.4081/gh.2022.1073

32. Portuguese Republic. PORTAL DIPLOMÁTICO. *REPRESENTAÇÃO PERMANENTE DE PORTUGAL JUNTO DA UNIÃO EUROPEIA*. (2020). Available at: https://ue.missaoportugal.mne.gov.pt/pt/portugal/sobre-portugal/dados-gerais (Accessed February 02, 2022).

33. INE. Indicadores de geografia física. (2013). Available at: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&userLoadSave=Load&userTableOrder=9959&tipoSeleccao=1&contexto=pq&selTab=tab1&submitLoad=true (Accessed June 27, 2024).

34. INE. População residente (N.º) por Local de residência (resultados preliminares Censos2021) e Sexo; Decenal. (2021). Available at: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0012380&contexto=bd&selTab=tab2 (Accessed June 27, 2024).

35. Leston Bandeira M, Azevedo A, Sousa Gomes C, Tomé LP, Mendes MF, et al. (2014). "Dinâmicas Demográficas e Envelhecimento da População Portuguesa, 1950-2011 Evolução e Perspectivas", Fundação Francisco Manuel dos Santos, Dezembro, Lisboa.

36. DGS. Data science for social good Portugal. covid19pt-data. (2021). Available at: https://github.com/dssg-pt/covid19pt-data (Accessed June 27, 2024).

37. Mateus ALP, Otete HE, Beck CR, Dolan GP, Nguyen-Van-Tam JS. Effectiveness of travel restrictions in the rapid containment of human influenza: a systematic review. *Bull World Health Organ*. (2014) 92:868–880D. doi: 10.2471/BLT.14.135590

38. INE. Densidade populacional (N.º/ km²) por Local de residência (NUTS - 2013) e Tipologia de áreas urbanas. Estimativas anuais da população residente. (2020). Available at: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&contecto=pi&indOcorrCod=0008337&selTab=tab0&xlang=pt (Accessed June 27, 2024).

39. Tuli S, Tuli S, Tuli R, Gill SS. Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing. *Elsevier*. (2020) 11:100222. doi: 10.1016/j.iot.2020.100222

40. Bhopal R. Concepts of epidemiology: Integrating the ideas, theories, principles and methods of epidemiology. *2nd* ed. New York (USA): Oxford University Press (2008). p. 420.

41. UNISDR. Living with risk: A global review of disaster reduction initiatives. Geneva, Switzerland: United Nations (2004). 429 p.

42. Borrego-Salcido C, Juárez-del-Toro R, Fonseca-Zendejas AS. The waves and cycles of COVID-19 pandemic: a phase synchronization approach. *AJS*. (2023) 52:25–38. doi: 10.17713/ajs.v52i3.1450

43. Shaldehi SJ, Dastjerdi DA. On intrinsically ergodicity of synchronized systems. *J Stat Manag Syst*. (2019) 22:447–58. doi: 10.1080/09720510.2018.1563957

44. Oliveira S, Ribeiro AI, Nogueira J, Rocha J. Simulating the effects of mobility restrictions in the spread of SARS-CoV-2 in metropolitan areas in Portugal. *PLoS One*. (2022) 17:e0274286. doi: 10.1371/journal.pone.0274286

45. Silva M, Betco I, Capinha C, Roquette R, Viana CM, Rocha J. Spatiotemporal dynamics of COVID-19 infections in mainland Portugal. *Sustain For*. (2022) 14:10370. doi: 10.3390/su141610370

46. Viana MC, Rocha J. Evaluating dominant land use/land cover changes and predicting future scenario in a rural region using a memoryless stochastic method. *Sustain For*. (2020) 12:4332. doi: 10.3390/su12104332

47. Roberts M, Andreasen V, Lloyd A, Pellis L. Nine challenges for deterministic epidemic models. *Epidemics*. (2015) 10:49–53. doi: 10.1016/j.epidem.2014.09.006

48. Daley DJ, Gani J. Epidemic modelling: An introduction. Cambridge studies in mathematical biology. Cambridge: Cambridge University Press (1999).

49. Kumar N, Oke J, Nahmias-Biran B. Activity-based epidemic propagation and contact network scaling in auto-dependent metropolitan areas. *Sci Rep*. (2021) 11:22665. doi: 10.1038/s41598-021-01522-w

50. Chin WCB, Bouffanais R. Spatial super-spreaders and super-susceptibles in human movement networks. *Sci Rep*. (2020) 10:18642. doi: 10.1038/s41598-020-75697-z

51. Allen LJS. An introduction to stochastic epidemic models. In: F Brauer, DriesscheP van den and J Wu, editors. Mathematical epidemiology. Berlin, Heidelberg: Springer Berlin Heidelberg; (2008). p. 81–130.

52. Hyndman RJ, Koehler AB, Ord JK, Snyder RD. Forecasting with exponential smoothing. *Spring*. (2008) 4:15. doi: 10.1007/978-3-540-71918-2

53. Brown RG. Statistical forecasting for inventory control. New York: McGraw-Hill (1959).

54. Holt CC. Forecasting seasonals and trends by exponentially weighted moving averages. *Int J Forecast*. (2004) 20:5–10. doi: 10.1016/j.ijforecast.2003.09.015

55. Winters PR. Forecasting sales by exponentially weighted moving averages. *Manag Sci*. (1960) 6:324–42. doi: 10.1287/mnsc.6.3.324

56. Hyndman RJ, Athanasopoulos G. Forecasting: Principles and practice. 2nd ed. exponential smoothing. Melbourne, Australia: OTexts (2018).

57. Akaike H. A new look at the statistical model identification. *IEEE Trans Autom Control*. (1974) 19:716–23. doi: 10.1109/TAC.1974.1100705

58. Archdeacon TJ. Correlation and regression analysis. Madison (USA): A historian's guide Univ of Wisconsin Press (1994). 352.

59. MUG K, Hill V, Ruis C, Dellicour S, Bajaj S, JT MC, et al. Spatiotemporal invasion dynamics of SARS-CoV-2 lineage B.1.1.7 emergence. *Science*. (2021) 373:889–95. doi: 10.1126/science.abj0113

60. Yang W, Deng M, Li C, Huang J. Spatio-temporal patterns of the 2019-ncov epidemic at the county level in Hubei province, China. *Int J Environ Res Public Health*. (2020) 17:1–11. doi: 10.3390/ijerph17072563

61. Wells CR, Sah P, Moghadas SM, Pandey A, Shoukat A, Wang Y, et al. Impact of international travel and border control measures on the global spread of the novel 2019 coronavirus outbreak. *Proc Natl Acad Sci*. (2020) 117:7504–9. doi: 10.1073/pnas.2002616117

62. Ferreira CP, Marcondes D, Melo MP, Oliva SM, Peixoto CM, Peixoto PS. A snapshot of a pandemic: the interplay between social isolation and COVID-19 dynamics in Brazil. *Patterns*. (2021) 2:100349. doi: 10.1016/j.patter.2021.100349

63. Jia JS, Lu X, Yuan Y, Xu G, Jia J, Christakis NA. Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature*. (2020) 582:389–94. doi: 10.1038/s41586-020-2284-y

64. Hamidi S, Sabouri S, Ewing R. Does density aggravate the COVID-19 pandemic? *J Am Plan Assoc*. (2020) 86:495–509. doi: 10.1080/01944363.2020.1777891

65. Hamidi S, Ewing R, Sabouri S. Longitudinal analyses of the relationship between development density and the COVID-19 morbidity and mortality rates: early evidence from 1,165 metropolitan counties in the United States. *Health Place*. (2020) 64:102378. doi: 10.1016/j.healthplace.2020.102378

66. Marques TS. O mosaico territorial do risco ao contágio e à mortalidade por covid-19 em portugal continental. *Finisterra*. (2020) 115:19–26. doi: 10.18055/Finis20383

67. Moniz M, Soares P, Nunes C. COVID-19 transmission dynamics: a space-and-time approach. *Port J Public Heal*. (2020) 38:4–10. doi: 10.1159/000515535

68. Barlacchi G, Perentis C, Mehrotra A, Musolesi M, Lepri B. Are you getting sick? Predicting influenza-like symptoms using human mobility behaviors. *EPJ Data Sci*. (2017) 6:27. doi: 10.1140/epjds/s13688-017-0124-6

69. Engebretsen S, Engø-Monsen K, Aleem MA, Gurley ES, Frigessi A, de Blasio BF. Time-aggregated mobile phone mobility data are sufficient for modelling influenza spread: the case of Bangladesh. *J R Soc Interface*. (2020) 17:20190809. doi: 10.1098/rsif.2019.0809

70. Amdaoud M, Arcuri G, Levratto N, Succurro M, Costanzo D. Geography of COVID-19 outbreak and first policy answers in European regions and cities. *Appl Theatr Res*. (2020):58.

71. Gushulak BD, MacPherson DW. Population mobility and infectious diseases: the diminishing impact of classical infectious diseases and new approaches for the 21st century. *Clin Infect Dis*. (2000) 31:776–80. doi: 10.1086/313998

72. Grassly NC, Fraser C. Mathematical models of infectious disease transmission. *Nat Rev Microbiol*. (2008) 6:477–87. doi: 10.1038/nrmicro1845

73. Jorge DCP, Oliveira JF, Miranda JGV, Andrade RFS, Pinho STR. Estimating the effective reproduction number for heterogeneous models using incidence data. *R Soc Open Sci*. (2022) 9:220005. doi: 10.1098/rsos.220005

74. Andersen LM, Harden SR, Sugg MM, Runkle JD, Lundquist TE. Analyzing the spatial determinants of local Covid-19 transmission in the United States. *Sci Total Environ*. (2021) 754:142396. doi: 10.1016/j.scitotenv.2020.142396

75. Zhang CH, Schwartz GG. Spatial disparities in coronavirus incidence and mortality in the United States: an ecological analysis as of may 2020. *J Rural Health*. (2020) 36:433–45. doi: 10.1111/jrh.12476

76. Kang D, Choi H, Kim JH, Choi J. Spatial epidemic dynamics of the COVID-19 outbreak in China. *Int J Infect Dis*. (2020) 94:96–102. doi: 10.1016/j.ijid.2020.03.076

77. Mourao P, Bento R. Explaining COVID-19 contagion in Portuguese municipalities using spatial autocorrelation models. *Rev Galega Econ*. (2021) 30:1–12. doi: 10.15304/rge.30.1.6984

78. Sun F, Matthews SA, Yang TC, Hu MH. A spatial analysis of the COVID-19 period prevalence in U.S. counties through June 28, 2020: where geography matters? *Ann Epidemiol*. (2020) 52:54–59.e1. doi: 10.1016/j.annepidem.2020.07.014

79. Almendra R, Santana P, Costa C. Spatial inequalities of COVID-19 incidence and associated socioeconomic risk factors in Portugal. *Boletín la Asoc Geógrafos Españoles*. (2021) 91. doi: 10.21138/bage.3160

80. Mollalo A, Mohammadi A, Mavaddati S, Kiani B. Spatial analysis of COVID-19 vaccination: a scoping review. *Int J Environ Res Public Health*. (2021) 18:12024. doi: 10.3390/ijerph182212024

81. Middya AI, Roy S. Geographically varying relationships of COVID-19 mortality with different factors in India. *Sci Rep*. (2021) 11:7890. doi: 10.1038/s41598-021-86987-5

82. Gould PR. The slow plague: A geography of the AIDS pandemic Wiley (1993).

83. Cromley EK, McLafferty SL. GIS and public health Cambridge, Massachusetts (USA): Wiley-Blackwell. (2011). 248.

84. Ladyman J, Wiesner K. What is a complex system? Yale (USA): Yale University Press. (2020). 184.

85. Tobler WR. A computer movie simulating urban growth in the Detroit region. *Econ Geogr*. (1970) 46:234–40. doi: 10.2307/143141

86. Brinks V, Ibert O. From Corona virus to Corona crisis: the value of an analytical and geographical understanding of crisis. *Tijdschr Econ Soc Geogr*. (2020) 111:275–87. doi: 10.1111/tesg.12428

87. Mollalo A, Vahedi B, Rivera KM. GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. *Sci Total Environ*. (2020) 728:138884. doi: 10.1016/j.scitotenv.2020.138884

88. Brauer F. Compartmental models in epidemiology. In: F Brauer, DriesscheP van den and J Wu, editors. Mathematical epidemiology. Berlin, Heidelberg: Springer Berlin Heidelberg; (2008). p. 19–79.