Spatial-temporal monitoring risk analysis and decision-making of COVID-19 distribution by region

Leandro Pereira*

ISCTE – Instituto Universitário de Lisboa,
WINNING Lab,
Lisbon, Portugal
Email: leandro.pereira@iscte-iul.pt
*Corresponding author

Jorge Correia, José Sequeiros and José Santos

WINNING Lab,
Rua Fernão Lopes, 409 – 2º Esq 4150-318,
Porto, Portugal
Email: jorge.correia@winning.pt
Email: jose.augusto.sequeiros@winning.pt
Email: jose.santos@winning.pt

Carlos Jerónimo

ISCTE – Instituto Universitário de Lisboa,
WINNING Lab,
Lisbon, Portugal
Email: Carlos.Miguel.Jeronimo@iscte-iul.pt

Abstract: The purpose of this study is to model, map, and identify why some areas present a completely different dispersion pattern of COVID-19, as well as creating a risk model, composed of variables such as probability, susceptibility, danger, vulnerability, and potential damage, that characterises each of the defined regions. The model is based on a risk conceptual model proposed by Bachmann and Allgöwer in 2001, based on the wildfire terminology, analysing the spatial distribution. Additionally, a model based on population growth, chaotic maps, and turbulent flows is applied in the calculation of the variable probability, based on the work of Bonasera (2020). The results for the Portuguese case are promising, regarding the fitness of the said models and the outcome results of a conceptual model for the epidemiological risk assessment for the spread of coronavirus for each region.

Keywords: COVID-19; spatial-temporal; risk analysis; chaos theory; gravity model.

Reference to this paper should be made as follows: Pereira, L., Correia, J., Sequeiros, J., Santos, J. and Jerónimo, C. (2022) 'Spatial-temporal monitoring risk analysis and decision-making of COVID-19 distribution by region', *Int. J. Simulation and Process Modelling*, Vol. 18, No. 1, pp.23–35.

Biographical notes: Leandro Pereira is an Assistant Professor with Habilitation in Management at ISCTE Business School. He received his PhD in Project Management. He is also the CEO and Founder of Winning Scientific Management. He is also the former President of Business Case Institute, PMI Portugal Chapter and Training Specialist of the Court of Auditors. As CEO, he receives from Best Teams Leaders the award of Best Team Leader and CEO of Portugal in 2017 in Portugal. He is also a PMP from PMI and ROI certified. As researcher, he published more than 100 scientific publications and ten books. As a student, he received the best student award from University of Minho. He is an international expert in strategic management, project management, benefits realisation management, and problem solving.

Jorge Correia is a Winning partner, he is in charge of the Winning Porto office. With over 19 years of experience, most of it is in consulting and project management in large national and multinational organisations. He has extensive experience in managing various customer-facing transformation projects in different sectors (retail/FMCG, manufacturing, banking, healthcare, textile, mining, oil and gas, telecommunications, mobile communications, information systems). He has strong analytical skills and great knowledge and awareness of how to solve complex and multifaceted problems, with several years of experience working with multicultural teams and extensive travels in countries around the world (Europe, Asia, Africa and USA).

José Sequeiros is a Manager of Winning Scientific Management, coordinates and develops a team of consultants and trainers in national projects. Certified by IPMA®, he has extensive experience in consulting projects. He is able to develop and deliver trainings and workshops, he also has experience in this area. With a background in Biomedical Engineering and Industrial Engineering and Management, he has extensive knowledge of process optimisation methodologies, quality and operational excellence. He has participated in strategy, innovation, digital transformation and public policy projects across multiple industries. He has participated in integrated projects in various sectors, such as public administration, health, industry, services, among others.

José Santos is a PhD student at ISCTE Business School. He is also a Manager from Winning Incentive Grantes. He is an international expert in quantitative research models and has a huge experience in research projects. He is developing a research on how to capture economic value from investment projects.

Carlos Jerónimo is the Executive Director of the Executive Master in Program and Project Management at INDEG-ISCTE and is an invited Assistant Professor in several curricular units. He is also a founding partner of Winning Scientific Management, where he coordinates and develops a team of consultants and trainers in national and international projects, being responsible for the sale, technical innovation and quality assurance of the execution of the services provided. He received a degree in Computer and Telematics from the University of Aveiro, PhD in Strategic Management at ISCTE, has developed his entire career in the area of consulting in implementation, customisation and design of methods and systems applicable to management support.

1 Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) outbreak has been ongoing since it was first reported in December 2019 in China, and it rapidly took pandemic proportions (Søreide et al., 2020). This widespread of the coronavirus or COVID-19 virus could be compared to the spreads of fires in California (Bonasera et al., 2020; Jia et al., 2020). The start occurs in one focal point and quickly spread over larger regions until it becomes difficult to stop. As of June 2021, the number of cases and deaths exponentially incremented reaching a total of 174 million confirmed cases and 3.8 million deaths (Worldometer, 2021).

Mathematical models are a powerful tool that proved their importance in previous diseases outbreaks, conducive to the understanding of the dynamics of disease (Hazarika and Gupta, 2020; Lyra et al., 2020; Nogueira et al., 2020). They provide useful predictions about the transmission of the disease and the effectiveness of possible control measures. This kind of information is crucial for public health policymakers (Jia et al., 2020; Kang et al., 2020).

The SARS-CoV-2 virus had the first confirmed case in mainland Portugal in March 2020. This virus had an initial spread more centralised in the North region. The distribution and initial concentration of the infected have been the subject of various assumptions and empiricism. However, no study indicates or quantifies the reason for this geographic distribution quantitatively, that is, using complex models that allow determining and modelling a correct geographic distribution based on demographics or on statistically significant variables that add statistical value to the model. The interest in this dichotomy between the various NUTS III of mainland Portugal is, from a scientific point of view, crucial to understand the spread of the pandemic at a national level and to be able to take mitigating measures that allow the reduction or control of the pandemic dispersion. The purpose of this study is to model, map, and identify why these areas present a completely different dispersion pattern, as well as creating a risk model, composed of variables such as probability, susceptibility, danger, vulnerability, and potential damage, that characterises each of the defined regions.

2 Literature review

Mathematical epidemiology is a subfield of epidemiology that is based on the characteristics of biological phenomena transcribing them into models that will be solved by analytical and simulation processes. In the mathematical modelling of epidemics, two types of models are considered: deterministic and stochastic (Chen, 2015; Singh et al., 2018).

Deterministic models: given the initial conditions, the epidemic process is generally described through differential equations that model the infectious process within a dynamical system. In deterministic models it is possible to control the factors that intervene in the dynamics of the process, and, therefore, it is possible to predict concrete results according to these factors (Philippe and Mansi, 1998). There is no ambiguity in the results. Stochastic models: are probabilistic, with the state variable being an inherent property to every individual in the population. This approach may involve more detail about the properties of the system, which can be elaborated based on the properties and actions of everyone in the population and its structure of contacts (Uehara et al., 2012). In stochastic models, it is impossible to control the factors that intervene in the dynamics of the phenomenon, and consequently, it is not possible to obtain unique results. Each possible outcome has an associated probability. Stochastic models depend on random variations such as the risk of disease exposure and the contagion period. These models are mainly used when these variations are considered important, as is the case in small populations.

2.1 Risk conceptual model

In 2001, a risk conceptual model was proposed (Bachmann and Allgöwer, 2001) based on the wildfire terminology, analysing the spatial distribution of wildland fire risk using a geographic information system. This model is composed of several variables, which can be adapted, as well, to the epidemiological context. Not because of the spatial behaviour but because the development and assumption of the different components of the model which allow the implementation of a more comprehensive risk model covering several different aspects of a risk situation development.

2.2 Probability

Probability expresses the likelihood that a given event will occur and thus, it can be understood as an indicator of the uncertainty of the occurrence of this event. In a classical approach, it is understood that all events, not being conditioned to the previous existence of others, have the same possibility of occurring and therefore an equal probability. In conditioned probabilities, it is understood that a given event has a given probability of occurring, conditioned on the probability that a previous event has occurred (Calapez et al., 2021).

2.3 Susceptibility

Susceptibility expresses the propensity of a given area or territorial unit to be affected by the phenomenon studied, evaluated based on its intrinsic properties. A territorial unit will be susceptible as it is more affected or enhances the occurrence and development of the phenomenon. In the case of COVID-19, a given area will be more susceptible the better it allows the propagation of the pandemic (Weissman et al., 2020).

2.4 Danger

The danger is equivalent to what is called a hazard in Anglo-Saxon literature. The danger is, according to Varnes' definition (1984), the probability of occurrence of potentially destructive phenomena, in each time interval and each area. This notion of danger encompasses two dimensions: time and space. Therefore, it encompasses the two components described above, probability, whose calculation can be based on the existing history for the event, and susceptibility, which addresses aspects related to the territory for which the phenomenon is being studied.

2.5 Vulnerability

Vulnerability expresses the degree of loss to which a given element is subject to the occurrence of the treated phenomenon. Vulnerability is expressed on a scale that varies between zero – no damage occurs – and one – the damage is total, destroying the element at risk (Rivera-Izquierdo et al., 2020).

2.6 Risk

In the literature, the mathematical expression of risk is often found as the product of danger and vulnerability, R = P x V (Bachmann and Allgöwer, 2002). One difficulty this approach raises is that it cannot adequately differentiate the actual loss of different elements with the same vulnerability. Looking at COVID-19, and as an example, a small area may have a greater vulnerability than a considerably larger area and is therefore subject to a greater degree of infection. That is, based on R = P x V if we assume that the risk is higher in areas with higher population density will seek to identify a contingency plan for these areas first and placing the remaining areas with a lower risk scale. However, the chances of infection in areas with lower density may correspond to a greater attractiveness of dispersion. At this point the introduction of an additional variable, the economic value variable (number of infected people x potential number of admissions), is useful. Potential harm is thus the product between vulnerability and the value of the element at risk.

Recalling the definition presented by Bachmann and Allgöwer (2002), the risk is the probability that

COVID-19 will occur in a specific location, under certain circumstances, and its expected consequences, characterised by impacts on the affected objects. Based on this definition and transposing the same conceptual framework to the pandemic, internationally accepted in other domains, risk will be understood here as the product between danger and potential damage (Figure 1).

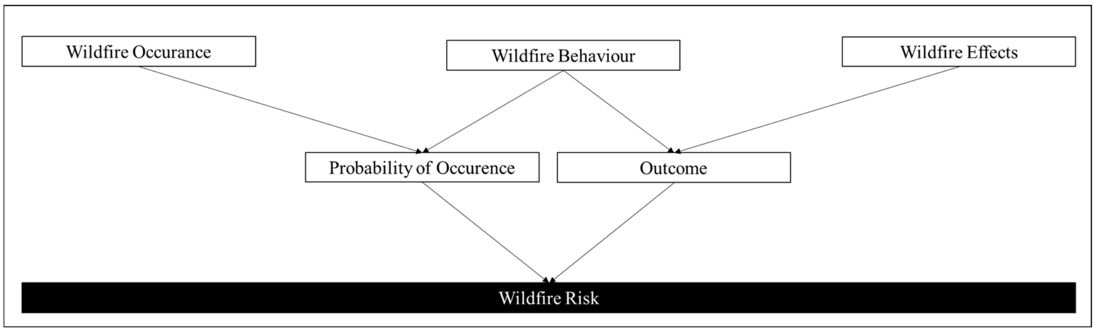
The import of these concepts to an area such as epidemiology implies the use of more complex mathematical models as integrators of various factors.

Modern chaos theory has been applied successfully to several science fields which have some common features: a small perturbation, grows exponentially with a defined coefficient, the Lyapunov exponent, and finally saturates (Bonasera and Zhang, 2020; Zheng and Bonasera, 2020).

$$(Bonasera\ and\ Zhang,\ 2020)$$

$$N(d) = \frac{d_o d_\infty}{do + d_\infty e^{-\lambda d}} \tag{1}$$

Figure 1 Risk methodology and wildfire research



Source: Andreas Bachmann and Allgöwer (2001)

As Bonasera (2020) puts it, equation (1) is the logistic equation as a model of population growth. It is the solution of a simple first-order non-linear ordinary differential equation, in which d0 indicates the moment that a small perturbation occurs, it grows exponentially with a coefficient λ, the Lyapunov exponent, and it finally saturates to a value d∞ >> d0 (Bonasera et al., 2020; Bonasera and Zhang, 2020). This is an adequate model for the parameterisation of the probabilistic component of the risk model presented by Bachmann and Allgöwer (2001).

The gravity model is a framework borrowed from transportation theory. It can also be used to model the spread of infectious diseases (Kraemer et al., 2019; Vespignani et al., 2020). The geographic spread of infectious pathogens may be driven by infected individuals traveling between areas and on the local characteristics such as population density and contact patterns, among others.

Like the gravity model, the radiation model considers the origin and destination of the trips made, which allows calculating the time and distance, and the drawn from other populations within the same radius, allowing it to create patterns. This model assumes that every location has a certain level of competitiveness and attractiveness.

Equation (2) (gravity model) and equation (3) (radiation model) represent the predicted human movements between each pair of places, so the human mobility estimates are reflective of the general fluxes of the population. In other words, the susceptibility of a given location to be affected by the phenomenon.

(*Kraemer et al.*, 2019)

$$T_{i,j} = k\frac{N_i^{\alpha} N_j^{\beta}}{d_{i,j}^{\gamma}} \tag{2}$$

(*Kraemer et al.*, 2019)

$$T_{i,j} = T_i \frac{N_i N_j}{\left(N_i + S_{i,j}\right)\left(N_i + N_j + S_{i,j}\right)} \tag{3}$$

In 2019, this model was validated on data from the 2014-2016 Ebola virus disease outbreak in West Africa (Kraemer et al., 2019) and in 2020, again, by Kraemer, the effect of human mobility and control measures on the COVID-19 epidemic in China (Vespignani et al., 2020).

2.7 *Theme for the investigation*

The solution proposed for this study is based on two aspects, deterministic and probabilistic. This strand is framed in two quantitative models, chaos theory, and gravity model to find a trend or pattern that allows an in-depth study of measurements of geometric distance, density, and complex and dynamic systems on the SARS-CoV-2 virus as well as creating a risk model. Chaos theory will be used to model how a small disturbance can grow exponentially and then saturate to a finite value, a mechanism like the dissemination of COVID-19. On the other hand, the gravity model simulates the dispersion of COVID-19 based on the centroids of the regions, tending to be the most populated places, with the places with the greatest circulation of transport, namely airports and international railway stations. Due to the population dynamics, namely the pedestrian behaviours (Mykoniatis et al., 2021) the inclusion of this model is mainly aimed at including the spatial and geographic variables in the spread of the pandemic. This model aims to reproduce the observed timing and spread of the pandemic at a regional level, considering variables as transmission, population size, vulnerability, and distance.

2.8 *Critical analysis*

The spatial distribution of COVID-19, despite being studied by several authors including Bonasera and Zhang (2020), Jia et al. (2020) and Santos (2020) only indicates the current distribution of the number of COVID-19 infections, not calculating or using predictive models for the construction of a risk component composed of several variables. In terms of epidemic simulation models applied to COVID-19, Bai (2020) applied a system of first-order ordinary differential equations (ODEs) and spatial agent-based model (ABM) although without having into account the chaotic movement of people along the different stages of the several quarantines imposed. Thus, this risk model proposal, based in the conceptual frame of Bachmann and Allgo (2001), aims to consolidate issues not addressed in other studies, enabling the prediction of risk variation in each region as a function of the evolution predicted by the model using the conjunction of a mathematical and a statistical model.

3 Research objectives/research questions

Based on the adoption of a deterministic and stochastic philosophy, the purpose of finding a trend or standard that allows an in-depth study of measurements of geometric distance, density, and complex and dynamic systems on the SARS-CoV-2 virus. These models are complementary to each other and will allow a more comprehensive and holistic view of the pandemic situation in Portugal. That is, the conjugation of these models will allow the understanding, or at least a significant approximation, to the development of a risk model with a high level of accuracy for understanding the dynamics of COVID-19 propagation at the national level.

4 Research methodology

COVID-19 virus has shown different variations since the beginning of the pandemic. To have a brighter view of its impact, the government resorted to the method of epidemiological waves temporal space. According to the World Health Organization, the term 'wave', when related to the virus, corresponds to the variation of the cases (Hazarika and Gupta, 2020). We can state that a wave ended when the virus was brought under control and cases had fallen substantially and, for the next wave to start, it is necessary to have a sustained rise in infections. Therefore, a wave can be defined by two main characteristics: its upwards and downwards periods; and by the increase in an upward period or the decrease in a downward period to be substantial by sustaining over a period to distinguish them from an uptick, a downtick, reporting errors, or volatility in new cases.

In Portugal, it is possible to identify three different waves since the beginning of the pandemic. The first started in March 2020 and it lasted until the beginning of May. This first was characterised by the appearance of this virus in Portugal, because of a shoe fair in Italy. The second wave, dated from October to the beginning of December, was marked by a significant increase in the indicators, such as the number of cases and people in the UCI, leading to another lockdown. The third and final wave started in December and ended at the end of January 2021. This was the one that had the most impact in Portugal with all the indicators achieving their maximum and bringing the country to the brink of collapse. By that time, Portugal had achieved a maximum of 12,000 cases a day (urworldindata.org, 2021).

4.1 Proposed risk conceptual model

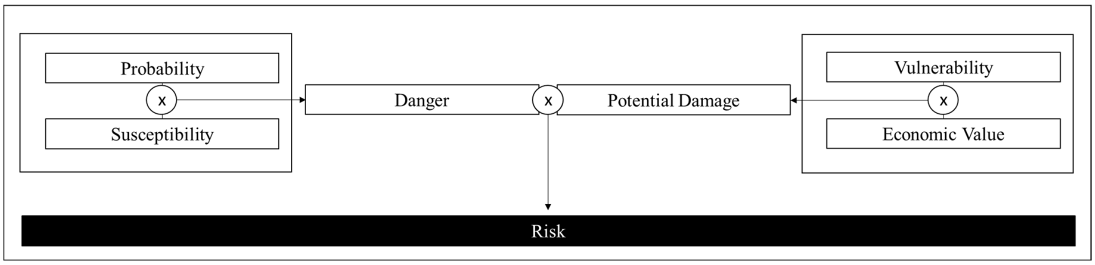
Based on the work of Bachmann and Allgöwer (2001), the proposed risk conceptual model is as in Figure 2. In the proposed model, the risk can be calculated by multiplying the potential damage by the danger. In their work, they defined forest fire risk as the probability of a wildfire to occur at a specified location and under given circumstances and its expected outcome as defined by the impacts on the affected objects. The migration from the initial premise to the proposed model is based on this same concept. It tries to comprehend the probability of a positive case to occur at a specified location and under given circumstances and its expected outcome as defined by the impacts on the affected people.

In the model, the danger is calculated by the probability of the event and how susceptible the area is to that event, while the potential damage is calculated by the economic value multiplied by the vulnerability. The main innovation of this model is related to how each of its basic components is obtained. The use of chaos theory and gravity model to model the probability and susceptibility components, respectively, is shown as the main novelty.

The evolution of COVID-19 transmission has been widely researched in the past year. Bonasera and Zhang, (2020) and Zheng and Bonasera (2020) found not only a way to model the evolution of the number of COVID cases, but also the probability of a test coming up positive. The models were originally tested for the Italian population, therefore both models were tested for the Portuguese reality using data collected from official data from Portuguese institutions, namely the Directorate-General for Health (DGS, 2021). After testing if the model fit the Portuguese data, the model was used to calculate the probability of a COVID-19 test from each area be positive on a given day.

This study used the NUTS III, Nomenclature of Territorial Units for Statistical Purposes, which is a hierarchical system used to divide the territory into regions, according to the Regulation (EC) No. 1059/2003 of the European parliament and the Council of 26 May 2003. These references are applied by the European Union and Eurostat to elaborate regional statistics and to define and elaborate local politics and funds distribution, respectively. When using this classification as a reference, it was possible to identify more regionally specific data, such as the COVID-19 tests performed, positive tests, virus effects, among others reported daily, important for the study as well as spatial detail which will allow a more concrete statistical analysis. If the authors have chosen, the NUT II it will englobe a low level of detail for the proposed work. In the opposite, level of municipality, will create a complex entropy due to not uniform data distribution among the several municipalities which could envisage the model. Thus, using a NUTS III in this study will provide a more comprehensive spatial data distribution which can be replicated to same geographical areas in other European countries due to the same geographical division.

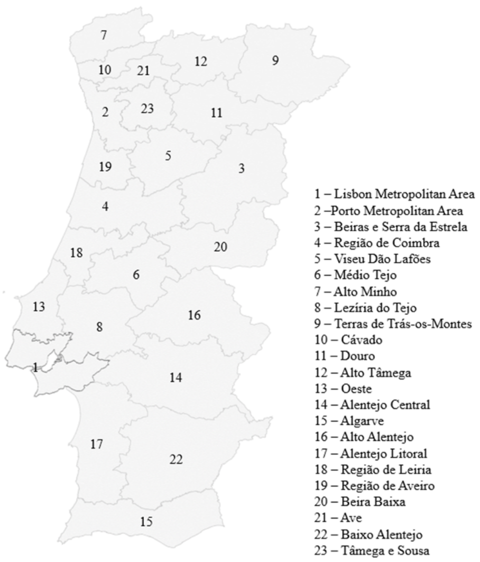
Figure 2 Proposed epidemiological risk conceptual model for COVID-19



Source: Self-elaborated

To calculate susceptibility, a model was created based on the gravity model. Including as focus points all airports and train stations with international connections, the first step was to calculate the mean distance of each area to these focal points. The closer the area to all the focal points, the more susceptible it was to being affected by COVID, as these focal points are entries for new transmission chains. Since COVID-19 is transmitted by people, the susceptibility metric is given by the mean distance to the focal points adjusted by the population density of each area. The values of population density to each area were retrieved from PORDATA (2021a). Once all susceptibility values were calculated, they were normalised for values between one and zero. While it was important to normalise the scale, the area less susceptible should not be ignored, therefore zero was replaced with a very small value – 10–5.

Figure 3 Identification of mainland Portuguese NUTS III



The danger parcel of the model is given by the susceptibility multiplied by the probability, since the first was already normalised and the second is a value between zero and one, there is no need to normalise the result. Furthermore, the probability is the only factor used allowed to be null, which means that danger can also be null if the probability of a test coming up positive for that area was found zero by the probability model.

When analysing the effects of COVID-19, human life is the main affected asset, and therefore, the economic value of each area is given by the total population of the area. The values used for the total population were collected from PORDATA (2021b), which represent the total number of people living in the area at the end of 2019.

As found by many authors, including Götzinger and Santiago-García (2020), Goujon et al. (2020), Meena et al. (2020) Nogueira et al. (2020), Rivera-Izquierdo et al. (2020) and Shahid et al. (2020), COVID-19 has been affecting more older people, not only are they more likely to develop serious symptoms but are also more likely to die from it. This means that areas, where the population is older, are more vulnerable to the disease. This vulnerability was calculated by comparing the population distribution by age group of each area with the country mean. If the area has more elders than the country's average, then it will increase its vulnerability. On the other hand, if there are more children or teenagers then the vulnerability will decrease. Finally, the values calculated were normalised and the zero values were replaced with 10–5.

The potential damage of each area is given by the normalised value of the total population multiplied by the vulnerability of the area. Afterward, these values are normalised and multiplied by the danger already calculated to obtain the values of risk, which are once more transformed in scale from zero to one. All the variables are summarised in Table 1.

Figure 4 represents by pseudocode, the form of computational implementation using the Miniconda version 4.9.2 and Python version 3.8 tools.

Figure 4 Pseudocode for calculating relative risk

Algorithm 1: Pseudocode for calculating relative risk

```
Result: Spatio-Temporal Monitoring Risk Analysis and decision-making of Covid-19
        Distribution in Portugal by NUTS II
initialization;
while NUTS III area to analyse do
    calculate Probability (based on Equation 1);
    calculate Susceptibility (based on Equation 2 and 3);
    calculate Vulnerability (based on population distribution by age group of each area
      with the country mean);
    calculate Economic Value (based on total population of the area);
    calculate Relative Risk;
    create heat map with Relative Risk
end
```

Source: Self-elaborated

Table 1 Summary table with the different variables, their description, and calculation method

Variable	*Description*	*Calculation method*
Probability	Probability of a test coming up positive	Positives divided by the total number of tests
Susceptibility	The closer the area to all the focal points, as airports and international train stations, the more susceptible of being affected by COVID	Euclidean distance function to calculate distances between centroids and focal points
Vulnerability	Comparing the population distribution by age group of each area with the country means	Population distribution data by age group, calculate the difference in the distribution of the entire country, normalise the vulnerability values
Economic value	Number of individuals living in a certain region	Total population of the area
Danger	The older the population and higher the probability, the more dangerous is the region.	Susceptibility multiplied by the probability
Potential damage	The normalised value of the total population multiplied the vulnerability of the area	Population distribution data, calculate the difference in the distribution of the entire country, normalise the vulnerability values
Risk	The probability that COVID-19 will occur in a specific location, under certain circumstances, and its expected consequences, characterised by impacts on the affected objects	Potential damage multiplied by danger

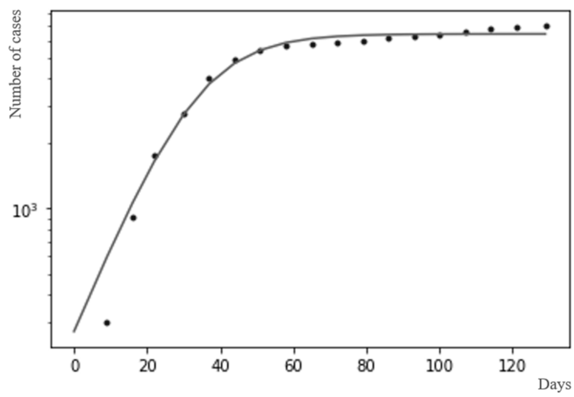
Source: Self-elaborated

5 Results and discussion

5.1 Risk conceptual model

Table 2 aggregates all the results obtained for the parameterisations of the different components of the risk model, by region, on May 13, 2021. The table sorts the relative risk by region from highest to least, namely in the Lisbon metropolitan area to Tâmega and Sousa. Predictably, the regions of the metropolitan areas of Lisbon and Porto are those that present a higher relative risk, 1 and 0.631, respectively. In contrast, Baixo Alentejo and Tâmega and Sousa are the regions with the lowest relative risk, mainly due to their location and vulnerability. Given the variables in question, such as resident population and distance to transport focal points, the two largest Portuguese cities have the highest relative values. On the study date, and regarding the probability of obtaining a positive test, the metropolitan area of Lisbon also had the highest values (0.265), followed by the regions Viseu dao Lafões and Coimbra (0.265 and 0.249), places with many active cases. On the other hand, Alto Tâmega and Tâmega e Sousa have the lowest probability values (0.132 and 0.154, respectively). Beira Baixa and Baixo Alentejo are the regions with the lowest susceptibility, due to the distance of their centroids and focal points of transportation in Portugal. Also, as expected, the regions of the metropolitan areas of Lisbon and Porto have the highest values. Alto Tâmega, Terras and Trás-os-Montes e Beiras e Serra da Estrela are the regions with the highest vulnerability (1, 0.854 and 0.733, respectively) meaning that they are the regions with the oldest population in the country, unlike the regions of Aveiro, Ave, and Tâmega e Sousa.

Figure 5 Number of positive cases for Alentejo Litoral



Source: Self-elaborated

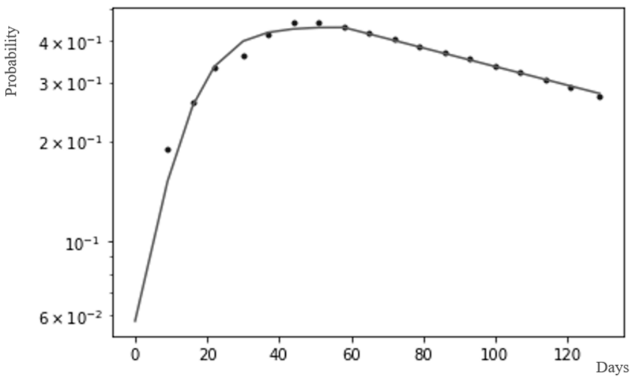
The first step was to test the fitness of the model proposed by Bonasera et al. (2020), Bonasera and Zhang (2020) and Zheng and Bonasera (2020). The model calculates the number of positives, and the probability of a test is positive for individual waves of the disease. Therefore, this study focused on the third Portuguese wave that started after Christmas. It was possible to model the total number of positive COVID-19 cases of each area with very high accuracy, the lowest R^2 obtained was 0.985 for Alentejo Litoral (Figure 5). The model for the probability of a test coming out positive also adjusted very well to the data, the lowest R^2 was 0.966 achieved for the metropolitan area of Lisbon (Figure 6). In Table 3 are the equation (1) parameter values obtained for the different Portuguese mainland regions.

Table 2 Results obtained for the different components of the proposed risk model, by region, on May 13, 2021

Number	*NUTS III*	*Susceptibility*	*Probability*	*Economic value*	*Vulnerability*	*Danger*	*Potential damage*	*Relative risk*
1	Lisbon metropolitan area	9.98E-01	2.65E-01	1.00E+00	1.15E-01	2.65E-01	1.00E+00	1.00E+00
2	Porto metropolitan area	9.95E-01	1.69E-01	6.04E-01	1.90E-01	1.68E-01	9.96E-01	6.31E-01
3	Beiras e Serra da Estrela	9.14E-01	2.35E-01	7.39E-02	7.33E-01	2.15E-01	4.70E-01	3.82E-01
4	Região de coimbra	9.62E-01	2.49E-01	1.52E-01	2.76E-01	2.40E-01	3.63E-01	3.29E-01
5	Viseu Dão Lafões	9.36E-01	2.65E-01	8.79E-02	4.00E-01	2.48E-01	3.04E-01	2.85E-01
6	Médio Tejo	9.22E-01	1.98E-01	8.12E-02	4.42E-01	1.83E-01	3.11E-01	2.15E-01
7	Alto Minho	9.55E-01	2.40E-01	8.05E-02	3.11E-01	2.29E-01	2.17E-01	1.88E-01
8	Lezíria do Tejo	8.52E-01	2.24E-01	8.31E-02	2.92E-01	1.91E-01	2.10E-01	1.52E-01
9	Terras de Trás-Os-Montes	7.25E-01	1.83E-01	3.75E-02	8.54E-01	1,33E-01	2,78E-01	1,39E-01
10	Cávado	9.87E-01	2.10E-01	1.41E-01	1.10E-01	2.07E-01	1.35E-01	1.05E-01
11	Douro	9.32E-01	2.18E-01	6.66E-02	2.36E-01	2.03E-01	1.37E-01	1.05E-01
12	Alto Tâmega	7.63E-01	1.32E-01	3.01E-02	1.00E+00	1.01E-01	2.61E-01	9.95E-02
13	Oeste	9.41E-01	2.04E-01	1.25E-01	1.05E-01	1.92E-01	1.14E-01	8.24E-02
14	Alentejo Central	4.70E-01	2.14E-01	5.32E-02	4.19E-01	1.00E-01	1.93E-01	7.33E-02
15	Algarve	9.61E-01	1.63E-01	1.53E-01	8.26E-02	1.57E-01	1.10E-01	6.50E-02
16	Alto Alentejo	3.77E-01	2.02E-01	3.65E-02	6.27E-01	7.62E-02	1.98E-01	5.71E-02
17	Alentejo Litoral	5.59E-01	1.86E-01	3.27E-02	4.96E-01	1.04E-01	1.40E-01	5.51E-02
18	Região De Leiria	9.02E-01	1.90E-01	9.94E-02	8.76E-02	1.71E-01	7.55E-02	4.87E-02
19	Região de Aveiro	9.50E-01	1.99E-01	1.27E-01	4.52E-02	1.89E-01	4.98E-02	3.55E-02
20	Beira Baixa	2.80E-01	1.61E-01	2.80E-02	7.29E-01	4.49E-02	1.77E-01	3.00E-02
21	Ave	9.87E-01	1.96E-01	1.44E-01	2.44E-02	1.93E-01	3.04E-02	2.22E-02
22	Baixo Alentejo	1.00E-04	2.40E-01	4.06E-02	4.98E-01	2.40E-05	1.75E-01	1.59E-05
23	Tâmega e Sousa	9.83E-01	1.54E-01	1.45E-01	1.00E-05	1.51E-01	1.45E-06	2.20E-07

Source: Self-elaborated

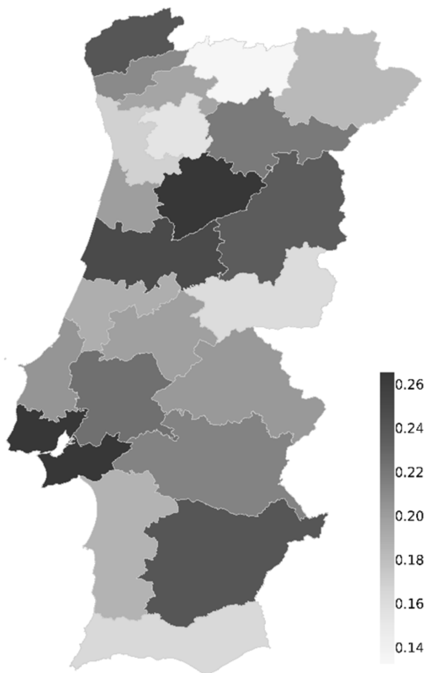
Figure 6 Probability for M.A. of Lisbon



Source: Self-elaborated

Once the parameters for each probability model were known it was possible to calculate the probability of each test for the present and future dates. Comparing the results obtained with those registered by Bonasera, roughly a year earlier, these are very close. In his work, Bonasera, says that for the different regions of Italy, the probability of testing positive for the virus as a function of time at 120 days will vary approximately between 0.1 and 0.45. The average value obtained for the regions in Portugal is 0.2, with a maximum of 0.265 referring to Viseu Dao Lafões and a minimum of 0.132 in Alto Tâmega, Figure 7.

Figure 7 Probability distribution, by region



Source: Self-elaborated

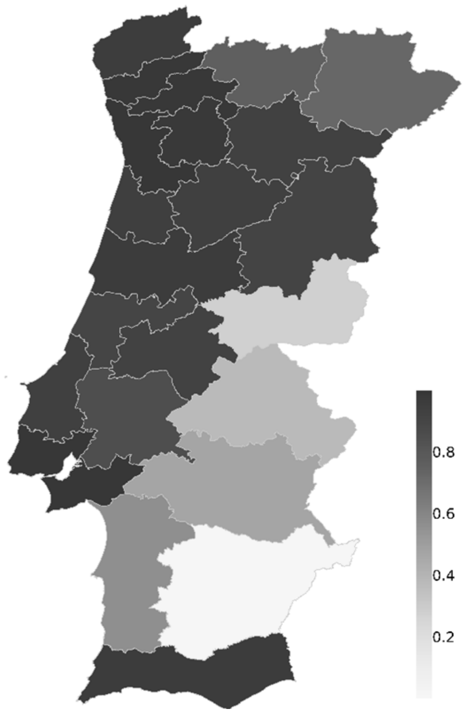
Table 3 Fitness values for the probability parameter for the different Portuguese mainland regions

Number	*NUTS III*	*D0*	*d∞*	*λ*	*d*
1	Lisbon metropolitan area	0.06634	0.44063	0.13830	57.39568
2	Porto metropolitan area	0.008363	0.27998	0.49941	51.32681
3	Beiras e Serra da Estrela	0.037866	0.42375	0.28606	53.98518
4	Região de Coimbra	0.05265	0.41113	0.17981	58.24238
5	Viseu Dão Lafões	0.05414	0.50427	0.21761	48.11530
6	Médio Tejo	0.02559	0.35983	0.32474	42.89124
7	Alto Minho	0.04168	0.42617	0.17460	50.82159
8	Lezíria do Tejo	0.04642	0.37081	0.16175	58.04667
9	Terras de Trás-Os-Montes	0.01155	0.34102	0.46523	50.36391
10	Cávado	0.00490	0.37706	0.62805	43.14844
11	Douro	0.034343	0.38332	0.31338	53.71805
12	Alto Tâmega	0.00302	0.25392	0.61109	34.42030
13	Oeste	0.043665	0.33410	0.19072	58.69298
14	Alentejo Central	0.01037	0.42999	0.49797	33.68746
15	Algarve	0.02160	0.25883	0.29582	54.53307
16	Alto Alentejo	0.02748	0.38462	0.30978	46.19382
17	Alentejo Litoral	0.03489	0.26899	0.14814	59.50167
18	Região De Leiria	0.02562	0.33280	0.32152	49.28826
19	Região De Aveiro	0.01330	0.32570	0.43044	48.63915
20	Beira Baixa	0.02098	0.30217	0.35030	49.23719
21	Ave	0.01292	0.35320	0.46704	48.00785
22	Baixo Alentejo	0.02250	0.38114	0.30685	58.11607
23	Tâmega e Sousa	0.01256	0.24104	0.40628	57.82926

Note: according to equation (1).

Source: Self-elaborated

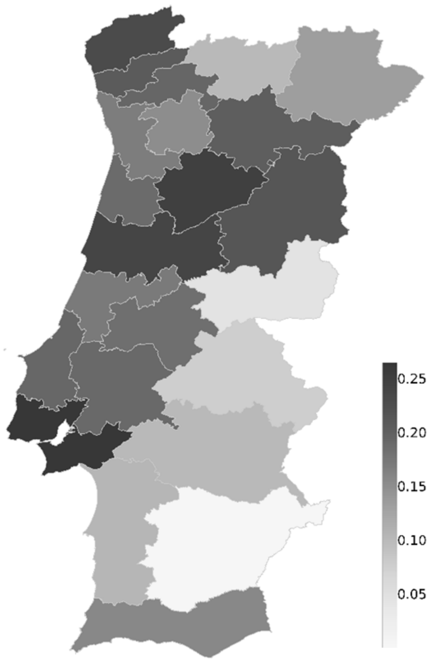
Figure 8 Susceptibility distribution, by region



Source: Self-elaborated)

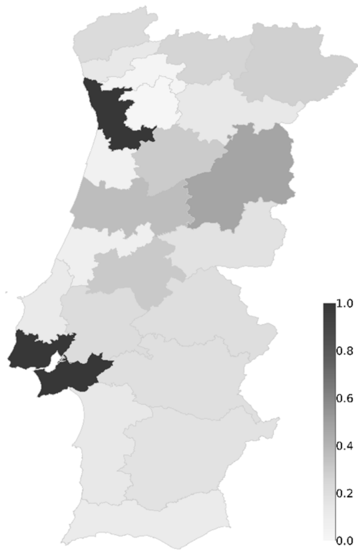
The susceptibility of each area was calculated according to the processes described before, and it was possible to identify Lisbon's metropolitan area as the most susceptible area closely followed by the Porto's metropolitan, and Baixo Alentejo as the least susceptible, Figure 8. These results are aligned to what could be expected, the metropolitan areas not only have the highest population densities but also include airports and train stations with international connections. On the other hand, Baixo Alentejo not only is further from the international connections used as focus points but also has a very low population density. Afterward, it was possible to calculate the danger of each area, as can be seen in Figure 9. While the areas in most or least danger are the ones most and least susceptible, the relationship is not true for the remaining areas. For instance, the metropolitan area of Porto, which was the second most susceptible is only the 14th in most danger due to its lower probability.

Figure 9 Danger distribution



Source: Self-elaborated

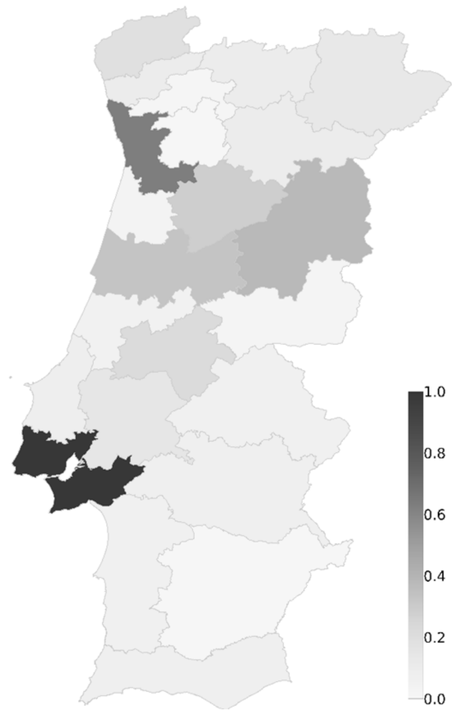
Figure 10 Potential damage distribution



Source: Self-elaborated

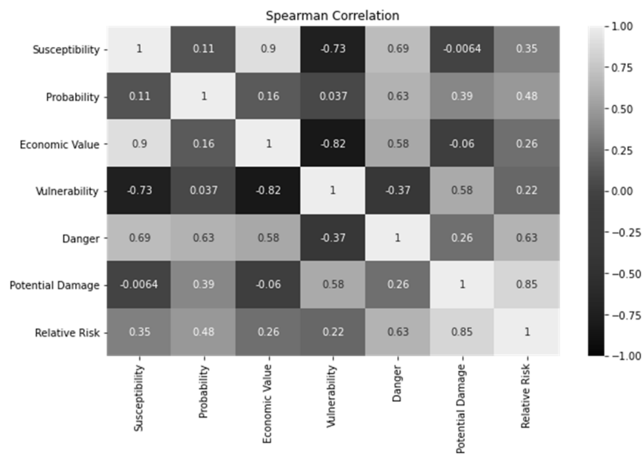
Figure 10 shows the potential damage of each area, calculated according to the description given in the methodology section. Once again, this map shows what was already expected given that the Portuguese population, especially the younger generation, is concentrated in the metropolitan areas. Finally, the risk values can be obtained by multiplying these two final factors (Figure 11). The final map shows the relative risk of each area for COVID-19 transmission, which enables the creation of localised measures pointing to the most vulnerable areas. The most susceptible areas to being infected by COVID-19 are the metropolitan areas and the Beiras region.

Figure 11 Relative risk distribution



Source: Self-elaborated

Figure 12 Spearman correlation for the parameters introduced in the proposed risk model

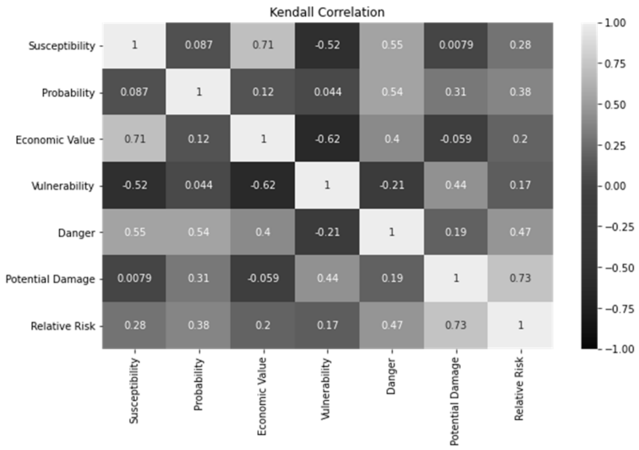


Source: Self-elaborated

5.2 Correlation indexes

Robust correlation measures can be used to construct multivariate covariance matrices, based on pairwise covariances (Puth et al., 2015). Non-parametric correlation estimators as the Kendall and Spearman correlation are widely used in the applied sciences. They are often said to be robust, in the sense of being resistant to outlying observations. They are examples of a non-parametric rank-order correlation that does not make any assumptions about the distribution of the data (Croux and Dehon, 2010). They also measure the monotonic relationship between two variables (Puth et al., 2015). A higher absolute value of Spearman's or Kendall's correlation coefficient indicates that there is a monotonic (but not necessarily linear) relationship between the variables (Puth et al., 2015). The correlation between the different parameters of the risk model has been tested. Figure 12 and Figure 13 show the obtained results for Spearman and Kendall correlation, respectively.

Figure 13 Kendall correlation for the parameters introduced in the proposed risk model



Source: Self-elaborated

Pearson's correlation presents better results than Kendall's correlation. As described above, these correlations describe some monotonic relationships between parameters, namely in the pair susceptibility and economic value, potential damage and relative risk, with a positive relationship, and the pairs vulnerability and susceptibility and vulnerability and economic value with a negative relationship. Coefficients close to 0 represent non-correlation between parameters. The correlations verified make sense within the pandemic context and the model itself. In terms of monotonic growth, the proximity to mobility centres is also reflected in the proportion of the population, as well as potential damage and relative risk. In the inverse field, vulnerability has a negative monotonic relationship with the parameters referring to susceptibility and economic value, in the sense that, in practice, the regions furthest away from the economic centres are more aged.

6 Conclusions

The COVID-19 pandemic was and still is analysed by its waves. This not only allows us to understand when it has reached its maximum but also positioned it in terms of events and festivities. The need for a more robust risk analysis model, which allows for the application of more localised measures and actions, proved to be extremely important during the different waves of COVID-19 seen in Portugal, especially since it differs from region to region, as shown with NUTS III.

A model based on work carried out in the 1990s and 00s on risk models for the assessment of wildland fires was proposed, with changes in the calculation of their components. These changes allow the application of more sophisticated models, such as the gravity model, radiation model, and the determination of the risk applied to this context, which is closer to the evolution of the pandemic.

The application of the chaotic model to calculate the probability had interesting results, with an average fit (R^2) in the order of 0.9804. The development of the curve and the model itself is closer to that verified empirically with the accentuated growth in the number of positive cases until reaching a plateau of stabilisation.

We believe that the inclusion of factors, such as the presence of airports and train stations with international tickets, population and age indicators adds strength to the analysis carried out and the application of measures more focused on regional needs and contexts. Therefore, the areas that were identified as being riskier are the metropolitan areas, for their traffic and international flow, and Beiras, mostly because of its elder population. Given the analysis carried out with these correlations, it is important to emphasise that the analysis of relative risk should not be done in isolation, but together with other parameters to complement the information. Different positive or negative variations may affect the negative risk value, hindering the implementation of effective and efficient measures locally, given geographic and demographic differences.

7 Limitation of the study and future work

The proposed model does not include the incorporation of information such as the total number of vaccinated and by age group, migratory and tourist flows, or phases of deconfinement. As future work, it is proposed the inclusion of these factors.

Acknowledgements

This article is a result of the project 'Análise Estatística e Modelação do Risco associado à Pandemia COVID-19' (AAC 15/SI/2020), supported by Lisbon Portugal Regional Operational Programme (LISBOA 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF).

References

Bachmann, A. and Allgöwer, B. (2001) 'A consistent wildland fire risk terminology is needed', *Fire Management Today*, Vol. 61, No. 4, pp.28–33, https://doi.org/10.2307/j.ctvx078wk.11.

Bachmann, A. and Allgöwer, B. (2002) 'Uncertainty propagation in wildland fire behaviour modelling', *International Journal of Geographical Information Science*, Vol. 16, No. 2, pp.115–127, https://doi.org/10.1080/13658810110099080.

Bai, S. (2020) 'Simulations of COVID-19 spread by spatial agent-based model and ordinary differential equations', *International Journal of Simulation and Process Modelling*, Vol. 15, No. 3, pp.268–277.

Bonasera, A. and Zhang, S. (2020) 'Chaos, percolation and the coronavirus spread', *Frontiers in Physics*, April, Vol. 8, pp.1–5, https://doi.org/10.3389/fphy.2020.00171.

Bonasera, A., Bonasera, G. and Zhang, S. (2020) 'The coronavirus spread: the Italian case', *European Physical Journal Plus*, Vol. 135, No. 6, pp.1–12, https://doi.org/10.1140/epjp/s13360-020-00494-6.

Calapez, T., Melo, P., Andrade, R. and Reis, E. (2021) *Estatística Aplicada*, Vol. 1, Silabo.

Chen, D. (2015) 'Modeling the spread of infectious diseases: a review', *Analyzing and Modeling Spatial and Temporal Dynamics of Infectious Diseases*, December 2014, pp.19–42, https://doi.org/10.1002/9781118630013.ch2.

Croux, C. and Dehon, C. (2010) 'Influence functions of the Spearman and Kendall correlation measures', *Statistical Methods and Applications*, Vol. 19, No. 4, pp.497–515, https://doi.org/10.1007/s10260-010-0142-z.

DGS (2021) *Ponto de Situação Atual em Portugal* [online] https://COVID19.min-saude.pt/ponto-de-situacao-atual-em-portugal/ (accessed 11 June 2021).

Götzinger, F. and Santiago-García, B. (2020) 'COVID-19 in children and adolescents in Europe: a multinational, multicentre cohort study', *The Lancet Child and Adolescent Health*, Vol. 4, No. 9, pp.653–661, https://doi.org/10.1016/S2352-4642(20)30177-2.

Goujon, A., Natale, F., Ghio, D., Conte, A. and Dijkstra, L. (2020) 'Age, gender and territory of COVID-19 infections and fatalities', *BMJ Open*, Vol. 10, https://doi.org/10.2760/642022.

Hazarika, B.B. and Gupta, D. (2020) 'Modelling and forecasting of COVID-19 spread using wavelet-coupled random vector functional link networks', *Applied Soft Computing Journal*, Vol. 96, p.106626, https://doi.org/10.1016/j.asoc.2020.106626.

Jia, J.S., Lu, X., Yuan, Y., Xu, G., Jia, J. and Christakis, N.A. (2020) 'Population flow drives spatio-temporal distribution of COVID-19 in China', *Nature*, Vol. 582, No. 7812, pp.389–394, https://doi.org/10.1038/s41586-020-2284-y.

Kang, Y., Gao, S., Liang, Y., Li, M., Rao, J. and Kruse, J. (2020) 'Multiscale dynamic human mobility flow dataset in the U.S. during the COVID-19 epidemic', *Scientific Data*, Vol. 7, No. 1, pp.1–13, https://doi.org/10.1038/s41597-020-00734-5.

Kraemer, M.U.G., Golding, N., Bisanzio, D., Bhatt, S., Pigott, D. M., Ray, S.E., … Reiner, R.C. (2019) 'Utilizing general human movement models to predict the spread of emerging infectious diseases in resource poor settings', *Scientific Reports*, Vol. 9, No. 1, pp.1–11, https://doi.org/10.1038/s41598-019-41192-3.

Lyra, W., do Nascimento, J.D., Belkhiria, J., de Almeida, L., Chrispim, P.P.M. and de Andrade, I. (2020) 'COVID-19 pandemics modeling with modified determinist SEIR, social distancing, and age stratification, the effect of vertical confinement and release in Brazil', *PLoS ONE*, Vol. 15, No. 9, pp.1–17, https://doi.org/10.1371/journal.pone.0237627.

Meena, J., Yadav, J., Saini, L., Yadav, A. and Kumar, J. (2020) 'Clinical features and outcome of sars-cov-2 infection in children: a systematic review and meta-analysis', *Indian Pediatrics*, Vol. 57, No. 9, pp.820–826.

Mykoniatis, K., Angelopoulou, A. and Smith, T.G. (2021) 'Assessing escalator pedestrian traffic dynamics amid COVID-19 pandemic using hybrid simulation', *International Journal of Simulation and Process Modelling*, Vol. 16, No. 3, pp.237–246.

Nogueira, P.J., de Araújo Nobre, M., Costa, A., Ribeiro, R.M., Furtado, C., Bacelar Nicolau, L., … Vaz Carneiro, A. (2020) 'The role of health preconditions on COVID-19 deaths in Portugal: evidence from surveillance data of the first 20293 infection cases', *Journal of Clinical Medicine*, Vol. 9, No. 8, p.2368, https://doi.org/10.3390/jcm9082368.

Philippe, P. and Mansi, O. (1998) 'Nonlinearity in the epidemiology of complex health and disease processes', *Theoretical Medicine and Bioethics*, Vol. 9, No. 6, pp.591–607, https://doi.org/10.1023/A:1009979306346.

PORDATA (2021a) *Portuguese Population Density, According to the Census* [online] https://www.pordata.pt/en/Municipalities/Population+density++according+to+the+Census-591 (accessed 14 May 2021).

PORDATA (2021b) *Portuguese Resident Population, Estimates at December 31st: Total and by Age Group* [online] https://www.pordata.pt/en/Municipalities/Resident+population++estimates+at+December+31st+total+and+by+age+group-137 (accessed 14 May 2021).

Puth, M.T., Neuhäuser, M. and Ruxton, G.D. (2015) 'Effective use of Spearman's and Kendall's correlation coefficients forassociation between two measured traits', *Animal Behaviour*, Vol. 102, pp.77–84, https://doi.org/10.1016/j.anbehav.2015.01.010.

Rivera-Izquierdo, M., Del Carmen Valero-Ubierna, M., R-delAmo, J.L., Fernández-García, M.Á., Martínez-Diz, S., Tahery-Mahmoud, A., Rodríguez-Camacho, M., Gámiz-Molina, A.B., Barba-Gyengo, N., Gámez-Baeza, P., Cabrero-Rodríguez, C., Guirado-Ruiz, P.A., Martín-Romero, D.T., Láinez-Ramos-Bossini, A.J., Sánchez-Pérez, M.R., Mancera-Romero, J., García-Martín, M., Martín-delosReyes, L. M., Martínez-Ruiz, V., Lardelli-Claret, P., … Jiménez-Mejías, E. (2020) 'Sociodemographic, clinical and laboratory factors on admission associated with COVID-19 mortality in hospitalized patients: a retrospective observational study', *PloS One*, Vol. 15, No. 6, p.e0235107, https://doi.org/10.1371/journal.pone.0235107 (accessed 14 May 2021).

Santos, J.P. (2020) *COVID-19 Spatial Distribution*, https://doi.org/10.13140/RG.2.2.35688.70405.

Shahid, Z., Kalayanamitra, R., McClafferty, B., Kepko, D., Ramgobin, D., Patel, R., … Jain, R. (2020) 'COVID-19 and older adults: what we know', *Journal of the American Geriatrics Society*, Vol. 68, No. 5, pp.926–929, https://doi.org/10.1111/jgs.16472.

Singh, M., Marathe, A., Marathe, M.V. and Swarup, S. (2018) 'Behavior model calibration for epidemic simulations', in *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS'18), International Foundation for Autonomous Agents and Multiagent Systems*, Richland, SC, pp.1640–1648.

Søreide, K., Hallet, J., Matthews, J.B., Schnitzbauer, A.A., Line, P.D. and Lai, P.B.S. (2020) 'Immediate and long-term impact of the COVID-19 pandemic on', https://doi.org/10.1002/bjs.11670.

Uehara, T., Nagase, Y. and Wakeland, W. (2012) 'System dynamics implementation of a model of population and resource dynamics with adaptation', *Proc.12th Biennial Conf. Int'l Soc. for Ecological Economics*, Rio de Janeiro, Brazil, June.

urworldindata.org (2021) *Portugal: What is the Daily Number of Confirmed Cases?* [online] https://ourworldindata.org/coronavirus/country/portugal#confirmed-cases (accessed 13 December 2021).

Varnes, D.J. (1984) *Landslide Hazard Zonation: A Review of Principles and Practice*, Natural Hazards, UNESCO, Paris.

Vespignani, A., Tian, H., Dye, C., Lloyd-Smith, J.O., Eggo, R. M., Shrestha, M., … Leung, G.M. (2020) 'Modelling COVID-19', *Nature Reviews Physics*, Vol. 2, No. 6, pp.279–281, https://doi.org/10.1038/s42254-020-0178-4.

Weissman, G.E., Crane-Droesch, A., Chivers, C., Luong, T., Hanish, A., Levy, M. Z., … Halpern, S.D. (2020) 'Locally informed simulation to predict hospital capacity needs during the COVID-19 pandemic', *Annals of Internal Medicine*, https://doi.org/10.7326/m20-1260.

Worldometer (2021) *COVID-19 Coronavirus Pandemic* [online] https://www.worldometers.info/coronavirus/?utm_campaign=homeAdvegas1? (accessed 9 June 2021).

Zheng, H. and Bonasera, A. (2020) 'Chaos, percolation and the coronavirus spread: a two-step model', *European Physical Journal Plus*, Vol. 135, No. 10, https://doi.org/10.1140/epjp/s13360-020-00811-z.