Modeling a Mumps Outbreak through Spatially Explicit Agents

J. Margarida Simoes[a,b,1]

[a] CASA - Centre for Advanced Spatial Analysis
1-19 Torrington Place, WC1E 6BT London, UK

[b] E-Geo - Centro de Estudos de Geografia e Planeamento Regional
Avenida de Berna 26-C, 1069-061 Lisboa, Portugal

Abstract: Human-environment systems, such as the ones where epidemics take place, are characterized by heterogeneity, nonlinear relationships, and hierarchical structures that give rise to difficulties in understanding the system behavior. A good approach to this kind of problem is to start from a general understanding of the low-level processes and elements, and generate aggregate system behavior by simulating the individual entities in the system.
In this study it was developed an Agent Based Model (ABM), that puts together a movement and an infection model. The focus was on relaxing the assumption of random mixing of the population and on adopting an irregular space, with an heterogeneous distribution of individuals. The model was applied to simulate a mumps outbreak in Portugal (1996), and it produced a good estimative of the spatial pattern of infection.
For running the model it was developed a program in C++, with a built in Geographical Information System (GIS) that provides spatial display and analysis functionalities.
Key Words: Agent Based Modeling; Geographic Information Systems; Epidemics; Simulation; Mumps

[1]E-mail: j.simoes@ucl.ac.uk, URL: http://www.casa.ucl.ac.uk/joanamargarida

1 Introduction

Recently, virus epidemics such as Severe Acute Respiratory Syndrome (SARS) had called the attention of the governments, public health officials and the public in general, for the importance of predicting the patterns of viral infection. Besides the unquestionable loss of human life, decreased worker productivity as a result of illness costs industry millions of dollars, every year [1]. Also, models of the spread of infectious diseases should be useful, generally, in the analysis of other diffusion phenomena such as innovation or cultural diffusion patterns [3].
In epidemic modeling, the *movement* process is as important as the *infectious* process itself. Once in real life it is not likely to exist any population where the mixing is completely random [4], this study was primarily focused on developing a realistic, structured, movement model. In this context the spatial component of the model was as important as the temporal one, and therefore the relevance of using *Geographical Information System* (GIS) technology. Besides building such a model and implementing it, another objective was to apply the model to a real case using available data, and to validate it. The case study was a mumps outbreak that occurred in continental Portugal, between 1996 and 1997, due to a vaccination failure [8].
In the next sections, we will describe the theoretical model and show and discuss some results of its application to the case study. As we see the tool developed in the context of this work as an important achievement, there will be a quick reference to the program in a section about material and methods.

2 Theoretical Model

In science in general, the *bottom up* approach emerges the macroscopic behavior of a system trough the interaction of its constituents, without being disturbed by the details of these interactions [2]. It includes techniques such as *Cellular Automata* (CA), *Small World* Graphs (SW) and *Agent Based Modeling* (ABM). The model developed in the context of this study, started as a Site Exchange Cellular Automata (SECA) [9] and developed later to a ABM based on network concepts.
Figure 1 shows the typical structure of a SECA, which was the structure followed in this model. Each discrete time step is subdivided in two stages: movement and infection. The movement stage is ruled by a movement model and the infection stage is ruled by an infection model. These two models are described with more detailed on the next subsections.

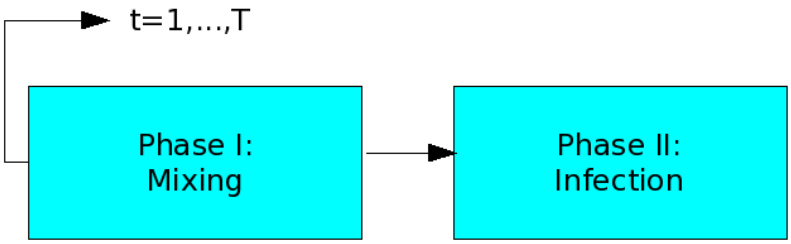


Figure 1: SECA model: each time step is subdivided in two sequential phases: mixing and infection. These correspond to two subsets of rules.

2.1 The Movement Model

The movement model relies on a geographic space that is split into geographical significant units, called *regions* (ς). These regions, that we exemplify on Figure 2, provide the basis for a set of different scales of movement: *neighborhood*, *intra-region*, *inter-region* and *random*.

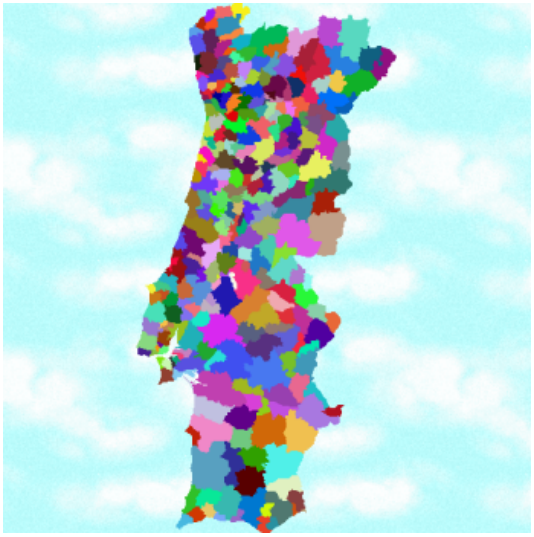


Figure 2: In the simulations described in this paper, the ς were based on the Portuguese regions *Concelhos*. In this map we see *Concelhos* represented with distinct colors;

Elaborating more on each kind of movement, we can say that neighborhood movement

($D1$) represents the movement of a individual in a street or a block, involving activities such as staying at home and going to the pub; intra region movement ($D2$) is the movement inside the city, for instance going to work or shopping; inter region movement ($D3$) is traveling to neighboring cities, for instance visiting relatives, or for taking part in leisure activities. Finally, random movement ($D4$) links together distant parts of the network and is related to another kind of social graph: the *small world* network [10]. In Figure 3 we can see the different scales involved on the movement model.

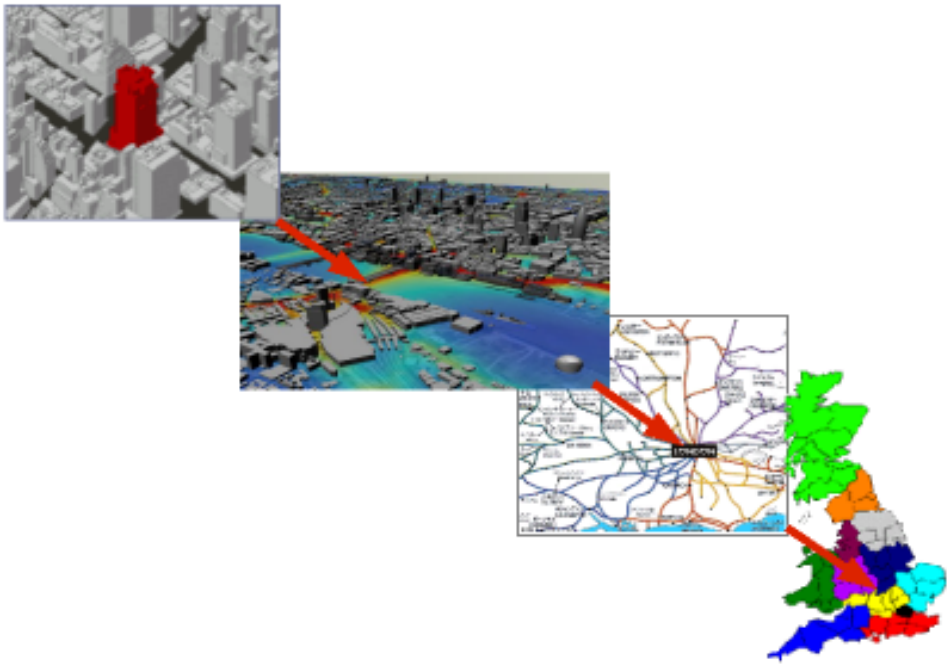


Figure 3: In this figure we can see the different scales involved in the movement model. From top to bottom, we see a neighborhood, an intra region scale, an inter region scale and a random scale.

These types of movement, which are associated to different social activities and have different distance ranges, are weighted and combined, to emerge the movement network. Using a bottom up approach, the global structure of the network emerges from the displacement of each individual a, according to equation (1).

$$a_{(i,j)}^{(t+1)} = a_{(i,j)}^{(t)} + d \tag{1}$$

The stochastic variable d has a probabilistic distribution, according to equation (2).

$$d = (P_1)D_1 + (P_2)D_2 + (P_3)D_3 + (P_4)D_4 = \sum_{x=1}^{4}(P_x)D_x = 1 \tag{2}$$

2.1.1 The Infection Model

The infection model is based on the classical SIR model (Susceptible-Infectious-Removed), formulated by Kermack and McKendrick in 1927 [7], that expresses the relations between the different population categories, in terms of differential equations. These categories are:

- Susceptibles (S) - individuals capable of acquiring the disease (in this model the whole population) disease;
- Infected (I) - individuals who can transmit the disease;
- Removed (R) - individuals who are either dead, recover or become immune from the disease;

In the ABM developed in this work we adopt a variation of the SIR model called SEIR, that we see in equations (3), (4), (5) and (6). This model introduces a fourth category very relevant in childhood diseases such as mumps: the *latent* or exposed (E). This category includes the individuals that acquired the disease but are not infectious yet.

$$\frac{dS}{dt} = -\beta SI \tag{3}$$

$$\frac{dE}{dt} = \beta SI - \theta E \tag{4}$$

$$\frac{dI}{dt} = \theta E - \alpha I \tag{5}$$

$$\frac{dR}{dt} = \alpha I \tag{6}$$

The contact parameter β regulates the transition from susceptible to latent and the parameters θ and α regulate the transition from exposed to infectious, and infectious to removed. From the point at which the individual gets infected till it becomes infectious there, is a *latent period* (that includes the symptomatic and assymptomatic period), and

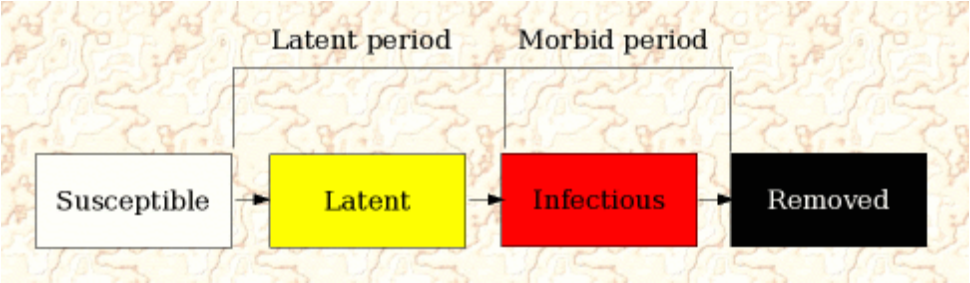


Figure 4: In this diagram we show the epidemic state changes through time, being the states the different epidemic categories.

from the point that it gets infectious till it is removed, there is a *morbid period*. Figure 4 shows the sequence of state transitions and how we define the latent and morbid periods.

We model the contact process of mumps as exclusive function of the physical proximity. Therefore the contact is directly determined by the movement model: to occur infection it is only necessary to have an immediate proximity between a susceptible and a infective and we called this distance parameter, infection radius.

3 Material and Methods

This model was implemented as a computer program, written in C++, encapsulating all the GIS functionalities through programming. The ability to deal with spatial information in an efficient way and to provide relevant ways of displaying it, was thought to be a key feature in the development of the program. The final product is flexible enough to deal with different geographical settings, different initial conditions, and different combinations of parameters in an easy way. Figure 5 shows a screenshot of this program.

More information about the software is available at:
`http://virgil.casa.ucl.ac.uk/ joana/cv/spemod.htm`

4 Simulation of a Mumps Outbreak

The case study for this model is a epidemic outbreak of mumps that took place in continental Portugal, in 1996. In figure 6 we can see where Portugal is located in the context

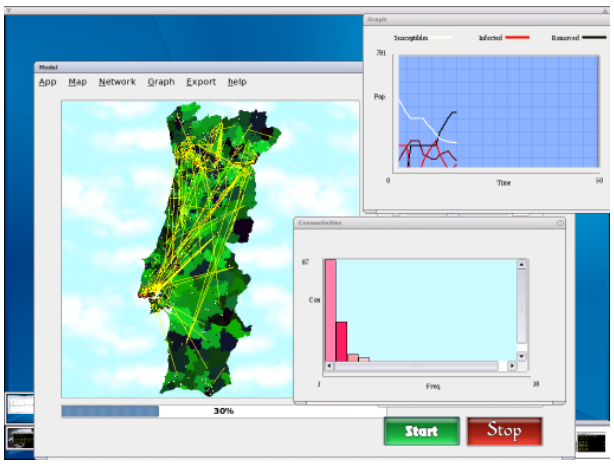


Figure 5: Snapshot of the program, running on a Linux desktop. In this image we can see the different types of information (map, graphics) displayed by the software, during a simulation.

of Europe.

The mumps vaccine was introduced in the 60s and prior to this, the disease affected many children aged 5-9 years and usually before 15. By 1987, a 98 % reduction in mumps incidence was reported worldwide, as a direct consequence of a massive effort of vaccination[2] However, an epidemic outbreak took place in Portugal in 1996-97, an episode that [6] and [8] attribute to vaccination failure. The initial conditions for the simulations reproducing this outbreak are the susceptible and infectious population in the beginning of the epidemic outbreak. The data available for the simulations is a series of reported cases of mumps, from 1989 to 1999 in continental Portugal. This dataset uses as spatial basis the *Concelho*, which is part of the official administrative division basis from the National Statistics Institute (INE). To estimate the susceptible population, it was used the census data from INE, also disaggerated at *Concelho* level (Figure 7).

Before showing the results it is important to note that the dataset may be incomplete and that it may have introduction errors. Also, some assumptions had to be made in order to run the model, as the totality of data required is not available and the model can not deal with the full dataset. The limitations in the number of individuals supported by the program lead to a distortion between the number of suspceptibles and infectious. Figure 8 shows the initial infected cases in the dataset and in the model which are different in magnitude, but have roughly the spatial distribution.

[2] `http://www.cdphe.state.co.us/dc/Epidemiology/manual/School_Guidelines.pdf`

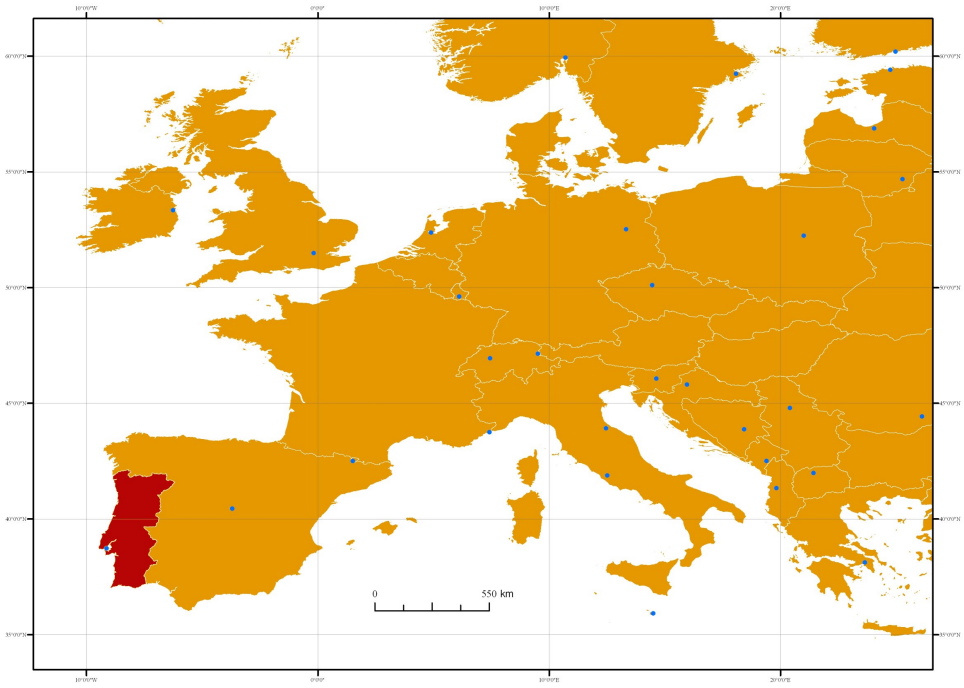


Figure 6: This is a framing map of the study area, Portugal (in red), in the context of Europe.

Figure 9 shows the values used for the infection parameters described in section 2.1.1 (page 5) and Figure 10 shows the weights considered for emerging the movement network, as described in section 2.1 (page 3).

4.0.2 The Results

One hundred simulations were run, with the conditions and dataset described in the previous subsection. The correlation coefficient between the epidemic size in each time step and the observed dataset varies between 0.88 and 0.94 and the average correlation is 0.92. In Figure 11 we show the distribution of the epidemic size in the dataset and the average output of the model, during the 100 simulations. The scales are very different and there is an under estimation of the higher values which can be explained by the biased initial conditions refer on the previous subsection.

Figure 12 shows the distribution of affected individuals in the dataset and in the model, at the end of the outbreak. Although the magnitudes are very different, we can identify in both scenarios the same clusters of affected individuals.

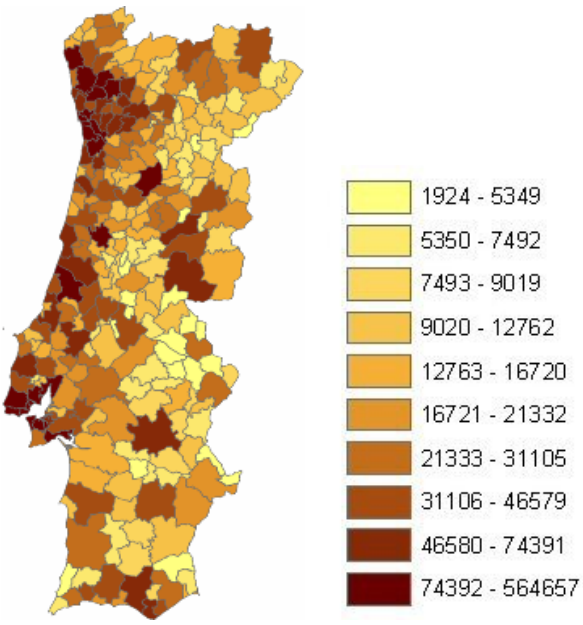


Figure 7: Population distribution, according to the Census from 1991. This map show the number of habitants per *Concelho*.

To remove the magnitude difference between the two series, they were standardized with a classification based on the quartiles. The lower and upper quartiles were evaluated according to equations (7) and (9) and the second quartile, median, used (8).

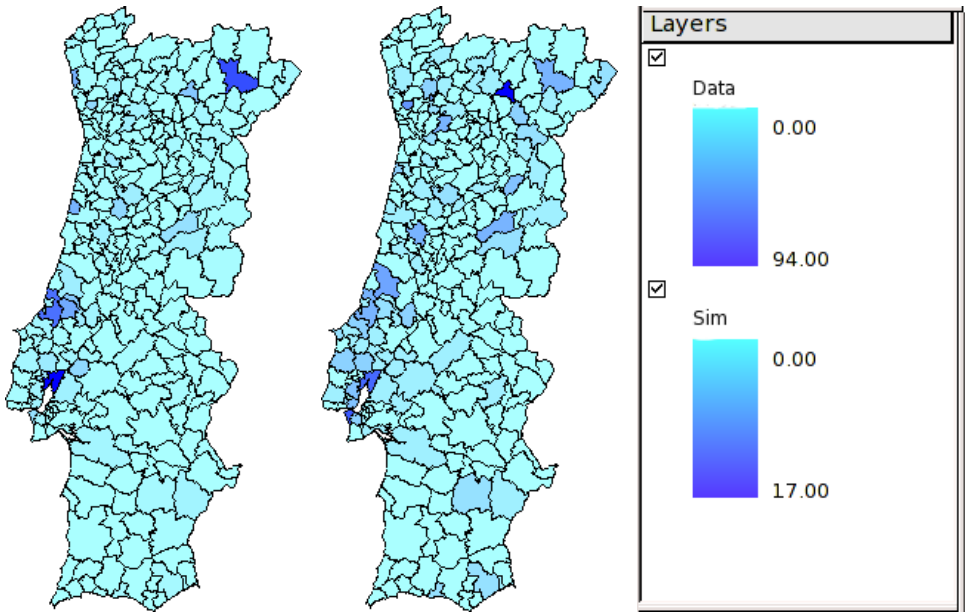
$$Q_1 = \frac{(N+1)}{4}, \tag{7}$$

$$Q_2 = \frac{(N+1)}{2}, \tag{8}$$

$$Q_3 = 3\frac{(N+1)}{4}, \tag{9}$$

In figure 13 we show the comparison between the quartile class in each region, in the dataset and in the simulations and in figure 14 we show the normalized distribution of affected individuals in the dataset and in the model.

As we can see, after the normalization the simulation results are much closer to the observed data. To evaluate the estimative of the spatial distribution of infection, we



(a) (b)

Figure 8: Initial Conditions: a) in the dataset (1993) and b) in the simulations (t=0).

elaborated a map with the quotient between the value in the dataset and the value in the simulation for each region (Figure 15). We can observe that in many cases, the values are 1, and the quotient is never less than 0.25.

5 Summary and Outlook

From what was observed on the previous section, we can conclude that the model produces good estimatives of the spatial distribution of affected individuals although the magnitude the of epidemic size is biased by the initial conditions.
It is important to note that this model deals with a lot of random components and thus we can face a great variability on the results. For this reason, it is important to assure that we can rely on average values of a large number of simulations run over time, that is, we have to test if the values present a normal distribution and if they are converging to an average. The tests revealed that although the values generated by the model do not follow a standard normal distribution, its levels of skewness and kurtosis are fairly acceptable, and they present a normal behavior of the measures of central tendency, as

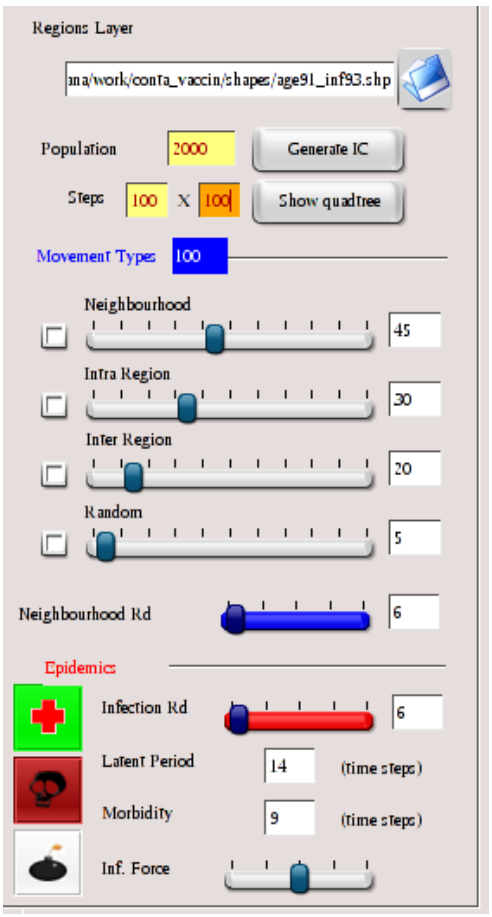


Figure 9: Choice of parameters for the simulations, shown through the user interface of the program.

well as a low dispersion around the mean. These results validate the significance of the simulations, and also the use of average values to represent the series.

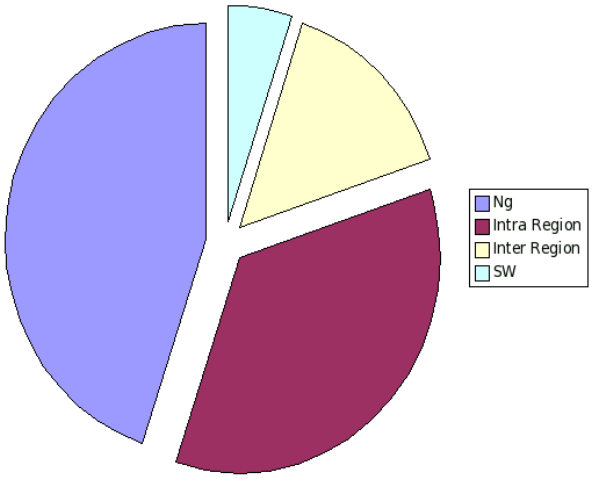


Figure 10: This chart shows the distribution of probabilities for the different movement components, which was adopted throughout this work. These are experimental values based on a distance decay law: shorter range movements have a higher probability to occur than larger range movements; the random component, which is associated to small worlds theory, has a very small probability.

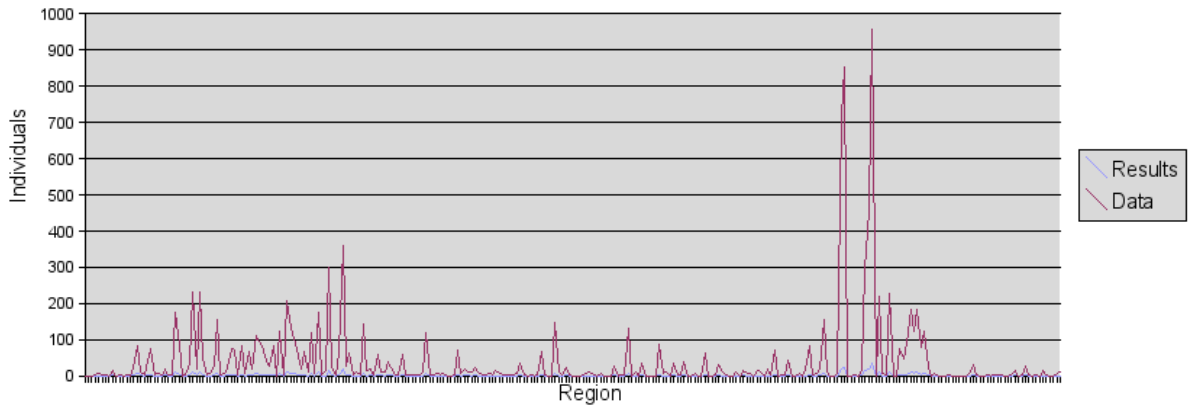
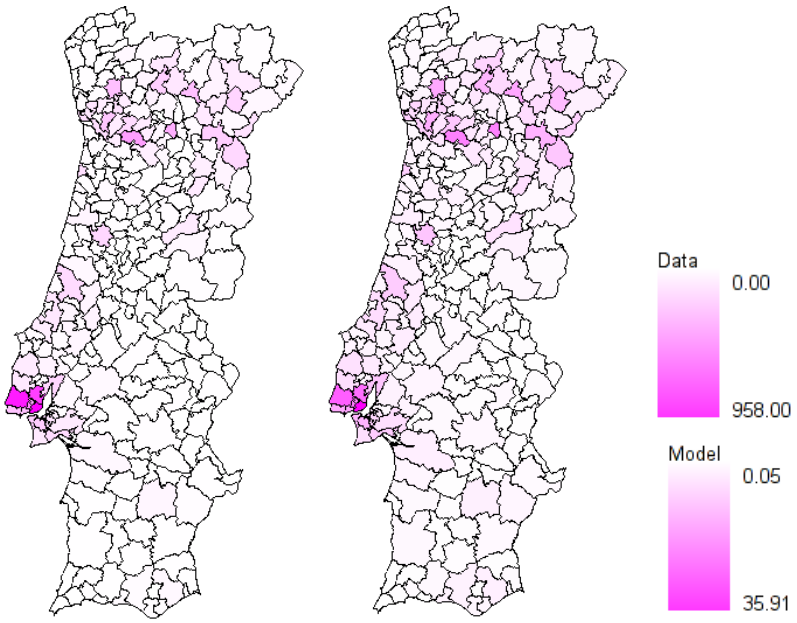


Figure 11: Comparison between epidemic size per region, in the dataset and in the model.



(a) (b)

Figure 12: This figure shows the distribution of the affected individuals at the end of the outbreak, a) in the dataset and b) in the simulations;

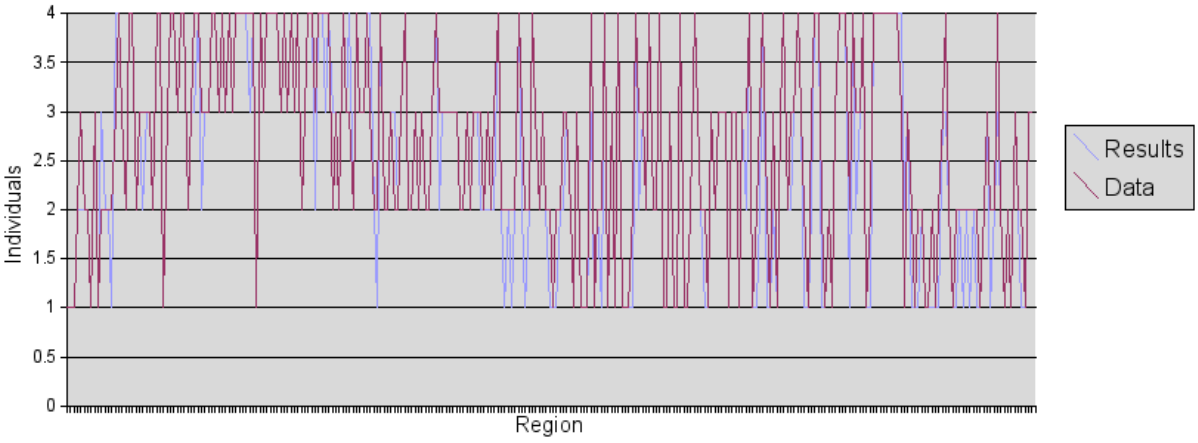


Figure 13: Comparison between the quartile class of the affected individuals by region, in the dataset and in the model

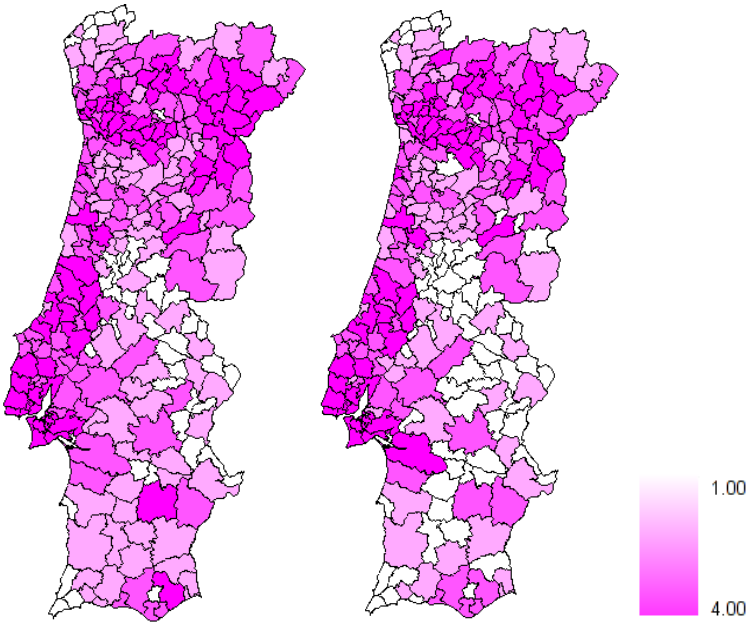


Figure 14: Distribution of the affected individuals classified in quartiles, at the end of the outbreak, a) in the dataset and b) in the simulations;

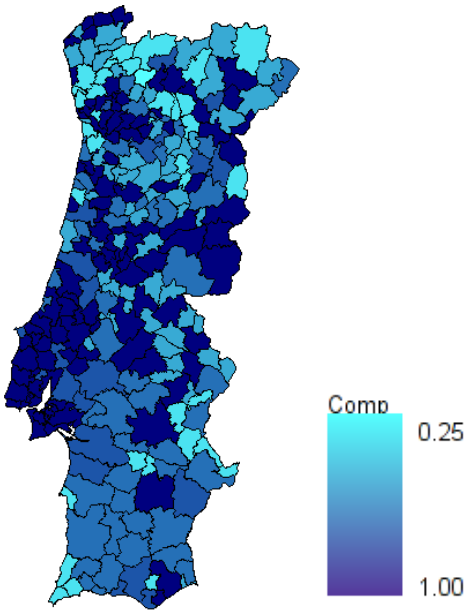


Figure 15: This map shows the quotient between observed and simulated values per region.

Acknowledgments

I would like to thank Carmo Gomes, from FCUL, for providing me the dataset I use in these simulations, my supervisor Michael Batty for his support and encouragement, and finally, the Portuguese Foundation for Science and Technology (FCT) for the scholarship they awarded me.

References

[1] Ching Fu, S. & Stacey, G. M., 2003, Epidemic modelling using cellular automata, in *Proceedings of the First Australian Conference on Artificial Life*, (Canberra), pp. 43–57.

[2] Chopard, B. & Droz, M., 1998, *Cellular Automata Modelling of Physical Systems* (Cambridge University Press, UK).

[3] Cliff, A.D., Hagget, P., Ord, J.K. & Versey, G.R. Spatial Diffusion, 1981, *An Historical Geography of Epidemics in an Island Community* (Cambridge University Press, Cambridge).

[4] Deijfen, M., 2000, *Epidemics on Social Network Graphs* (Master Thesis presented on the Department of Mathematics, Stockholm University, Sweden).

[5] Diekmann, O. & Heesterbeek, J. 2000, *Mathematical Epidemiology of Infectious Diseases* (John Wiley & Sons, USA).

[6] Dias, J., Cordeiro, M. Afzal, M. & Freitas, M., 1996, *Mumps epidemic in Portugal despite high vaccine coverage - preliminary report* (Report No. 1, Eurosurveillance, Portugal), pp. 25–28.

[7] Epstein, J. M., 1997, *Nonlinear Dynamics, Mathematical Biology and Social Science* (Addison-Wesley Publishing Company Inc., Reading).

[8] Gonalves, G., Araujo, A. & Monteiro Cardoso, M.L., 1998, *Outbreak of mumps associated with poor vaccine efficacy* (Report No. 3, Eurosurveillance, Portugal), pp. 119–121.

[9] Mansilla, R. & Gutierrez, J., 2000, Deterministic site exchange cellular automata model for the spread of diseases in human settlements, (e-print `http://www.citebase.org/cgi-bin/fulltext?format=application/pdf&identifier=oai:arXiv.org:nlin/0004012`).

[10] Newman, M., 1999, *Small worlds: the structure of social networks* (Report No. 99-12-080, Santa Fe Institute Technical Report).