
ASSESSING THE SPATIO-TEMPORAL SPREAD OF COVID-19 VIA COMPARTMENTAL MODELS WITH DIFFUSION IN ITALY, USA, AND BRAZIL

A PREPRINT

Malú Grave

Dept. of Civil Engineering
COPPE/Federal University of Rio de Janeiro
P.O. Box 68506, RJ 21945-970, Rio de Janeiro, Brazil
malugrave@nacad.ufrj.br

Alex Viguerie

Department of Mathematics
Gran Sasso Science Institute
Viale Francesco Crispi 7, L'Aquila, AQ 67100, Italy
alexander.viguerie@gssi.it

Gabriel F. Barros

Dept. of Civil Engineering
COPPE/Federal University of Rio de Janeiro
P.O. Box 68506, RJ 21945-970, Rio de Janeiro, Brazil
gabriel.barros@coc.ufrj.br

Alessandro Reali

Dipartimento di Ingegneria Civile ed Architettura
Università di Pavia
Via Ferrata 3, Pavia, PV 27100, Italy
alereali@unipv.it

Alvaro L. G. A. Coutinho

Dept. of Civil Engineering
COPPE/Federal University of Rio de Janeiro
P.O. Box 68506, RJ 21945-970, Rio de Janeiro, Brazil
alvaro@nacad.ufrj.br

February 16, 2021

ABSTRACT

The outbreak of COVID-19 in 2020 has led to a surge in interest in the mathematical modeling of infectious diseases. Such models are usually defined as compartmental models, in which the population under study is divided into compartments based on qualitative characteristics, with different assumptions about the nature and rate of transfer across compartments. Though most commonly formulated as ordinary differential equation (ODE) models, in which the compartments depend only on time, recent works have also focused on partial differential equation (PDE) models, incorporating the variation of an epidemic in space. Such research on PDE models within a Susceptible, Infected, Exposed, Recovered, and Deceased (SEIRD) framework has led to promising results in reproducing COVID-19 contagion dynamics. In this paper, we assess the robustness of this modeling framework by considering different geometries over more extended periods than in other similar studies. We first validate our code by reproducing previously shown results for Lombardy, Italy. We then focus on the U.S. state of Georgia and on the Brazilian state of Rio de Janeiro, one of the most impacted areas in the world. Our results show good agreement with real-world epidemiological data in both time and space for all regions across major areas and across three different continents, suggesting that the modeling approach is both valid and robust.

Keywords COVID-19 · Compartmental models · Diffusion-reaction · Partial differential equations · Adaptive mesh refinement and coarsening

1 Introduction

The COVID-19 pandemic has caused an extraordinary global disruption, both in terms of human lives and economic damage. To limit the spread of contagion, governments have taken unprecedented measures. Such actions have included, but not been limited to, quarantines, curfews, lockdowns, and domestic and international travel suspensions. While undoubtedly effective and considered necessary by many experts, these measures also carry a high cost in terms of economic impact and mental and physical health, among others. In part, such restrictions are motivated by the lack of reliable data and models on this disease’s transmission in time and space, forcing cautious (and not always entirely rational) responses from the authorities and population. More than ever, these events demonstrate the need for tools to predict the spatio-temporal dynamics of contagion. It is important to highlight that such tools should be not only accurate, but, given the difference in the quality of available data in distinct regions of the world (or even within the same country), they should also be robust with respect to the data used for parameter identification. In this context, model validation - also against data possibly coming from different countries - should play an important role.

The mathematical modeling of epidemics is a well-established field, and has been applied with success for many important cases. Examples of past applications include Dengue [1], Zika [2], HIV [3], SARS [4], Malaria [5], Ebola [6], and many others. However, the urgency of the COVID-19 pandemic has motivated the need for more research in this area, with several models for this pandemic outbreak being presented in recent months [7, 8, 9, 10, 11, 12, 13, 14, 15, 16].

Such models typically follow a *compartmental* framework, in which the population under study is divided into compartments with assumptions about the nature and time rate of transfer from one compartment to another [17, 8]. The large majority of the proposed compartmental models are composed of a system of ordinary differential equations (ODEs) in time. Though such ODE models are simple to formulate, analyze, and solve numerically, they do not naturally account for individuals’ movements from one region to another. To incorporate spatial variations, one may define regional compartments corresponding to different geographic areas and introduce coupling terms to account for movements across regions. Such approaches have also been applied to COVID-19 in, for example, [18, 10, 19]. ODE models may be used in combination with other approaches as well, such as with a neural network [20] or an agent-based computational framework [21, 22, 23]. While all these approaches are effective, they lack the ability to model spatial transmission at the continuous level, which may have some advantages.

To this end, one may instead resort to a compartment model based on partial differential equations (PDEs). Such models are decidedly less common than ODE ones, in particular, due to their increased difficulty and time involved in both implementation and numerical solution. However, such an additional cost is counterbalanced by the fact that PDE models naturally (and accurately) incorporate spatial information, allowing for a continuous description of spatio-temporal dynamics, with the potential to account for geographical features, population heterogeneity, and multi-scale dynamics with relative ease. While quite uncommon, PDE models have been already used with success in the past to model epidemics [24, 25, 26].

For the specific case of COVID-19, PDE models were applied in [7, 8, 27, 14], in which a compartmental SEIRD model (*susceptible, exposed, infected, recovered, deceased*) was used. A modified version of this model, with the addition of a quarantined compartment, was applied in [16]. Such modeling framework has shown promising results, being able to recreate observed data and make predictions with reasonable accuracy, but these successful applications have been so far obtained on a limited scale, with results restricted to single, specific regions. Therefore, despite the clearly high potential shown by this modeling approach, important questions remain regarding its *robustness*. In particular, it is still to be proven the model capability to guarantee reliable results for a wide range of areas, which may exhibit significantly different geographical, behavioral, and demographic characteristics. Furthermore, even if the model may demonstrate the ability to represent different regions accurately, it is unclear to what extent parameters must be specifically tuned to obtain accurate results for a given case. In fact, ideally, a robust model should be applicable to a wide range of settings, requiring relatively little specific parameter tuning to produce acceptable results for a given case.

Within this context, the current work aims at evaluating the robustness of the model used in [7, 8, 27] by studying in detail its application to a series of different cases. In order to accomplish this task, the model is applied to the following three cases: The Italian region of Lombardy, the U.S. state of Georgia, and the Brazilian state of Rio de Janeiro. These are not only the homelands of the authors, but they do represent regions of three distinct continents with very different features in terms of culture, population, and mobility, as well as government response to the pandemic. To deeply test its robustness to different region-specific characteristics, the model is applied across each case with limited parameter tuning and case-specific adjustments, and it is shown to be capable of producing consistently reliable results, in particular for predictions relative to short to mid-time spans, which are relevant for timely political decisions.

The remainder of this work is organized as follows: Section 2 introduces the governing equations and the relevant parameters, as well as the adopted terminology and notation. This section is also used to briefly discuss considerations regarding the numerical solution of the associated discrete problem. In the subsequent sections, the model is then tested

on the cases of Lombardy, Georgia, and Rio de Janeiro, and for each case extensive discussion and analysis of results are provided. Finally, the paper is concluded by a summary of the main findings and suggestions for future research directions.

2 Governing equations

We present the governing equations following the continuum mechanics framework first shown in [7], as opposed to more traditional notations common in mathematical and biological references. We consider a system that may be decomposed into N distinct populations: $u_1(\mathbf{x}, t)$, $u_2(\mathbf{x}, t)$, ..., $u_N(\mathbf{x}, t)$. Let $\Omega \in R^2$ be a simply connected domain of interest and denote its boundary as $\partial\Omega = \Gamma_D \cup \Gamma_N$. Let $[0, T]$ denote a generic time interval. The transient nonlinear diffusion-reaction system of equations written in compact vector notation then reads as follows:

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{A} + \mathbf{B}(\mathbf{u})) \mathbf{u} - \nabla \cdot (\boldsymbol{\nu} \nabla \mathbf{u}) - \mathbf{f} = 0 \text{ in } \Omega \times [0, T] \quad (1)$$

$$\mathbf{u} = \mathbf{u}_D \text{ in } \Gamma_D \times [0, T] \quad (2)$$

$$(\boldsymbol{\nu} \nabla \mathbf{u}) \cdot \mathbf{n} = \mathbf{h} \text{ in } \Gamma_N \times [0, T], \quad (3)$$

where $s(\mathbf{x}, t)$, $e(\mathbf{x}, t)$, $i(\mathbf{x}, t)$, $r(\mathbf{x}, t)$, and $d(\mathbf{x}, t)$ denote the densities of the *susceptible*, *exposed*, *infected*, *recovered*, and *deceased* populations, respectively. We additionally denote the *cumulative number of infected* as $c(\mathbf{x}, t)$ and the sum of the living population as $n(\mathbf{x}, t)$; i.e., $n(\mathbf{x}, t) = s(\mathbf{x}, t) + e(\mathbf{x}, t) + i(\mathbf{x}, t) + r(\mathbf{x}, t)$. We then use the compact vector notation to represent these quantities as: $\mathbf{u} = [s, e, i, r, d]^T$.

The tensors \mathbf{A} , \mathbf{B} and $\boldsymbol{\nu}$, and the vector \mathbf{f} are defined according to the particular dynamics of the system in question. In most applications, $\boldsymbol{\nu} = \boldsymbol{\nu}(\mathbf{x}, t)$, that is, diffusion is time-dependent, heterogeneous, and anisotropic; however, this is not strictly necessary. In addition to the boundary conditions (2), (3), we must also specify an initial condition $\mathbf{u}(\mathbf{x}, 0) = \mathbf{u}_0$. We define the *total population* $U_i(t)$ of a given compartment $u_i(\mathbf{x}, t)$ as:

$$U_i(t) = \int_{\Omega} u_i(\mathbf{x}, t) d\Omega \quad (4)$$

for $i = 1, 2, \dots, N$.

To adequately define a model of COVID-19 contagion, we make the following assumptions, as discussed in [7, 27]:

- We consider only mortality due to COVID-19;
- We do not consider new births;
- We assume that a portion of exposed individuals will never develop symptoms (asymptomatic cases), and hence will move directly from the exposed compartment to the recovered compartment;
- We assume that both pre-symptomatic and asymptomatic (the exposed compartment), as well as symptomatic (the infected compartment) individuals may spread the disease;
- There is a latency period after exposure and before the development of symptoms;
- Movement is proportional to the population size; i.e., we expect greater diffusion to occur in heavily populated regions;
- There is no movement among the deceased population.

Under the above assumptions, the *frequency-dependent* system of equations reads:

$$\frac{\partial s}{\partial t} + \frac{\beta_i}{n} si + \frac{\beta_e}{n} se - \nabla \cdot (n \boldsymbol{\nu}_s \nabla s) = 0 \quad (5)$$

$$\frac{\partial e}{\partial t} - \frac{\beta_i}{n} si - \frac{\beta_e}{n} se + (\alpha + \gamma_e)e - \nabla \cdot (n \boldsymbol{\nu}_e \nabla e) = 0 \quad (6)$$

$$\frac{\partial i}{\partial t} - \alpha e + (\gamma_i + \delta)i - \nabla \cdot (n\nu_i \nabla i) = 0 \quad (7)$$

$$\frac{\partial r}{\partial t} - \gamma_e e - \gamma_i i - \nabla \cdot (n\nu_r \nabla r) = 0 \quad (8)$$

$$\frac{\partial d}{\partial t} - \delta i = 0, \quad (9)$$

where β_i and β_e denote the transmission rates between symptomatic and susceptible individuals and asymptomatic and susceptible individuals, respectively (units days⁻¹), α denotes the incubation period (units days⁻¹), γ_e corresponds to the asymptomatic recovery rate (units days⁻¹), γ_i the symptomatic recovery rate (units days⁻¹), δ represents the mortality rate (units days⁻¹), and $\nu_s, \nu_e, \nu_i, \nu_r$ are the diffusion parameters of the different population groups as denoted by the sub-scripted letters (units km² persons⁻¹ days⁻¹). Note that all these parameters can be considered time and space-dependent.

We may express the model (19)-(23) in the general form given by equation (1) by defining the tensors \mathbf{A} , \mathbf{B} , $\boldsymbol{\nu}$, and the vector \mathbf{f} in the following way:

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \alpha + \gamma_e & 0 & 0 & 0 \\ 0 & -\alpha & \gamma_i + \delta & 0 & 0 \\ 0 & -\gamma_e & -\gamma_i & 0 & 0 \\ 0 & 0 & -\delta & 0 & 0 \end{bmatrix} \quad (10)$$

$$\mathbf{B} = \begin{bmatrix} 0 & \frac{\beta_e}{n_e} s & \frac{\beta_i}{n_i} s & 0 & 0 \\ 0 & -\frac{\beta_e}{n} s & -\frac{\beta_i}{n} s & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (11)$$

$$\boldsymbol{\nu} = \begin{bmatrix} \nu_s & 0 & 0 & 0 & 0 \\ 0 & \nu_e & 0 & 0 & 0 \\ 0 & 0 & \nu_i & 0 & 0 \\ 0 & 0 & 0 & \nu_r & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (12)$$

$$\nu_k = \begin{bmatrix} \nu_{xx}^k & \nu_{xy}^k \\ \nu_{yx}^k & \nu_{yy}^k \end{bmatrix} \text{ with } k = s, e, i, r \quad (13)$$

$$\mathbf{f} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \quad (14)$$

Assuming additionally that the region of interest is isolated, we then prescribe the following homogeneous Neumann boundary conditions:

$$\nabla s \cdot \mathbf{n} = 0 \quad (15)$$

$$\nabla e \cdot \mathbf{n} = 0 \quad (16)$$

$$\nabla i \cdot \mathbf{n} = 0 \quad (17)$$

$$\nabla r \cdot \mathbf{n} = 0, \quad (18)$$

or simply $(\boldsymbol{\nu} \cdot \nabla \mathbf{u}) \cdot \mathbf{n} = 0$.

We note that one may also consider the following model related to (5)-(9):

$$\frac{\partial s}{\partial t} + \beta_i \left(1 - \frac{A}{n}\right) si + \beta_e \left(1 - \frac{A}{n}\right) se - \nabla \cdot (n\nu_s \nabla s) = 0 \quad (19)$$

$$\frac{\partial e}{\partial t} - \beta_i \left(1 - \frac{A}{n}\right) si - \beta_e \left(1 - \frac{A}{n}\right) se + (\alpha + \gamma_e)e - \nabla \cdot (n\nu_e \nabla e) = 0 \quad (20)$$

$$\frac{\partial i}{\partial t} - \alpha e + (\gamma_i + \delta)i - \nabla \cdot (n\nu_i \nabla i) = 0 \quad (21)$$

$$\frac{\partial r}{\partial t} - \gamma_e e - \gamma_i i - \nabla \cdot (n\nu_r \nabla r) = 0 \quad (22)$$

$$\frac{\partial d}{\partial t} - \delta i = 0. \quad (23)$$

The model (19)-(23) is known as the *density-dependent* formulation and was studied in [8, 7]. The difference between the models can be seen in equations (5)-(9) and (19)-(23). In (5)-(9), the contact terms are normalized by the living population, whereas such a normalization does not occur in (19)-(23). In the *frequency-dependent* formulation (5)-(9), this normalization implies that the contagion is independent of population density, while this is not the case in the density-dependent formulation, as the name may suggest [28, 29]. Both models may be valid and deliver accurate results, depending on the physical situation, and we will show computations performed with both formulations in the present work.

The other major difference between (5)-(9) and (19)-(23) is the presence of the Allee effect A . This term has been used extensively in other settings, with the form used above inspired directly by applications in cancer modeling [30, 31]. The Allee term serves to reduce transmission in areas where the population density is below a certain threshold A , by bringing the population in the exposed compartment to the susceptible compartment in such regions. Consequently, in those areas, the population in compartments e and i tends to zero, which eventually cancels out the transfer term where the Allee term participates. In [8], this term served to accentuate the contagion in major urban regions while reducing it in less-populated areas, consistent with the observed physics. We briefly note also that as s , i , and e are all less than n by definition, we do not expect blowup of this term, even for very small n .

For the numerical solution of (5)-(9), (19)-(23), we discretize in space using a Galerkin finite element formulation. The resulting systems of equations are stiff, leading us to employ implicit methods for time integration. We apply the second-order Backward Differentiation Formula (BDF2) in all cases, which offers second-order accuracy while remaining unconditionally stable. We additionally make use of an adaptive mesh refinement and coarsening strategy (AMR/C), allowing us to resolve multiple scales. One may find more details about the adopted methods in [7, 27].

All simulations have been performed using the `libMesh`, a C++ FEM open-source software library for parallel adaptive finite element applications [32]. `libMesh` also interfaces with external solver packages like PETSc [33] and Trilinos [34]. It is an excellent tool for programming the finite element method and can be used for one-, two-, and three-dimensional steady and transient simulations. This library provides native support for AMR/C, thus providing a natural environment for the present study. The main advantage of `libMesh` is the possibility of focusing on implementing the specific features of the modeling without worrying about adaptivity and code parallelization. Consequently, the effort to build a high performance computing code tends to be minimized.

3 Simulation of the COVID-19 virus spread in some of the world's hot-spots

In the following section, we demonstrate the robustness of the models (5)-(9), (19)-(23) by simulating the COVID-19 outbreak for three different geographic regions: Lombardy (Italy), the U.S. state of Georgia, and the state of Rio de Janeiro (Brazil). These regions cover three different continents, and each exhibits very different geographical features, population characteristics, economic, social, cultural factors, and government responses. By achieving good results for each case, we aim at showing that the modeling approach shown here is robust to general cases and does not require extensive modification or specific parameter tuning to achieve reasonable results for a given case.

The three cases shown were chosen for a number of reasons. Notably, they represent the authors' home regions, so there is an obvious interest from this point-of-view. Beyond this point, however, the regions also have specific characteristics that make them remarkable case studies. Italy is one of the hardest-hit regions in Europe, and Lombardy represented

the epicenter of the outbreak in Italy, as well as the first area in a Western country to exhibit community spread. During the first wave of the epidemic in Spring 2020, Lombardy was disproportionately impacted compared to other regions in Italy, in which the epidemic was less severe. It remains the hardest-hit region in Italy, even after the second wave of infections has been less geographically restricted. As Lombardy was studied specifically in [8, 7], it also represents a natural choice for validating the new code and the adaptive-meshing strategy implemented in the present work.

The U.S. state of Georgia is a logical choice for several reasons. The state has 159 counties, second to only Texas among U.S. states. However, the average county size for Georgia is less than half of Texas, giving the state a good county-by-county spatial resolution, and allowing us to evaluate the spatial accuracy of the model in reasonable detail. Further, the state was the first among U.S. states to reopen in late April, and has since taken a large number of policy measures, in both relaxing and increasing restrictions, providing us a large number of decisions to incorporate and analyze in our model.

Rio de Janeiro is also a sensible choice for many of the same reasons as Georgia. At the time of writing, Brazil is the third-most severely impacted country in the world in terms of total cases, and the second one in terms of total deaths. Rio de Janeiro has the largest number of deaths per inhabitant within Brazil, making its simulation worthwhile on severity alone. Aside from this aspect, the area also exhibits interesting geographical features, notably its natural harbor and large population centers on either side. Additionally, there is good availability of mobility data, which we have incorporated into the diffusion parameters. The use of such data is logical for a model of this type, and its consideration represents an interesting and important component for this modeling approach.

3.1 Lombardy, Italy

In our first numerical test, we seek to reproduce, and possibly improve, the simulations of Lombardy, Italy, as shown in [8, 7]. This model was already shown in [8] to predict the first wave of the COVID-19 outbreak in Lombardy with good accuracy. We note that for this simulation we use the density-dependent model with the Allee term (19)-(23). While using the same parameters as shown in [8, 7], we incorporate two major improvements. First, we employ the adaptive mesh refinement and coarsening strategy shown in [27]. Second, rather than using Backward-Euler time-stepping, we utilize a BDF2 scheme, as [35, 7] suggest that such a scheme is better suited to the system of equations in question. Thus, the primary motivation of this simulation is the validation of the current code and algorithmic framework.

The AMR/C procedure uses a local error estimator to drive the refinement and coarsening procedure, considering the error of an element relative to its neighbor elements in the mesh. The elements are flagged depending on a refining (r_f) and a coarsening (c_f) fraction. The refinement level is limited by a maximum h -level (h_{max}). In this case, for the AMR/C procedure, we set $h_{max} = 1$, $r_f = 0.95$, $c_f = 0.01$. We apply the adaptive mesh refinement every 4 time-steps. The original mesh has 10,987 linear triangular elements, and, after refinement, the minimum spatial resolution is about 1 kilometer. We initially refine the domain in one level, and the time-step is defined as $\Delta t = 0.25$ days.

In Fig. 1 we compare the results obtained here with those from [7]. We observe the same qualitative dynamics; the epidemic follows the same path, moving from the southern areas of the region northward into the major metropolitan areas, and the Milan area in particular. The changes to the meshing and time-stepping schemes result in somewhat lower contagion, particularly in the areas in and around the city of Milan in the western part of the region. Examining the progression of the mesh adaptation in Fig. 1, we see that the AMR/C algorithm successfully follows the transmission path, changing in time to adapt to the evolving geography of the epidemic in time. We briefly note that when using the current code with Backward-Euler and a fixed mesh (not shown), the results were identical to those obtained in [8]; thus, we can conclude that the differences between the solution depicted in the present work and that shown in [8] represent improvements resulting from the differences in meshing and time-stepping used in the present one. From this, we conclude that the current code and algorithmic framework is validated.

3.2 Georgia, USA

From now on we use the frequency-dependent model (5)-(9) together with adaptive mesh refinement and the BDF2 time-stepping scheme. In the next test, we aim at reproducing the COVID-19 outbreak in the U.S. state of Georgia.

3.2.1 Model Construction

We have first obtained the map of the state of Georgia (GA) along with the county boundaries in shapefile format from https://arc-garc.opendata.arcgis.com/datasets/dc20713282734a73abe990995de40497_68. To triangulate the GA region, we follow the steps given by [14]:

- Load the GA map file in the freely available QGIS software [36];

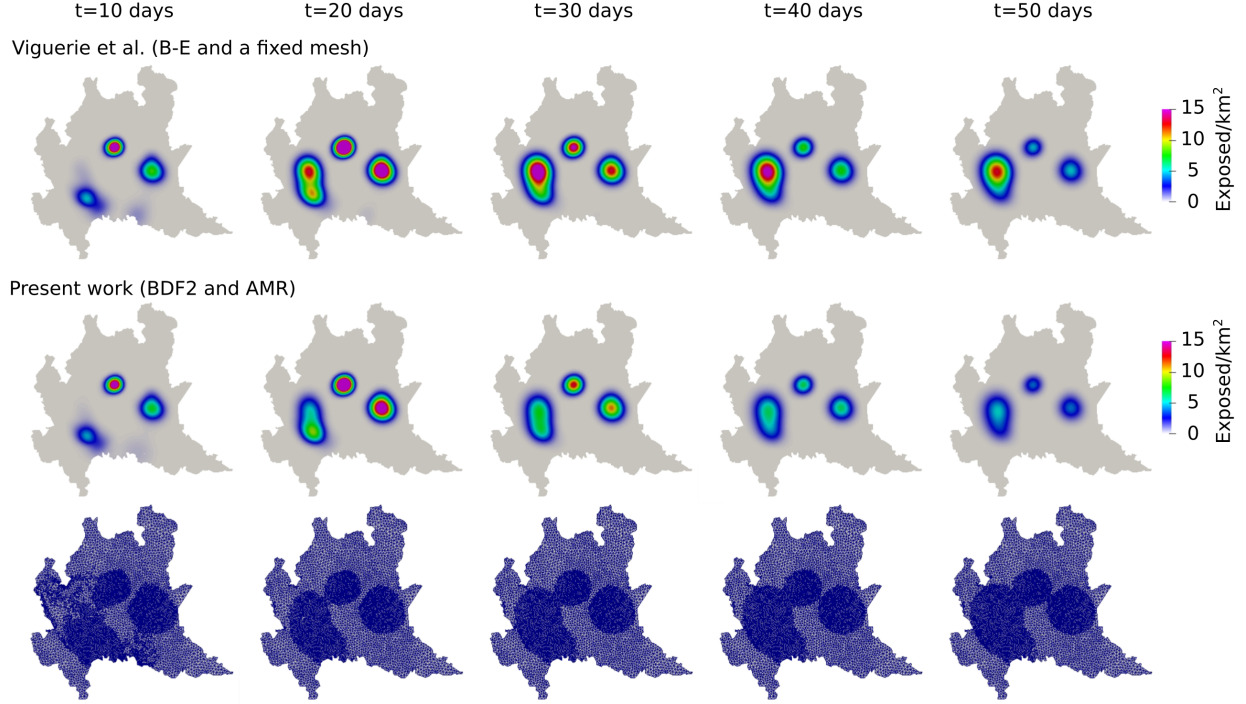


Figure 1: Spread of the COVID-19 virus spread at Lombardy, Italy. Comparison between the results obtained in [8], in which the time integration was done using the Backward-Euler method with a fixed mesh, and the present work, in which we use the BDF2 method and AMR.

- Coarse grain the outer boundary segments using the Simplify tool in QGIS. The original map has some very small length segments, which may create problems in triangulation or result in a very fine mesh;
- Obtain the vertices using the Extract Vertices tool in QGIS and save the vertices layer using the save layer as option. Select As XY in the Graphical category and save the file in a .csv format;
- Prepare a Gmsh input file using the vertices' file for triangulation.

In Figure 2, we show the generated grid for the state of Georgia. The original mesh has 32,056 linear triangular elements, and after refinement, the minimum spatial resolution is about 2 kilometers. We initially refine the domain in one level. For the AMR/C procedure, we set $h_{max} = 1$, $r_f = 0.95$, $c_f = 0.01$. We apply the adaptive mesh refinement every 4 time-steps. The time-step is defined as $\Delta t = 0.25 \text{ days}$.

We define the initial infected population accordingly to the data provided by the Johns Hopkins University [37]. We define the beginning of the simulation on 25 March 2020 and simulate 250 days. The exposed population is more difficult to estimate, as they consist of asymptomatic and pre-symptomatic individuals. Here, we consider 10 times the number of infected [38]. The susceptible population is based on the estimation of the population of each county, given by the U.S. Census Bureau [39]. The populations are divided by the area of each county and distributed on the 159 areas of the mesh as $people/km^2$. In Figs. 3, 4, and 5 we show the initial conditions.

The parameters of the simulation are defined as: $\alpha = 1/8 \text{ day}^{-1}$, $\gamma_i = 1/24 \text{ day}^{-1}$, $\gamma_e = 1/6 \text{ day}^{-1}$, $\delta = 1/160 \text{ day}^{-1}$. These values are based on available data from the literature regarding the mortality, incubation period, and recovery time for infected and asymptomatic patients [8].

The contact rate and diffusion are estimated based on the policies adopted in the Georgia state during the period of the simulation (<https://coronavirus.jhu.edu/data/state-timeline/new-deaths/georgia/>). This leads to the time-dependent values as shown in Fig. 6.

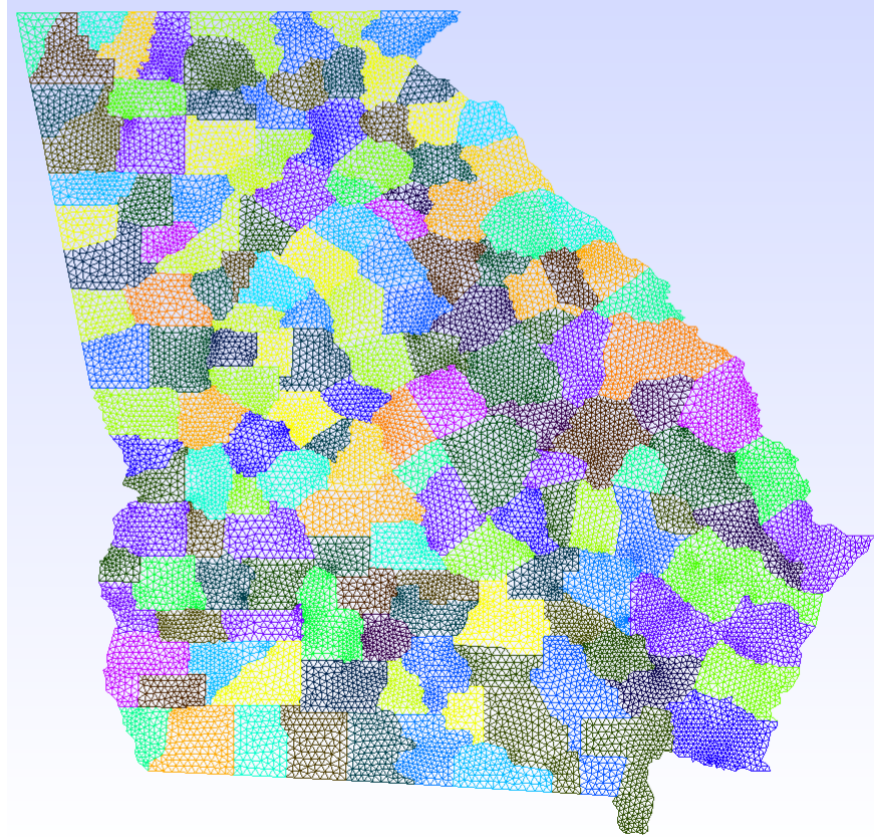


Figure 2: Map of the state of GA partitioned into 159 internal counties.

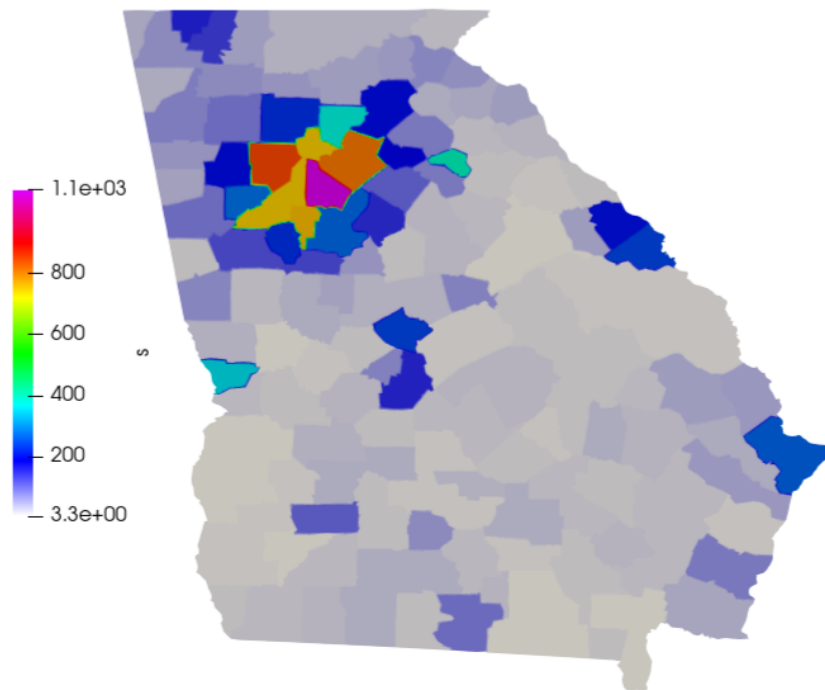


Figure 3: Initial susceptible population (*people/km²*).

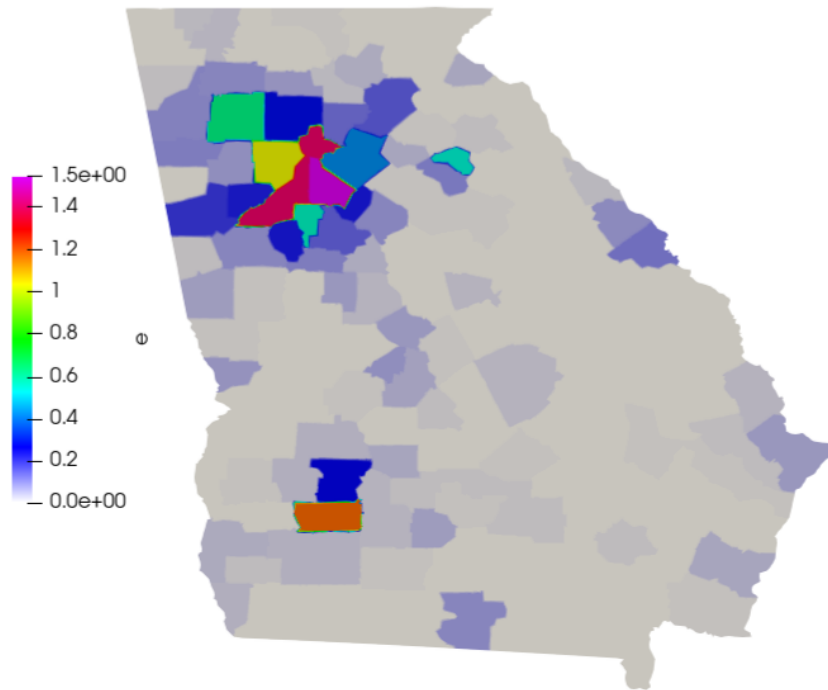


Figure 4: Initial exposed population (*people/km²*).

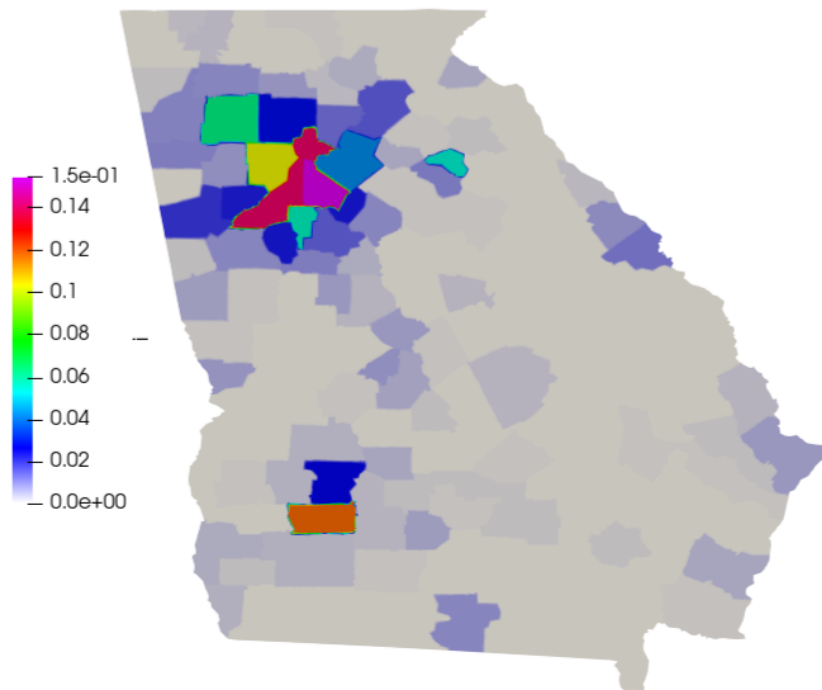


Figure 5: Initial infected population (*people/km²*).

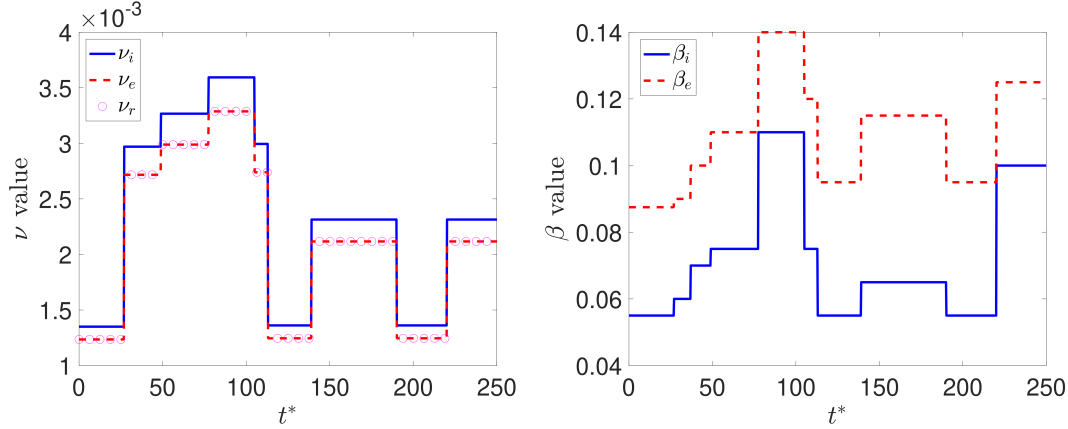


Figure 6: Diffusion (left) and contact rate (right) in time for Georgia.

3.2.2 Results

To compare our model results with the data available, we compare the number of deaths caused by the COVID-19. First, we show the values of the whole state. Then, we integrate the values of some counties and plot them in time (Fig. 8 and 9). Fig. 10 also shows some snapshots of the simulation with the populations and mesh at different time-steps.

Referring to the entire state in Fig. 7, we observe that the model accurately captures both qualitative and quantitative dynamics of the epidemic with good accuracy. The difference in relative L^2 norm between the measured and simulated deaths is 5.75%, confirming the agreement. We note that a small discrepancy between the simulated and measured data occurs at the end of the simulated period; this is caused by a series of deaths being added retroactively to the data, with the actual date of these deaths being unknown.

Turning our attention to the individual counties, we see more variation in the results. We start by noting that the initial dynamics (first 60 days) is predicted well nearly everywhere, with a divergence between the simulated and measured deaths becoming more apparent as the epidemic progresses. This divergence generally indicates that such models' predictive power may decrease in time, as one may reasonably expect. Nonetheless, we do observe encouraging results over the entire simulated time frame, particularly in certain areas. We first assess counties comprising the Atlanta metropolitan area. The heavily populated counties of DeKalb, Gwinnett, Clayton, and Cobb show very good agreement with the simulated data. The most populous county in Georgia, Fulton County, is over-predicted. On the fringe of the Atlanta metropolitan area, Forsyth, Clarke, and Cherokee counties were also over-predicted.

Outside of the Atlanta area, we found a general underestimation of the outbreak severity in the Bibb, Richmond, and Chatham counties. As the "first wave" of infections is well-simulated, this likely represents a shortcoming of the diffusive model. In particular, after the initial re-openings, additional source terms may need to be included in certain areas to model the effects of travelers arriving from elsewhere, as our model does not account for nonlocal movements. In contrast, Dougherty County, which was heavily impacted early in the epidemic, shows the opposite trend, and the outbreak during the later stages of the simulated period is less severe than as predicted by the model. This shows that the model is sensitive to the initial conditions, and spatially-variable parameters, in particular contact rates, may be needed to account for such phenomena properly.

3.3 Rio de Janeiro, Brazil

The last simulation aims at reproducing the COVID-19 outbreak in the state of Rio de Janeiro, Brazil. Rio de Janeiro has the second number of deaths and infected in Brazil, but the largest number of deaths/inhabitant as per January 2021¹

3.3.1 Model Construction

We have obtained the map of the state of Rio de Janeiro (RJ) along with the county boundaries in shapefile format from ftp://geofftp.ibge.gov.br/organizacao_do_territorio/malhas_territoriais/

¹<https://www.statista.com/statistics/1107109/brazil-coronavirus-deaths-state/>

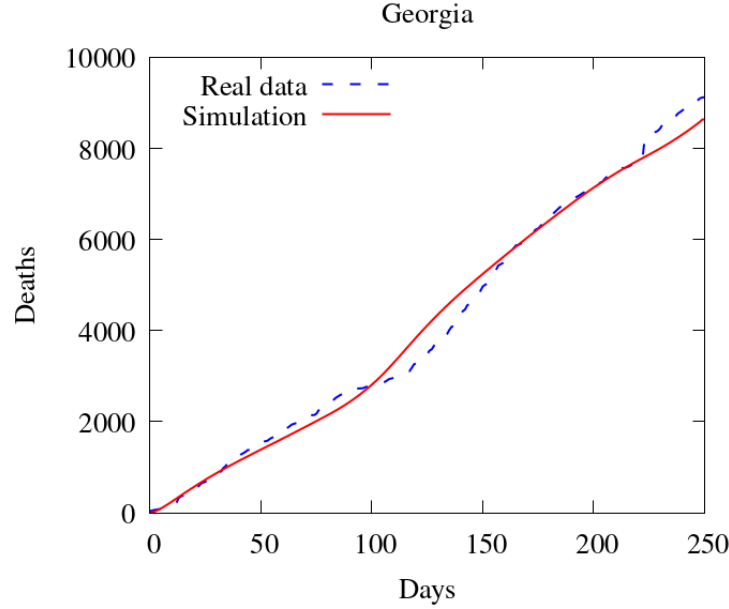


Figure 7: Comparison between simulation and real data of deaths at Georgia (total).

malhas_municipais/municipio_2017/UFs/. To triangulate the RJ region, we follow the same steps explained in the previous section.

In Fig. 11, we show the grid generated for the state of Rio de Janeiro. We eliminated some islands to facilitate the simulation. The original mesh has 11,632 linear triangular elements, and after refinement, the minimum spatial resolution is about 500 meters. We initially refine the domain in one level. For the AMR/C procedure, we set $h_{max} = 1$, $r_f = 0.95$, $c_f = 0.01$. We apply the adaptive mesh refinement every 4 time-steps. The time-step is defined as $\Delta t = 0.25$ days.

We define the initial infected population accordingly to the data provided by the Brazilian Ministry of Health, collected by <https://covid19br.wcota.me/>. We define the beginning of the simulation on 25 March 2020 and simulate 180 days. However, since there is a delay until people develop symptoms, another delay in receiving the test results, and another one for the case being disclosed, we consider that the initial infected population is equal to the case numbers provided for each county on 1 April 2020. As in the case of Georgia, we initialize the exposed population as 10 times the number of infected [38]. The susceptible population is based on the estimation of the population of each county, given by the Brazilian Institute for Geography and Statistics². The populations are divided by the area of each county and distributed on the 92 areas of the mesh as *people/km²*.

In Figs. 12, 13, and 14 we show the initial conditions. The population of the state of Rio de Janeiro is concentrated near the capital and the metropolitan region.

The parameters of the simulation are defined as: $\alpha = 1/7$ day⁻¹, $\gamma_i = 1/24$ day⁻¹, $\gamma_e = 1/6$ day⁻¹, $\delta = 1/160$ day⁻¹. These values are based on available data from the literature regarding the mortality, incubation period, and recovery time for infected and asymptomatic patients [8].

The contact rate is estimated based on the reproduction number estimation given in <https://perone.github.io/covid19analysis/>. We define $\beta_e = \beta_i = 0.215$ days⁻¹, and we multiply this value by the time function given by Fig. 15b, which is a simplification of the behavior of the reproduction number between 25 March 2020 and 21 September 2020. The diffusion coefficient is based on the social distancing estimation given by <https://coronavirus.ufrj.br/covidimetro/>, which presents the perception of average weekly social isolation based on the movement of cell phones in the state of RJ. The diffusion behaves opposite to the social isolation, i.e., when one increases, the other one decreases. Therefore, we set $\nu_s = \nu_e = \nu_r = 1 \times 10^{-3}$ and $\nu_i = 1 \times 10^{-5}$ km²persons⁻¹days⁻¹, and scale these values by a time-dependent function (Fig. 15a) that tries to represent the social isolation during the 180 days of simulation.

²<https://www.ibge.gov.br/cidades-e-estados/rj.html>

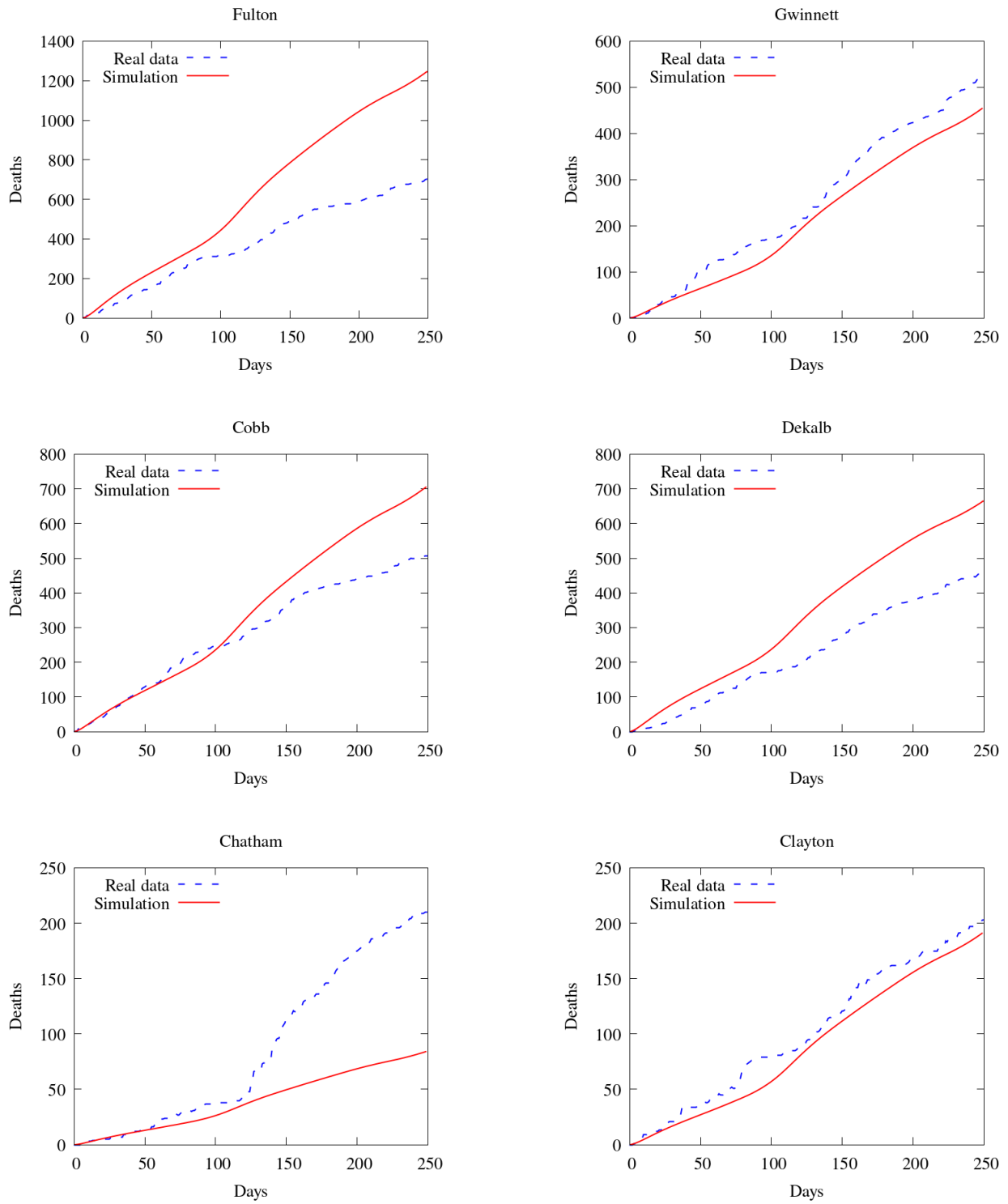


Figure 8: Comparison between simulation and real data of deaths at Georgia (per county) - part 1.

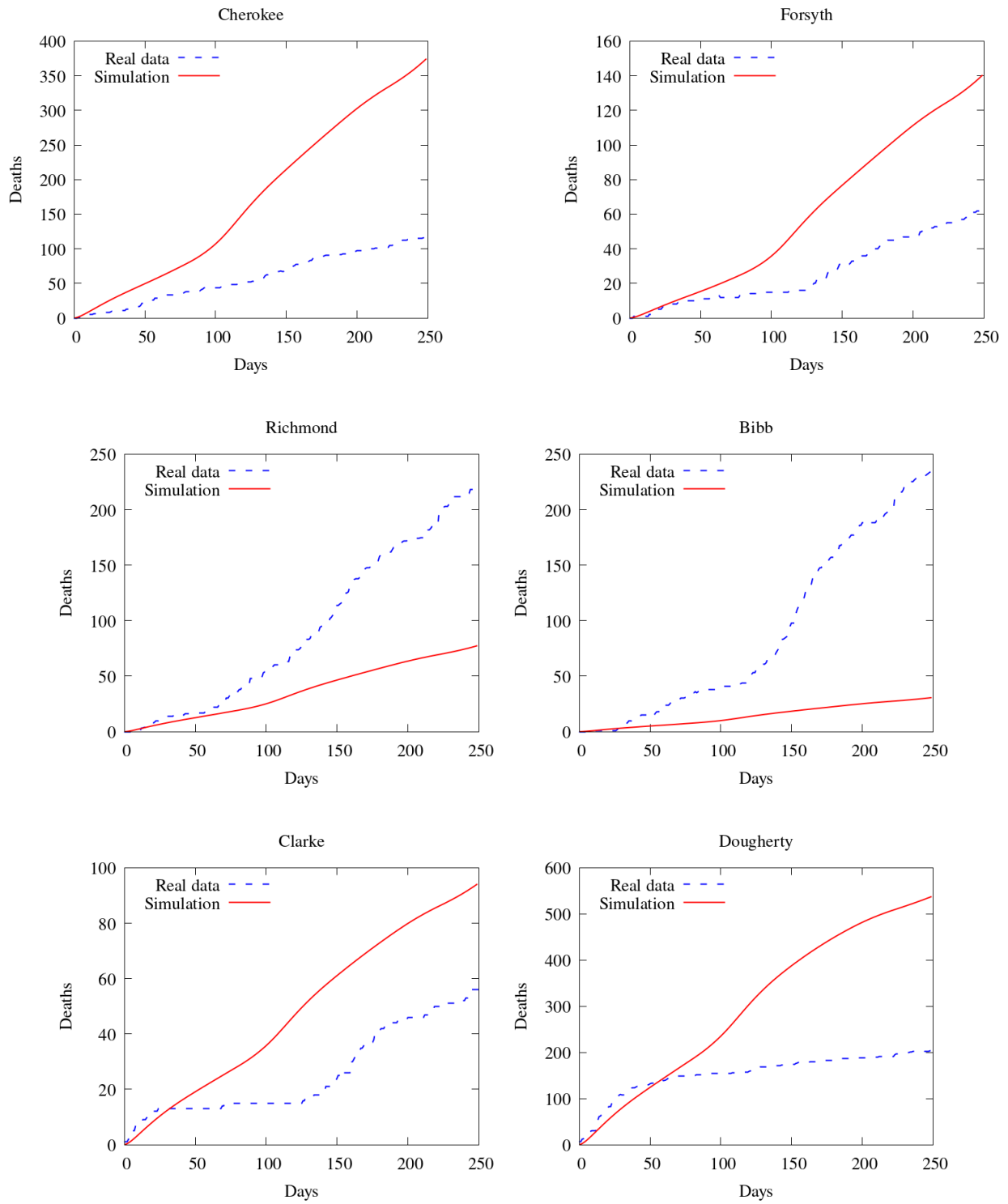


Figure 9: Comparison between simulation and real data of deaths at Georgia (per county) - part 2.

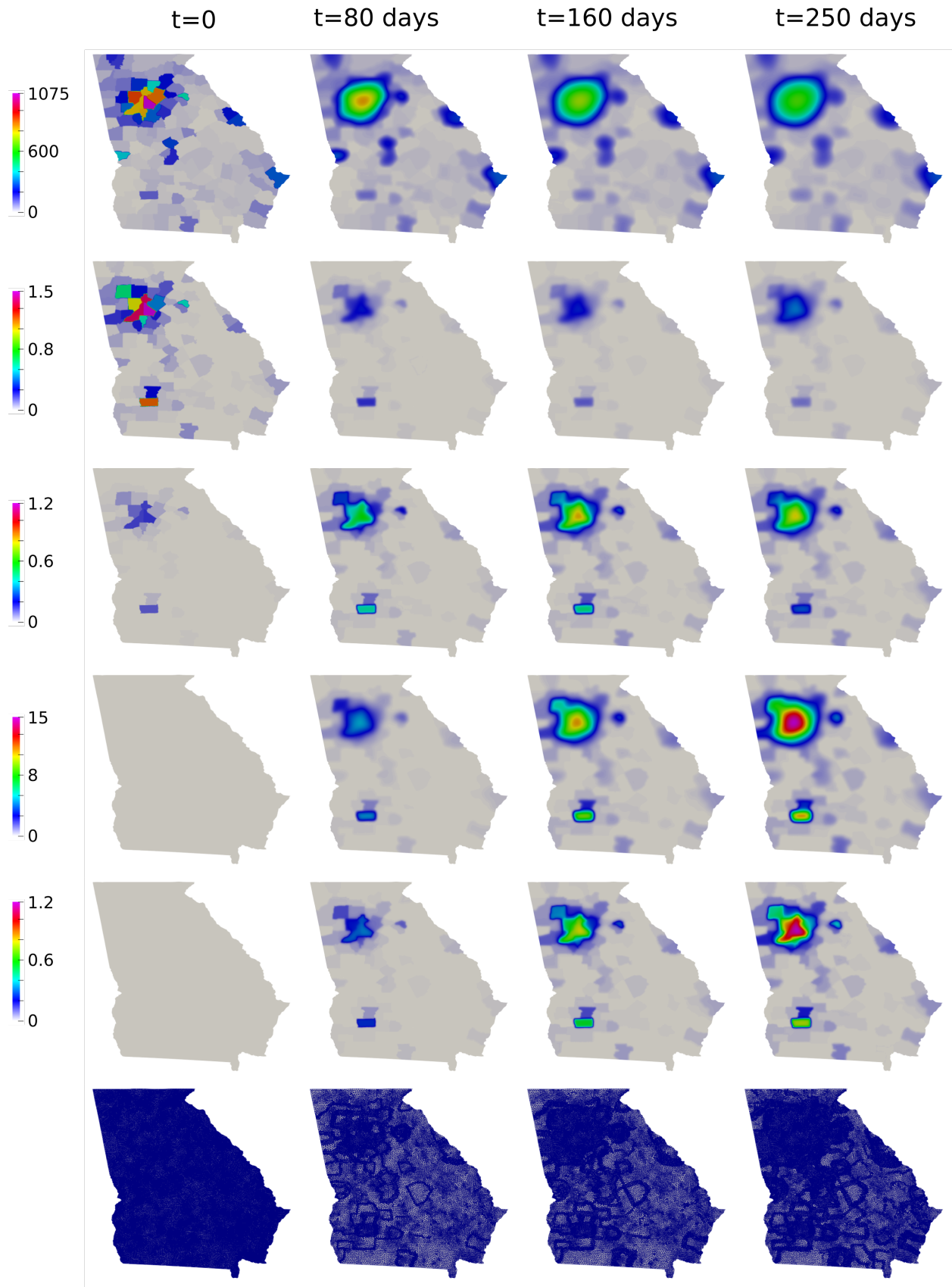


Figure 10: Snapshots of Georgia simulation. Top to bottom: Susceptible, Exposed, Infected, Recovery and Deceased population (*people/km²*), followed by the mesh at different time-steps.

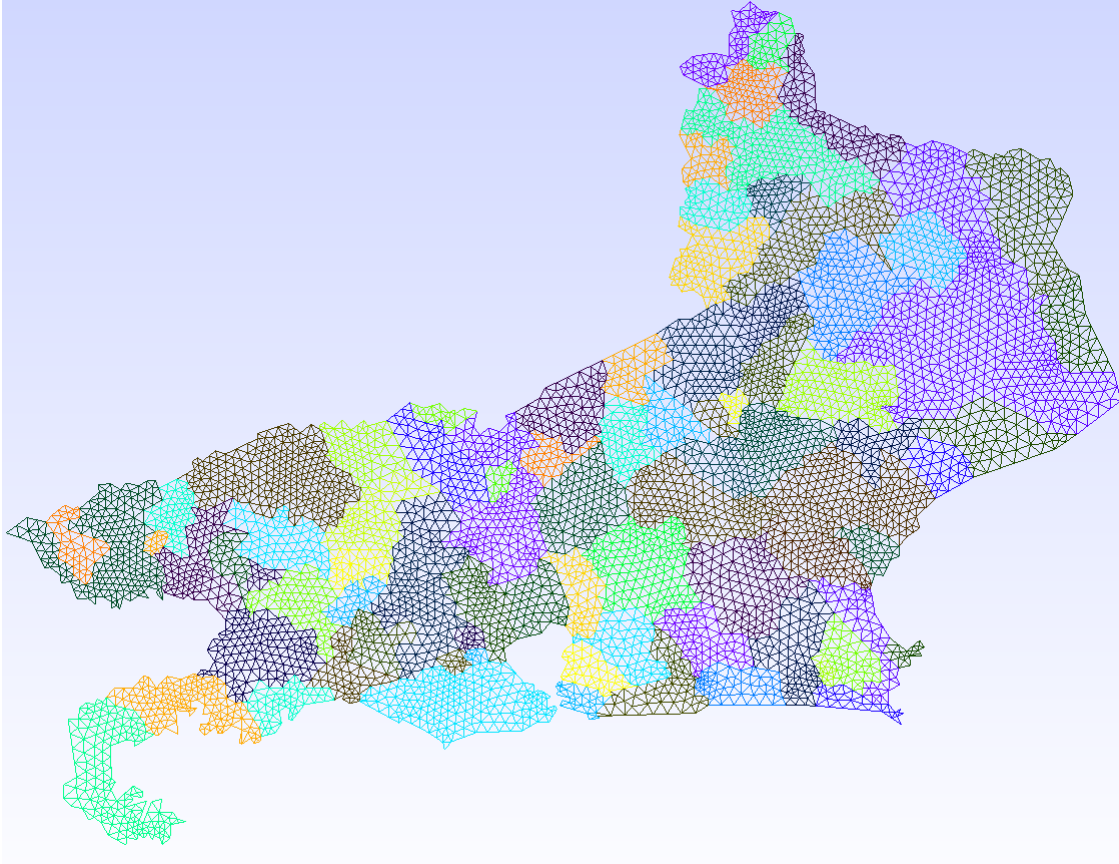


Figure 11: Map of the state of RJ partitioned into 92 internal counties.

3.3.2 Results

To compare the results of our model with the data available, we compare the number of deaths caused by the COVID-19. First, we show the values of the whole state (Fig. 16). Then, we integrate the values of some counties and plot them in time (Figs. 17 and 18). Fig. 19 also shows some snapshots of the simulation with the populations and mesh at different time-steps.

As in the previous case in Georgia, most selected counties show good agreement with the available data. Notably, over the entire region, the relative difference in L^2 norm between the observed and simulated fatalities is only 10.06%. Over the first several weeks, the results are in good agreement nearly everywhere, and, in the areas around the city of Rio de Janeiro, we capture the dynamics very accurately over the entire considered time interval. In addition to the county of Rio de Janeiro, we observe particularly good agreement in the highly populated areas of São Gonçalo, Duque de Caxias, Belford Roxo, and Itaboraí. In the areas of Petrópolis, Volta Redonda, and Nova Iguaçu, we obtain somewhat less accurate, but still nonetheless reasonable results. In these latter cases, we observe good agreement over the initial phases of the dynamics, with some differences later in the time interval. These deviations are unsurprising, as one may naturally expect increased difficulty further in time.

We do observe some notable instances of overestimation in Niterói and underestimation in Cabo Frio and Campos dos Goytacazes. In the latter cases, there were no initial infected or exposed population in this simulation, making diffusion the only possible way for the virus to reach the areas. Given the large distances between these areas and the hotspots near the city center of Rio de Janeiro (157 km to Cabo Frio and 279 km to Campos dos Goytacazes), as well as large sparsely-populated areas between them, population-weighted diffusion was not able to properly represent these dynamics. The model best captures spatial dynamics over relatively short distances overpopulated regions. This represents a shortcoming of the model, and additional terms, such as source/sink terms, fractional diffusion operators, or bilaplacian diffusion terms, may be required to properly account for such nonlocal dynamics. In the case of Niterói, the overprediction may result from an overestimation of the initial exposed population or failure to take into account specific local policies. This is similar to what we observed in Georgia, in which certain particular regions had

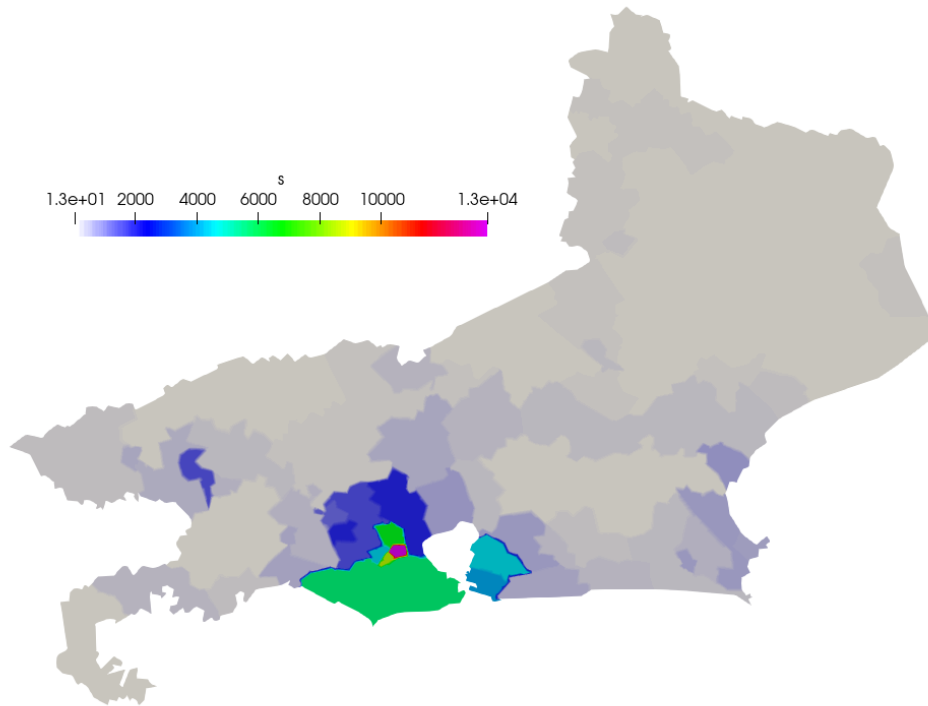


Figure 12: Initial susceptible population ($people/km^2$).

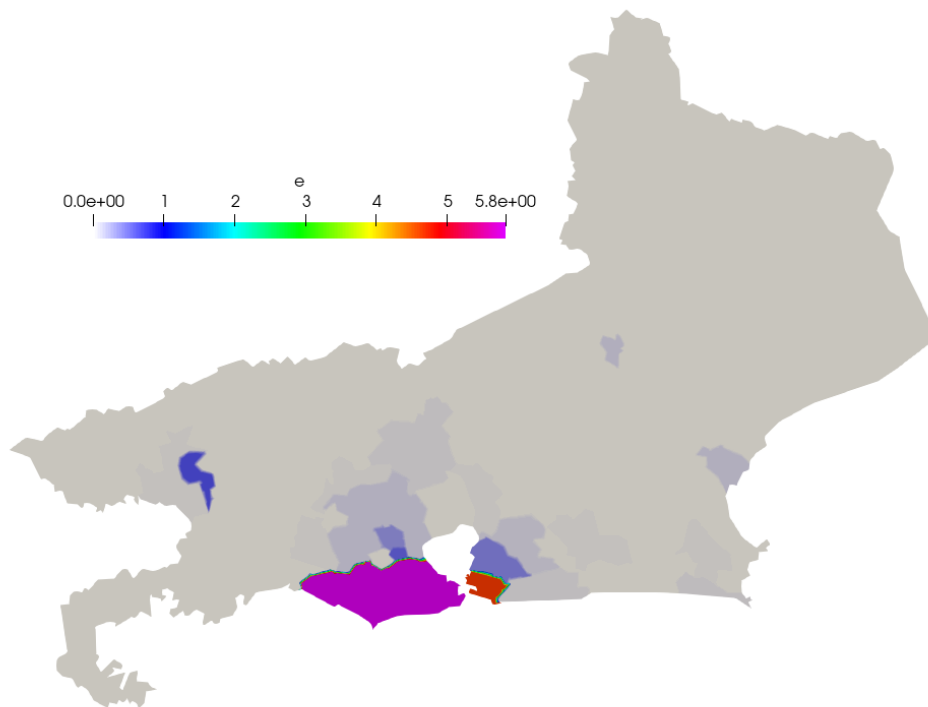


Figure 13: Initial exposed population ($people/km^2$).

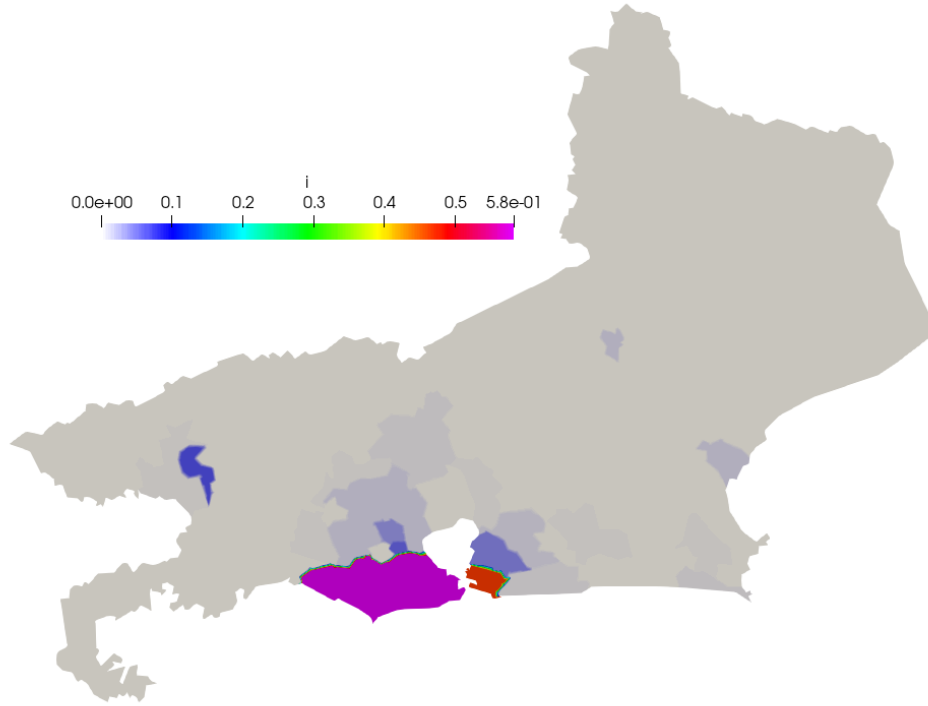
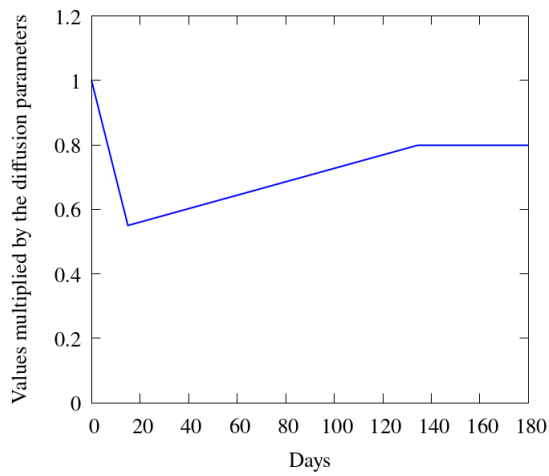
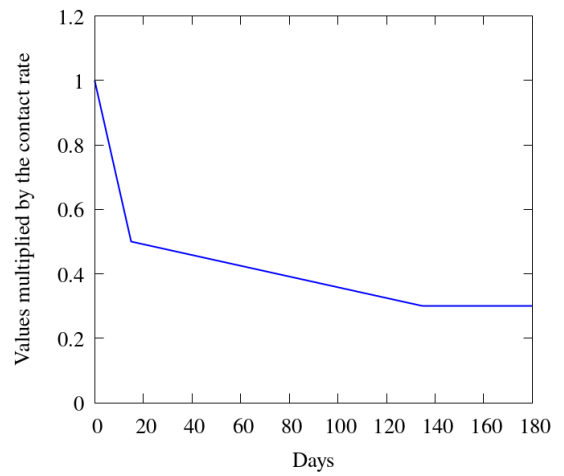


Figure 14: Initial infected population (*people/km²*).



(a) Function for ν .



(b) Function for β .

Figure 15: Functions that multiplies the diffusion (left) and contact rate (right) in time for Rio de Janeiro.

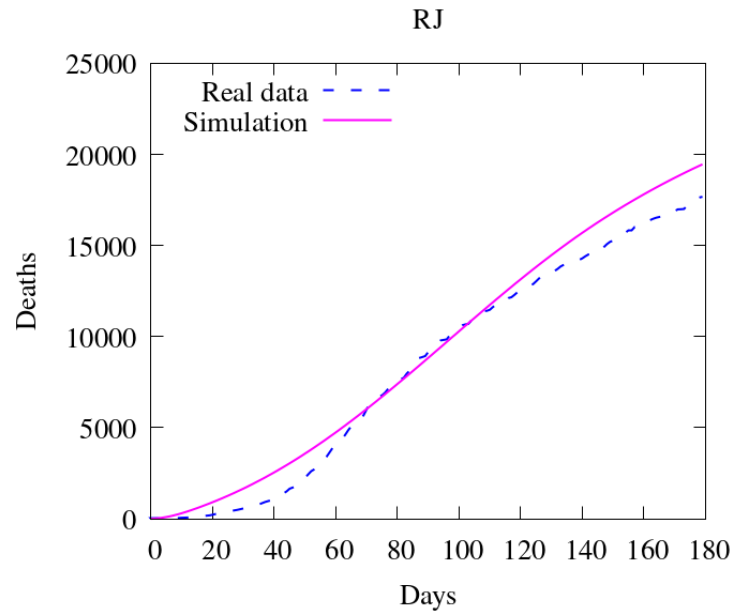


Figure 16: Comparison between simulation and real data of deaths at Rio de Janeiro state (total).

dynamics that deviated from the measured data. However, as in Georgia’s case, we observe very good agreement when considering the totality of the region, with particular areas less well-represented due to the natural limitations of our modeling approach. In these instances, such cases provide clear directions in which we may improve the modeling framework.

4 Conclusions and future work

In this work, we have established the robustness of the SEIRD model introduced in [8] and further extended in [7, 27, 14] by examining both its frequency- and density-dependent formulations and applying it to three settings over different continents: the region of Lombardy, Italy (density-dependent), the U.S. state of Georgia, and the state of Rio de Janeiro, Brazil (frequency-dependent). All these regions have very different geographic characteristics, population density patterns, policy response measures, and cultural contexts generally; despite this, the models were able to adequately reproduce the spatio-temporal contagion dynamics with minimal parameter differences and very little tuning. This provides strong evidence towards the robustness of the model, as it can produce adequate results across different settings without requiring extensive parameter fitting and learning.

The results obtained showed good qualitative agreement generally across all regions; however, they show clear room for improvement. In the cases of Georgia and Rio de Janeiro, we find that the initial conditions have a large influence, even over the long-term dynamics of the system. This is particularly apparent in the case of Dougherty County in Georgia, where an early outbreak led to a long-term overestimation of contagion, and in Campos dos Goytacazes and Cabo Frio in Rio de Janeiro, where a relative lack of early cases led to an under-prediction of the long term contagion effects. This is likely due to the diffusion formulation limitations, which cannot correctly account for non-local mobility across different areas, such as returning travelers. Such effects could be incorporated in a number of ways by including more sophisticated source/sink terms, fractional diffusion operators, or bilaplacian terms. In Lombardy, we observe an under-prediction of contagion in less-populated regions; this appears to be characteristic of the density-dependent formulation, and the frequency-dependent models applied in Georgia and Rio de Janeiro do not suffer from this problem. Across all cases, the same parameter values were used everywhere. While this represents evidence towards robustness and is in some sense a positive aspect, results could, of course, be further improved by using different definitions in different regions, something explored briefly in [8] and in more detail in [14]. Such approaches may incorporate more sophisticated methods, including optimization and machine-learning techniques. These shortcomings represent natural directions for future work in this area. Nonetheless, we believe that our results clearly establish the viability and robustness of this PDE modeling framework, suggesting this as a suitable path for future research, and ultimately may perhaps prove useful to public health decision-makers.

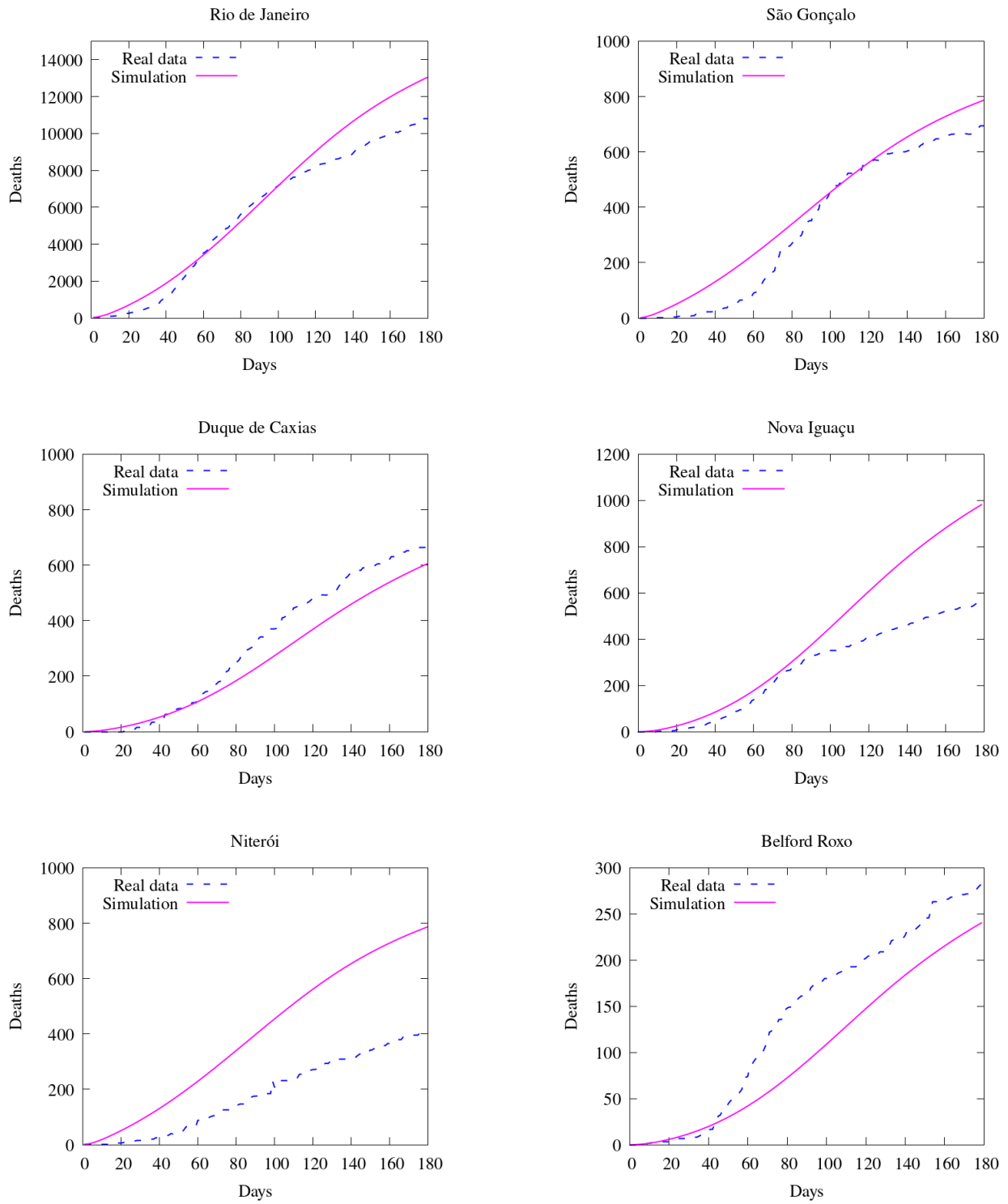


Figure 17: Comparison between simulation and real data of deaths at RJ (per county) - part 1.

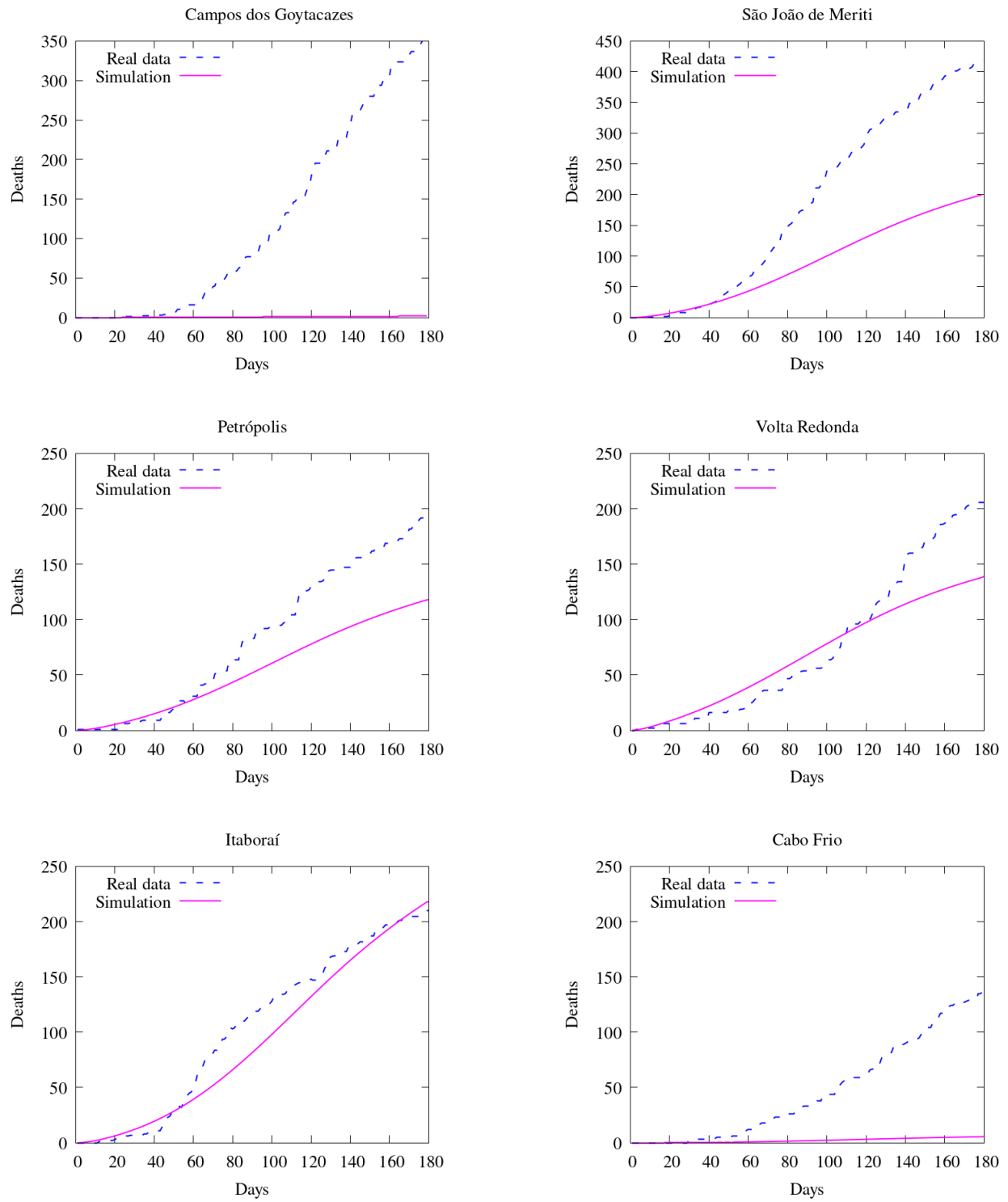


Figure 18: Comparison between simulation and real data of deaths at RJ (per county) - part 2.

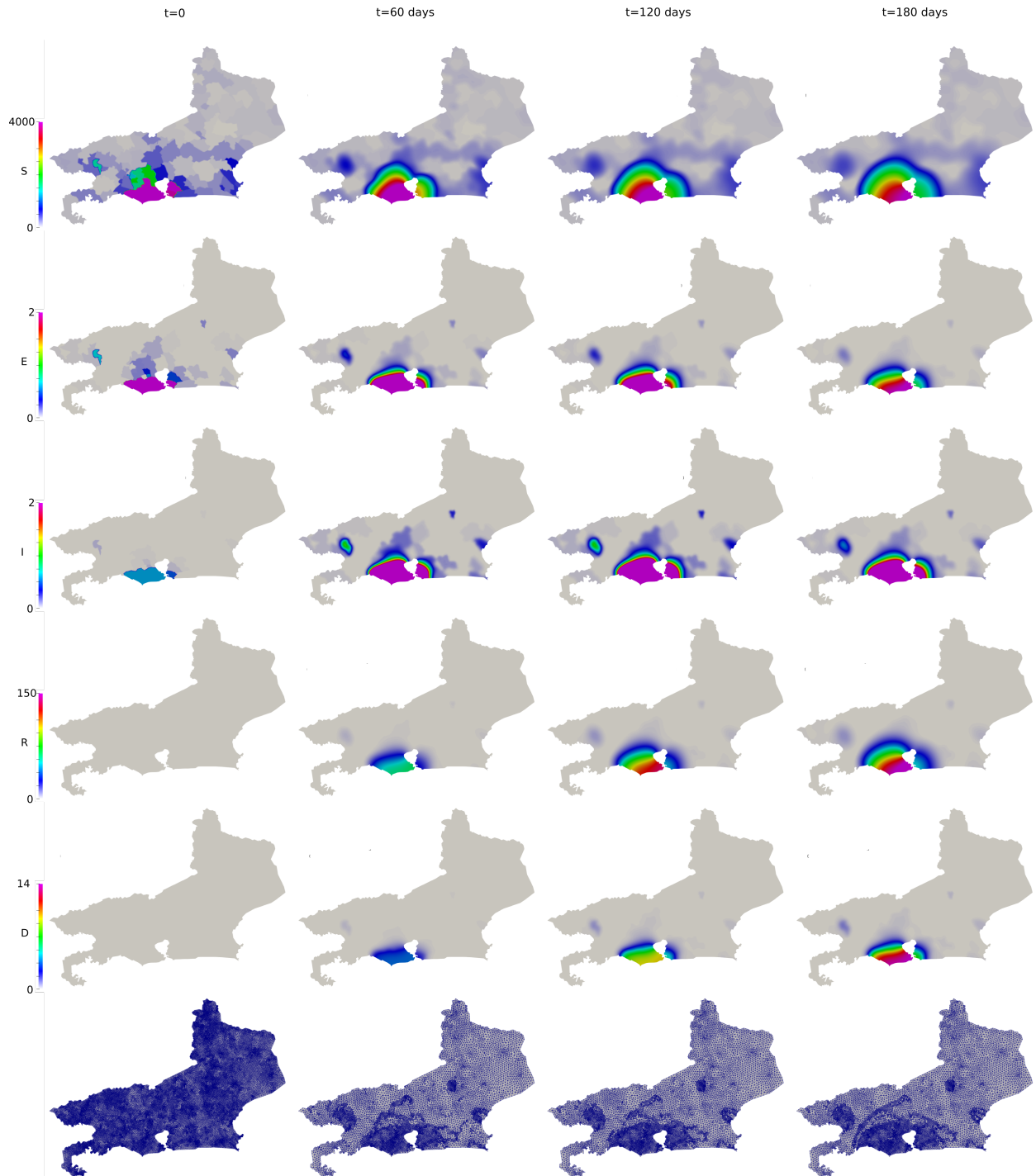


Figure 19: Snapshots of Rio de Janeiro simulation. Top to bottom: Susceptible, Exposed, Infected, Recovery and Deceased population (*people/km²*), followed by the mesh at different time-steps.

5 Acknowledgements

This research was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 and CAPES TecnoDigital Project 223038.014313/2020-19. This research has also received funding from CNPq and FAPERJ. A. Reali was partially supported by the Italian Ministry of University and Research (MIUR) through the PRIN project XFAST-SIMS (No. 20173C478N).

References

- [1] Richard A Erickson, Steven M Presley, Linda JS Allen, Kevin R Long, and Stephen B Cox. A dengue model with a dynamic aedes albopictus vector population. *Ecological Modelling*, 221(24):2899–2908, 2010.
- [2] Eber Dantas, Michel Tosin, and Americo Cunha Jr. Calibration of a seir–sei epidemic model to describe the zika virus outbreak in brazil. *Applied Mathematics and Computation*, 338:249–259, 2018.
- [3] Zindoga Mukandavire, Prasenjit Das, Christinah Chiyaka, and Farai Nyabadza. Global analysis of an hiv/aids epidemic model. *World Journal of Modelling and Simulation*, 6(3):231–240, 2010.
- [4] Chris Dye and Nigel Gay. Modeling the sars epidemic. *Science*, 300(5627):1884–1885, 2003.
- [5] Alemayehu Midekisa, Gabriel Senay, Geoffrey M Henebry, Paulos Semuniguse, and Michael C Wimberly. Remote sensing-based time series models for malaria early warning in the highlands of ethiopia. *Malaria Journal*, 11(1):165, 2012.
- [6] Pheny E Lekone and Bärbel F Finkenstädt. Statistical inference in a stochastic epidemic seir model with control intervention: Ebola as a case study. *Biometrics*, 62(4):1170–1177, 2006.
- [7] Alex Viguerie, Alessandro Veneziani, Guillermo Lorenzo, Davide Baroli, Nicole Aretz-Nellesen, Alessia Patton, Thomas E Yankeelov, Alessandro Reali, Thomas JR Hughes, and Ferdinando Auricchio. Diffusion–reaction compartmental models formulated in a continuum mechanics framework: application to covid-19, mathematical analysis, and numerical study. *Computational Mechanics*, pages 1–22, 2020.
- [8] Alex Viguerie, Guillermo Lorenzo, Ferdinando Auricchio, Davide Baroli, Thomas JR Hughes, Alessia Patton, Alessandro Reali, Thomas E Yankeelov, and Alessandro Veneziani. Simulating the spread of covid-19 via spatially-resolved susceptible-exposed-infected-recovered-deceased (seird) model with heterogeneous diffusion. *Applied Mathematics Letters*, 111:106617, 2021.
- [9] Julien Arino and Stéphanie Portet. A simple model for covid-19. *Infectious Disease Modelling*, 5:309–315, 2020.
- [10] Giulia Giordano, Franco Blanchini, Raffaele Bruno, Patrizio Colaneri, Alessandro Di Filippo, Angela Di Matteo, and Marta Colaneri. Modelling the covid-19 epidemic and implementation of population-wide interventions in italy. *Nature Medicine*, pages 1–6, 2020.
- [11] José M Carcione, Juan E Santos, Claudio Bagaini, and Jing Ba. A simulation of a covid-19 epidemic based on a deterministic seir model. *Frontiers in Public Health*, 8:230, 2020.
- [12] Diego Tavares Volpatto, Anna Claudia Mello Resende, Lucas Anjos, Joao Vitor Oliveira Silva, Claudia Mazza Dias, Regina Cerqueira Almeida, and Sandra Mara Cardoso Malta. Spreading of covid-19 in brazil: Impacts and uncertainties in social distancing strategies. *medRxiv*, 2020.
- [13] Juliane F Oliveira, Daniel CP Jorge, Rafael V Veiga, Moreno S Rodrigues, Matheus F Torquato, Nivea B da Silva, Rosemeire L Fiaccone, Luciana L Cardim, Felipe AC Pereira, Caio P de Castro, et al. Mathematical modeling of covid-19 in 14.8 million individuals in bahia, brazil. *Nature communications*, 12(1):1–13, 2021.
- [14] Prashant K Jha, Lianghao Cao, and J Tinsley Oden. Bayesian-based predictions of covid-19 evolution in texas using multispecies mixture-theoretic continuum models. *Computational Mechanics*, 66(5):1055–1068, 2020.
- [15] Ivan Korolev. Identification and estimation of the seird epidemic model for covid-19. *Journal of econometrics*, 220(1):63–85, 2020.
- [16] Fleurianne Bertrand and Emilie Pirch. Least-squares finite element method for a meso-scale model of the spread of covid-19. *Computation*, 9(2), 2021.
- [17] Fred Brauer, Carlos Castillo-Chavez, and Zhilan Feng. *Mathematical models in epidemiology*. Springer, 2019.
- [18] Marino Gatto, Enrico Bertuzzo, Lorenzo Mari, Stefano Miccoli, Luca Carraro, Renato Casagrandi, and Andrea Rinaldo. Spread and dynamics of the covid-19 epidemic in italy: Effects of emergency containment measures. *Proceedings of the National Academy of Sciences*, 117(19):10484–10491, 2020.

- [19] Francesc Aràndiga, Antonio Baeza, Isabel Cordero-Carrión, Rosa Donat, M Carmen Martí, Pep Mulet, and Dionisio F Yáñez. A spatial-temporal model for the evolution of the covid-19 pandemic in spain including mobility. *Mathematics*, 8(10):1677, 2020.
- [20] Valerio La Gatta, Vincenzo Moscato, Marco Postiglione, and Giancarlo Sperli. An epidemiological neural network exploiting dynamic graph structured data applied to the covid-19 outbreak. *IEEE Transactions on Big Data*, 2020.
- [21] TI Zohdi. An agent-based computational framework for simulation of global pandemic and social response on planet x. *Computational Mechanics*, 66(5):1195–1209, 2020.
- [22] Petrônio CL Silva, Paulo VC Batista, Hélder S Lima, Marcos A Alves, Frederico G Guimarães, and Rodrigo CP Silva. Covid-abs: An agent-based model of covid-19 epidemic to simulate health and economic effects of social distancing interventions. *Chaos, Solitons & Fractals*, 139:110088, 2020.
- [23] Navid Mahdizadeh Gharakhanlou and Navid Hooshangi. Spatio-temporal simulation of the novel coronavirus (covid-19) outbreak using the agent-based modeling approach (case study: Urmia, iran). *Informatics in Medicine Unlocked*, 20:100403, 2020.
- [24] Joshua P Keller, Luca Gerardo-Giorda, and Alessandro Veneziani. Numerical simulation of a susceptible–exposed–infectious space-continuous model for the spread of rabies in raccoons across a realistic landscape. *Journal of Biological Dynamics*, 7(sup1):31–46, 2013.
- [25] Mi-Young Kim. Galerkin methods for a model of population dynamics with nonlinear diffusion. *Numerical Methods for Partial Differential Equations: An International Journal*, 12(1):59–73, 1996.
- [26] Robert Stephen Cantrell and Chris Cosner. *Spatial ecology via reaction-diffusion equations*. John Wiley & Sons, 2004.
- [27] Malú Grave and Alvaro L. G. A. Coutinho. Adaptive mesh refinement and coarsening for diffusion-reaction epidemiological models, 2020.
- [28] Madan K Oli, Meenakshi Venkataraman, Paul A Klein, Lori D Wendland, and Mary B Brown. Population dynamics of infectious diseases: a discrete time model. *Ecological Modelling*, 198(1-2):183–194, 2006.
- [29] Peter H Thrall, Arjen Biere, and Marcy K Uyenoyama. Frequency-dependent disease transmission and the dynamics of the silene-ustilago host-pathogen system. *The American Naturalist*, 145(1):43–62, 1995.
- [30] Kaitlyn E Johnson, Grant Howard, William Mo, Michael K Strasser, Ernesto ABF Lima, Sui Huang, and Amy Brock. Cancer cell population growth kinetics at low densities deviate from the exponential growth model and suggest an allee effect. *PLoS biology*, 17(8):e3000399, 2019.
- [31] Marcello Delitala and Mario Ferraro. Is the Allee effect relevant in cancer evolution and therapy? *AIMS Mathematics*, 5(6):7649–7660, 2020.
- [32] B. S. Kirk, J. W. Peterson, R. H. Stogner, and G. F. Carey. libmesh: a c++ library for parallel adaptive mesh refinement/coarsening simulations. *Journal Engineering with Computers*, 22(3):237–254, 2006.
- [33] Satish Balay, Shrirang Abhyankar, Mark F. Adams, Jed Brown, Peter Brune, Kris Buschelman, Lisandro Dalcin, Alp Dener, Victor Eijkhout, William D. Gropp, Dmitry Karpeyev, Dinesh Kaushik, Matthew G. Knepley, Dave A. May, Lois Curfman McInnes, Richard Tran Mills, Todd Munson, Karl Rupp, Patrick Sanan, Barry F. Smith, Stefano Zampini, Hong Zhang, and Hong Zhang. PETSc Web page. <https://www.mcs.anl.gov/petsc>, 2019.
- [34] The Trilinos Project Team. The Trilinos Project Website.
- [35] Malú Grave, José J. Camata, and Alvaro L. G. A. Coutinho. A new convected level-set method for gas bubble dynamics. *Computers & Fluids*, 209:104667, 2020.
- [36] QGIS Development Team. *QGIS Geographic Information System*. Open Source Geospatial Foundation, 2009.
- [37] Ensheng Dong, Hongru Du, and Lauren Gardner. An interactive web-based dashboard to track covid-19 in real time. *The Lancet infectious diseases*, 20(5):533–534, 2020.
- [38] Ruy Freitas Reis, Bárbara de Melo Quintela, Joventino de Oliveira Campos, Johnny Moreira Gomes, Bernardo Martins Rocha, Marcelo Lobosco, and Rodrigo Weber dos Santos. Characterization of the covid-19 pandemic and the impact of uncertainties, mitigation strategies, and underreporting of cases in south korea, italy, and brazil. *Chaos, Solitons & Fractals*, page 109888, 2020.
- [39] United States Census Bureau. <https://www.census.gov/quickfacts/GA>, 2020 (accessed April 1 2020).