

Mathematical Modeling and Georeferenced Forecasting for the COVID-19 at the State of RS, Brazil

J. C. MARQUES^{1*}, A. DE CEZARO², D. C. L. ALVES³, P. V. A. B. LISBÔA⁴,
MATHEUS LAZO⁵ and D. A. SALES⁶

Received on December 22, 2021 / Accepted on December 20, 2022

ABSTRACT. In this contribution, we present a predictive tool developed to help in the management of the evolution of the COVID-19 pandemic situation in Rio Grande do Sul (RS) - Brazil. This tool is the result of georeferenced data analysis, mathematical modeling, and parameter calibration for the dynamics of a SIR-type model defined on a spatial structure that allows distinct subpopulations to interact, similar to the controlled distancing (A_l for $l = 1, \dots, 21$) groupings proposed by the RS government and public health authorities. The predictive analysis, updated biweekly, provides three distinct scenarios per month (milder, average, and severe) and is made available as WebSIGs (Geographic Information System - GIS). The forecast of the average scenario for each A_l group is the result of a simulation of the proposed SIR-type dynamics with calibrated parameters derived from an augmented Lagrangian maximum a posteriori estimation and data on the number of infected cases made available by the RS Health Secretariat. The milder and severe scenarios are obtained from the average scenario, with changes in the contagion rates of each A_l group. When compared to the number of infections reported in each A_l group, the modeling predictions for a biweekly time window (the first two weeks) were quite satisfactory, with errors ranging from 0% to 5.13%, gradually increasing over time. Therefore, we suggest a biweekly re-calibration of the parameters and corresponding forecasts as a wise strategy.

Keywords: COVID-19 forecast, SIR-type model, parameter calibration, georeferenced data.

1 INTRODUCTION

Since the first reported case of COVID-19, there has been a great mobilization of the scientific community to understand the dynamics of SARS-CoV-2 (the virus that causes COVID-19 [19])

*Corresponding author: Joice Chaves Marques – E-mail: joicec.marques@hotmail.com

¹IMEF/FURG, Rio Grande, RS, Brazil – E-mail: joicec.marques@hotmail.com <https://orcid.org/0000-0003-2137-2164>

²IMEF/FURG, Rio Grande, RS, Brazil – E-mail: adrianocezaro@furg.br <https://orcid.org/0000-0001-8431-9120>

³IMEF/FURG, Rio Grande, RS, Brazil – E-mail: dclealalves@gmail.com <https://orcid.org/0000-0002-5255-123X>

⁴IMEF/FURG, Rio Grande, RS, Brazil – E-mail: paulovictor_fjv@hotmail.com <https://orcid.org/0000-0002-3465-9853>

⁵IMEF/FURG, Rio Grande, RS, Brazil – E-mail: matheuslazo@furg.br <https://orcid.org/0000-0001-9741-9411>

⁶IMEF/FURG, Rio Grande, RS, Brazil – E-mail: dsales@furg.br <http://orcid.org/0000-0002-3496-5711>

and how the virus spreads. In this scenario, the increased demand for drugs, and the rapid development of vaccines, become the focus of most researchers. While known and tested drugs have shown little or no efficacy, and mass vaccination of the entire world population is unthinkable in the short term due to the time required for manufacturing, distribution, and application, non-pharmacological procedures have remained an important alternative to tightening pandemic effects. In this context, mathematical modeling has been an indispensable tool to formulate forecasts and to give assistance to public leaders in the making of public policies to slow the spread of the disease, e.g., [1, 11, 14] and references therein.

On the other hand, it was also clear that science is not enough to contest the effects of the pandemic without an efficient way of spreading the relevant information. In this sense, the use of geotechnologies has also proved to be of great value, especially Digital Cartography and Geographic Information Systems (GIS) as tools for georeferencing and spatial representation of the occurrences of COVID-19. Several initiatives have used digital maps and cybercartography for the cartographic organization of data, didactically illustrating the density of the epidemic and, mainly, supporting decision-making in the effective application of control devices to reduce the transmission of the virus [2, 9]. Such initiatives, both in the academic sphere and in territorial management, follow an extensive tradition of Health Cartography as a technical-scientific instrument to support epidemiological analysis.

The present study aims to add to the understanding, prediction, dissemination to society, and support for decision-making regarding the dynamics and spread of the COVID-19 pandemic by proposing a spatially structured SIR model integrated into georeferencing tools. We present a predictive tool developed to help in the management of the evolution of the COVID-19 pandemic in the Rio Grande do Sul (RS) - Brazil. The work consists of three steps as follows: 1) Organization and treatment of the reported data of infected people, in a georeferenced manner, which serves as an entry in mathematical modeling (initial conditions and data for parameters calibration for the proposed SIR type model), as detailed in Section 2; 2) Mathematical modeling and parameter calibration of COVID-19 dynamics by a SIR model with interacting multiple populations, as proposed in Sections 3 and 4; 3) Organization of COVID-19 spread scenarios in WebSIGs set-up as described in Section 4.

2 ORGANIZATION AND DATA OBSERVATION

Online digital mapping, also known as cybercartography, is an informational solution for publishing georeferenced data, making it more accessible to other users [10]. The methodology consists of three stages: the organization and processing of data provided by the State Health Department; mathematical modeling; and the creation of the WebSIG (Figure 1).

The creation of the geographic information plan requires data only from official sources. The 2019 vector base in the *shapefile* format provided by the Brazilian Institute of Geography and Statistics (IBGE) was admitted as a municipal area¹. The regionalization of the modeling, in the

¹Available in <https://www.ibge.gov.br/>

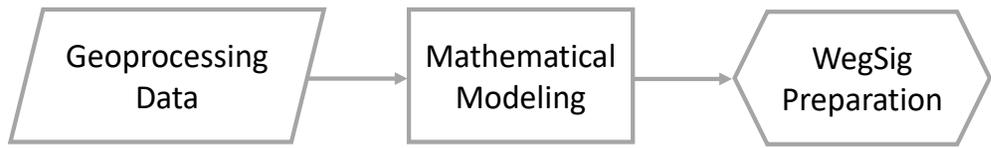


Figure 1: Flowchart representative of the stages involved in the work.

form of regional blocks or clusters, had as reference the information provided by the portal of planning, state budget, and health management of RS, which were distributed into twenty-one clusters provided by the controlled distancing model of the state, corresponding to the 30 health regions (see [6, Fig. 2]).

To this end, the thirty Health Regions of RS, which subsidize actions in the sphere of planning and infrastructure of the State's health, served as the basis for the construction of the twenty-one groups provided for by the Controlled Distancing Model of RS (Data control COVID-19 CORONAVÍRUS, 2020) and in this work the same criteria were adopted. According to the epidemiological action plan, which includes the classification flags of the degree of risk and the respective social distancing protocols, such groups were delimited using criteria of health and economic activity.

Each cluster is composed of one or more health regions that aggregate, respectively, a fraction of the 497 municipalities that make up the state of RS. In this study, the same special criteria proposed by the Controlled Distance Model of RS were adopted, facilitating access to the registry of reported infected people, data made available daily by the Painel Coronavirus – RS², as well as comparisons between the epidemiological actions and the results obtained by the modeling over the past few months. Because the data is cumulative for each day, a subtraction was performed with the count of the cases from the previous day to obtain the new cases for each day.

2.1 Organization and processing of geographic data: COVID-19 positive cases reported

The methodologies of organization, acquisition, and processing of data on the number of infected cases by Sars-Cov-2 officially reported by the Health Department of RS, available in electronic means, are presented in this work between the months of March 2020 and June 2021. In addition, we demonstrated how such a set of data is used to calibrate the parameters for a mathematical modeling of the spread of the disease in the 21 health groups of the state, which results in a set of predictive data for three distinct scenarios for the spread of COVID-19, in RS over a period of 28 days from the initial date of the data of positive cases reported. Such a set of predictive data can be used by health authorities as an auxiliary means in the decision-making for measures to control and combat the spread of COVID-19 and these are available for access through interactive maps *WebSIGs* with the spatial distribution of the projected cases. In this manuscript, we present a small sample of the collected and stored data as well as the projected data, the complete set of

²Available in <https://ti.saude.rs.gov.br/covid19/>

which is in the repository <https://github.com/exactum-furg> or as complementary material in [6]. In a simplified procedure, the processes described above can be represented by the following flowchart:

First, we performed the creation process of the Geographic Information Plan of RS, which corresponds to the first step of the flowchart presented in Figure 1. IBGE 2021 data in the form of a shapefile from an automatic recovery system was used for this purpose. This information was used as the basis of the territorial and demographic division of the municipalities of RS and reorganization in the form of regional blocks or clusters, covering the 21 health groupings provided by the social-distancing plan developed by the State Government, based on the information provided by the Portal of Planning, Budget, and Health Management of RS. In [6, Fig. 2] it is possible to notice this geographical division.

Subsequently, the reported data of positive cases for COVID-19 from each of the 21 groupings was accessed daily. The positive case data accumulated is downloaded and systematized in an Excel spreadsheet in a georeferenced way. The georeferencing of the data allows us to link the characteristics of each of the 21 health groupings that are used as input data for mathematical modeling. Such characteristics of the health groups include the population, the area, the name of the group, the code of the regions that make-up each group and the number reported of infected cases (number of positive cases of COVID-19 reported). See the details in Table 1. According to these data, it is possible to calculate important variables for the study, such as: new cumulative cases of the week, population density, as well as the total number of infected cases reported for each group.

The last step is the preparation of the acquired data of reported COVID-19 positivity cases for use in mathematical modeling. To perform it, a conversion factor/pixel to standardize the output spatial size of the cell (spatial resolution) of data *rasters* (vector format) generated by converting formats is used. For the standardization of this factor, we adopted the combined relationship between the centroids of the polygonal shape of each cluster and its maximum area, thus resulting in a conversion factor of 0.1, which results in a spatial resolution of 10×10 km cell/pixel. This resolution best matched the geometry of the health groups, preserving their original format [15]. This process allows the organization of the database of people infected by COVID-19 collected from the Coronavirus panel, for each of the 21 health clusters, available in table format and, by processing, feed files of the vector type (shapefiles) that are converted into arrays with coordinate registration (raster). Finally, the database is transformed into ASCII format for later use as initial conditions and data for mathematical modeling and parameter calibration.

It is worth noting that every day at 6 p.m., data on the number of reported positive COVID-19 infections is accessed and downloaded from the Coronavirus panel. Afterward, these are grouped by the weekly cumulative, to correct any issues related to the data reported at the weekends. As a result, for each new date defined as the basis, the number of cases will be presented week by week, beginning with the first week.

Table 1: Data organization according to the planning regions of the state of Rio Grande do Sul, $t_0^1 = 24/03/21$ and $t_0^2 = 07/04/21$.

Health region	Groupings	Area km^2	Population
R01, R02	A ₁ -Santa Maria	26535.20	270865
R03	A ₂ -Uruguaiana	41475.58	395831
R04, R05	A ₃ -Capão da Canoa	8766.41	458551
R06	A ₄ -Taquara	5896.09	236699
R07	A ₅ -Novo Hamburgo	1337.51	902777
R08	A ₆ -Canoas	2733.80	727179
R09	A ₇ -Guaíba	11093.46	323594
R10	A ₈ -Porto Alegre	2893.78	2125191
R11	A ₉ -Santo Ângelo	15396.88	265271
R12	A ₁₀ -Cruz Alta	8805.18	178743
R13	A ₁₁ -Ijuí	726798	229464
R14	A ₁₂ -Santa Rosa	5774.34	415175
R15 R20	A ₁₃ -Palmeira das Missões	11052.07	363139
R16	A ₁₄ -Erechim	6628.44	299323
R17, R18, R19	A ₁₅ -Passo Fundo	19552.53	655264
R21	A ₁₆ -Pelotas	34902.26	998250
R22	A ₁₇ -Bagé	15192.96	152731
R23, R24, R25, R26	A ₁₈ -Caxias do Sul	18880.02	1247004
R27	A ₁₉ -Cachoeira do Sul	12063.63	202303
R28	A ₂₀ -Santa Cruz do Sul	7467.65	634963
R29, R30	A ₂₁ -Lajeado	4922.09	340656
Groupings	Counties	Reported at t_0^1	Reported at t_0^2
A ₁	32	765	487
A ₂	11	464	344
A ₃	23	97	338
A ₄	8	189	118
A ₅	15	775	628
A ₆	18	848	836
A ₇	19	240	284
A ₈	6	642	595
A ₉	24	125	184
A ₁₀	13	193	189
A ₁₁	20	209	105
A ₁₂	22	152	94
A ₁₃	52	270	180
A ₁₄	33	127	90
A ₁₅	62	477	482
A ₁₆	22	216	403
A ₁₇	6	65	206
A ₁₈	49	555	667
A ₁₉	12	159	68
A ₂₀	13	438	349
A ₂₁	37	534	213

2.1.1 Spatial resolution

The choice of spatial resolution went through two important criteria for the generation of results: a) The size of the pixels in relation to the geometry (area) of the Health Regions, that is, whether the pixel quantity satisfactorily respects the contour of the Regions of Health, b) result processing time.

For the first criterion (pixels \times geometry), it was evident that the 30 km resolution overgeneralized (vector simplification criteria) the polygons that form the health clusters, mischaracterizing them. Thus, the pixels of resolutions of 10 km and 5 km better represent the geometries of the clusters, that is, a generalization that does not deform the geometry of the clusters.

The last criterion takes into account the processing time of each scenario, with the difference between the 5 km resolutions and the 10 and 30 km resolutions being evident. While the 10 km resolution takes 4 min to process all the data from the projected cases, the 5 km resolution takes approximately 40 min, taking more computational time for processing. Whereas the modeling was performed using a notebook with Inter® Core™ i7-7500U CPU @ 2.70GHz x 4, with 7.6GiB of memory and Mesa Inter® HD Graphic 620 (KBL GT2) video graphics.

In view of the criteria, the 10 km resolution presented better results in general without significant loss of area or boundary deformation. This is responsible for good coefficients in the cross-analysis; it represents the geometric features of the health region clusters and a relatively low processing time for the amount of data analyzed.

3 MATHEMATICAL MODELING

The georeferenced forecasting for COVID-19 proposed here assumes that the constant population N is subdivided into a spatially structured network with a size corresponding to the size of the state of RS sub-divisions as described in Section 2. Indeed, this network is composed of 68×83 sites (regions). This network is organized so that each cluster is composed of a certain number of sites established in the data organization step described in Section 2. It should be noted that this number depends on the precision adopted in the rasterization process and on the grouping area. According to the Coupled Map Network (CMN) theory, e.g. [18] and references therein, each site $\mathbf{x} = (x_i, x_j)$ of this network is represented by integer coordinates $x_i = is, x_j = js$, where s is the size of the site, for $i \in \{1, \dots, 63\}, j \in \{1, \dots, 83\}$.

The total population $N_{i,j}$ of each (i, j) site is normalized with respect to the population density and assumed constant, such that the total population in the lattice satisfies $\sum_{i=1}^{68} \sum_{j=1}^{83} N_{i,j} = 1$. Such assumptions have implications for the magnitude of the model parameters (see Table 2). Furthermore, the model is considered as a SIR-type compartmental model, e.g., [12, 13, 14] and references therein, where the total population of each site $N_{i,j}$ is divided into the susceptible fractions $S_{i,j}(t)$, infected $I_{i,j}(t)$ and removed $R_{i,j}(t)$, at each time instant $t \geq 0$, following a compartmental SIR-type model [12, 13].

Table 2: Calibrated parameters from the COVID-19 infected people reported in $t_0 = 24/03/21$ and $t_0 = 07/04/21$, obtained from algorithm ((4.6)) for the corresponding health clusters. The relative errors between predicted and reported cases for the two weeks following the day t_0 for the average scenario are $ER_1\%$ and $ER_2\%$. For the severe scenario, $ERP_2\%$ is the relative error of the second subsequent week from the day t_0 .

t_0	A_j	$\beta_{i,j}$	$\beta_{i,\hat{j}}$	$\gamma_{i,j}$	$ER_1\%$	$ER_2\%$	$ERP_2\%$
24/03/21	A_1	0.2420	0.0017	0.0774	0.37	0.60	0.26
	A_2	0.3580	0.0025	0.1161	0.09	1.73	1.36
	A_3	0.7800	0.0052	0.2516	3.26	2.41	1.95
	A_4	0.4300	0.0029	0.1387	0.35	0.71	0.18
	A_5	0.2920	0.0019	0.0942	0.13	2.78	2.23
	A_6	0.3399	0.0023	0.1096	0.32	0.05	0.15
	A_7	0.4900	0.0033	0.1581	0.36	0.17	0.82
	A_8	0.5830	0.0039	0.1890	0.08	0.15	0.48
	A_9	0.4990	0.0033	0.1610	0.66	1.95	1.51
	A_{10}	0.2860	0.0019	0.0922	0.64	5.3	5.1
	A_{11}	0.3290	0.0022	0.1061	0.69	6.56	6.91
	A_{12}	0.3550	0.0024	0.1145	0.18	2.24	1.43
	A_{13}	0.3710	0.0025	0.1197	1.38	3.12	4.67
	A_{14}	0.4000	0.0027	0.1290	0.06	0.47	0.54
	A_{15}	0.3750	0.0025	0.1219	0.12	5.13	5.18
	A_{16}	0.5460	0.0036	0.1760	2.19	0.86	1.07
	A_{17}	0.5390	0.0036	0.1739	0.98	2.39	2.54
	A_{18}	0.5280	0.0035	0.1706	0.36	1.77	2.41
	A_{19}	0.4100	0.0027	0.1322	1.34	1.13	1.17
	A_{20}	0.3040	0.0020	0.0981	0.32	3.91	2.74
	A_{21}	0.3500	0.0023	0.1130	0.57	1.75	2.72
07/04/21	A_1	0.365	0.0024	0.1177	0.35	29.63	15.30
	A_2	0.5200	0.0044	0.1677	0.08	7.98	21.66
	A_3	0.392	0.0026	0.1264	0.27	1.76	10.18
	A_4	0.574	0.0038	0.1852	0.20	5.87	11.26
	A_5	0.341	0.0023	0.11	0	44.51	29.58
	A_6	0.3000	0.0020	0.0968	0.23	30.95	18.89
	A_7	0.4200	0.0028	0.1355	0.14	3.34	9.48
	A_8	0.3420	0.0023	0.1103	0.18	28.52	15.21
	A_9	0.3900	0.0026	0.1258	0.24	4.97	7.19
	A_{10}	0.2290	0.0015	0.0739	0.65	30.79	21.19
	A_{11}	0.4490	0.0030	0.1448	0.35	8.45	20.54
	A_{12}	0.42889	0.0028	0.1383	1.73	3.25	15.46
	A_{13}	0.3290	0.0022	0.1061	0.63	18.89	6.93
	A_{14}	0.4830	0.0032	0.1558	1.38	29	38.96
	A_{15}	0.3850	0.0025	0.1241	0.05	0	11.56
	A_{16}	0.4080	0.0027	0.1316	0.05	6.95	5.93
	A_{17}	0.3200	0.0021	0.1032	0.55	11.33	0.51
	A_{18}	0.4060	0.0027	0.1309	0.07	11.77	1.74
	A_{19}	0.5910	0.0039	0.1906	0.61	27.11	39.15
	A_{20}	0.3330	0.0022	0.1074	0.15	13.44	78.25
	A_{21}	0.3990	0.0026	0.1287	0.11	5.22	19.98

We assume that the dynamics of evolution occurs through the coupled system (compartmental model of the SIR type with the interaction of multiple populations and without vital dynamics in the $i \in \{1, \dots, 68\}$ and $j \in \{1, \dots, 83\}$ sites) given by

$$\begin{aligned} \dot{S}_{i,j}(t) &= -S_{i,j}(t) \left(\beta_{i,j}(t)I_{i,j}(t) + \sum_{(\hat{i}, \hat{j}) \in V_{i,j}} \beta_{\hat{i}, \hat{j}}(t)I_{\hat{i}, \hat{j}}(t) \right) \\ \dot{I}_{i,j}(t) &= S_{i,j}(t) \left(\beta_{i,j}(t)I_{i,j}(t) + \sum_{(\hat{i}, \hat{j}) \in V_{i,j}} \beta_{\hat{i}, \hat{j}}(t)I_{\hat{i}, \hat{j}}(t) \right) - \gamma_{i,j}(t)I_{i,j}(t) \\ \dot{R}_{i,j}(t) &= \gamma_{i,j}(t)I_{i,j}(t), \end{aligned}$$

The initial conditions for the model are obtained at each site (i, j) from the fraction of infected reported cases $I_{i,j}(t_0)$ (at the instant t_0), given by the relationship

$$S_{i,j}(t_0) = 1 - I_{i,j}(t_0) - R_{i,j}(t_0), I_{i,j}(t_0) \geq 0, R_{i,j}(t_0) \geq 0, \tag{3.1}$$

Furthermore, because the model does not account for reinfection, $R_{i,j}(t_0)$ is the accumulated proportion of individuals removed during the entire course of the pandemic up to time t_0 . Thus, $R_{i,j}(t_0)$ represents the population density in each (i, j) site that is not yet infected.

Next, we make some observations regarding the dynamics (3.1)-(3.1), where we take the opportunity to describe the parameters and the notation used. The dynamic (3.1) should be interpreted as follows: the probability of infection is due to contact between susceptible and infected individuals from the same site and is proportional to the effective contact rate $\beta_{i,j}(t)$, or by contact with infected individuals from neighboring sites, $V_{\hat{i}, \hat{j}}$ to $\hat{i} \in \{1, \dots, 68, : \hat{i} \neq i\}$ and $\hat{j} \in \{1, \dots, 83, : \hat{j} \neq j\}$, which is proportional to the rate $\beta_{\hat{i}, \hat{j}}(t)$. The rate of removals (recovered or dead) is given by $\gamma_{i,j}(t)$. If $\beta_{\hat{i}, \hat{j}}(t) = 0$ for every $\hat{i} \in \{1, \dots, 68, : \hat{i} \neq i\}$ and $\hat{j} \in \{1, \dots, 83, : \hat{j} \neq j\}$, then (3.1) reduces to the SIR model for the isolated population at each site, e.g., [14] and references therein. The central question is how the disease spreads through the network given the interaction with the neighborhoods $V_{\hat{i}, \hat{j}}$ and, in this case, the respective $\beta_{\hat{i}, \hat{j}}(t) \neq 0$. As a result, we will assume that the interaction occurs in an infinitely small time-step relative to the time of the dynamics t , with no migration. In other words, individuals from different populations interact and return to their reference sites faster than the time $t \rightarrow t + \Delta t$.

We ended this section by showing the well-posedness of the initial value problem (3.1)-(3.1).

Theorem 3.1. *Let $T > t_0 \geq 0$ be given. Assume the model parameters be continuous in $[t_0, T]$. Then, the model (3.1)-(3.1) has a unique positive solution in $[t_0, T]$. This means that its state variables remain non-negative for any trajectory initialized at non-negative initial conditions. Moreover, the solution is continuous with respect to the initial data and the model parameters in the interval $[t_0, T]$.*

Proof. The continuity of the parameters implies the continuity of the right hand side with respect to t . Furthermore, a direct calculation shows that all the coordinates of the Jacobian matrix on the right hand side of (3.1) are continuous. As a result, uniformly bounded in $[t_0, T]$. Then, the

mean value theorem implies that the right hand site of (3.1) is Lipschitz continuous with respect to the second variable. Hence, the assertions adhere to the general theory of the well-posedness of initial value problems, for example, e.g. [16]. \square

The results in Theorem 3.1 may possibly have an extension to $[t_0, \infty[$ (although we do not have a proof for it). Since we are interested in short-running forecasted scenarios of the dynamics (3.1)-(3.1), we restrict ourselves to finite time intervals in Theorem 3.1.

4 ON AN AUGMENTED LAGRANGIAN MAXIMUM A POSTERIORI (AL-MAP) STRATEGY FOR CALIBRATING THE MODEL PARAMETERS

There are three main unknown parameters to be calibrated in the model (3.1). Indeed, the contagious rate within the population $\beta_{i,j}$, the contagious rates for interacting populations $\beta_{i,j}^{\hat{}}$ and the removing rate $\gamma_{i,j}$.

It is worth mentioning that parameter calibration for the model (3.1) is an inverse problem, e.g., [7] and references therein.

Since daily reported of COVID-19 infected cases are sparse and subject to many uncertainties, mainly the sub-notifications during the weekends, we use the weekly cumulative of new COVID-19 reported cases $\mathbf{Y}_{i,j} = \{\mathbf{I}_{i,j}^{t_0}, \dots, \mathbf{I}_{i,j}^{t_0+k\tau}\}$ as observed data, where τ represents one week in the time scale³, for each of the (i, j) -sites.

In this contribution, we propose calibrating the vector of constant parameter $X_{i,j}^k(t) := (\beta_{i,j}^{t_0+k\tau}(t), \beta_{i,j}^{\hat{t}_0+k\tau}(t), \gamma_{i,j}^{t_0+k\tau}(t)) = (\beta_{i,j}^k, \beta_{i,j}^{\hat{k}}, \gamma_{i,j}^k)$, for $t \in [t_0, t_0 + \tau]$, using a constrained maximum a posteriori probability (MAP) estimate approach [1, 4]. It is given by the following constrained optimization problem

$$\hat{X}_{i,j}^k \in \operatorname{argmax}_X \Pi_{post}^k(X_{i,j}^k | \mathbf{I}_{i,j}^{t_0+k\tau}), \text{ subject to } C_{i,j}^{t_0+k\tau}(X_{i,j}^k) = 0 \tag{4.1}$$

where, $C_{i,j}^{t_0+k\tau}(X_{i,j}^k) = \hat{I}_{i,j}(t_0 + \tau) - \mathbf{I}_{i,j}^{t_0+k\tau}$, for $\hat{I}_{i,j}(t_0 + \tau) := I_{i,j}(\hat{X}_{i,j}^k)(t_0 + \tau)$ representing cumulative infected populations at the site (i, j) given by the unique solution of (3.1), with initial conditions (3.1) the set of values at $t = t_0$, as a function of the parameters $X_{i,j}^k$ at $t = t_0 + \tau$. $\Pi_{post}^k(X_{i,j}^k | \mathbf{I}_{i,j}^{t_0+k\tau})$ is the log-posterior density functional given by

$$\Pi_{post}^{t_0+k\tau}(X_{i,j}^k | \mathbf{I}_{i,j}^{t_0+k\tau}) \propto \Pi^k(\mathbf{I}_{i,j}^{t_0+k\tau} | X_{i,j}^k) + \Pi_{prior}^k(X_{i,j}^k | X_{i,j}^{k-1}), \tag{4.2}$$

for the log-likelihood density

$$\Pi^k(\mathbf{I}_{i,j}^{t_0+k\tau} | X_{i,j}^k) \propto \mathbf{I}_{i,j}^{t_0+k\tau} \log(\sigma \hat{I}_{i,j}^{t_0+k\tau}) - \sigma \hat{I}_{i,j}^{t_0+k\tau} - \log(\mathbf{I}_{i,j}^{t_0+k\tau}!). \tag{4.3}$$

³In the numerical simulations presented in this contribution, we use the data of five consecutive weeks. Hence, $k = 4$. In particular, in the simulations, where we use the $\mathbf{I}_{i,j}^{t_0}$ as the initial conditions and $\mathbf{I}_{i,j}^{t_0+\tau}$ for the first step in the parameter calibration and $\mathbf{I}_{i,j}^{t_0+2\tau}$ as the initial conditions and $\mathbf{I}_{i,j}^{t_0+3\tau}$ for the parameter re-calibration.

$\Pi^{t_0+k\tau}(\mathbf{I}_{i,j}^{t_0+k\tau}|X_{i,j}^k)$ is the logarithm of the Poisson distribution with mean given by $\sigma \hat{I}_{i,j}^{t_0+k\tau} := \sigma I_{i,j}(\hat{X}_{i,j})(t_0 + k\tau)$ the unique solution of (3.1)-(3.1), for $t = t_0 + k\tau$. The $\log(\mathbf{I}_{i,j}^{t_0+k\tau}!)$ is approximated by the Stirling’s formula:

$$\log(\mathbf{I}_{i,j}^{t_0+k\tau}!) \sim (1/2) \log(2\pi \mathbf{I}_{i,j}^{t_0+k\tau}) + \mathbf{I}_{i,j}^{t_0+k\tau} \log(\mathbf{I}_{i,j}^{t_0+k\tau}) - \mathbf{I}_{i,j}^{t_0+k\tau},$$

in the data time-series $\mathbf{Y}_{i,j}$.

Moreover, we use the Gaussian priors in the posterior densities

$$\Pi_{prior}^{t_0+k\tau}(X_{i,j}^k|X_{i,j}^{k-1}) \propto -\alpha \|X_{i,j}^k - X_{i,j}^{k-1}\|^2 \tag{4.4}$$

as the regularization term that accounts for the uncertainties in the data $\mathbf{I}_{i,j}^{t_0+k\tau}$, with $\alpha > 0$ the regularization parameter [7]. It is worth mentioning that the (MAP) is a log-concave estimate of an unknown quantity that equals the mode of the posterior distribution, resulting in a point estimate of an unobserved quantity based on empirical data. (MAP) estimations are closely related to the method of maximum likelihood (ML) estimation. However, (MAP) employs an augmented optimization objective that incorporates a prior distribution (that quantifies the additional information available through prior knowledge of a related event) over the quantity one wants to estimate. Hence, MAP estimation can be seen as a regularization of maximum likelihood estimation. In particular, the (MAP) estimation proposed in (4.1) can be interpreted as an iterated-Tikhonov regularized solution, e.g., [5, Section 2.2] or [8].

Due to the constraints in the optimization problem (4.1), the (MAP) estimation cannot guarantee that any maximizers of $\Pi_{post}^{t_0+k\tau}(X_{i,j}^k|\mathbf{I}_{i,j}^{t_0+k\tau})$ satisfies the constraints. For that fate, we propose the augmented Lagrangian (AL-MAP). It consists in the introduction of the augmented Lagrange functional L , which is formally defined by⁴

$$L_{i,j}^k(X_{i,j}^k; \lambda_{i,j}^k, \mu_{i,j}^k) := -\Pi_{post}^{t_0+k\tau}(X_{i,j}^k|\mathbf{I}_{i,j}^{t_0+k\tau}) + \langle \lambda_{i,j}^k, C_{i,j}^{t_0+k\tau} \rangle + \mu_{i,j}^k \|C_{i,j}^{t_0+k\tau}\|^2. \tag{4.5}$$

The Augmented Lagrangian replaces a constrained optimization problem by a series of unconstrained problems through the so-called duality theory [3]. In (4.5), $X_{i,j}^k$ is the primal variable ($\lambda_{i,j}^k, \mu_{i,j}^k$) are the dual variables. In particular, $\lambda_{i,j}^k$ can be interpreted as a “generalized” Lagrange multiplier and the scalar $\mu_{i,j}^k > 0$ is a penalty factor that allows one to establish a duality relation for problems of non-convex (concave) type as (4.1).

Next, we present a framework from which an algorithm can be derived for the constrained optimization problem (4.1), in terms of the proposed (AL-MAP) approach. The main ideas and necessary definitions can be found in [17, Cap. 11, Sec. K^*] or in [3, Section 4.4].

Theorem 4.1. *Let the general assumption on this contributions holds true. Then:*

- i) *There exists a pair $(\bar{\lambda}_{i,j}^k, \bar{\mu}_{i,j}^k)$ supporting an exact penalty representation as in Definition 11.60 [17].*

⁴Remember that $\max G = \min -G$, whenever a solution for the problem exist.

ii) The optimal solutions of the primal and augmented dual problems are characterized as saddle points of the augmented Lagrangian. In other words

$$\begin{aligned} \hat{X}_{i,j}^k &= \inf_X L_{i,j}^k(X_{i,j}^k; \bar{\lambda}_{i,j}^k, \bar{\mu}_{i,j}^k), \\ \bar{\lambda}_{i,j}^k &= \sup_{\lambda} L_{i,j}^k(\hat{X}_{i,j}^k; \lambda_{i,j}^k, \mu_{i,j}^k), \text{ for any } \mu_{i,j}^k \geq \bar{\mu}_{i,j}^k. \end{aligned} \tag{4.6}$$

iii) Let $(X_{i,j}^{k,0}, \lambda_{i,j}^{k,0}, \mu_{i,j}^{k,0})$ and appropriated initial guess given, with $\mu_{i,j}^{k,0}$ large enough. Then, the sequence $(X_{i,j}^{k,l}, \lambda_{i,j}^{k,l}, \mu_{i,j}^{k,l})$ generated as

$$\begin{aligned} X_{i,j}^{k,l} &= \inf_X L_{i,j}^{k,l}(X_{i,j}^k; \lambda_{i,j}^{k,l-1}, \mu_{i,j}^{k,l-1}), \\ \mu_{i,j}^{k,l} &= (1 + \xi^l) \mu_{i,j}^{k,l-1} \\ \lambda_{i,j}^{k,l} &= \lambda_{i,j}^k + \mu_{i,j}^{k,l} C_{i,j}^k(X_{i,j}^{k,l}), \end{aligned} \tag{4.7}$$

for positive of ξ^l , converges towards the saddle point of the augmented Lagrangian $L_{i,j}^k(X_{i,j}^k; \lambda_{i,j}^k, \mu_{i,j}^k)$.

Proof. Notice that augmented Lagrangian (4.5) is a proximal Lagrangian in the sense of [17, Exemple 11.57], for a convex function $-\Pi_{post}^{t_0+k\tau}(X_{i,j}^k | \mathbf{I}_{i,j}^{t_0+k\tau})$. Hence, the assertions on items i) and ii) follows from [17, Theorem 11.61] and [17, Theorem 11.59], respectively. Item iii) follows from the theory developed in [3, Section 4.4]. □

5 COVID-19 FORECASTED SCENARIOS FOR RS

In this section, we presented the COVID-19 infected population forecasted scenarios for RS, in one of the hostile phases of the pandemic, namely, between March and May of 2021, in which: a) - called average scenario - in which the forecasts are driven by the dynamics (3.1)-(3.1) with the parameters $\hat{X}_{i,j}^k$ calibrated from the (AL-MAP) algorithm (4.6) and b) based on changes in the parameters of the situation a), we determine milder scenarios (or optimistic) and more severe (or pessimistic) scenarios by decreasing or increasing the contagion rate $\hat{\beta}_{i,j}^k$ calibrated in a). These last two scenarios emulate low or high population adherence to control measures of COVID-19 dissemination.

The presented scenarios are obtained by following the steps below. The numerical routines are implemented by the authors in Octave using the *lsqnonlin* routine for the nonlinear minimization of the augmented Lagrangian.

- 1) Select date t_0 for the database (as in Section 2). Let the initial conditions (3.1) be given at t_0 . It corresponds to $k = 0$ in the corresponding data set $\mathbf{Y}_{i,j}$. Let $T > t_0$, where T is final time in which the predictive dynamics is considered. In our simulated scenarios T corresponds to 28 days.

- 2) Calibrate the model parameters with the algorithm (4.6), and then treat them as constant for $t_0 \leq t \leq T$.
- 3) While $t_0 \leq t \leq T$, the forecasts follow the discrete-time (3.1) dynamics corresponding to Euler's method, with step-size $h = 1/2$.
- 4) The forecast is assumed to follow the dynamic (3.1) as long as $t \leq T$ or until a new base (and a new t_0 , for example $t_0 \leftarrow t_0 + 2\tau$) becomes available. In any case, go back to item 1).

Then, the forecasted scenarios are georeferenced to the corresponding A_l health clusters in a process that is, basically, inverse to the data acquisition presented in Section 2. In the end, five distinct layers (in color scales) are generated in the health clusters map, representing, respectively, the number of infected people reported in the analyzed week and the number of infected predicted for the next 7, 14, 21 and 28 days, for each of the previously announced scenarios. These results are made available in a WebSIG (interactive and georeferenced format) on the project's website⁵.

5.1 Forecasted example

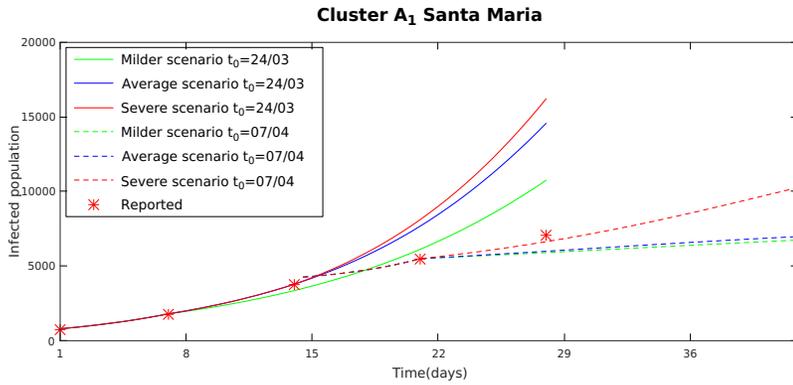
In this subsection, we present the results of predictions for the infected population by COVID-19 with our proposed strategy. It differs from the ones available on the project's website⁴, since it appears below as the graphics of the infected population dynamics of the clusters for the health clusters determined by the controlled distancing model proposed by the government of the state of RS. The parameters are calibrated from algorithm (4.6) based on reported COVID-19 infected people at two databases t_0 , in 24/03/21 and 07/04/21, respectively.

We only present the results from the clusters A_1 -Santa Maria, A_2 - Uruguaiana, A_8 - Porto Alegre, A_{15} -Passo Fundo, A_{16} -Pelotas and A_{18} - Caxias do Sul, as shown in Figures 2-3, respectively. Such selections represent clustering with high population density, such as the RS capital Porto Alegre and the industrialized region of Caxias do Sul; medium-populated clustering regions (Passo Fundo, Pelotas, and Santa Maria); and sparsely populated areas, such as Uruguaiana. Hence, it covers the most important population dynamics in the state of RS. It is worth mentioning that the simulations are performed for all the 21 clusters simultaneously.

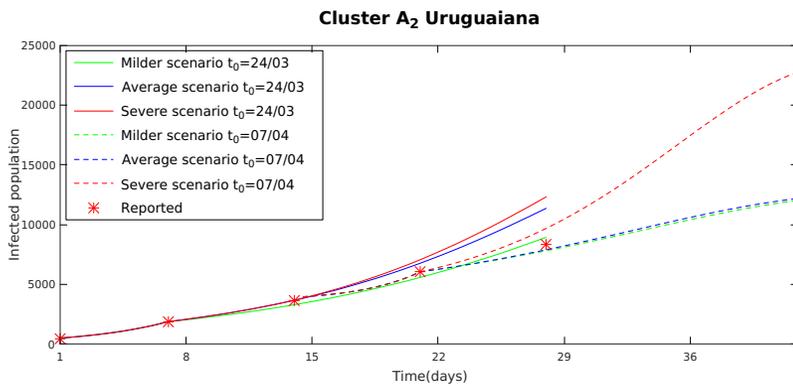
In Table 2, we present the calibrated parameters for the average scenario (the one obtained with the parameters calibrated using the observed data and algorithm (4.6)). The COVID-19 reported cases used as initial conditions and database in simulations are given in the data paper [6].

Figures 2-3 show the predictions of infected population from the simulations obtained from the calibrated parameters shown in Table 2. The continuous lines in Figures 2-3 correspond to the COVID-19 infected people predictions (for the 3 scenarios - (green, blue, red) correspond to (milder, average, severe) scenarios, respectively.) for the dates of 07, 14 and 21/04/21 with the parameters calibrated based on the reported infected cases on $t_0 = 24/03/21$ as initial conditions. The parameters are then re-calibrated based on the number of infected people reported in

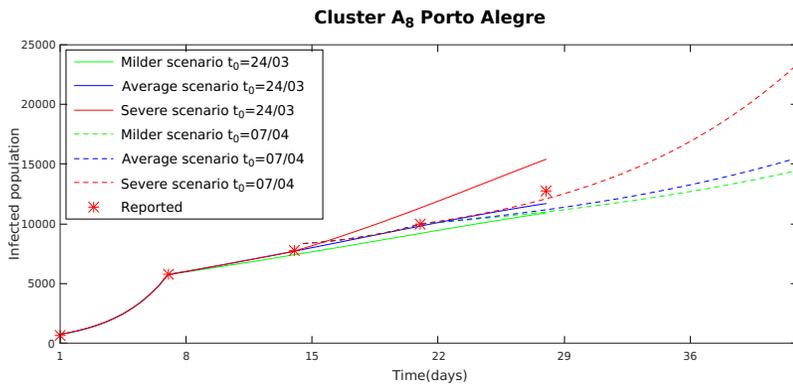
⁵Available in <https://exactum.furg.br/>



(a)

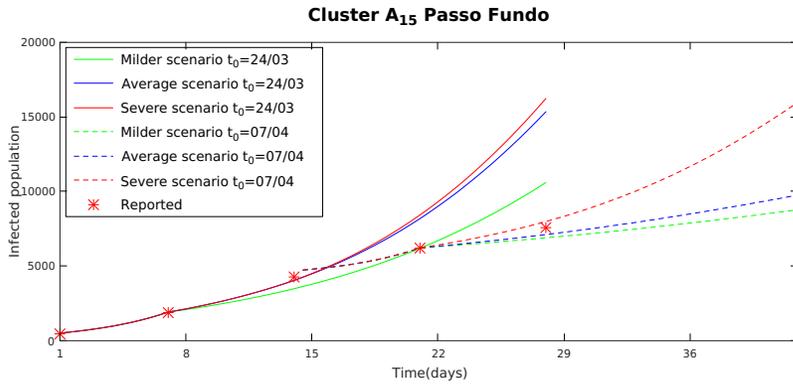


(b)

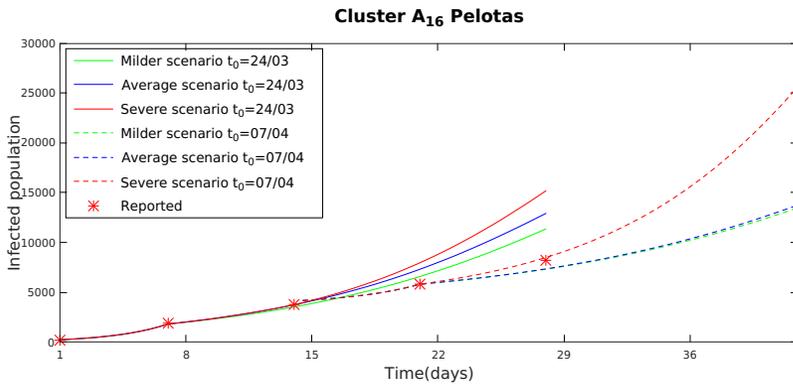


(c)

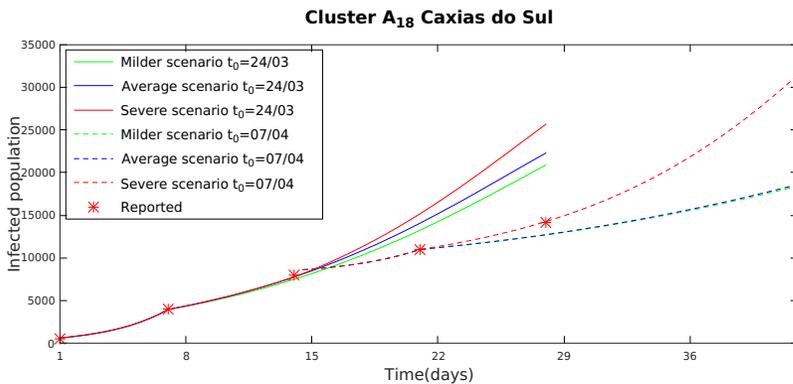
Figure 2: Infected forecast scenarios at each databased for the corresponding health clusters. The average scenario corresponds to the calibrated parameters presented in Table 2.



(a)



(b)



(c)

Figure 3: Infected forecast scenarios at each databased for the corresponding health clusters. The average scenario corresponds to the calibrated parameters presented in Table 2.

the week of the day $t_0 = 07/04/21$. The predictions corresponding to the 3 scenarios (with re-calibrated parameters as in Table 2 corresponding to $t_0 = 07/04/21$) are presented as the dashed curves in Figures 2- 3. It corresponds to predicted infected cases of COVID-19, for the dates 21/04, 28/04 and 5/05. The relative error $ER_j\%$, for the first two weeks after calibration, is presented in the Table 2.

Note that when the first prediction completes 14 days (continuous lines in 07/04), the errors of the results of the first base start to increase and the curves of the three scenarios deviate from the reported data. However, with the strategy of re-calibrating the parameters and setting new predictions every two weeks, when the “old” predictions start to be less reliable, we already have them replaced with the predictions based on the new database. Despite the many external factors that are not taken into account in the proposed COVID-19 modeling, the presented scenarios show a significant agreement with the reported cases for short time periods (around 15 days) - see the errors in Table 2). In particular, COVID-19 reported cases are enclosed by the forecasted scenarios envelope, in the time horizon of 15 days.

One final observation regarding the results presented in Figures 2-3 is that, for some clusters, the milder scenario describes the data for the long time horizon better than the one predicted by the average scenario (the one corresponding to the calibrated parameters - see last column in Table 2). It is because data reported from the end of March/21 to the beginning of May/21 corresponds to the period in which RS went through the most severe restriction (black flag) of economic activity since the beginning of the COVID-19 pandemic.

6 CONCLUSIONS AND FUTURE DIRECTIONS

In this contribution we present the results of the development of a predictive tool with the capability to help manage the evolution of the COVID-19 pandemic in the state of RS, whose interactive form can be found on the EXACTUM website. This tool is the result of the analysis of georeferenced data supported by the mathematical modeling dynamics of the SIR-type model with a spatial structure and interaction of multiple populations (model 3.1), corresponding to the controlled distance clusters, thus encompassing a forecast for the state as a whole.

We propose an augmented Lagrangian maximum a posteriori estimation method for calibrating the model parameters, based on data from people infected with COVID-19 reported by health authorities. The obtained forecasts provide a good description of the evolution of the reported cases of newly infected (see Table 2 and Figures 2-3) for a short time horizon (15 days). Therefore, we suggest calibrating the parameters in a time window that corresponds to a maximum of two weeks. With this strategy, we verify that the number of infected for future dates remains enveloped by the predictions obtained by the three proposed scenarios.

Future developments of this approach include the possibility of extending the results to all Brazilian states. The difficulties of this extension lie in the need for a higher computing capacity than that used by the team in this project, where the computation was done on a personal computer.

Moreover, there is an urgent need for standardization in the way the reported data is handled by the health authorities of each state or by the federal ministry of health.

Finally, it is important to point out that the inclusion of other relevant information in the modeling, such as ICU bed occupancy forecast and, more recently, vaccination, is easily incorporated into the dynamics 3.1. In this way, extensions in this sense are also possible.

Acknowledgments

We thank Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS), Brazil, Grant No. 20/2551-0000272-1. D.A. Sales acknowledges the support of CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brazil.

REFERENCES

- [1] V.V.L. Albani, R.M. Velho & J.P. Zubelli. Estimating, monitoring, and forecasting COVID-19 epidemics: a spatiotemporal approach applied to NYC data. *Sci Rep*, **11** (2021), 9089.
- [2] M.d.G. Albuquerque, J.R.d. Santos, A.F. Silveira, D.L. Bornia Junior, R.B. Nunes, T.B.R. Gandra, A.L.C.d. Silva, G.B.d. Miranda & A.R. Trevisol. Influence of socio-economic indicators and territorial networks at the spatiotemporal spread dynamics of Covid-19 in Brazil. *Sociedade & Natureza*, **33** (2021). doi:10.14393/SN-v33-2021-59688. URL <https://seer.ufu.br/index.php/sociedadennatureza/article/view/59688>.
- [3] D. Bertsekas. “Constrained optimization and Lagrange multiplier methods”. Computer Science and Applied Mathematics. Academic Press, New York (1982).
- [4] G. Chowell. Fitting dynamic models to epidemic outbreaks with quantified uncertainty: a primer for parameter uncertainty, identifiability, and forecasts. *Infectious Disease Modelling*, **2**(3) (2017), 379–398.
- [5] A. De Cezaro. “On a Parabolic Inverse Problem Arising in Quantitative Finance: Convex and Iterative Regularization”. PhD Theses- IMPA (2010).
- [6] A. De Cezaro & et. all. Georeferenced data from the COVID-19 pandemic for the state of Rio Grande do Sul: Reported positive cases and simulated forecasts. *Latin American Data in Science*, **1**(2) (2021), 49–62.
- [7] A. De Cezaro, A. Marinho de Oliveira & F. Travessini De Cezaro. “Identificação de parâmetros em Equações Diferenciais: Teoria e Aplicações”. UFPI (2012).
- [8] A. De Cezaro, O. Scherzer & J. Zubelli. Convex regularization of local volatility models from option prices: Convergence analysis and rates. *Nonlinear Analysis: Theory, Methods & Applications*, **75**(4) (2012), 2398–2415.
- [9] D.P. de Paula, D.H. Miranda de Medeiros, E. Lacerda Barros, R.G. Pinheiro Guerra, J.d.O. Santos, J.S. Queiroz Lima & R.M. Leite Monteiro. Diffusion of Covid-19 in the Northern Metropolis in Northeast Brazil: territorial dynamics and risks associated with Social Vulnerability. *Sociedade &*

- Natureza*, **32** (2020), 639–656. doi:10.14393/SN-v32-2020-56098. URL <https://seer.ufu.br/index.php/sociedadennatureza/article/view/56098>.
- [10] R. Ferreira. Atlas, Cibercartografia e Neogeografia. *Revista Portuguesa de Estudos Regionais*, (2016), 31–44.
- [11] J. Govaert. Coronavirus: UK changes course amid death toll fears. (2020). URL <https://www.bbc.com/news/health-51915302>.
- [12] H.W. Hethcote. The mathematics of infectious diseases. *SIAM review*, **42**(4) (2000), 599–653.
- [13] W. Kermack & A. Mckendrick. A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society of London. Series A, Containing papers of a Mathematical and Physical Character*, **115**(772) (1927), 700–721.
- [14] M.J. Lazo & A. De Cezaro. Why can we observe a plateau even in an out of control epidemic outbreak? A SEIR model with the interaction of n distinct populations for Covid-19 in Brazil (2020).
- [15] P. Longley. Geographical Information Systems: a renaissance of geodemographics for public service delivery. *Progress in Human Geography*, **29**(1) (2005), 57–63. doi:10.1191/0309132505ph528pr. URL <https://doi.org/10.1191/0309132505ph528pr>.
- [16] L. Perko. “Differential Equations and Dynamical Systems”. Springer (2001).
- [17] R. Rockafellar & R.B. Wets. “Variational analysis”, volume 317 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin (1998), xiv+733 p.
- [18] L. Rodrigues, M. Varriale, W. Godoy & D. Mistro. “Coupled Map Lattice Model for Insects and Spreadable Substances” (2014), 141-169 p. doi:10.1007/978-3-319-06877-0_7.
- [19] World Health Organization. WHO Coronavirus Disease (COVID-19) Dashboard (2020). URL <https://covid19.who.int.2020/7/29>.

