

Dinâmica da COVID-19 utilizando dados georreferenciados para o Estado do Rio Grande do Sul

Joice Chaves Marques¹
 IMEF/FURG, Rio Grande, RS
 Deivid Cristian Leal Alves²
 IMEF/FURG, Rio Grande, RS
 Paulo Victor de Araújo Brito Lisboa³
 IMEF/FURG, Rio Grande, RS
 Adriano De Cezaro⁴
 IMEF/FURG, Rio Grande, RS
 Matheus Lazo⁵
 IMEF/FURG, Rio Grande, RS
 Dinalva Aires de Sales⁶
 IMEF/FURG, Rio Grande, RS

Resumo Nesta contribuição apresentamos uma ferramenta preditiva desenvolvida para auxiliar no gerenciamento da evolução da pandemia de COVID-19 no estado do Rio Grande do Sul (RS). Resultado da análise de dados georreferenciados suportados pela dinâmica do modelo do tipo SIR (Suscetíveis, Infectados e Recuperados ou Removidos) definido em uma estrutura espacial e com interação de múltiplas populações que correspondem aos agrupamentos do distanciamento controlado (A_j , para $l = 1, \dots, 21$) propostos pelas autoridades governamentais e de saúde pública do RS. Os resultados preditivos, atualizados quinzenalmente, fornecem três cenários distintos (otimista, realista e pessimista), para um horizonte mensal e são disponibilizados na forma de WebSIGs (Sistema de Informação Geográfico on line) em [4]. Os cenários otimista e pessimista são obtidos a partir do cenário médio, com mudanças nas taxas de contágio de cada um dos A_l agrupamentos, que são calibradas a partir dos dados reportados do número de infectados disponibilizados pela Secretaria de Saúde (SES/RS). As previsões da modelagem mostram-se bastante satisfatórias para uma janela de tempo quinzenal (as primeiras duas semanas), cujos erros variam no intervalo de 0% á 5,13% dos infectados reportados em cada um dos agrupamentos A_l , aumentando gradativamente para um horizonte de tempo maior, chegando á 52% ao final dos 28 dias.

Palavras-chave. SARS-CoV-2, Modelo tipo SIR estruturado, Múltiplas populações interagindo, WebSIGs, Projeção

1 Introdução

Desde o surgimento dos primeiros casos de COVID-19 é notável a dedicação da comunidade científica no sentido de compreender a dinâmica e propagação do SARS-CoV-2 (vírus que causa a

¹joicec.marques@hotmail.com

²dclealalves@gmail.com

³paulovictor_f_jv@hotmail.com

⁴adrianocezaro@furg.br

⁵matheuslazo@furg.br

⁶dsales@furg.br

COVID-19 [8]). Nesse sentido, a procura por medicamentos e o rápido desenvolvimento de vacinas tem sido a tônica de grande parte dos pesquisadores. Enquanto os medicamentos conhecidos e testados mostraram-se com pouco (ou nenhuma eficácia) e a vacinação em massa de toda a população mundial é algo impensável no curto prazo, haja visto o tempo necessário para a fabricação, distribuição e aplicação da mesmas na população, restaram as medidas não-farmacológicas como única alternativa para a contenção dos efeitos da pandemia. Desde então, a modelagem matemática tem sido uma ferramenta indispensável para formular as previsões e auxiliar órgãos públicos na tomada de decisões em relação a medidas que evitem a rápida propagação do vírus [5,9]. Por outro lado, também ficou claro a que ciência não é o suficiente para combater os efeitos da pandemia, sem uma forma eficiente de disseminação das informações relevantes.

O presente estudo visa contribuir para a compreensão, previsão e servir de suporte para tomadas de decisões relativas a dinâmica e propagação da pandemia de COVID-19 para o estado do RS, baseado na divisão proposta para o distanciamento controlado (A_j , para $j = 1, \dots, 21$).

O trabalho consiste em três etapas encadeadas da seguinte forma: **A)** Organização e tratamento dos dados reportados de pessoas infectadas, de forma georreferenciadas, que servem como entrada na modelagem matemática (condições iniciais para o modelo do tipo SIR proposto na Seção 3, como detalhado na Seção 2; **B)** Modelagem matemática da dinâmica da COVID-19 por um modelo tipo SIR com interação de múltiplas populações, como proposto na Seção 3; **C)** Organização de cenários da disseminação da COVID-19 em formato de WebSIGs como descrito na Seção 4. Conforme detalhadas a seguir.

2 Organização e tratamento dos dados

A construção do plano geográfico de informação compreendeu apenas dados de fontes oficiais. Foi admitida como área territorial municipal a base vetorial 2019 no formato *shapefile* disponibilizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE). A regionalização da modelagem, na forma de blocos regionais ou agrupamentos, teve como referência as informações fornecidas pelo portal de Planejamento, Orçamento e Gestão de Saúde do estado do Rio Grande do Sul que foram divididas em vinte e um agrupamentos previstos pelo Modelo de distanciamento controlado do Estado [2].

3 Modelagem matemática

Na estrutura espacial do modelo SIR proposto nesse trabalho, consideramos uma rede bidimensional de tamanho correspondente ao recebido da etapa de georreferenciamento dos dados descritos na Seção 2. Rede, esta, composta de 68×83 sítios (nomenclatura associada a cada célula (pixel)). A espacialização do modelo está embasada na teoria de Rede de Mapas Acoplados (RMA) em que cada uma das regiões (sítios) $\mathbf{x} = (x_i, x_j)$ desta rede é representada por coordenadas inteiras (com $x_i = is$, $x_j = js$, onde s é o tamanho do sítio), para $i \in \{1, \dots, 63\}$, $j \in \{1, \dots, 83\}$. Essa rede é organizada de forma que cada agrupamento seja composto por determinado número de sítios estabelecido na etapa de organização dos dados descrita na seção 2. Cabe salientar que esse número depende da precisão adotada no processo de rasterização e da área do agrupamento. Em cada sítio (i, j) tem-se, para $i \in \{1, \dots, 68\}$ e $j \in \{1, \dots, 83\}$, a população total $N_{i,j}$ normalizada e assumida constante, de forma que a total na rede satisfaz $\sum_{i=1}^{68} \sum_{j=1}^{83} N_{i,j} = 1$. Ainda, o modelo é considerado como um modelo compartimental do tipo SIR, onde a população total de cada sítio $N_{i,j}$ é dividida nas frações de Suscetíveis $S_{i,j,t}$, Infectados ou Infeciosos $I_{i,j,t}$ e Removidos ou Recuperados $R_{i,j,t}$, no instante de tempo t .

Assumimos que a dinâmica de evolução se dá pelo sistema acoplado (modelo compartimental do tipo SIR com interação de múltiplas populações e sem dinâmica vital) dada da seguinte forma

$$\begin{aligned} \dot{S}_{i,j,t} &= -S_{i,j,t} \left(\beta_{i,j,t} I_{i,j,t} + \sum_{(\hat{i}, \hat{j}) \in V_{i,j}} \beta_{\hat{i}, \hat{j}, t} I_{\hat{i}, \hat{j}, t} \right) \\ \dot{I}_{i,j,t} &= S_{i,j,t} \left(\beta_{i,j,t} I_{i,j,t} + \sum_{(\hat{i}, \hat{j}) \in V_{i,j}} \beta_{\hat{i}, \hat{j}, t} I_{\hat{i}, \hat{j}, t} \right) - \gamma_{i,j,t} I_{i,j,t} \\ \dot{R}_{i,j,t} &= \gamma_{i,j,t} I_{i,j,t}. \end{aligned} \tag{1}$$

- 1) Escolhida a data t_0 para a base de dados (como na Seção 2) e dado $T > t_0$, onde T tempo final em que a dinâmica preditiva é considerada:
- 1.1) As condições iniciais para o modelo, em cada sítio (i, j) , são obtidas a partir da fração de casos reportados de infectados ou infecciosos I_{i,j,t_0} (no instante t_0 em que a base de dados é considerada), dadas pela relação

$$S_{i,j,t_0} = 1 - I_{i,j,t_0} - R_{i,j,t_0}, I_{i,j,t_0} \geq 0, R_{i,j,t_0} > 0, \tag{2}$$

onde R_{i,j,t_0} é a proporção acumulada de indivíduos removidos (recuperados ou mortos) durante todo o percurso da pandemia até o tempo t_0 , uma vez que o modelo não considera reinfeção. Dessa forma, S_{i,j,t_0} corresponde a densidade populacional ainda não infectada, armazenada em cada célula do agrupamento, como descrito na Seção 2.

- 2) Os parâmetros do modelo são ajustados (e considerados contantes para $t_0 \leq t \leq T$) com base nos dados da semana t_0 e na primeira semana posterior, como descrito na Seção 4.
- 3) Enquanto $t_0 \leq t \leq T$, as previsões seguem a dinâmica (1), a tempo-discreto obtida pelo método de Euler, dada por

$$\begin{aligned} S_{i,j,t+1} &= S_{i,j,t} - \left(S_{i,j,t} \left(\beta_{i,j,t} I_{i,j,t} + \sum_{(\hat{i}, \hat{j}) \in V_{i,j}} \beta_{\hat{i}, \hat{j}, t} I_{\hat{i}, \hat{j}, t} \right) \right) h, \\ I_{i,j,t+1} &= I_{i,j,t} + \left(S_{i,j,t} \left(\beta_{i,j,t} I_{i,j,t} + \sum_{(\hat{i}, \hat{j}) \in V_{i,j}} \beta_{\hat{i}, \hat{j}, t} I_{\hat{i}, \hat{j}, t} \right) - \gamma_{i,j,t} I_{i,j,t} \right) h \\ R_{i,j,t+1} &= R_{i,j,t} + \gamma_{i,j,t} I_{i,j,t} h, \end{aligned} \tag{3}$$

onde h é o tamanho do passo considerado no método numérico.

- 4) A previsão é mantida de acordo com a dinâmica (3) enquanto $t \leq T$ ou até que uma nova base (e um novo t_0) está disponível. Em qualquer caso, volta para o item 1).

A seguir tecemos algumas observações com relação a dinâmica (3), onde aproveitamos para descrever os parâmetros e a notação utilizada. A dinâmica (3) deve ser interpretada da seguinte maneira: a probabilidade de infecção se dá pelo contato entre indivíduos suscetíveis e infecciosos do mesmo sítio e é proporcional a taxa de contato efetivo $\beta_{i,j,t}$, ou pelo contato com indivíduos infecciosos dos sítios vizinhos, $V_{\hat{i}, \hat{j}}$ para $\hat{i} \in \{1, \dots, 68, : \hat{i} \neq i\}$ e $\hat{j} \in \{1, \dots, 83, : \hat{j} \neq j\}$, a qual é proporcional a taxa $\beta_{\hat{i}, \hat{j}, t}$. Ainda $\gamma_{i,j,t}$ é a taxa de removidos e, por fim, $\mu_{i,j,t}$ é a taxa de nascimentos e mortes, que assumimos serem iguais. Se $\beta_{\hat{i}, \hat{j}, t} = 0$ para todo $\hat{i} \in \{1, \dots, 68, : \hat{i} \neq i\}$

e $\hat{j} \in \{1, \dots, 83, : \hat{j} \neq j\}$, então (3) equivale a um modelo SIR para a população de cada sítio de maneira isolada, e.g. [6]. Uma questão fundamental é determinar a maneira com que a doença se espalha pela rede, dada a interação com as vizinhanças $V_{i,\hat{j}}$ e, neste caso com os respectivos $\beta_{i,\hat{j},t} \neq 0$. Para tal, assumiremos que a interação ocorre em um tempo infinitamente pequeno comparado ao tempo da dinâmica $t \rightarrow t + 1$, de forma que não exista migração. Em outras palavras, os indivíduos de populações distintas interagem e retornam a seus sítios de referência, no instante de tempo t . Os parâmetros do modelo são considerados de forma que o tempo t corresponda ao número de dias (em particular $\Delta t \cdot h = 1$). As previsões são realizadas para as quatro semanas posteriores a semana inicial da base, no instante de tempo t_0 obtida do banco de dados descritos na Seção 2.

Resultados de boa colocação, existência de pontos críticos e estabilidade, bem como outras características da solução para o modelo (1)-(2) são estabelecidos em [7].

4 Parâmetros, cenários simulados e georreferenciamento

A calibragem de parâmetros para o modelo (3) é um problema inverso (veja [3] e referências). Como os dados reportados estão sujeitos a muita incerteza, a maneira correta de calibrar os parâmetros exige a aplicação de métodos de regularização [3].

Nesta abordagem, assumiremos como situação hipotética que os dados reportados não estão sujeitos a tais incertezas (embora existam), e concentraremos nossa atenção em determinar cenários em que: a) as previsões representem os dados reportados (com base nos parâmetros de taxa de contágio $\beta_{i,j,t}$ que melhor fitam os dados obtidos na Seção 2) - Cenário médio - e b) com base em modificações nos parâmetros da situação a), determinar cenários mais brandos (ou otimista) e mais severos (ou pessimista) que descrevam a disseminação da COVID-19. Obtemos esses cenários através da variação da taxa de contágio $\beta_{i,j,t}$ estabelecida em a). Estes dois últimos cenários podem ser interpretados como uma baixa ou alta adesão da população as medidas de controle a disseminação da COVID-19.

Assumimos ainda que a interação $\beta_{i,\hat{j},t}$ entre populações distintas dentro da rede é proporcional a taxa de contágio dentro do próprio sítio $\beta_{i,j,t}$. Esta escolha é feita com base no fato que proporcionalidade é maior, para as regiões de maior densidade populacional (pois $\beta_{i,j,t}$ também é maior neste caso), como é o caso da região metropolitana de Porto Alegre e Caxias do Sul. Além disso, a taxa de recuperação (ou de remoção) também, claramente, depende da taxa de contágio (uma vez que o período de infecção é finito). Para as simulações que apresentaremos, assumimos que a $\gamma_{i,j,t} = \frac{\beta_{i,j,t}}{\sigma_{i,j}(t)}$ onde $\sigma_{i,j}(t)$ assume um valor calibrado⁷ para a primeira semana, ou seja, para $t_0 \leq t < t_0 + 14$ dias e $\sigma_{i,j}(t)$ muda nas demais semanas ($t_0 + 14 \leq t < T$ dias), resultando em $\gamma_{i,j,t}$ descrito na Tabela 1, para os agrupamentos que apresentaremos neste trabalho.

Por fim $\beta_{i,j,t}$ (para cada sítio) é calibrado como a média dos valores obtidos pelo algoritmo de filtro de partículas proposto em [1], adaptado para o modelo (3), após três passos internos. O espaço amostral inicial é constituído pelo número de infectados reportados e a taxa de contágio $\bar{\beta}_{i,j}$ como um *a priori*. Os parâmetros são recalibrados a cada nova entrada de dados da base, isto é, nos tempos t_0 , e depois consideradas constantes para todo $t_0 \leq t \leq T$, como no item 3).

Os resultados simulados são realocados de forma georreferenciada aos agrupamentos de saúde A_j correspondentes em um processo, basicamente, inverso ao da aquisição dos dados na Seção 2. Ao final geram-se cinco camadas distintas (em escalas de cores) no mapa de agrupamentos de saúde, representando, respectivamente, o número de infectados reportados na data analisada e o número de infectados projetados para os próximos 7, 14, 21 e 28 dias, para cada um dos cenários

⁷Relacionado ao tempo infecção/recuperação, temos que $\gamma_{i,j,t}$ está relacionado com $\beta_{i,j,t} = \beta_{i,j}(t - 14)$. Assim, $\sigma_{i,j}(t)$ é aumentado ou diminuído seguindo esta lógica.

previamente anunciados. Estes são disponibilizados em um formato interativo e georreferenciado no site do projeto acessível em [4].

Tabela 1: Parâmetros calibrados a partir das bases de dados reportados dos dias 24/03/21 e 07/04/21. $ER_1\%$ e $ER_2\%$ correspondem aos Erros Relativos percentuais da simulação com relação aos reportados, da primeira e segunda semana após a data base da calibragem, para o cenário médio.

Modelagem data base	A_j	$\beta_{i,j}$	$\beta_{i,j}$	$\gamma_{i,j}$	$ER_1\%$	$ER_2\%$
24/03/21	A_1	0,2400	0,00160	0,077412	0,37	0,60
	A_2	0,3600	0,00240	0,116129	0,09	1,73
	A_8	0,5860	0,00390	0,189032	0,08	0,15
	A_{15}	0,3780	0,00252	0,121935	0,12	5,13
	A_{16}	0,5459	0,00364	0,176096	2,19	0,86
	A_{18}	0,5290	0,00353	0,170645	0,36	1,77
07/04/21	A_1	0,3650	0,00243	0,117742	0,61	10,45
	A_2	0,5200	0,00437	0,167742	0,32	7,7
	A_8	0,3420	0,00228	0,110322	0,93	20,75
	A_{15}	0,3850	0,00257	0,124193	0,11	2,11
	A_{16}	0,4080	0,00272	0,131613	1,61	33,15
	A_{18}	0,4060	0,00271	0,130968	0,58	17,78

Na Figura 1, apresentamos os resultados das simulações para duas bases de dados usadas para calibrar os parâmetros na modelagens para os agrupamentos de saúde, conforme a Tabela 1. Vale ressaltar que os parâmetros foram calibrados com base no número de infectados reportados na semana do dia 24/03/21 (dia 01 das simulações), de forma a fitarem os dados reportados em 31/03/21. As curvas (correspondentes aos traços contínuos na Figura 1) corresponde as previsões (para os 3 cenários) de acordo com a dinâmica (3), projetados para as datas 07, 14 e 21/04/21. Os parâmetros são, então, recalibrados com base número de infectados reportados na semana do dia 07/04/21 (novo t_0 - dia 01 das simulações apresentadas de forma tracejadas na Figura 1), de forma a fitarem os dados reportados em 14/04/21. As previsões correspondentes aos três cenários (com parâmetros recalibrados) são apresentados como as curvas tracejadas na Figura 1, para as datas de 21/04, 28/04 e 05/05. O erro relativo percentuais $ER_j\%$, para as duas primeiras semanas após a calibragem, são apresentadas nas duas últimas colunas da Tabela 1.

Observemos que quando a primeira modelagem completa 14 dias (em 07/04) a nova base estabelecida para essa data tende a fitar “melhor” os dados futuros (Figura 1). Dessa forma, quando os erros dos resultados da primeira base começam aumentar e as curvas dos três cenários afastam-se dos dados reportados já temos a substituição por resultados provenientes da nova base. No entanto, a temporalidade de 15 dias para a recalibragem pode apresentar erros significativos da precisão das previsões futuras (veja colunas ER_j na Tabela 1), haja visto os inúmeros fatores externos que não são levados em conta na modelagem, como medidas de restrições ou relaxamento de circulação, entre outras. Por fim, é possível concluir que as previsões para todos os cenários apresentados fitam de maneira bastante significativa os dados reportados para a primeira semanas após a calibragem, mas, podem deixar a desejar em termos de precisão para horizontes de tempo maior.

Note ainda que em alguns agrupamentos, as simulações do cenário otimista fitam os dados de longo prazo, melhor que o cenário com dados calibrados (cenário médio). Isto deve-se muito ao fato de que os dados reportados do início de abril (a serem fitados) correspondem ao período em que o RS passou pela mais severa restrição (bandeira preta) das atividades econômicas desde o início da pandemia de COVID-19. Tal fenômeno se apresenta de maneira reversa, para os dados reportados no final do mês de abril, impactados pela reabertura gradual das atividades econômicas.

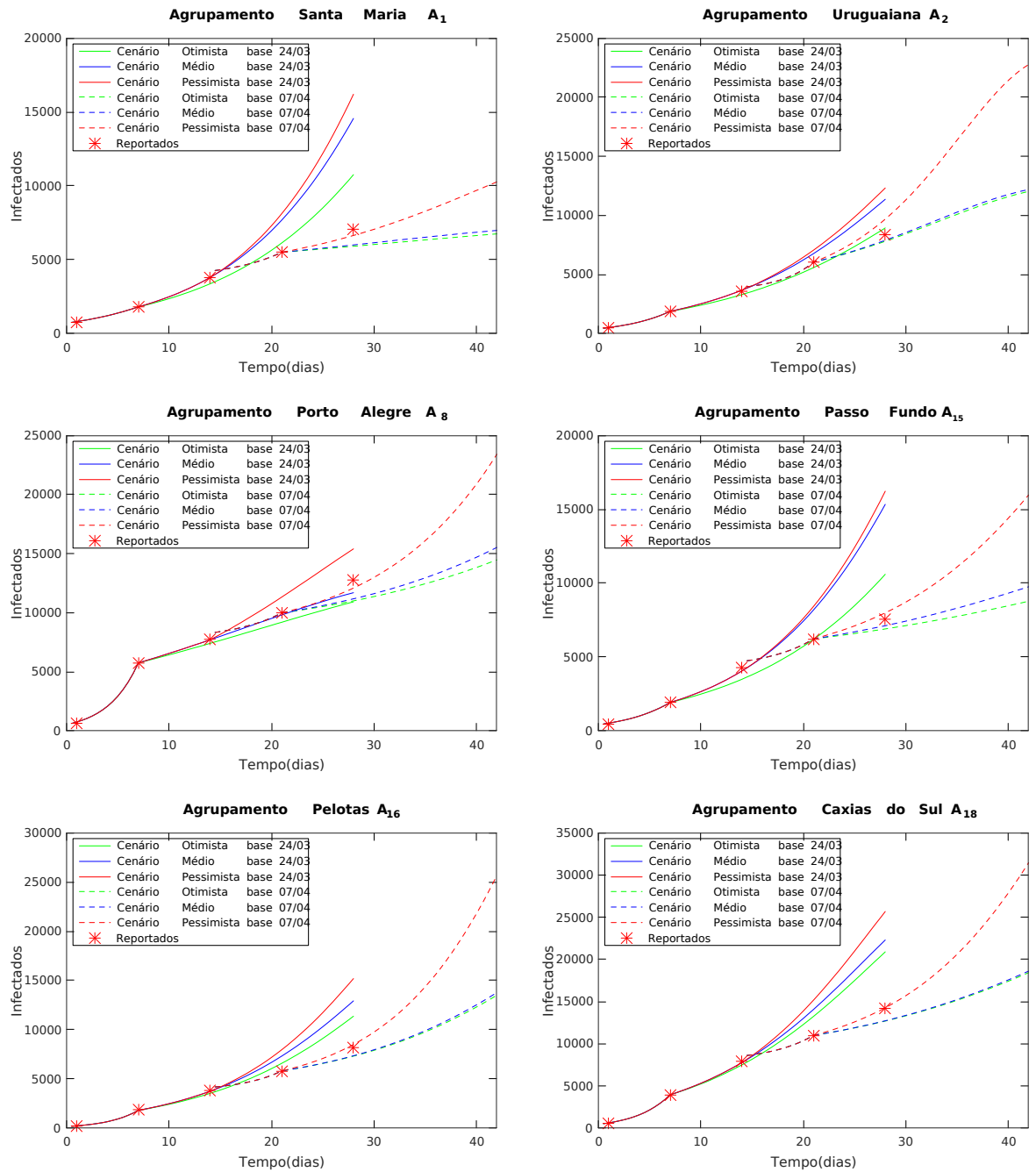


Figura 1: Número de infectados reportados e resultados das simulações (três cenários) para distintos agrupamentos de saúde em duas diferentes bases conforme legenda. Cenário médio corresponde aos parâmetros calibrados apresentados na Tabela 1.

5 Conclusões e aprimoramentos

Nesta contribuição apresentamos os resultados preliminares do desenvolvimento de uma ferramenta preditiva para auxiliar no gerenciamento da evolução da pandemia de COVID-19 no estado

do Rio Grande do Sul (RS), cuja forma interativa pode ser encontrada em [4]. Esta ferramenta é o resultado da análise de dados georreferenciados suportados pela dinâmica do modelo do tipo SIR com uma estrutura espacial e com interação de múltiplas populações (modelo (3)), correspondentes aos agrupamentos do distanciamento controlado, englobando assim uma previsão para o estado como um todo.

As previsões obtidas fornecem boa descrição da evolução dos casos reportados de novos infectados (ver Tabela 1 e Figura 1) para um horizonte de tempo curto (15 dias), mesmo com dados sob a influência de medidas restritivas. Dessa forma, sugerimos a calibragem dos parâmetros em uma janela temporal que corresponde a duas semanas, no máximo.

Ainda estamos melhorando a calibragem dos parâmetros pelo método de filtro de partículas derivado de [1], de forma a considerar a calibragem de paramétrica completa do modelo (3) a partir dos dados reportados e de dados de movimentação entre agrupamentos.

Como desdobramentos futuros da presente abordagem estão a possibilidade de extensão dos resultados para municípios do RS (e para o país como um todo).

Por fim é importante salientar que a inclusão de outras informações relevantes na modelagem, como previsão de ocupação de leitos de UTI são facilmente incorporáveis a dinâmica (3).

Referências

- [1] Calvetti, D. and Hoover, A. and Rose, J. and Somersalo, E. *Bayesian dynamical estimation of the parameters of an SE(A)IR COVID-19 spread model*, preprint, 2020.
- [2] Comitê de dados COVID-19 Coronavírus. *Regiões de saúde e seus municípios: modelo de distanciamento controlado do Rio Grande do Sul (RS)*. **Gov.RS**, 2020. Disponível em: < <https://distanciamentocontrolado.rs.gov.br> >. Acesso em: 01 de nov. de 2020.
- [3] De Cezaro, A. and Marinho de Oliveira, A. and Travessini De Cezaro, F. *Identificação de parâmetros em Equações Diferenciais: Teoria e Aplicações*. UFPI, Piauí, 2012.
- [4] Equipe exactum. *Projeções da dinâmica de infectados..* Disponível em: <<https://exactum.furg.br/mapas-interativos>>.
- [5] Govaert, J. *Coronavirus: UK changes course amid death toll fears..* **BBC NEWS**, 2020. Disponível em: < <https://www.bbc.com/news/health-51915302> >. Acesso em: 17 de mar. de 2020.
- [6] Hethcote, H. W. *The mathematics of infectious diseases*. SIAM review, 42, 4, 599-653, 2000.
- [7] Marques, J. C. and De Cezaro, A. and Lazo, M. *On a emerging plateau in a multi-population SIR model*, preprint, 2021.
- [8] Organization, W.H. *WHO Coronavirus Disease (COVID-19) Dashboard*. **World Health Organization**, 2020. Disponível em: < <https://covid19.who.int> >. Acesso em: 15 de out. de 2020.
- [9] Park, A. *Here's How Scientists and Public-Health Experts Recommend the U.S. Gets Back to 'Normal'*. **TIME**, 2020. Disponível em: < <https://time.com/5829387/coronavirus-reopening-science> >. Acesso em: 30 de ab. de 2020.