On the heterogeneous spread of COVID-19 in Chile

Danton Freire-Flores [a,b,1], Nyna Llanovarced-Kawles [a,b,1], Anamaria Sanchez-Daza [b,c,*], Álvaro Olivera-Nappa [a,b]

[a] Department of Chemical Engineering, Biotechnology, and Materials, Universidad de Chile, Beauchef 851, 8370448 Santiago, Chile
[b] Centre for Biotechnology and Bioengineering, Universidad de Chile, Beauchef 851, 8370448 Santiago, Chile
[c] Institute for Cell Dynamics and Biotechnology, Beauchef 851, 8370456, Santiago, Chile

ABSTRACT

Non-pharmaceutical interventions (NPIs) have played a crucial role in controlling the spread of COVID-19. Nevertheless, NPI efficacy varies enormously between and within countries, mainly because of population and behavioral heterogeneity. In this work, we adapted a multi-group SEIRA model to study the spreading dynamics of COVID-19 in Chile, representing geographically separated regions of the country by different groups. We use national mobilization statistics to estimate the connectivity between regions and data from governmental repositories to obtain COVID-19 spreading and death rates in each region. We then assessed the effectiveness of different NPIs by studying the temporal evolution of the reproduction number R_t. Analysing data-driven and model-based estimates of R_t, we found a strong coupling of different regions, highlighting the necessity of organized and coordinated actions to control the spread of SARS-CoV-2. Finally, we evaluated different scenarios to forecast the evolution of COVID-19 in the most densely populated regions, finding that the early lifting of restriction probably will lead to novel outbreaks.

© 2021 Elsevier Ltd. All rights reserved.

1. Introduction

The Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) is the seventh reported coronavirus that can infect humans [1,2]. As a consequence of the fast global spread and severe effects of the infectious disease caused by SARS-CoV-2, COVID-19 was declared a pandemic by the World Health Organization on 11 March 2020, resulting in 214 countries affected to date [3,4]. After more than one year of living in a pandemic world and despite scientific efforts, effective treatments for COVID-19 are not yet available. Even though several countries are deploying vaccination plans worldwide, uncertainties related to vaccine efficacy and uptake suggest the necessity of keeping non-pharmaceutical interventions (NPIs) in place for preventing excess deaths and the emergence of escape variants [5–7].

Since the pandemic began, researchers have been proposing different epidemiological models to evaluate and forecast the evolution of the disease. A significant part of those models derives from the renowned SIR model proposed by Kermack and McKendrick in 1927 [8], which compartmentalizes the population exposed to the virus in the variables susceptible (S), infected (I) and Recovered (R), whose interaction determines the evolution of the disease over time. Even though helpful, SIR models rely on several hypotheses that are rather hard to meet [9], most of them related to homogeneity [10]. SIR models assume a "perfect mixing", where individuals are equally likely to meet, and from a continuous perspective, fractions of them do all the time. Individuals are assumed to have the same transition rates, translating to the same probability of being infected and the same average time to recover. Furthermore, SIR models do not include a latent period, one of the signature characteristics of COVID-19 [11]. Direct extensions of the SIR model include extra compartments to represent variables of epidemiological interest, as SEIR models, differentiating exposed individuals that are not yet infectious, SIRD, differentiating deaths, and SEIRA models, considering asymptomatic carriers, among others [12–19].

Viral spread depends not only on its biological properties but also on the behavior and susceptibility of the population where it propagates [20]. Therefore, a more realistic analysis of the propagation of the disease in a heterogeneous population usually requires models that include the interaction between the different sub-populations [21,9,22]. These sub-populations interact dynamically with each other, and the spreading modes are not necessarily isotropic [23]. Furthermore, our understanding of viral spread is greatly determined by data availability and quality, which has been proven to have significant delays [24], weekly modulation [25], and even weekend effects [26,27].

D. Freire-Flores, N. Llanovarced-Kawles, A. Sanchez-Daza et al.

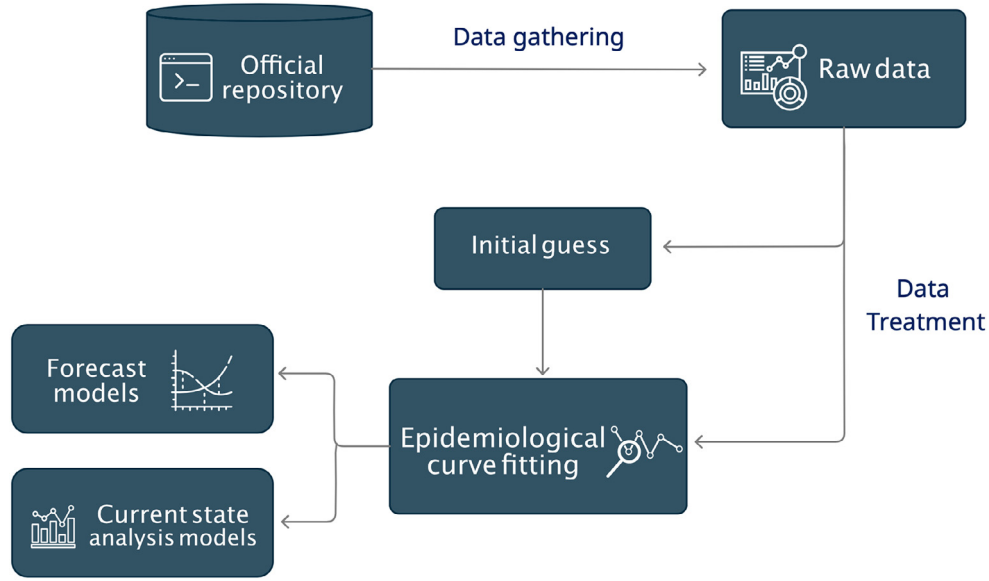*Chaos, Solitons and Fractals 150 (2021) 111156*

Fig. 1. Schematic representation of the inverse problem-solving pipeline.

The case of Chile is an example of countries where typical SIR models would not work; its geopolitical centralization isolates and scatters the different regions that behave as independent populations with varying rules of interaction [23,24,28]. Besides, the profound economic inequality as reported by the Gini's index of the world bank [29] among different social classes constitutes both physical and behavioral anisotropy for the spread of COVID-19. Chile has 16 regions with non-homogeneous connectivity, represented in the modeling by the interaction matrix (Φ), in which each module can be weighted. This matrix is not necessarily symmetric, it considers the fractions of the communities that effectively interact, and it is modified by the spreading rate (β) of each sub-population according to the restriction measures implemented by the government to every community, representing the effective interaction in a given period.

In this work, we modified the multi-group SEIRA model presented in [23] to study the COVID-19 spreading dynamics in Chile. We incorporate a compartment for the deaths and design a tailored parameter fitting methodology to fit the parameters governing the dynamics. We incorporate real-world data for estimating connectivity matrices and solve the inverse problem for parameter fitting using data from official sources. We assessed the effectiveness of NPIs by studying how parameters evolved in the different epidemiological periods. We also study the differences between data-driven and model-based estimates of the reproduction number to evaluate whether the observed dynamics are driven by the local behavior or by the coupling with other regions. By a Monte Carlo-inspired procedure [30], we estimated parameter variability and provided a statistically-based forecast of the pandemic's evolution in Chile. We also study three future scenarios assuming different levels of social distancing, which allow us to predict the effect of the various health policies implemented by the government; in consequence, authorities can take better decisions to prevent the spreading [31].

2. Methodology

2.1. Overview

We detail the implementation workflow of the proposed SEIRD multi-group model in Fig. 1. We first collect and pre-process official data of the spread of COVID-19 in Chile from local authori-

ties [32]. The interaction and mobility data required for modeling the connectivity between the sixteen regions of Chile is estimated from national statistics [33–35] and projections for 2020 [36]. We obtained default and initial parameter values for the model from regional reports and literature. Then, we study the temporal evolution of the spreading and death rate in each region, using a combination of the *simulated annealing* and gradient descent algorithms to minimize the error existing between the reported (raw) and simulated curves. Finally, using these values and based on different social distancing measures, various possible scenarios are presented for active infected and deceased cases within an established temporal horizon.

2.1.1. Interaction dynamics between different regions

We summarize the interaction structure between the different sub-populations considered in the SEIRD model in a connection matrix (Φ). In this matrix, entries $(\Phi)_{i,j}$ represent the fraction of individuals from the ith sub-population that moves to the jth. Note that Φ is not necessarily symmetric, as the flux of individuals moving from i to j may be different from the ones from j moving to i. An example of this situation is the migration to centralized regions, where a considerable fraction of individuals move to, but there is no considerable migration in the other way around. Therefore, $(\Phi)_{i,j} \neq (\Phi)_{j,i}$. In our model, Φ is also modulated by the local spreading rates β_i, thereby including governmental restrictions valid in the different periods considered.

For each region i, we estimated $(\Phi)_{i,j}$ as the fraction of individuals over fifteen years old from i temporally moving to j because of work or studies, based on national mobility data [33–36]. The displacement matrix Φ comprises factors e_i and C_{ij}; The value e_i represents the fraction of the population that leaves the region, defined as the number of people from region i that travels to another region, divided by its total population $n_{\text{hab},i}$:

$$e_i = \frac{\sum_{j \neq i} \text{individuals from Region } i}{n_i} \tag{1}$$

On the other hand, C_{ij} represents the fraction of those individuals leaving i that go to region j. Thus, we can estimate $(\Phi)_{ij}$ as:

$$(\Phi)_{ij} = e_i C_{ij}, \qquad (\Phi)_{ii} = 1 - \sum_{k \neq i} e_i C_{ik}. \tag{2}$$

Table 1
Parameters description and initial values range assignment for the inverse-problem solution.

Parameter	Description	Value or range	Units	Source
α	Asymptomatic ratio of the population	14	%	[32]
β_i	Spreading rate of the virus in region i	0.00 – 0.30	$days^{-1}$	[49]
γ	Recovery rate	1/14	$days^{-1}$	[50]
ϵ	Latent to infectious transition rate	1/5	$days^{-1}$	[51–53]
θ_i	COVID-19 induced death rate in region i	0.00 – 0.02	$days^{-1}$	[32]
ξ	Factor of behavioral virulence of asymptomatics	1	-	[54]

Given the geographical characteristics of Chile and its centralization, Regions do not have the same level of displacement between them (Table S2).

2.2. Model description

2.2.1. Equations

In our model, i represents one of the sixteen regions of Chile, and we represent the spreading dynamics of SARS-CoV-2 using differential equations in a SEIRD compartmental model where the temporal scale is measured in days. Susceptible individuals S_i can acquire the virus after an effective contagion from an infected individual and would be moved to the exposed compartment E_i. There, they would spend, on average, a time t_c (latent period) until they become infectious I_i. After an average time t_i (infectious period), infectious individuals are moved to a final recover R_i or death D_i compartment. The description, initial values, and sources for all the parameters in equations from (1) to (5) are shown in Table 1.

Assuming that natural birth and death rate of the population can be neglected when compared with the COVID-19 induced death rates –which also occur in much shorter timescales–, the system of differential equations describing the dynamics is:

$$\frac{dS_i}{dt} = -(1 + (\xi - 1)\alpha) \sum_{j=1}^{n} \beta_j \left(\frac{\Phi^{ij} S_i}{n_j^T} \right) \left(\sum_{k=1}^{n} \Phi^{kj} I_k \right) \tag{3}$$

$$\frac{dE_i}{dt} = (1 + (\xi - 1)\alpha) \sum_{j=1}^{n} \beta_j \left(\frac{\Phi^{ij} S_i}{n_j^T} \right) \left(\sum_{k=1}^{n} \Phi^{kj} I_k \right) \quad - \epsilon_i E_i \tag{4}$$

$$\frac{dI_i}{dt} = \epsilon_i E_i - (\gamma_i + \theta_i) I_i \tag{5}$$

$$\frac{dR_i}{dt} = \gamma_i I_i \tag{6}$$

$$\frac{dD_i}{dt} = \theta_i I_i \tag{7}$$

2.3. Data treatment

Raw data presented corresponds to that reported daily by the Government of Chile, available in Ministry of Science [32]. This work analysed official data reported daily by the Government of Chile [37], consisting in a total of 244 days until November the 2nd, 2020.

Regional-level daily data for each region i starts in March 3rd 2020 and include new daily infected cases, total infected cases, and total deaths (D_i). However, due to a reporting inconsistency generated when the official reporting guidelines changed (June 1st), datasets contained systematic errors. We corrected those errors as explained in Supplementary Section 3. Using these corrected datasets (available as supplementary dataset), we calculate total recovered (R_i) and active infected cases (I_i), defining also variables

accounting for the experimental time (T_{exp}) and number of inhabitants per region (n_{hab}).

For parameter fitting, we defined fifteen epidemiological periods for the pandemic progression in Chile, which limits matched relevant governmental measures detailed in Table 2 [38–40].

2.4. Initial guess and nested problems

The set of initial values for the parameters (Supplementary Table S3) are expected to reflect the behavior of the different population variables (SEIRD) in the different epidemiological periods, according to the confinement measures established by the government in every region. We perform a separate parameter fitting for every epidemiological period and could be extended to a higher number of periods, if necessary.

After obtaining the simulated curves for the SEIRD variables, using the values in Table S3, we define which parameters will be estimated by the simulated annealing algorithm, before proceeding to the formal parameter fitting. We considered the percentage of asymptomatic patients to be constant across regions, based on governmental reports [37]. Recovery and exposure rates are considered uniform at the country level as the Chilean health system have not been overwhelmed to date. Therefore, we obtain in the parameter fitting the local spreading and death rates (β_i and θ_i respectively). We also provide freedom to the initial conditions of the first epidemiological period, leaving them to be determined in the parameter fitting procedure.

We numerically obtain the solution of the SEIRD model (Eq. (1)) for every parameter combination evaluated throughout the fitting process. Subsequently, we quantify a measure proportional to the mean squared error (MSE) between simulated curves and raw data reported for each region, combined in the cost functional J^k being minimized in the kth epidemiological period. The optimization algorithm selected was a combination of *simulated annealing* and *gradient descent*.

Finally, the initial guess for initial condition of the SEIRD variables in each region was defined as follows: the initial value for the number of susceptible ($S_i(0)$) people is considered as the total population of the region, while $E_i(0) = I_i(0) = R_i(0) = D_i(0) = 0$, except in Maule Region, where $I_{Maule}(0) = 1$ and $S_{Maule}(0) = n_{hab} - 1$ because this region was where the first COVID-19 case in Chile was reported.

2.5. Parameter fitting strategy: resolution of inverse problem

We determined the set of parameters that best describes the observed national dynamics (with regional resolution) by minimizing a cost functional J^k for each of the k epidemiological periods $T_k = [t_i^k, t_f^k]$ as described in Table 2. This functional accounts for the total MSE (mean squared error) between the number of infected I, deaths D, and recovered R cases and raw data obtained from official repositories [37]. Structurally, J^k includes the three contributions:

$$J^k = J_I^k + J_R^k + J_D^k. \tag{8}$$

Table 2
Epidemiological periods.

Period	2020 dates	Remarkable government measures
I	Mar 3 – Apr 6	3/3 First case reported, 3/15 Mandating Schools closure
		3/16 Borders were closed, 3/18 "State of exception*" enacted
		3/21 First deceased person, 3/25 Lockdown in some Metropolitan Region areas
II	Apr 7 – Apr 21	4/19 Government call for a "new normality"
III	Apr 22 – May 6	4/23 School were indefinitely closed
IV	May 7 – May 21	5/13 Announcement of lockdown for 90% of Metropolitan Region
V	May 22 – Jun 5	Health Ministry defines new COVID-19 case-reporting criteria:
		- Expansion of the dead case criterion
		- Active contagious case since the onset of symptoms
		- Include positive PCRs in laboratory reports as active cases
		- Expansion of the dead case criterion
		- Active contagious case since the onset of symptoms
		- Include positive PCRs in laboratory reports as active cases
VI	Jun 6 – Jun 20	Change of Health Minister
		6/16 The "state of exception"* is extended for another 90 days
		6/17 31,412 previously dismissed cases were integrated into the official count
		6/18 Previously unnotified cases were added to the daily infectionslist due to an inclusion criterion change
VII	Jun 21 – Jul 05	Chile reached sixth place in the total number of confirmed cases worldwide
		7/4 Some communes of the Metropolitan Region completed 100 days in quarantine
VIII	Jul 6 – Jul 20	7/19 Announcement and enactment of the "Step by Step" lockdown release plan
IX	Jul 21 – Aug 4	7/28 Lockdown release for districts in the Metropolitan and Valparaíso Regions
X	Aug 6 – Aug 19	7/28 Transition step area broadened in the Metropolitan Region
XI	Aug 20 – Sep 3	8/28 Several districts of Bío-Bío and Maule Regions went back to lockdown
XII	Sep 4 – Sep 18	Communes of O'Higgins, Magallanes Regions, among others, move back to lockdown
XIII	Sep 19 – Oct 3	The state of exception due to catastrophe will remainin force by a presidential mandate for the next 79 days
XIV	Oct 4 – Oct 18	First schools reopening
XV	Oct 19 – Nov 2	10/25 Chile's 2020 National Plebiscite

*In which the government may transcend the rule of law in the name of the public good.

Because of differences in the number of inhabitants across regions, without explicitly correcting it, the error contributed by a small region will be less than that contributed by a large region. To avoid unrealistic solutions where small regions would be left aside because of the algorithm's blind drive to minimize the error, we included a weighting factor. Each contribution of each region was weighted by $w_X^{i,k}$, a factor directly proportional to the ratio between the total infected, recovered or deceased calculated cases for the all country in the last day of the epidemiological period, and the respective number of infected, recovered or deceased calculated cases in the i region the same day.

$$J_X^k = \sum_{i=1}^{16} \sum_{t \in [t_i^k, t_f^k]} w_X^{i,k} \left(X_i^{\text{mod}}(t) - X_i^{\text{data}}(t) \right)^2 \qquad (9)$$

$$w_X^{i,k} = w_X \left(\frac{\sum_{i=1}^{16} X_i^{\text{mod}}(t_f^k)}{X_i^{\text{mod}}(t = t_f^k)} \right)^{1.7}, \qquad (10)$$

where the 1.7 exponent in $w_X^{i,k}$ was empirically set for balancing the contributions of large and small regions, and the correction factor w_X connects the different contributions in functional J_X. Given the difference in magnitude of deceased in comparison with active infected and recovered cases we define $w_I = w_R = 1$ and $w_D = 10^2$.

We then find the optimal parameters for the fit as the argument of a minimization problem, numerically minimizing value of J_k for each epidemiological period. Noteworthy, each epidemiological period described in Table 2 has at least 14 data points ranging in times $[t_i^k, t_f^k]$ corresponding to daily levels of infection, recoveries, and deaths, thus constituting for each region (and per period) a total of 42 independent measures. On the other hand, we only have to fit two parameters per region and per epidemiological period (namely, β_i^k and θ_i^k), thus not risking overfitting.

The implementation of the mathematical model and resolution of the parameter-fitting inverse problem was performed in Matlab version R2018a.

2.6. Variability assessment and forecasting

To obtain reliable values for the selected parameters, Monte Carlo simulations [30] were performed according to the methodology described in Contreras et al. [41] ($N = 450$). In these simulations, we induce a $\pm 5\%$ white noise in the input data, and fit the parameters using this mildly noisy signal. Doing so, we aim to minimize the contribution of potential errors underlying the data. In that way, this experiment is also a sensitivity analysis of the method to the data. The inverse problem for curve-fitting is solved individually, resulting in distributions of parameters. We statistically obtained average population parameters and their variability from those distributions.

Once obtaining values for both parameters and their uncertainty, we evaluate different forecast scenarios for the evolution of COVID-19 pandemic Chile.

3. Results and discussion

3.1. Monte Carlo simulations

As a result of Monte Carlo multiple-simulations experiment performed we obtained distributions for both β_i and θ_i parameters in each region, and subsequently used them to numerically solve our model. Here we focus on active infected cases (I) and total deaths (D), since these we believe are of greater importance for policymakers.

Tables 3 and 4 show the median values obtained for β and θ, respectively, with a 95% confidence interval in the Metropolitan, Valparaíso and Bío-Bío Regions in each epidemiological period. Noteworthy, these regions concentrate the largest number of inhabitants in the northern, central, and southern zones of the country, adding up to almost 60% of the total population of Chile. We observe that the highest values of β were in the first epidemiological period, where no social distancing measures had been established, followed by the IV epidemiological period, for Metropolitan, Valparaíso, and Bío-Bío Regions — which was approximately

D. Freire-Flores, N. Llanovarced-Kawles, A. Sanchez-Daza et al.

Chaos, Solitons and Fractals 150 (2021) 111156

Table 3

β [$days^{-1}$] median values with 95% confidence intervals for each epidemiological period for Metropolitan, Valparaíso, and Bío-Bío Regions.

Epid. Period	Valparaíso		Metropolitan		Bío-Bío	
	Median	CI	Median	CI	Median	CI
I	0.98	[0.95–0.98]	1.03	[1.01–1.04]	1.10	[1.06–1.11]
II	0.05	[0.05–0.05]	0.14	[0.14–0.15]	0.01	[0.01–0.01]
III	0.11	[0.11–0.12]	0.29	[0.29–0.29]	0.01	[0.01–0.01]
IV	0.21	[0.21–0.22]	0.19	[0.19–0.20]	0.26	[0.25–0.28]
V	0.10	[0.10–0.10]	0.14	[0.14–0.15]	0.14	[0.14–0.15]
VI	0.09	[0.09–0.09]	0.06	[0.06–0.06]	0.08	[0.07–0.08]
VII	0.05	[0.05–0.05]	0.05	[0.05–0.05]	0.14	[0.14–0.15]
VIII	0.05	[0.05–0.06]	0.01	[0.01–0.01]	0.04	[0.04–0.04]
IX	0.05	[0.05–0.05]	0.02	[0.02–0.02]	0.06	[0.06–0.06]
X	0.07	[0.07–0.07]	0.12	[0.12–0.13]	0.10	[0.10–0.11]
XI	0.09	[0.09–0.09]	0.01	[0.01–0.01]	0.10	[0.10–0.12]
XII	0.06	[0.06–0.07]	0.08	[0.08–0.09]	0.05	[0.05–0.05]
XIII	0.03	[0.03–0.03]	0.03	[0.03–0.04]	0.07	[0.07–0.07]
XIV	0.06	[0.06–0.06]	0.08	[0.08–0.08]	0.07	[0.07–0.07]
XV	0.05	[0.05–0.05]	0.07	[0.07–0.07]	0.10	[0.10–0.10]

Table 4

θ [$days^{-1}$] median values with 95% confidence intervals for each epidemiological period for Metropolitan, Valparaíso and Bío-Bío Region.

Epid. Period	Valparaíso		Metropolitan		Bío-Bío	
	Median	CI	Median	CI	Median	CI
I	0.0029	[0.0027–0.0033]	0.0031	[0.0027–0.0033]	0.0031	[0.0027–0.0033]
II	0.0024	[0.0023–0.0027]	0.0024	[0.0023–0.0027]	0.0010	[0.0009–0.0012]
III	0.0039	[0.0036–0.0044]	0.0033	[0.0028–0.0033]	0.0020	[0.0018–0.0022]
IV	0.0028	[0.0027–0.0033]	0.0027	[0.0027–0.0027]	0.0012	[0.0009–0.0014]
V	0.0019	[0.0018–0.0022]	0.0018	[0.0018–0.0018]	0.0020	[0.0018–0.0022]
VI	0.0018	[0.0018–0.0022]	0.0018	[0.0018–0.0018]	0.0008	[0.0005–0.0008]
VII	0.0019	[0.0018–0.0022]	0.0027	[0.0027–0.0027]	0.0010	[0.0009–0.0014]
VIII	0.0018	[0.0018–0.0022]	0.0027	[0.0027–0.0027]	0.0009	[0.0009–0.0009]
IX	0.0028	[0.0027–0.0033]	0.0027	[0.0027–0.0031]	0.0009	[0.0009–0.0012]
X	0.0020	[0.0018–0.0022]	0.0018	[0.0018–0.0022]	0.0009	[0.0009–0.0013]
XI	0.0030	[0.0027–0.0033]	0.0018	[0.0018–0.0022]	0.0009	[0.0009–0.0014]
XII	0.0020	[0.0018–0.0022]	0.0018	[0.0018–0.0022]	0.0009	[0.0009–0.0014]
XIII	0.0020	[0.0018–0.0022]	0.0019	[0.0018–0.0022]	0.0019	[0.0018–0.0022]
XIV	0.0031	[0.0027–0.0033]	0.0020	[0.0018–0.0021]	0.0012	[0.0009–0.0014]
XV	0.0032	[0.0027–0.0033]	0.0019	[0.0018–0.0022]	0.0021	[0.0018–0.0022]

three weeks after the government announced "safe return to work" or "*the new normality*" in the Metropolitan Region −. We observe a change point in the general population behavior after the announcement, which was delayed to the IV epidemiological period because of the disease timeline (latency, incubation, recovery time) and significant delays in testing [24]. A higher rate of infections subsequently spread from Metropolitan Region to the regions with the highest rate of transfers: Valparaíso and Bío-Bío. In each epidemiological period, β_i values do not present a high variability, which is reflected in the narrow 95% confidence intervals. The periods in which there is a higher variability correspond to those with higher β values, so the number of infections was triggered and the fitting becomes more challenging.

These analyses are extended to each of the 13 remaining regions of the country, reporting median values of β_i and θ_i ($i = 1...16$) for each region and epidemiological period (Supplementary Tables S4 and S5). Raw data (parameter values for each realization of the Monte Carlo simulation) for all regions and epidemiological periods are reported as separate Supplementary Files for β and θ. The death rate θ_i should not be mistaken by the Infection Fatality Ratio (IFR), as the latter represents the fraction of individuals who die after being infected. If we focus on the infectious compartment I_i, the transitions rates to recovery or death are given by γ and θ_i, being $\gamma = 1/14$ at least one order of magnitude larger than θ_i (cf. Table 4). Estimating the IFR as $\frac{\theta_i}{\gamma + \theta_i}$, and using the inferred values of theta for all epidemiological periods and regions, we obtain a median IFR of 1.56%, which agrees well with official data [42]. A more detailed analysis is provided in the Supplementary Materials, Section S4.

The NPI (non-pharmaceutical interventions) agenda of the government was different for each region, and the main criterion for enacting them was the reported new cases. We observe that the spreading rate β remained high in some regions during the first epidemiological periods before decreasing, due to the time required for evidencing the effects of lockdowns and delays associated with the disease progression. In those regions reporting the first COVID-19 cases in Chile, the spreading rate β decreased faster because of the earlier establishment of NPIs. An example is the O'Higgins Region, which has the lowest average value for the fifteen epidemiological periods of spreading rate. Regarding the death rate θ, it remained relatively low in all regions since the hospital capacity at the country level for the correct care of critical patients was not exceeded, and also due to the efficiency in the inter-regional transfer of patients.

3.2. COVID-19 regional spreading dynamics

Using the different sets of parameters obtained in the Monte Carlo experiment, the simulated curves for the active infected cases (I) and deaths cases (D) are projected in each region over time with a two-level calculated confidence interval (60% and 95%). Figure 2 shows both the simulations and raw data for the Metropolitan, Valparaíso, and Bío-Bío Regions.
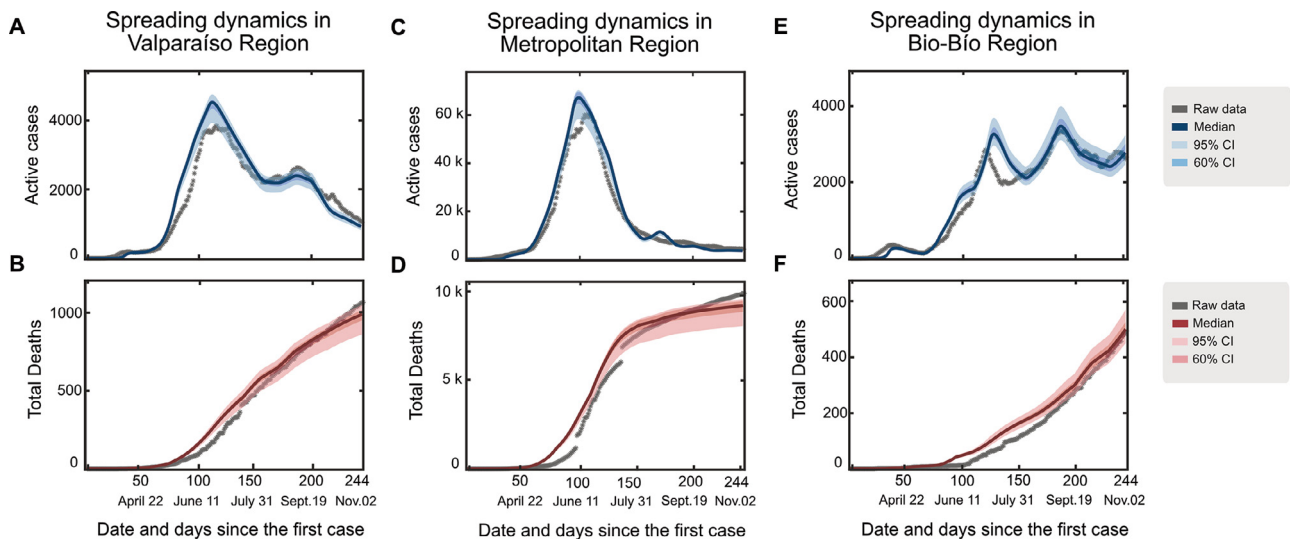
Fig. 2. Simulated curves of active infected (top) and deceased (bottom) cases for the Valparaíso (A,B), Metropolitan (C,D) and Bío-Bío (E,F) Regions, along with 60% and 95% confidence intervals. Raw data for the evaluated time-frame is also presented (grey dots).

We observe that simulated curves are in good agreement with raw data, which always remain within the confidence intervals and present consistent trends. The color intensity of these zones varies depending on the confidence intervals percentage (95% and 60%, light and dark shadows, respectively), calculated from the β and θ distributions obtained in the Monte Carlo experiment.

We observe a peak in the active cases in the Metropolitan Region between June and July, which corresponds to an increase in the number of deaths cases on the same period. Raw data for the active infected cases in the Valparaíso and Bío-Bío Regions has two peaks. The first one corresponds to the date where active cases in the Metropolitan Region were at their peak, so due to the high number of transfers between these regions, these active cases act as vectors of infection in the destination regions, thus increasing infected cases and therefore deceased. In contrast, the second one is related to the fact that the measures adopted by the government are sectorized for each Region. For Valparaíso and Bío-Bío Regions, the isolation measures were relaxed before than for the Metropolitan Region, resulting in a second increase rate of the active infected cases. Curves for the other regions are presented in the supplementary material (Figure 1 to Figure 13, Supplementary). As the time-frame set for each epidemiological period does not necessarily match the temporality of the different measures enacted in each region, some parts of the raw data drift from the median in certain epidemiological periods. Thus, it is crucial to consider the actual spreading dynamics between geographical Regions to represent more appropriately the scenario in each Region. This could also be solved by further reducing the number of days per period but risking the possibility of overfitting to raw data.

The proposed SEIRD model was able to adjust well to the data in both low-population and heavily populated regions, showing the relevance of the 1.7 factor described in the *parameter fitting strategy* section. Higher values for this adjustment factor result in a better adjustment in the small regions, but at the cost of a less rigorous adjustment in the larger regions. In contrast, with values lower than 1.7, the opposite occurs. In Figure 3, we present the results of a sensitivity analysis of the fitted set to perturbations of ($\pm20\%$) in the parameters obtained in for last epidemiological period.

We observe that, when making these variations in each of the 32 parameters obtained (16 β and 16 θ) the calculated error increases, thus confirming that the parameters obtained by solving the inverse problem effectively minimize the error existing between the simulated curves and the experimental ones.
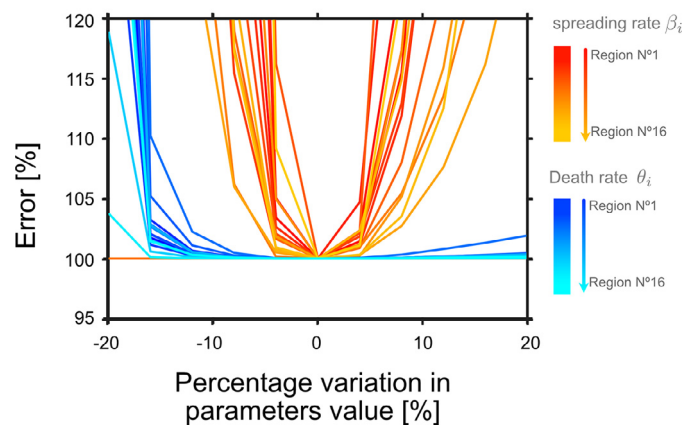

Fig. 3. Sensitivity analysis for the last epidemiological period. The red scale represents the variations in β values, from the first region (red) to the last (yellow), while the blue scale represents the variations in θ values, from the first region (blue) to the last one (cyan). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.3. On the values of R_t in coordinated government measures

The Effective Reproduction Number R_t represents the number of persons a single infected individual might infect, in a population that is aware of the disease representing the viral spread rate of the virus, and varies depending on the policies implemented by the government, such as quarantines [43]. Based on the parameters obtained from the simulations, it is possible to calculate the value of R_t as the ratio between the spreading rate β and the recovery plus deaths rate ($\gamma + \theta$), as demonstrated in Cintrón-Arias et al. [44]. This approach allows us to obtain the effective reproduction number driven only by the local population's behavior, decoupled from other regions:

$$R_t = \frac{\beta_i}{\theta_i + \gamma_i}. \tag{11}$$

We obtain a data-driven value for the observed R_t, R_t^{obs}, adapting the methodology presented in Contreras et al. [45], Medina-

D. Freire-Flores, N. Llanovarced-Kawles, A. Sanchez-Daza et al.
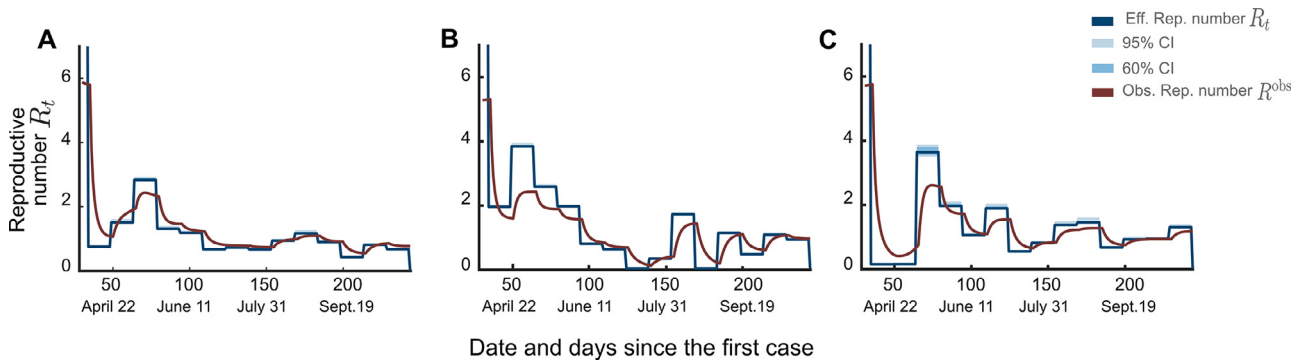
Chaos, Solitons and Fractals 150 (2021) 111156

Fig. 4. R_t values calculated from raw data (red) and simulated data (blue) for the Valparaíso (A), Metropolitan (B) and Bío-Bío (C) Regions with 60% and 95% confidence interval. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Ortiz et al. [46]:

$$R_t^{obs} = \frac{\Delta I}{\Delta R + \Delta D} + 1 \qquad (12)$$

In particular, both values do not necessarily need a match, as they represent -slightly- different quantities. The observed R_t^{obs} represents whether the overall trend is the spread or containment of the disease, purely driven by data, and is affected by testing and tracing governmental plans [47]. More extensive testing will uncover unnoticed infection chains, and also increase the numbers as the uncovering would be faster than the spread of the disease. On the other hand, the effective reproduction number R_t accounts for local trends on contagion, disregarding whether those cases would be noticed or remain uncovered throughout the disease timeline.

In Fig. 4 the R_t values obtained from official sources and simulation results in the Metropolitan, Valparaíso, and Bío-Bío Regions are presented, in conjunction with an 95% confidence interval. As is shown, both R_t values reflect a similar behavior. However, they are not identical, probably because there is a difference between the simulated and raw data due to the contribution of the inter-region movements in the infection rate, which agrees with the results obtained in the inverse problem-solving.

These results for R_t show the correlation between the imposed social distancing measures and the values of the Effective Reproduction Number, in the different regions. This value is higher in the III epidemiological period for the three presented regions and subsequently decreases as more restrictive measures are declared. The case of the Magallanes region (Figure 26, Supplementary Material) is of particular interest, in the absence of proper coordination around social distancing measures, a high R_t number is observed in mid-August and consequently a second outbreak of infected assets, more significant than the first one.

Projecting these R_t values into the future, a decrease in the virus spreading rate and a reduction of the number of active infected would be expected, since a $R_t < 1$ implies a slow rate of spread and the outbreak size would decay exponentially.

3.4. Model forecasting

To evaluate the impact of governmental interventions, we simulated different scenarios aiming to project the contagion trends observed if those interventions did not take place. In Fig. 5, we present projections of infected cases in the Metropolitan Region, if no more restrictions were imposed after the II and III epidemiological periods.

Due to the importance of establishing containment measures as soon as possible to avoid an excessive spread of the virus, the first three epidemiological periods were chosen to evaluate immediate quarantine (I) and non-imposition of measures (II and III) scenarios.

It is observed that in the scenario I the number of active infected cases reaches a peak of approximately 1500 (Fig. 5, C) cases, 30 days after the first reported case. The spreading rate (β) (Table 3) associated with this scenario corresponds to 0.05, registered in the VII epidemiological period in which the metropolitan region was in total quarantine. On the other hand, in scenarios II and III active infections peak over a million cases between days 150 - 200 (Figure 5, B). Scenario III is the one that presents a higher number of infections due mainly to two factors: the number of active infected cases at the beginning of the period and the spreading rate. In both factors, the values are higher in the third period than in the second. In addition to the spreading rate (β) (Table 3) in the third period being 0.29, while in the second period it is 0.14, a higher initial number of actively infected would facilitate the spread of the virus in the presence of a high spreading rate, explaining the behavior of the scenario II and III curves.

Using the same strategy, we analysed different future scenarios based on the projected restrictions for each region. On July 19th, the Government of Chile announced the beginning of the "Step by Step" plan to lift the current social distancing measures gradually. This plan involved the gradual opening in the different districts and regions of the country based on the contagion rate present in each one, the percentage of occupancy of ICU (intensive care unit) beds in hospitals, and the rate of PCR test positivity, among others. The steps are:

1. Quarantine: People cannot leave their homes.
2. Transition: People is allowed to leave the house with restrictions, only on weekdays.
3. Preparation: Individuals are free to move, but group gatherings are not permitted.
4. Initial Opening: Group gatherings are permitted, with a restricted number of people.
5. Advanced Opening: free group gatherings are permitted.

Each region that enters a new stage of the plan presents greater freedom of movement and, therefore, the number and intensity of contacts between people (and thereby the spreading rate β) could increase. Consequently, a *weighting factor* is assigned to each stage of the plan. This factor multiplies the β value of the corresponding region to projecting the effects that the "Step by Step" program will have on the extrapolation in the SEIRD variables (Table S6) [48].

Figure 6 present three different scenarios projected for the Metropolitan Region. These scenarios are projected in a time window until the end of 2020:

- Scenario 1: The Metropolitan Region remain in stage 1 of quarantine.
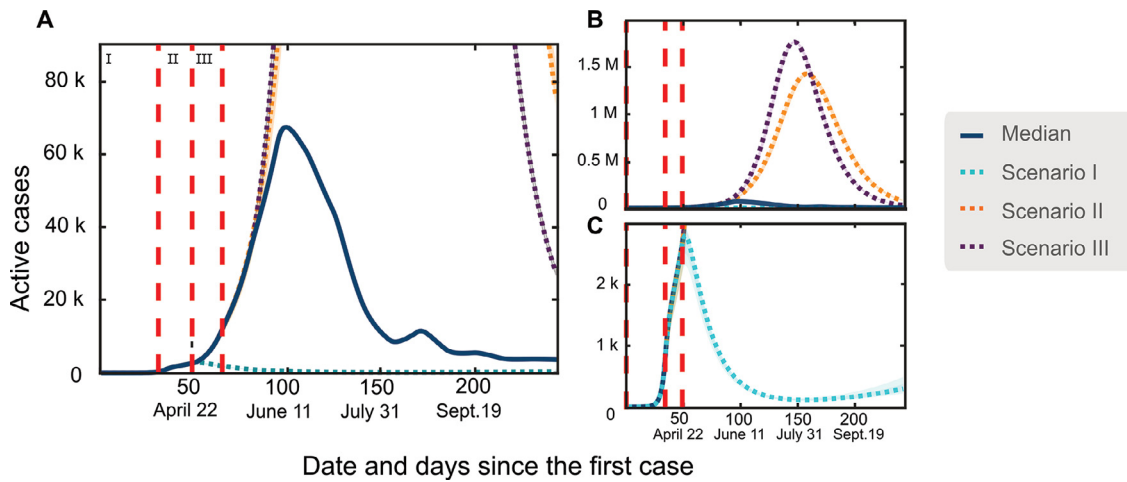- Scenario 2: Current scenario. The Metropolitan Region advances to 4th phase of preparation.

Fig. 5. Different simulated scenarios for active infected cases in the Metropolitan Region. A: Scenario I (cyan curve) corresponds to establishing a total quarantine from the first epidemiological period. In contrast, scenarios II (orange curve) and III (purple curve) represent the cases where no greater measures of social distancing took place as of periods II and III, respectively. The vertical dashed red lines indicate the date of periods I, II and III. Figures B and C correspond to different scales from figure A to visualize the scope of the curves. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
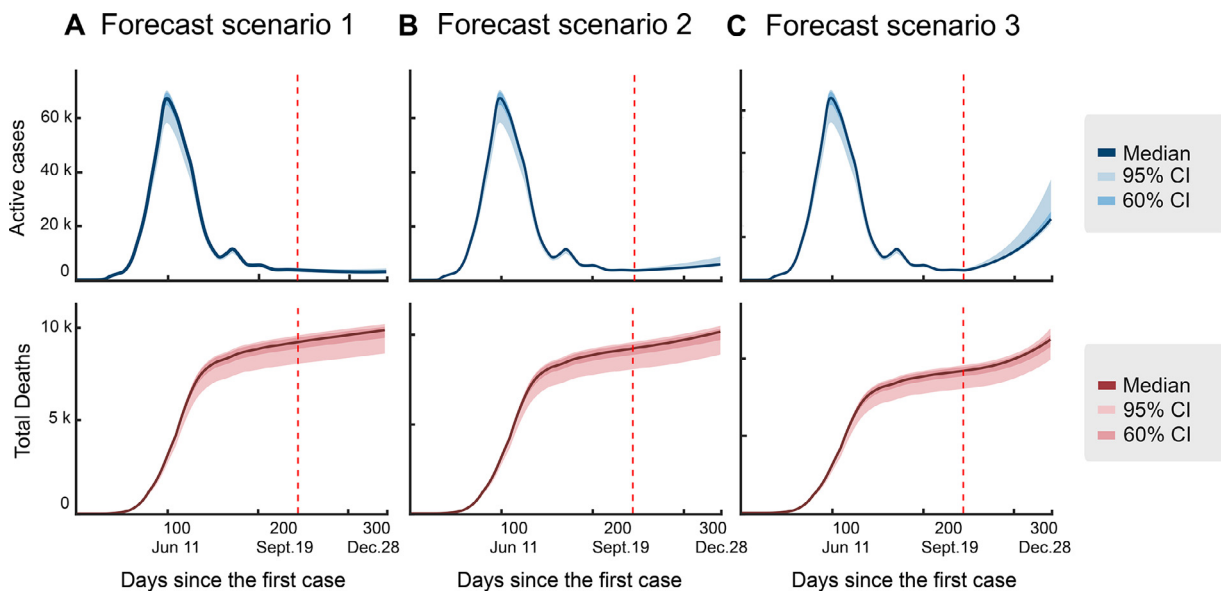

Fig. 6. Different scenarios proposed for the Metropolitan Region. The first scenario (A) represents the evolution of the active infected (top) and death (bottom) cases maintaining a lockdown from November 2nd (vertical red line) until the end of 2020, while the second (B) and third scenario (C) represent the current situation and a massive reopening respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

- Scenario 3: Limit scenario, in which the spreading rate β increases to such an extent that large outbreaks and the second wave of massive infections occur.

Based on the projections, we would expect a sustained decrease in case numbers if the quarantined regime continues. However, in the current scenario (initial opening phase), a second peak in contagion is expected. Nevertheless, this wave would be lower in magnitude compared with the one observed in the first half of 2020, and therefore the health system would also be able to handle it. Finally, in the third scenario (where a massive opening is proposed), we observe a more pronounced increase in case numbers, which will eventually exceed the number of infections observed at the beginning of the pandemic. Noteworthy, this is more or less the current (March 2021) national trend (see, *e.g.* [4]). In this scenario, it would be uncertain to predict occupancy levels of ICU beds in hospitals as well as requests for mechanical respirators. Nonetheless, we would expect a saturation of the public health system.

4. Conclusions

COVID-19 spreading dynamics depends on multiple factors, including the biological/epidemiological aspects of SARS-CoV-2, human behavior, governmental interventions, and heterogeneities among the affected population. Several research groups have statistically analysed these factors during pandemic progression, and we carefully included them in our modelling. The multi-group SEIRD model used considers the heterogeneous distribution and dynamic displacement of the Chilean population, grouped in 16 regions, and also a timeline of the different NPIs enacted by the government.

Following the presented results, our model shows to be efficient to the adjustment of the raw data, overcoming challenges as discontinuities and high variability in it, generating simulations that are easy to interpret and project with narrow confidence intervals. The measures imposed for a particular region can affect other regions, to a greater or lesser extent depending on their interaction, thus highlighting the need for coordinated governmental actions

to control the spread of COVID-19. We have shown that the multigroup SEIRD model presented in this work is a useful tool to represent the contribution of each region in a heterogeneously populated country and is helpful to forecast in the short term the evolution of the different population groups.

Declaration of Competing Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

CRediT authorship contribution statement

Danton Freire-Flores: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing - original draft, Writing - review & editing. **Nyna Llanovarced-Kawles:** Conceptualization, Formal analysis, Investigation, Data curation, Writing - original draft, Visualization, Writing - review & editing. **Anamaria Sanchez-Daza:** Conceptualization, Writing - original draft, Writing - review & editing, Project administration. **Álvaro Olivera-Nappa:** Project administration, Funding acquisition, Writing - review & editing.

Acknowledgment

Appendix A. List of symbols

X	Arbitrary variable for representing a generic fraction
n_i	Base number of members class i
n_i^T	Effective number of members class i
α	Asymptomatic ratio of the population
ξ	Extra factor of behavioral virulence of asymptomatic patients
Φ^{ij}	Fraction of class i in class j
p_i	Immunity ratio of newborns of class i
Λ_i	Net population growth rate i
d_i	Per capita base death rate of class i
β_i	Spreading rate of the virus in class i
ϵ_i	Inverse of the incubation time in class i
γ_i	Recovery rate of class i
θ_i	Pathogen induced death rate in class i
Φ	Interaction matrix
T_k	kth epidemiological period
t_i^k, t_f^k	First and last day of the kth epidemiological period
e_i	Fraction of people from region i travelling to other region
C_{ij}	Fraction of those individuals leaving region i that travel to region j
$w_X^{i,k}$	Weighting empirical factor for the X_i variable in the kth epidemiological period

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.chaos.2021.111156.

References

[1] Chan JFW, Kok KH, Zhu Z, Chu H, To KKW, Yuan S, Yuen KY. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. Emerg Microbes Infect 2020;9:221–36. doi:10.1080/22221751.2020.1719902. PMID: 31987001.

[2] Wu J, Yuan X, Wang B, Gu R, Li W, Xiang X, et al. Severe acute respiratory syndrome coronavirus 2: from gene structure to pathogenic mechanisms and potential therapy. Front Microbiol 2020;11:1576. doi:10.3389/fmicb.2020.01576.

[3] OMS. COVID-19: cronología de la actuación de la OMS. 2020. URL: https://www.who.int/es/news/item/27-04-2020-who-timeline--covid-19. Accessed: 2020-11-27.

[4] Worldometers.info. COVID-19 coronavirus pandemic, reported cases and deaths by country, territory, or conveyance. 2020. URL: https://www.worldometers.info/coronavirus/#countries. Accessed: 2021-03-25.

[5] Bauer S, Contreras S, Dehning J, Linden M, Iftekhar E, Mohr SB, Olivera-Nappa A, Priesemann V. Relaxing restrictions at the pace of vaccination increases freedom and guards against further COVID-19 waves in Europe. arXiv preprint arXiv:2103.06228.

[6] Contreras S, Priesemann V. Risking further COVID-19 waves despite vaccination. Lancet Infect Dis 2021. doi:10.1016/S1473-3099(21)00167-5.

[7] Moore S, Hill EM, Tildesley MJ, Dyson L, Keeling MJ. Vaccination and non-pharmaceutical interventions for COVID-19: a mathematical modelling study. Lancet Infect Dis 2021. doi:10.1016/S1473-3099(21)00143-2.

[8] Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. In: Proceedings of the royal society of London. Series A, Containing papers of a mathematical and physical character, vol. 115; 1927. p. 700–21.

[9] Hethcote HW, Thieme HR. Stability of the endemic equilibrium in epidemic models with subpopulations. Math Biosci 1985;75:205–27.

[10] Tolles J, Luong T. Modeling epidemics with compartmental models. Jama 2020;323:2515–16.

[11] He X, Lau EH, Wu P, Deng X, Wang J, Hao X, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat Med 2020;26:672–5.

[12] Boujakjian H. Modeling the spread of Ebola with SEIR and optimal control. SIAM Undergrad Res Online 2016;9:299–310.

[13] Contreras S, Dehning J, Mohr SB, Spitzner FP, Priesemann V. Low case numbers enable long-term stable pandemic control without lockdowns. medRxiv 2020.

[14] Cooper I, Mondal A, Antonopoulos CG. A sir model assumption for the spread of COVID-19 in different communities. Chaos Solitons Fractals 2020;139:110057.

[15] Dottori M, Fabricius G. Sir model on a dynamical network and the endemic state of an infectious disease. Physica A 2014;434. doi:10.1016/j.physa.2015.04.007.

[16] Gondim J. A., Tanaka T. Y.. Seird model in heterogenous populations: The role of commuting and social inequalities in the covid-19 dynamics. arXiv preprint arXiv:2008.06718.

[17] Lalwani S, Sahni G, Mewara B, Kumar R. Predicting optimal lockdown period with parametric approach using three-phase maturation SIRD model for COVID-19 pandemic. Chaos Solitons Fractals 2020:109939.

[18] Postnikov EB. Estimation of COVID-19 dynamics "on a back-of-envelope": does the simplest sir model provide quantitative parameters and predictions? Chaos Solitons Fractals 2020;135:109841.

[19] Weiss HH. The sir model and the foundations of public health. Mater Matematics 2013:0001–17.

[20] Sun Z, Zhang H, Yang Y, Wan H, Wang Y. Impacts of geographic factors and population density on the COVID-19 spreading under the lockdown policies of china. Sci Total Environ 2020;746:141347. doi:10.1016/j.scitotenv.2020.141347.

[21] Hethcote HW. The basic epidemiology models: models, expressions for R0, parameter estimation, and applications. In: Mathematical understanding of infectious disease dynamics. World Scientific; 2009. p. 1–61.

[22] Kong L, Wang J, Han W, Cao Z. Modeling heterogeneity in direct infectious disease transmission in a compartmental. Model Int J Environ Res Public Health 2016;13:253.

[23] Contreras S, Villavicencio HA, Medina-Ortiz D, Biron-Lattes JP, Olivera-Nappa A. A multi-group SEIRA model for the spread of COVID-19 among heterogeneous populations. Chaos Solitons Fractals 2020:109925.

[24] Contreras S, Biron-Lattes JP, Villavicencio HA, Medina-Ortiz D, Llanovarced-Kawles N, Olivera-Nappa A. Statistically-based methodology for revealing real contagion trends and correcting delay-induced errors in the assessment of COVID-19 pandemic. Chaos Solitons Fractals 2020;139:110087. doi:10.1016/j.chaos.2020.110087.

[25] Dehning J, Zierenberg J, Spitzner FP, Wibral M, Neto JP, Wilczek M, Priesemann V. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. Science 2020;369. doi:10.1126/science.abb9789.

[26] Bird S, Nielsen B. Now-casting of COVID-19 deaths in english hospitals. University of Oxford; 2020.

[27] Constantinesco A, Israel-Jost V, Choquet P. Spectral analysis of the daily evolution of deaths due to COVID-19 in France and in the world shows a weekend effect: myth or reality? medRxiv 2020. doi:10.1101/2020.06.23.20135442.

[28] Aroca P, González P, Valdebenito R. The heterogeneous level of life quality across chilean regions. Habitat Int 2017. doi:10.1016/j.habitatint.2017.06.010. In Press.

[29] World Bank. Gini index (world bank estimate) - chile. 2020. URL: https://data.worldbank.org/indicator/SI.POV.GINI?locations=CL. Accessed: 2020-10-07.

D. Freire-Flores, N. Llanovarced-Kawles, A. Sanchez-Daza et al.

Chaos, Solitons and Fractals 150 (2021) 111156

[30] Metropolis N, Ulam S. The Monte Carlo method. J Am Stat Assoc 1949;44:335–41. doi:10.1080/01621459.1949.10483310 . PMID: 18139350.

[31] Chimmula VKR, Zhang L. Time series forecasting of COVID-19 transmission in canada using LSTM networks. Chaos Solitons Fractals 2020:109864.

[32] Ministry of Science. COVID19 data. 2020. URL: https://github.com/MinCiencia/Datos-COVID19. Accessed: 2020-11-06.

[33] National Statistics Institute (INE). Censo 2002. 2002. URL: https://www.ine.cl/docs/default-source/censo-de-poblacion-y-vivienda/publicaciones-y-anuarios/2002/sintesiscensal-2002.pdf. Accessed: 2020-05-06.

[34] National Statistics Institute (INE). Censo 2012. 2012. URL: https://www.ine.cl/docs/default-source/censo-de-poblacion-y-vivienda/comites-y-notas-tecnicas/informe-comisión-investigadora-censo-2012/auditor. Accessed: 2020-05-06.

[35] National Statistics Institute (INE). Censo 2017. 2017a. URL: https://www.censo2017.cl/descargas/home/sintesis-de-resultados-censo2017.pdf, Accessed: 2020-05-06.

[36] National Statistics Institute (INE). Proyecciones de población. 2017b. URL: https://www.ine.cl/estadisticas/sociales/demografia-y-vitales/proyecciones-de-poblacion. Accessed: 2020-05-06.

[37] Ministry of Health. Cifras oficiales. 2020a. URL: https://www.gob.cl/coronavirus/cifrasoficiales/, Accessed: 2020-05-31.

[38] Biblioteca del Congreso Nacional de Chile. Qué es el estado de excepción constitucional de catástrofe?2020. URL https://www.bcn.cl/obtienearchivo?id=documentos/10221.1/78781/7/ESTADOEXCEPCION2-W800.png/. Accessed: 2020-11-05.

[39] Tercera L. Línea de tiempo: los hitos que han marcado los 100 días del coronavirus en Chile. 2020. URL: https://www.latercera.com/que-pasa/noticia/linea-de-tiempo-los-hitos-que-han-marcado-los-100-dias-del-coronavirus-en-chile/QPJKFMV3Y5CCDC7T4B7J4ZZLGY/, Accessed: 2020-11-07.

[40] Ministry of Health. Informe epidemiologico COVID-10. 2020c. URL: https://www.minsal.cl/nuevo-coronavirus-2019-ncov/informe-epidemiologico-covid-19/, Accessed: 2020-11-05.

[41] Contreras S, Medina-Ortiz D, Conca C, Olivera-Nappa A. A novel synthetic model of the glucose-insulin system for patient-wise inference of physiological parameters from small-size OGTT data. Front Bioeng Biotechnol 2020;8:195.

[42] Health Ministry of Chile. Casos confirmados en Chile COVID-19. 2020. Accessed: 2020-10-07, https://www.minsal.cl/nuevo-coronavirus-2019-ncov/casos-confirmados-en-chile-covid-19/.

[43] Perasso A. An introduction to the basic reproduction number in mathematical epidemiology. ESAIM Proc Surv 2018;62:123–38.

[44] Cintrón-Arias A, Castillo-Chávez C, Bettencourt L, Lloyd A, Banks H. The estimation of the effective reproductive number from disease outbreak data. Math Biosci Eng MBE 2009;6:261–82. doi:10.3934/mbe.2009.6.261.

[45] Contreras S, Villavicencio HA, Medina-Ortiz D, Saavedra CP, Olivera-Nappa A. Real-time estimation of Rt for supporting public-health policies against COVID-19. Front Public Health 2020. doi:10.3389/fpubh.2020.556689.

[46] Medina-Ortiz David, Contreras Sebastián, Barrera-Saavedra Yasna, Cabas-Mora Gabriel, Olivera-Nappa Álvaro. Country-wise forecast model for the effective reproduction number RT of coronavirus disease. Frontiers in Physics 2020;8:304. doi:10.3389/fphy.2020.00304.

[47] Contreras S, Dehning J, Loidolt M, Zierenberg J, Spitzner FP, Urrea-Quintero JH, et al. The challenges of containing SARS-CoV-2 via test-trace-and-isolate. Nat Commun 2021;12:1–13.

[48] Gobierno de Chile. Situación comunal Chile. 2020. URL: https://www.google.com/maps/d/u/0/viewer?mid=1y7LnezVdq7ZgFiB-cC9WizNJaA2ybDk3&ll=-33.50053550410916%2C-68.86242306277813&z=5. Accessed: 2020-10-01.

[49] Hu H, Nigmatulina K, Welkhoff P. The scaling of contact rates with population density for the infectious disease models. Math Biosci 2013;244. doi:10.1016/j.mbs.2013.04.013.

[50] Ministry of Health. Criterios para determinar a un paciente COVID-19 sin riesgo de contagio. 2020b. URL: https://www.minsal.cl/criterios-para-determinar-a-un-paciente-covid-19-sin-riesgo-de-contagio/, Accessed: 2020-11-08.

[51] Bar-On YM, Flamholz A, Phillips R, Milo R. Science forum: SARS-CoV-2 (COVID-19) by the numbers. Elife 2020;9:e57309.

[52] Lauer S, Grantz K, Bi Q, Jones F, Zheng Q, Meredith H, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. Ann Internal Med 2020;172. doi:10.7326/M20-0504.

[53] Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, Shaman J. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). Science 2020;368:489–93.

[54] Nikolai L, Meyer C, Kremsner P, Velavan T. Asymptomatic SARS coronavirus 2 infection: invisible yet invincible. Int J Infect Dis 2020;100. doi:10.1016/j.ijid.2020.08.076.