The effective reproduction number of COVID-19 in Chile, by commune and week

# Introduction

*The recent COVID-19 pandemic has generated an increasing need for indicators of the speed of spread of the pandemic. Among these, one of the most popular indicators is the effective reproduction number or effective R, which describes how many people an infectious person infects on average during their time in that condition, in a population where some people may have protection against the disease [ 1]. In Chile, the effective R is periodically calculated by the Center for Mathematical Modeling (CMM) and can be visualized at the regional level (Chile is composed of 16 regions) through a convenient dashboard.*[[1]](#footnote-1).

On the other hand, in July 2020, the Chilean government initiated the *Paso a Paso* (*Step-by-Step)* plan. The official government website describes the plan as follows:[[2]](#footnote-2)

*“The Step-by-Step Plan is a gradual strategy to face the pandemic according to the health situation of each particular area. It consists of 5 stages or gradual steps, ranging from Quarantine to Advanced Opening, with specific restrictions and obligations. The advance or retreat from one particular step to another is subject to epidemiological indicators, care network, and traceability. "*

Implementing this plan requires timely epidemiological information at a communal level (each region is made up of a variable number of communes, with 346 communes in total). However, the disaggregation of indicators such as the effective R tends to stop at the regional level. This is because the higher the level of disaggregation, the higher the level of imprecision in the estimates of the effective R. The degree of disaggregation under discussion can be better understood with some figures: according to the projections of the Chilean National Institute of Statistics for 2020, the population of Chile is 19,458,173, the median population of the regions is 763,975, and the median population of the communes is 20,071.

Indeed, the greater imprecision of any estimator of the effective R at the community level inhibits its use as a final product. However, these statistics can still be helpful if they are used as an input to a larger model. Suppose, for example, that we want to evaluate the benefits of the plan step-by-step using data with weekly-communal resolution. In such a case, using the effective-R of each commune-week as the dependent variable may make sense even if they are only considered as noisy estimators of the true effective-R. Therefore, we believe that it is valuable to have more disaggregated effective R estimates (particularly at the weekly-communal level) if they are used carefully.

This article aims to provide effective R estimates at the communal-weekly level for the Chilean case, using different methodologies, and comparing its strengths and weaknesses with a focus on the country under study. To our knowledge, these are the first estimates of its kind for Chile. Of course, it is impossible to review each effective-R estimation methodology available in the literature, so we concentrate on those presented in [2]. Here is a brief description of the methods included in that article: The Systrom method [3] assumes that both the new and the effective-R cases are random variables. Then, the method proposes a Bayesian state-space model where the new cases are interpreted as observations and the effective-R cases as states. Once the model is formalized, the effective Rs can be inferred using any Bayesian inference software (e.g., Stan). The Cislaghi method [4] assumes that the effective-R can be reasonably approximated by the percentage growth of the active cases, as long as these are previously smoothed. The Joint Research Center (JRC) method [5] assumes that the effective-R can be reasonably approximated as the solution of a differential equation that relates the effective-R to the number of new cases, and then solves for the effective-R under the assumption that it cannot change much in short periods. The method of the Robert Koch Institute [6] is like that of Cislaghi, differing only in how they smooth the series of new cases. Finally, the Wallinga-Lipsitch method [7] assumes that the effective-R can be reasonably approximated as the solution of a differential equation that relates the effective-R to the generation time (i.e., the time between infection events in a couple infected-infected).

Indeed, there are many other methodologies to estimate effective-R. Our selection contains the three most popular types of methodologies: methods based on Bayesian models that interpret effective-R as a latent variable, methods based on smoothing the new cases observed, and methodologies based on differential equations.

The rest of this article is organized as follows. Section 2 provides a detailed description of the methods considered, their differences, and data available in Chile for their implementation. Section 3 presents the results associated with each methodology applied to the Chilean data. Section 4 discusses the performance of each method, mentioning the strengths and weaknesses of each when applied to Chilean data. Finally, section 5 presents a brief conclusion.

# 2. Methodology

## 2.1. Estimators of effective R

In epidemiology, there are various estimators of the speed of the spread of diseases. Two very popular estimators are the basic reproduction number, , and the effective reproduction number, effective-R. The basic reproduction number describes how many people an infectious person infects on average during their time being infectious, in a population where no one is supposed to have any protection against the disease. It can be calculated as , where is the probability of transmission per contact, is the contact rate (number of contacts between individuals per unit of time), and is the length of the infectious period. On the other hand, the effective reproduction number describes how many people an infectious person infects on average during their time being infectious in a population where some people may have protection against the disease. It is important to note that is constant, while the effective R can vary over time. In this article, the focus is on the effective R.

In what follows, we will briefly summarize some popular estimators of effective R. For pedagogical purposes, we will begin by explaining the method for estimates at the national daily level, understanding that any hyperparameter defined in terms of days will be rounded to weeks when using a model with weekly resolution. For more information about these estimators, see [2].

### 2.1.1. Systrom method

Let be the number of new cases during the th day and let be the associated effective R. Systrom [3] proposes to infer from following a Bayesian approach. For this, he assumes that:

where is the reciprocal of the serial interval, i.e., the average time between the appearance of symptoms within a chain of infections, which is about seven for covid-19; and is a model parameter. In the original incarnation of the proposal, is estimated using maximum likelihood, so it is not strictly Bayesian. It is also possible to assign a relatively loose value (in our simulations, 0.1) or some prior distribution (Gamma, for instance). However, in our simulations, we have found that reasonable results are only achieved if the prior distribution is proper and relatively informative.

Regardless of whether is fixed or random, the latent effective-R must be learned from the data. Continuing with their Bayesian approach, the authors propose to estimate its posterior mean, i.e., the conditional expectation of given the sample. In practice, they approximate with , where each is an arti is an artificial random variable drawn (approximately) from the distribution e. In our implementation, we generated this artificial sample using MCMC (see [8] for a review). Specifically, we use the Hamiltonian Monte Carlo algorithm implemented in Stan [9].

### 2.1.2. Cislaghi's method

Let be the number of new cases during the -th day, and let be the associated effective-R. Cislaghi [4] proposes to approximate as , where , where is a smoothing parameter and represents the incubation period. Recommended values for and are and .

### 2.1.3. JRC method

During the initial phase of the pandemic, the reproduction number characterizes the exponential growth in the susceptible ones. The change in new cases is expressed as:

,

where is the number of susceptible people, is the number of infectious people, is the population size, and is the length of the infectious period. Assuming that , we have:

The European Commission's Joint Research Center (JRC) method [5] assumes that the effective R does not change much between short enough periods. Specifically, let and be two close moments in time. Then, replacing by at these two times points and taking logarithms, we have that:

Hence, if , then

Usually, is fixed in seven days.

### 2.1.4. RKI method

Let be the number of new cases. The method of the Robert Koch Institute (RKI) [6] estimates as

where is the generation period (i.e., the time between infection events in an infected-infected couple), set by this institute at 4 days for covid-19; although there are also versions that replace by 7 or even 14.

### 2.1.5. Wallinga-Lipsitch method

Before introducing the Wallinga-Lipsitch method [7], it is necessary to introduce the Euler-Lotka equation. The best way to understand this equation is in the context of a continuous-time population dynamics models. Thus, consider a model of this nature (concentrated only in women), where represents the number of births per unit of time, is the fraction of individuals surviving to age , and is the per-capita rate of births for mothers of age . Then, we have:

Now, suppose that . Then, we have:

which is the Euler-Lotka equation. In practice, we can interpret as the rate of production of female daughters by a woman of age , considering her survival and fertility rates.

Until now, the whole discussion has been relative to population dynamics, but if we reinterpret births as contagions, we have that, under this model, the effective R should be given by

The rate can be normalized as a distribution. Specifically, as

Replacing this expression in the Euler-Lotka equation, we have

However, if we interpret births as contagions, then can be interpreted as probability density function associated with the generation period. In this way, we conclude:

where is the moment-generating function of the generation period, . This function can always be estimated from a sample of generation periods, but the generated periods themselves must be estimated from the observed cases. In practice, each of these latent variables are estimated using a different Poisson model. Specifically, at each time period *t,* we fit a Poisson model using the subsample for some fixed integer *s.* Then, we estimate the generation at time *t* using the well-known properties of the Poisson distribution. Of course, we can only do this for *t* > 1. Therefore, this methodology cannot estimate the effective-R at the beginning nor the end of the period under study.

### 2.1.6. Martínez-Beneito et al. method

Let be the number of counts in the th commune, , during the th epidemiological week, ; let be the associated effective reproductive number and let the corresponding population size. Then, Martínez-Beneito et al. [10] model as

where is a matrix with each row representing a basis function (e.g., B-splines), and is a random matrix modeled as the sum of a random effect and the CAR distribution of Besag. Using the components of this model, and in parallel with Cori et al. [12], the authors propose the following estimator of :

where denotes the infectivity rate that quantifies, for each lag time, the probability of observing an -weeks difference between two cases, one primary and one secondary case of the disease, being an upper bound for the subindex . Following Martínez-Beneito et. al. [10], we use a shifted Gamma prior with mean 4.7 and standard deviation 2.9 and set to 25 (days). Once the are computed at the day level, they are summed and normalized to be used at the weekly level. The model is then fitted to each region independently.

Before continuing, we make 2 cautionary notes. The first note is about the spatial process employed. While the authors employed the CAR process of Leroux et al. [11], we found better results using the BYM model [13]. The second note is about the software employed for the estimation of the model parameters. Martínez-Beneito et. al. fit the model using WinBUGS, i.e., relying on MCMC methods. Although this is a reliable approach, it requires a lot of computing power. Because of this, we fit this model using INLA [14].

## 2.2. Differences between assumptions

To make a comparison between the different methods, it is necessary to investigate the underlying methodology. The different assumptions are listed under the description of each method. In general, the methods can be divided into two main types, based on an underlying assumption about the spread of the epidemic or if the epidemic spreads with a constant value in the generation time or the incubation period. The methods that use an underlying assumption about the spread of the epidemic are the JRC method, the Wallinga & Lipsitch method, and the Systrom method. Methods that do not use any underlying epidemic model and assume that the number of new cases will reproduce with a constant value over the generation or incubation time are the Robert Koch Institute method and the Cislaghi method.

## 2.3. Implementation

Most of the methods presented in Subsection 2.1 have a straightforward implementation. The only exception is the method of Systrom. This is because there is no obvious way of simulating a sample from the distribution of conditionally on the sample. In our implementation, we roughly simulate this sample using the Hamiltonian Monte Carlo algorithm (see [8] for a review) implemented in Stan [9]. We implemented all these methodologies in R, and left the code available at [https://github.com/COVID0248/covid-re/R/targets.R](https://github.com/COVID0248/covid-re), Inside this R script, the function that generates the effective-R using Systrom’s method is called get\_r\_systrom(), the function that generates the effective-R using Cislaghi’s method is called get\_r\_systrom(), and so on.

## 2.4. Data

In this work, we use the series of new cases due to covid according to commune and epidemiological week published by the Ministry of Science in the GitHub repository <https://github.com/MinCiencia/Datos-COVID19>. Specifically, we use product 15 (new cases according to the onset of symptoms).

# 3. Results

The file re.csv, available at <https://github.com/COVID0248/covid-re>/data, contains the estimated effective R, according to week, commune and method. In each case, the number of cases according to the onset of symptoms was used. Each parameter defined initially in days was rounded to weeks. In those communes where the number of cases was temporarily 0, it was rounded to 1 to avoid numerical problems when evaluating logarithms with the JRC method. The code to replicate re.csv is in the same repository. See the Appendix for more details on how to replicate re.csv.

Below are some graphs to visualize how the effective R varies for each method.

## 3.1. Spatial dimension

An easy way to appreciate the result of each method in its spatial dimension is with a map reporting the value of a statistic (e.g., a quantile) for each commune. Figures 1-3 do just this, using quantiles 10, 50, and 90, respectively. As can be seen, all the methods estimate a reasonable median Re (between 0.8 and 1.2).

Methods such as that of the Robert Koch Institute and Cislaghi have a clear bias. For values above one, they underestimate , while for values below one, they overestimate . Additionally, the Wallinga-Lipsitch method is among the fastest methods to detect changes in the effective R.

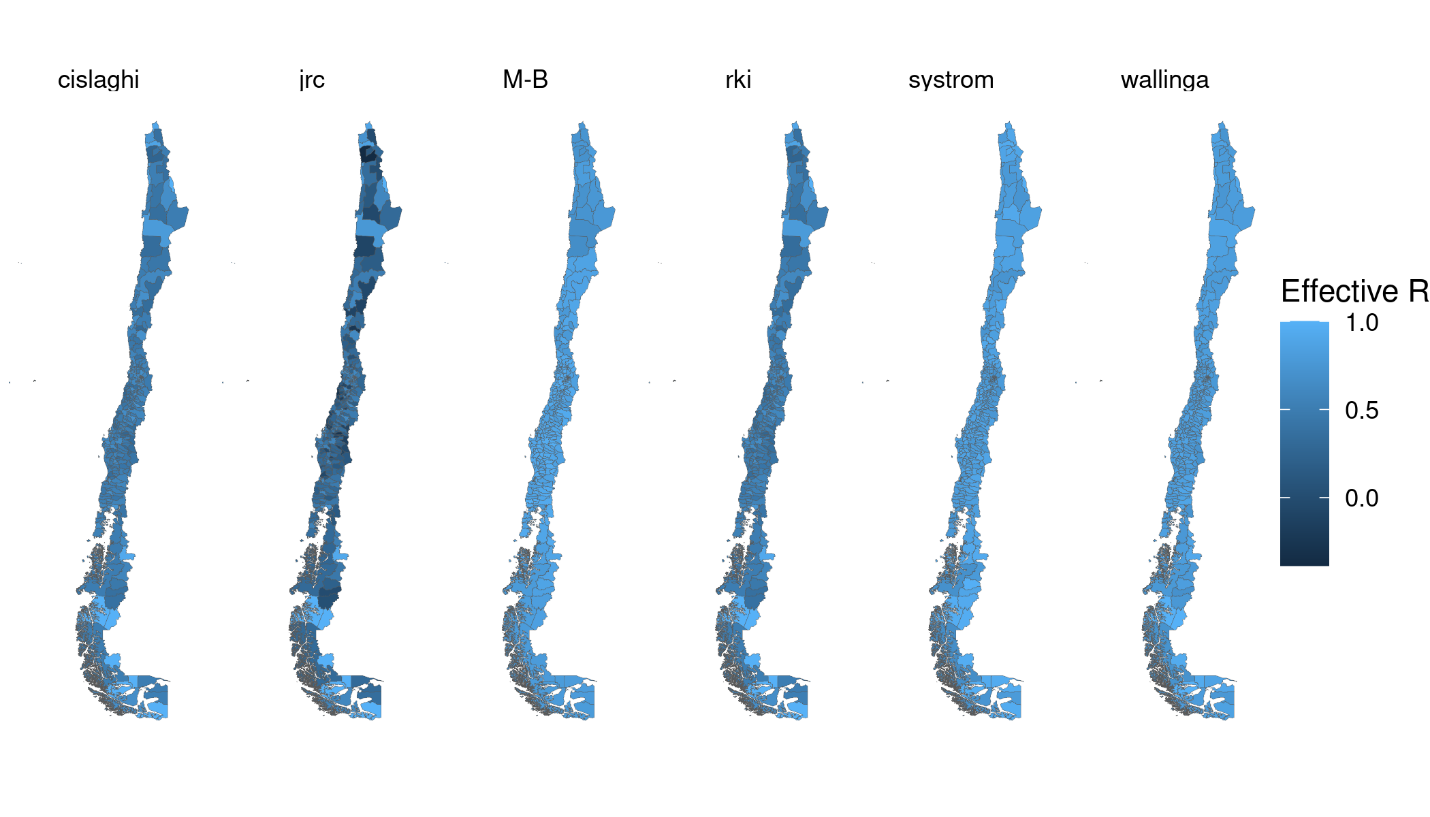


Figure 1. 10th percentile of the effective R, by commune and method.

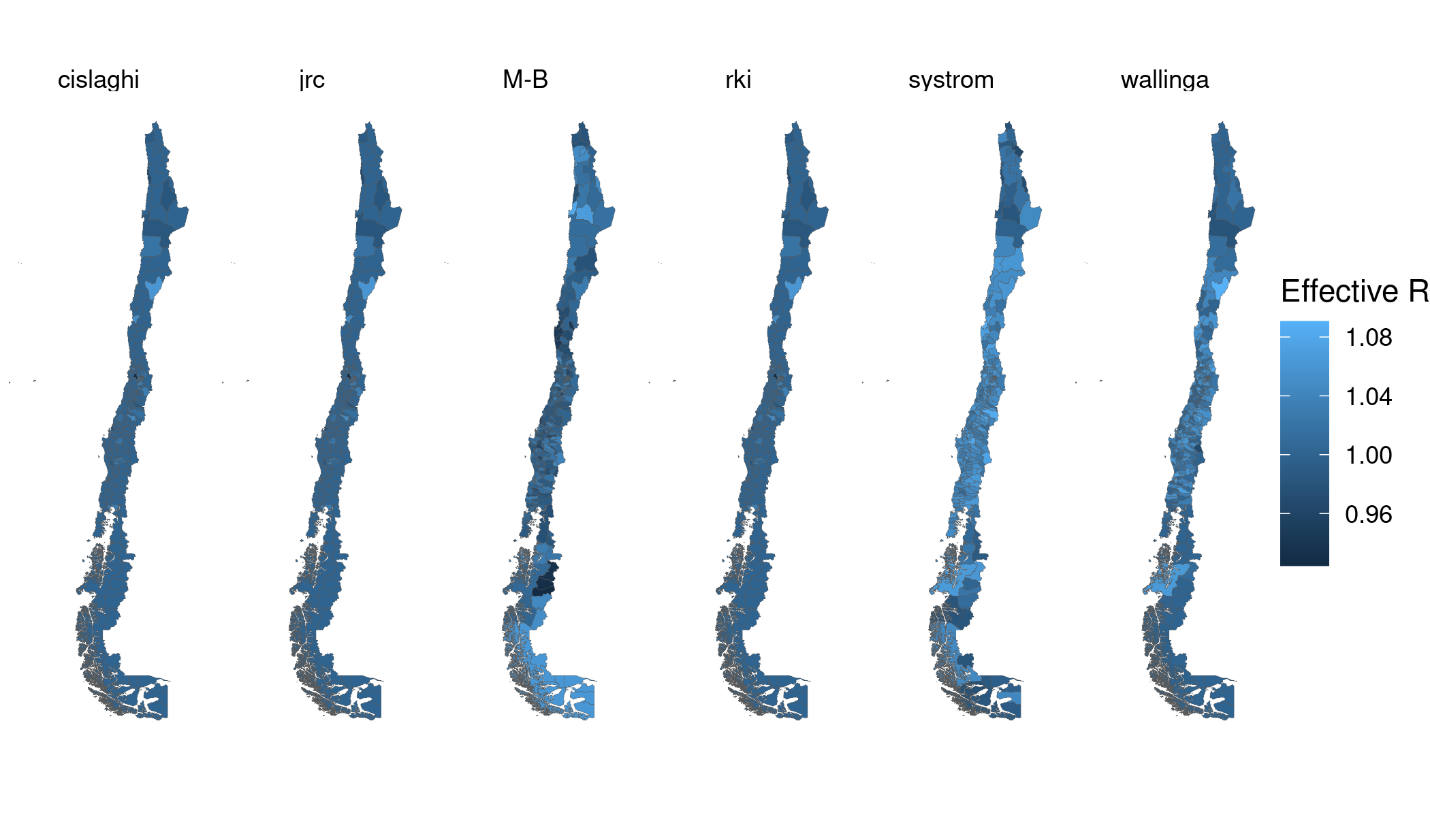


Figure 2. 50th percentile of the effective R, by commune and method.

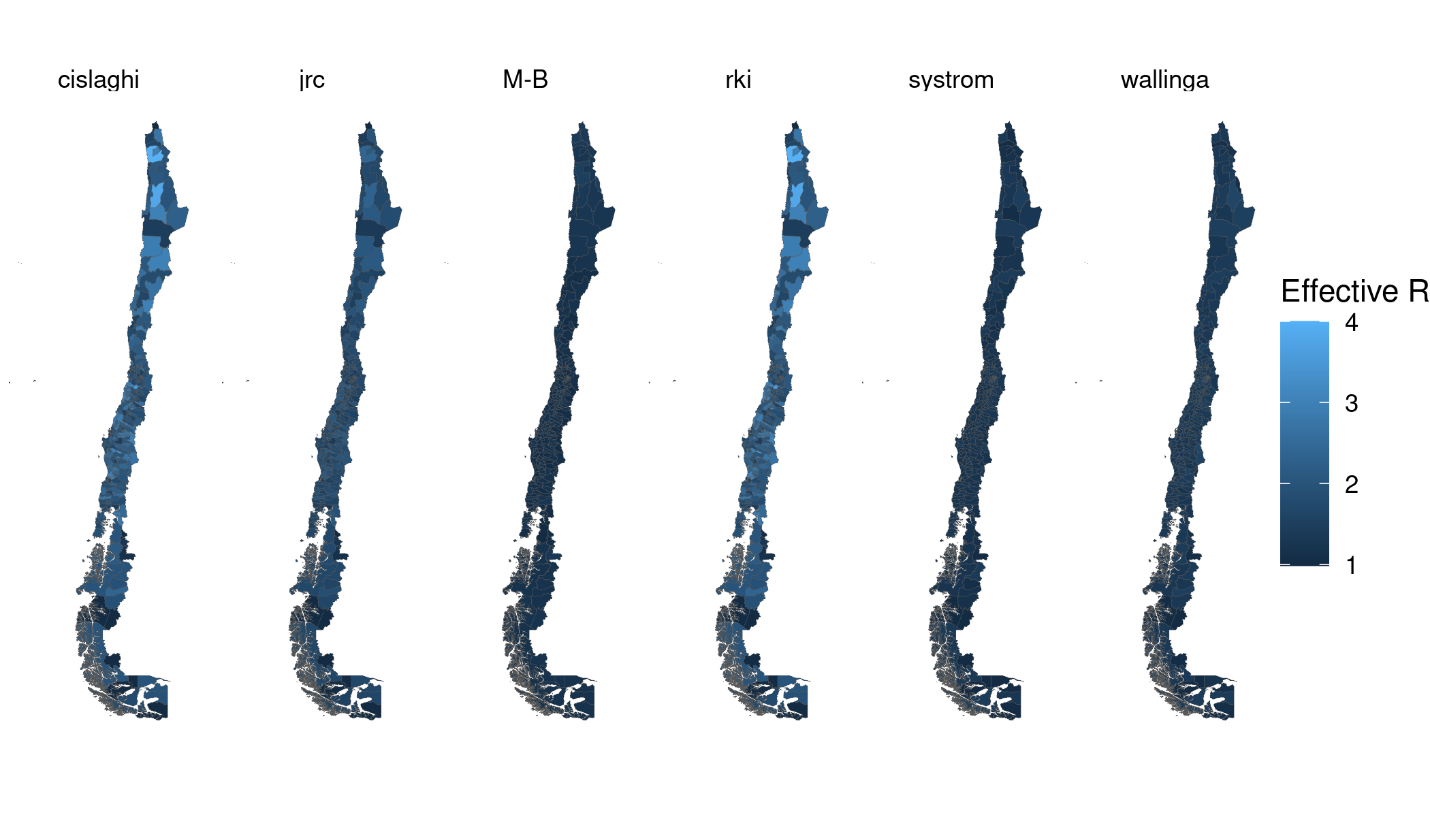
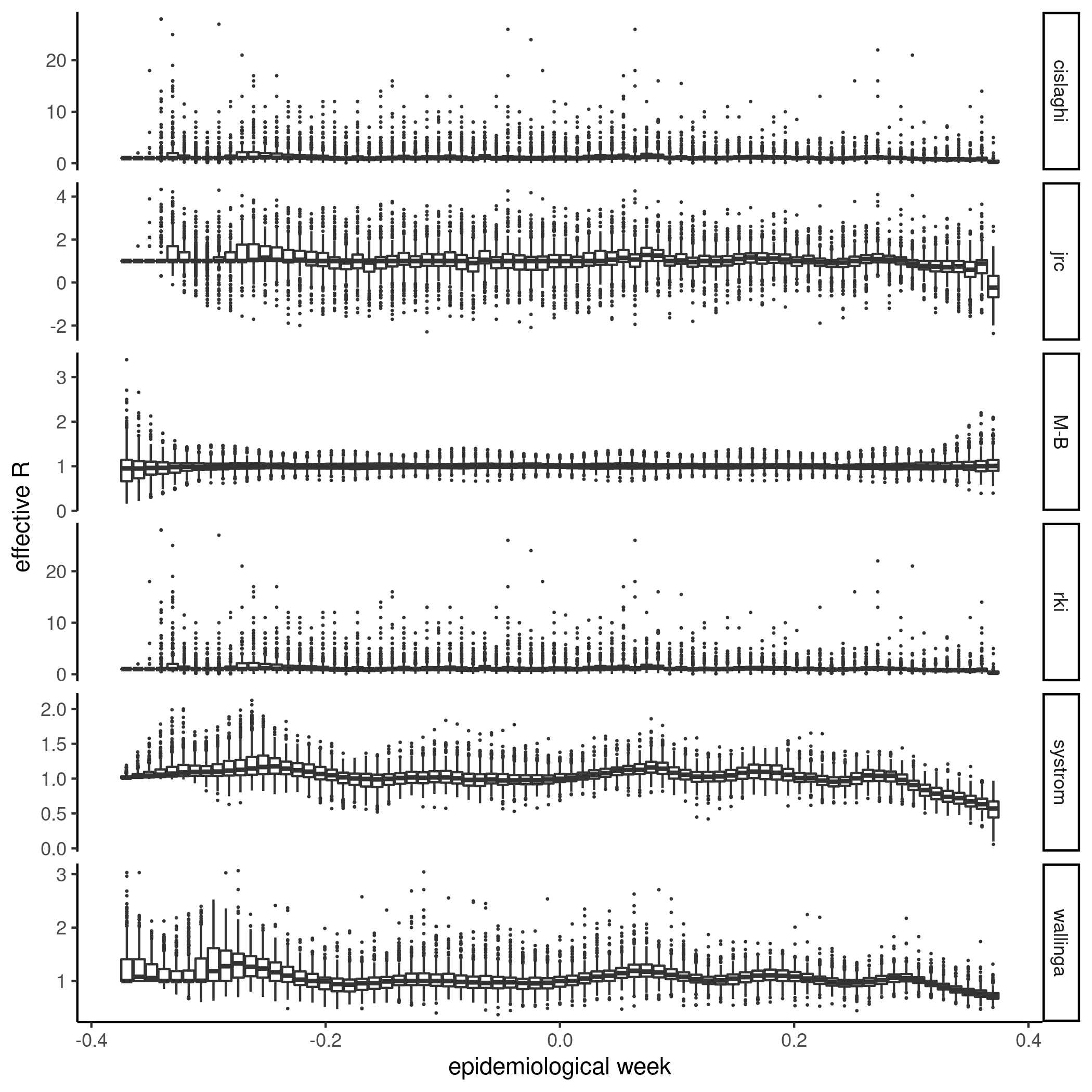


Figure 3. 90th percentile of the effective R, by commune and method.

## 3.2. Temporal dimension

A simple way to appreciate the result of each method in its temporal dimension is with a caterpillar plot that has the epidemiological week on the horizontal axis and presents a boxplot of the effective R on the vertical axis. Figure 4 does just this, according to the method. As can be seen, the JRC method has severe problems at the end of the series, and the Cislaghi and RKI methods have a significantly higher dispersion than the rest.

Figure 4. Boxplot of the effective R, by method and epidemiological week.

Of course, this graph alone is insufficient since it does not consider that there are particularly large or emblematic communes where a more precise calculation of the effective R is required. To correct the above, in Figure 5, we show the evolution of the effective R, according to the method and epidemiological week, in four regional capitals: Santiago, Valparaíso, Concepción, and Rancagua. As can be seen, the Wallinga-Lipsitch Method is the most stable. On the other hand, the JRC method is so unstable that it can even return negative values.

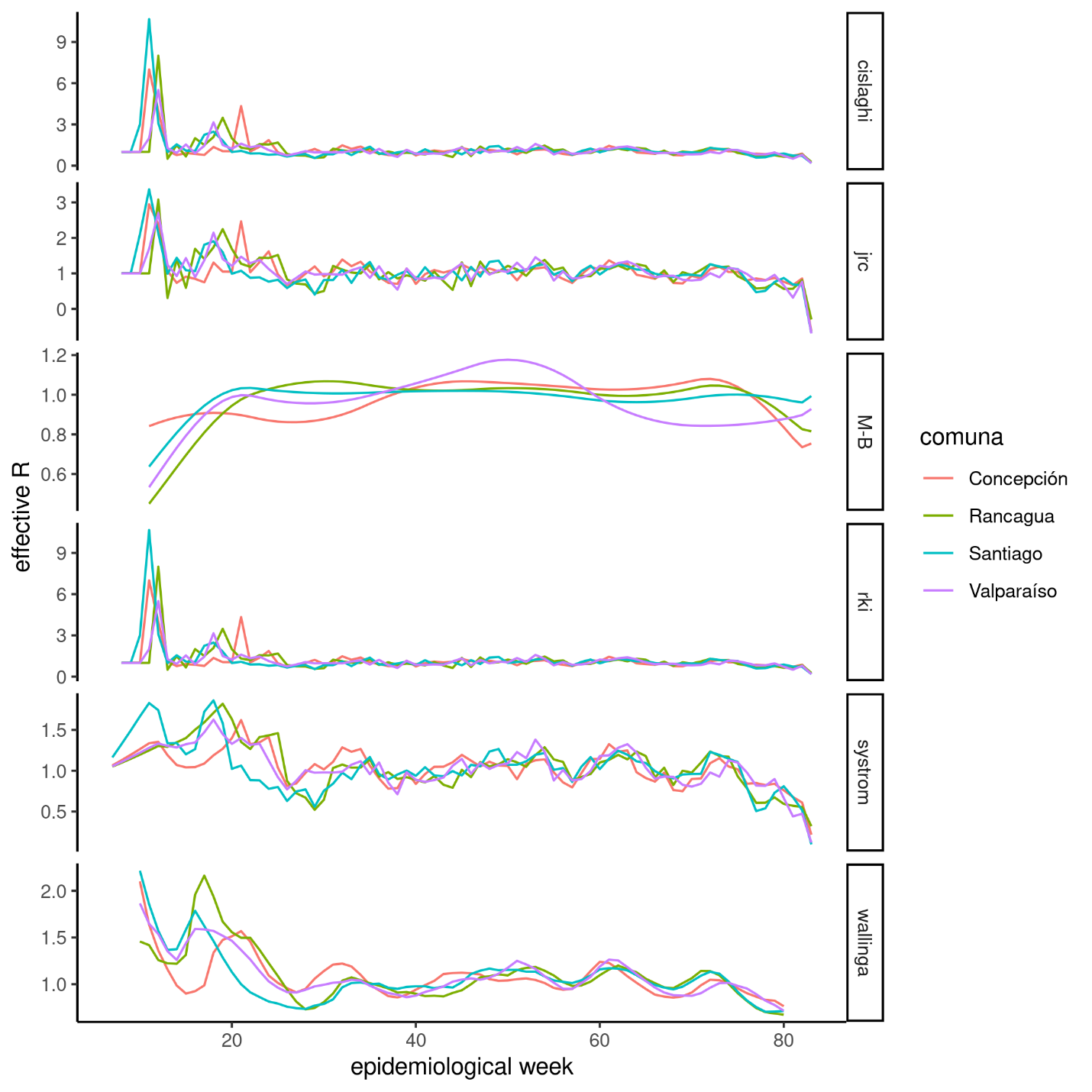


Figure 5. Effective R by method and epidemiological week in 4 regional capitals.

## 3.3. Critical Opinion

## Systrom method

The method is relatively stable even for small communes. However, there are occasions in which the effective R is unacceptably imprecise and may even yield negative values. This behavior is because the formulation of the Bayesian model does not prevent it in the prior specification.

## Cislaghi method

The method can be adapted to weekly data by setting incubation time as an approximation. However, in small communes, we can see estimated values that are difficult to defend (for example, we observe an effective R of 12 in “Los Muermos” during the 25th epidemiological week, although this is relatively small commune). For reference, the maximum and minimum values were 31 and 0.0096.

## JRC Method

The method can be adapted to weekly data by setting as an approximation of the incubation time and . However, in small communes, we can see values that are difficult to defend (for example, we observe an effective R of 3.5 in “Los Muermos” during the 25th epidemiological). For reference, the maximum and minimum values were 4.43 and -3.64.

## Robert Koch Institute method

The method is almost indistinguishable from the Cislaghi method if weekly data is used.

## Wallinga-Lipsitch method

The method gives much fewer volatile results than the competitors, although it depends on the prior distribution defined for the generation time. For the moment, the prior recommended in the consulted document has been used, adjusted to deal with weekly data instead of daily.

# 4. Possible usages of the effective R estimates

In order to illustrate how an estimated effective-R at the communal-weakly level can be used an input for answering interesting question for public policy, we consider the problem of estimating the relationship between communal effective-R and the lagged phase in the Step-by-Step plan. In order to assess this relationship, we consider the following mixed model:

where represents the effective-R during the *t-*th epidemiological week at the *i-*th commune in the *r-*th region, is the associated vector of covariates, and is an unknown parameter vector. Among the covariates, we included variables accounting for: the political importance of the commune (e.g., being a provincial or regional capital), its international connectivity (e.g., having an port or an airport), the fraction of its population between 20 and 64 years, its socio-demographic development (measured by 2 proxy variables: a socio-economic development index created by the Chilean Observatory of Public Health, and a rurality index created by the Ministry of Social Development), the lagged fraction of its population that has received 1 and 2 doses of any vaccine, the lagged fraction of its neighbor population in quarantine, the lagged number of PCR exams (as a fraction of the commune population), and the lagged phase in the Step-by-Step plan[[3]](#footnote-3). The lag order of each variable was selected independently, in such a way that the Akaike information criteria is minimized (assuming that the lag order of the other variables is fixed to 1). Table 1 shows estimated coefficients associated with the phase in the Step-by-Step plan:

|  |  |  |
| --- | --- | --- |
| Variable | Estimated coefficient | t-statistic |
| phase 2 (lag 1) | 0.0673 | 11.5 |
| phase 3 (lag 1) | 0.161 | 26.2 |
| phase 4 (lag 1) | 0.255 | 35.4 |

As expected, the less restrictive the phase, the greater the effective R; and these effects are all statistically significant.

# 5. Conclusion

The methods for estimating the effective R by commune and week produced a reasonable median effective R. Of course, the variability of these estimators at the commune level makes it impossible to use them as a final product. However, because of how they are constructed, they are still valid as a response variable in a regression model.

In case we have to choose one, we recommend using the Wallinga-Lipsitch method if stability is desired. If interpretability is desired, the Cislaghi, RKI, or JRC methods are a good choice. However, caution should be taken not to use communes where they are unacceptably imprecise, as these latter indicators have an interpretation even if they are not accepted as a good measure of the effective R.

Systrom's method has potential, but the model should be adapted to guarantee a priori that the effective Rs are positive. Even if the prior specification of the model does not have any conjugacy property, it is still estimable in software such as [Stan](https://mc-stan.org/), if the model of each commune is estimated separately.

# References

[1] K. J. Rothman, S. Greenland, and T. L. Lash. Modern epidemiology. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins, 2008.

[2] T. Asikainen and A. Annunziato. Effective reproduction number estimation from data series. Technical report. Joint Research Centre (European Commission), 2020. doi: [10.2760/036156](http://dx.doi.org/10.2760/036156).

[3] http://systrom.com/blog/the-metric-we-need-to-manage-covid-19/

[4] C. Cislaghi. Un cruscotto per monitorare l’evoluzione dell’epidemia. 2020. url: [https://www.scienzainrete.it/articolo/cruscotto-monitorare-levoluzione- dellepidemia/cesare-cislaghi/2020-04-09](https://www.scienzainrete.it/articolo/cruscotto-monitorare-levoluzione-%20dellepidemia/cesare-cislaghi/2020-04-09) (visited on 04/09/2021).

[5] Robert Koch Institut. Epidemiologisches Bulletin 17. 2020. url: <https://www.rki.de/DE/Content/Infekt/EpidBull/Archiv/2020/Ausgaben/17_20.html> (visited on 04/09/2021).

[6] K. Systrom. The Metric We Need to Manage COVID-19. 2020. url: <http://systrom.com/blog/the-metric-we-need-to-manage-covid-19/> (visited on 06/03/2020).

[7] J Wallinga and M Lipsitch. “How generation intervals shape the relationship between growth rates and reproductive numbers”. In: Proceedings of the Royal Society B: Biological Sciences 274.1609 (2007), pp. 599–604. doi: 10.1098/rspb.2006.3754. url: https://royalsocietypublishing.org/doi/abs/10.1098/rspb.2006.3754.

[8] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). Bayesian Data Analysis.

[9] Stan Development Team. YEAR. Stan Modeling Language Users Guide and Reference Manual, 2.21.0. [https://mc-stan.org](https://mc-stan.org/)

[10] Martinez-Beneito, M. A., Mateu, J., and Botella-Rocamora, P. (2020). Spatio-temporal small area surveillance of the covid-19 pandemics. arXiv e-prints, page arXiv:2011.03938

[11] Leroux, B. G., Lei, X., and Breslow, N. (2000). Estimation of disease rates in small areas: A new mixed model for spatial dependence. pages 179–191. Springer New York.

[12] Cori, A., Ferguson, N. M., Fraser, C., and Cauchemez, S. (2013). A new framework and soft-ware to estimate time-varying reproduction numbers during epidemics. American Journal ofEpidemiology, 178:1505–1512.

[13] Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. Annals of the Institute of Statistical Mathematics, 43:1–20.

[14] Rue, H., Lindgren, F., Simpson, D., Martino, S., Teixeira, E., Bakka, H., Riebler, A., and Fuglstad, G-A (2018) INLA: Full Bayesian Analysis of Latent Gaussian Models Using Integrated Nested Laplace Approximations.

# Appendix

# Replication Codes

All the codes needed to replicate this report are in the repository https://github.com/COVID0248/covid-re. To rebuild our database with the estimated effective R at the communal-weekly level, follow these steps:

1. Clone the repository:

git clone https://github.com/COVID0248/covid-re

1. Open R and set the project directory as your reference directory.
2. Restore the virtual environment of the project:

* renv::restore(prompt = FALSE)

Note that the installation of the sf library will only be successful if you have its dependencies. For more information, visit the [library’s website](https://r-spatial.github.io/sf/). Also note that the rstan library uses many artifacts (i.e., dependencies that are not other R packages), so it is difficult to ensure its automatic installation on any computer (at least, not without Docker). In case rstan fails, it is recommended to reinstall it manually. For more information, visit the [library website](https://mc-stan.org/users/interfaces/rstan).

1. Rebuild the DB.DD. with effective R’s:

* targets::tar\_make(r)

1. Extract the estimations as a tibble from the built project:

* dfr <- targets::tar\_read(r)

1. If you also want to reproduce the graphics in this document, you can proceed in the same way. For example, if you want to reproduce the graph with the 10th percentile of the effective R in each commune, run the following code:

* targets::tar\_make(plot\_r\_p10\_en)  
  targets::tar\_read(plot\_r\_p10\_en)

If you want to render the charts with the 50th and 90th percentiles, just replace plot\_r\_p10 with plot\_r\_p50\_en or plot\_r\_p90\_en, as appropriate. The targets associated with the 5 figures in the main document are:

1. plot\_r\_p10\_en.
2. plot\_r\_p50\_en.
3. plot\_r\_p90\_en
4. plot\_r\_bp\_en.
5. plot\_r\_ts\_en.

Hence, a reader can reproduce Figure 5 by running the following code:

targets::tar\_make(plot\_r\_ts\_en)  
targets::tar\_read(plot\_r\_ts\_en)

1. Available at <https://covid-19vis.cmm.uchile.cl/geo> [↑](#footnote-ref-1)
2. Consulted in <https://www.gob.cl/coronavirus/pasoapaso/>. [↑](#footnote-ref-2)
3. Most of these variables are available at <https://github.com/MinCiencia/Datos-COVID19>. We use products 7 (# of PCR tests), 15 (# of new cases), 74 (phase in the Step-by-Step plan), 80 (people with their first vaccine), and 29 (quarantine status), [↑](#footnote-ref-3)