Discrete random SIR models for Covid-19 pandemic in Uruguay

Enrique M. Cabaña

Former professor at Universidad de la República
Montevideo, Uruguay

Abstract

Two models for Covid evolution are discussed: both are discrete, based on Markov chains with random thinnings.

The simpler one, Rvac2021, adds the effect of vaccination and explores the use of Google mobility records for short-term predictions.

The other one exploits the availability of data including the individual histories of Covid-19 patients, during a short initial period of the epidemic.

Both models can be run in an interactive internet platform, with automatic access to data taken from the epidemic evolution and the vaccination process in Uruguay.

1- COVID in Uruguay from March 2020 to September 2021: Rvac2021

2- The discrete SIR model: a sum of individual Markov chains

   - The simplest SIR model

3- A model for Covid and the effect of vaccination

   - Classes or states of the model
   - The model random dynamics
   - Graphical output
   - Critical examination of results for simple extrapolation of β
   - Short-term prediction of the contagious rate based on social mobility
   - A counterfactual discussion about lethality

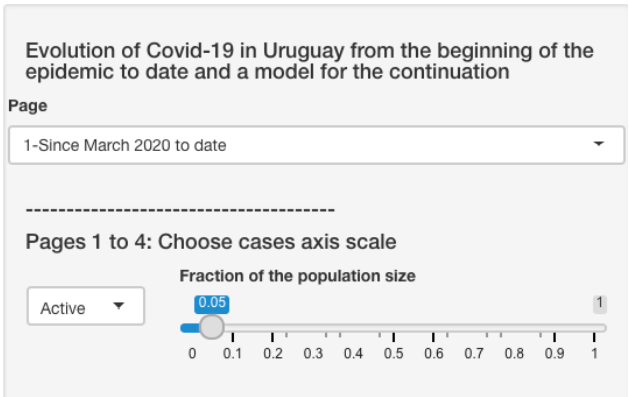
4- A data-adapted model for the beginning of the pandemic

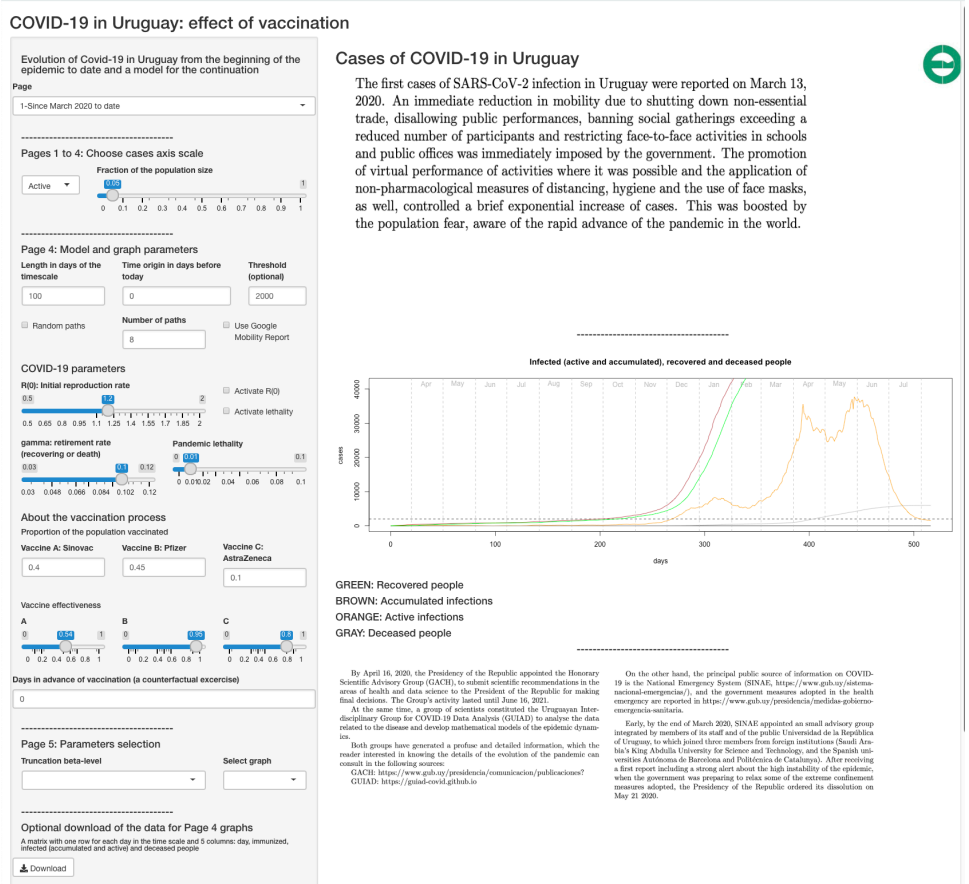
5- Adding a hidden state to Rvac and estimating the underreport

Let us first introduce the interactive *shiny* application *Rvac2021* `https://emcabana.shinyapps.io/Rvac2021` and use it to offer a bird's eye-view of the COVID in Uruguay from March 2020 to date.

The application starts with a screen that contains graphs of the amount of active infections and of the cumulative numbers of people infected with, immunized against and deceased by Covid-19 in the Country, with a population of approximately 3 450 000 inhabitants.

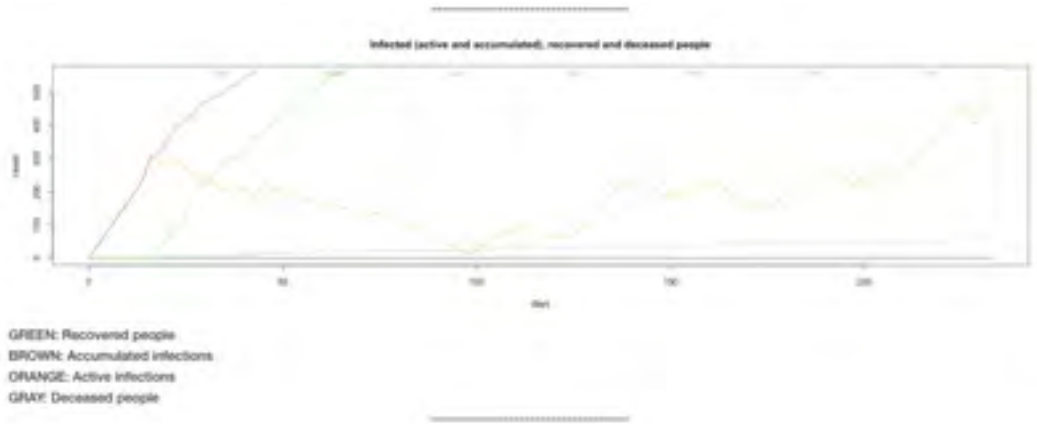
Since the amounts of these classes are very different, the user can conveniently choose the ordinate scale. Next figure shows the whole page with the scale adapted to display the active infections, and details the pages and scale selectors.





Remark. Days are counted from March 13, 2020, first day with detected infections in the country.

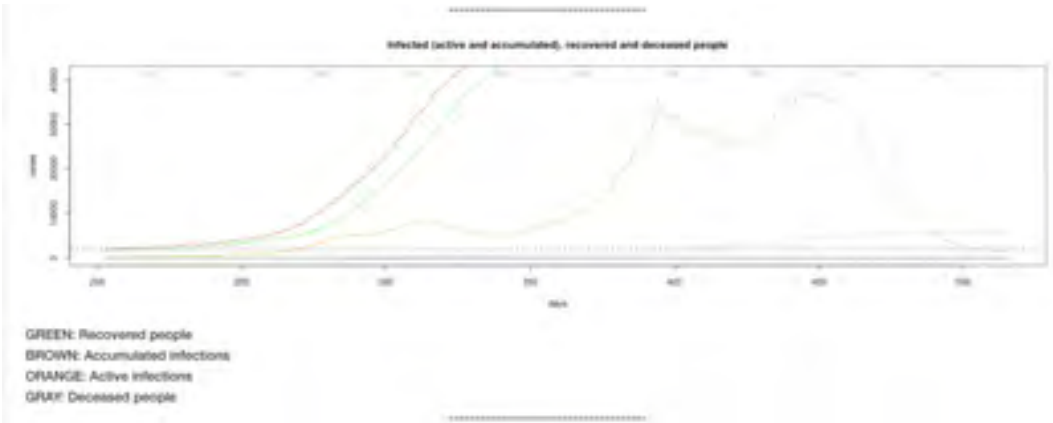
The next two pages of Rvac2021 show a striking different degree of success in managing the epidemic during the past and the present years.



During most of 2020 the daily number of new infections allowed the epidemic to be controlled by TETRIS[1], and it was not until the last months of the year that an increase in mobility joined with the government adoption of insufficient containment measures caused the situation to get out of control.

[1] TEst-TRace-ISolate strategy.

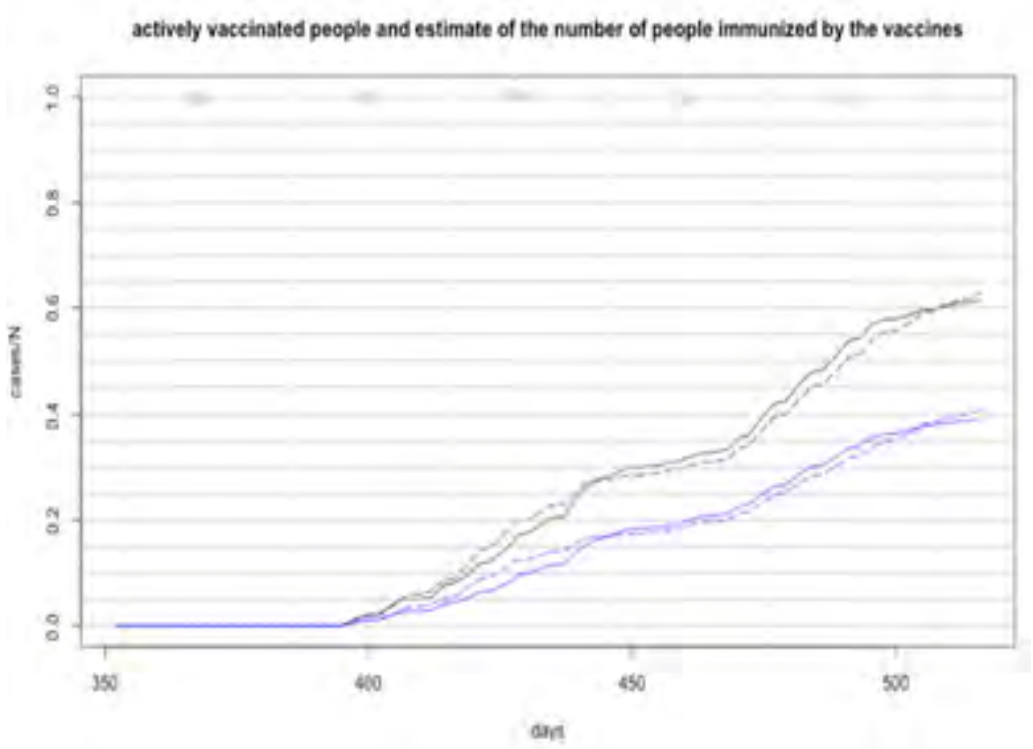
An intense vaccination plan started on February 27, 2021.
As of March 2021, in parallel with one of the best levels of vaccination against COVID-19 of Latin America, there was an exponential increase in new positive cases detected each day.



The number of active cases and deaths increased so much that Uruguay registered the highest COVID records in the world, measured as proportions with respect to the population size[2].

[2] More that 1000 active infections per 100000 inhabitants by the end of May

The graph at the bottom of page 3 describes the process of vaccination: the black lines indicate the number of people that received the second dose 15 or more days before, and the blue lines estimate the amount of immunized people.



The classical SIR model[3] partitions a population of constant size N into classes according to the health status of each member. In the simplest case, the classes are

S: susceptible people that can acquire the illness by contagion,

I: active infected and infectious people going through the disease

R: removed people, either recovered or deceased,

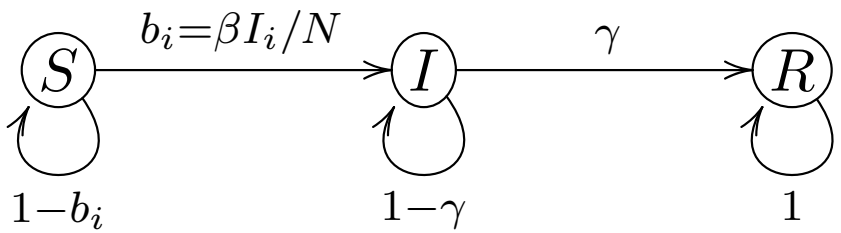
and the equations linking the sizes $S(t), I(t), R(t)$ at time t are

$$\frac{dS(t)}{dt} = -\beta S(t)I(t)/N, \frac{dI(t)}{dt} = (\beta S(t)/N - \gamma)I(t), \frac{dR(t)}{dt} = \gamma I(t).$$

From these equations, the interpretation of the parameters is plain.

[3] Kermack, W. O. and McKendrick, A. G. *A contribution to the mathematical theory of epidemics* Proceedings of the Royal Society of London Series A, **115**,700-721, 1927

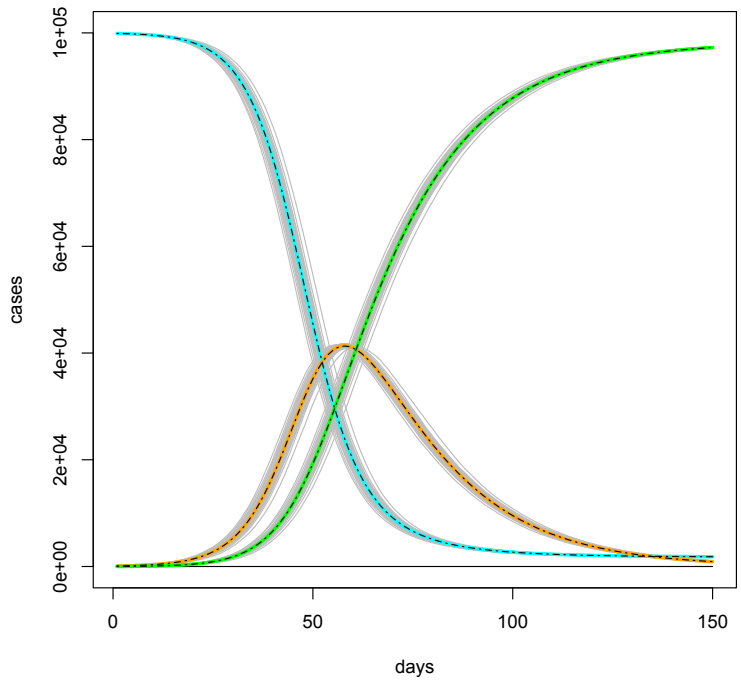
A random discrete version, both in time and sizes, is obtained by assuming that each member of the population describes independently a random path of the Markov chain:



The number of people in each class $X_i := (S_i, I_i, R_i)$ is a Markov chain with transitions $SI_i \sim \text{Bin}(S_i, b_i)$ from S to I and $IR_i \sim \text{Bin}(I_i, \gamma)$ from I to R.

By replacing each random transition SI_i and IR_i by the conditional expectations given $(X_h)_{h\leq i}$, namely $S_i b_i$ and $I_i\gamma$, a deterministic approach to the expectation of the random paths is obtained. Next figure shows that the approximation is very good.

Deterministic paths in color (S: cyan, I: orange, R: green), 50 random paths in gray background and 10000 random paths averages in dotted black lines. The parameters, arbitrarily chosen, are $N = 100000, \beta = 0.2, \gamma = 0.05, R_0 = 0, I_0 = 100$.



Our model for COVID is the subject of Page 4. The classes are the following:

S - Susceptible: people with no active vaccine inoculation that have not been documented as infected with SARS-CoV-2. (One person is said to be actively vaccinated 14 days or more after receiving the second dose of the vaccine. The model disregards the effect of the first dose of the vaccine).

SV^{ν}- ($\nu \in \{A, B, C\}$): People who remain not immunized after been actively inoculated with vaccine ν.

I^{ν}- ($\nu \in \{A, B, C\}$): People actively inoculated with vaccine ν, who suffer an active infection.

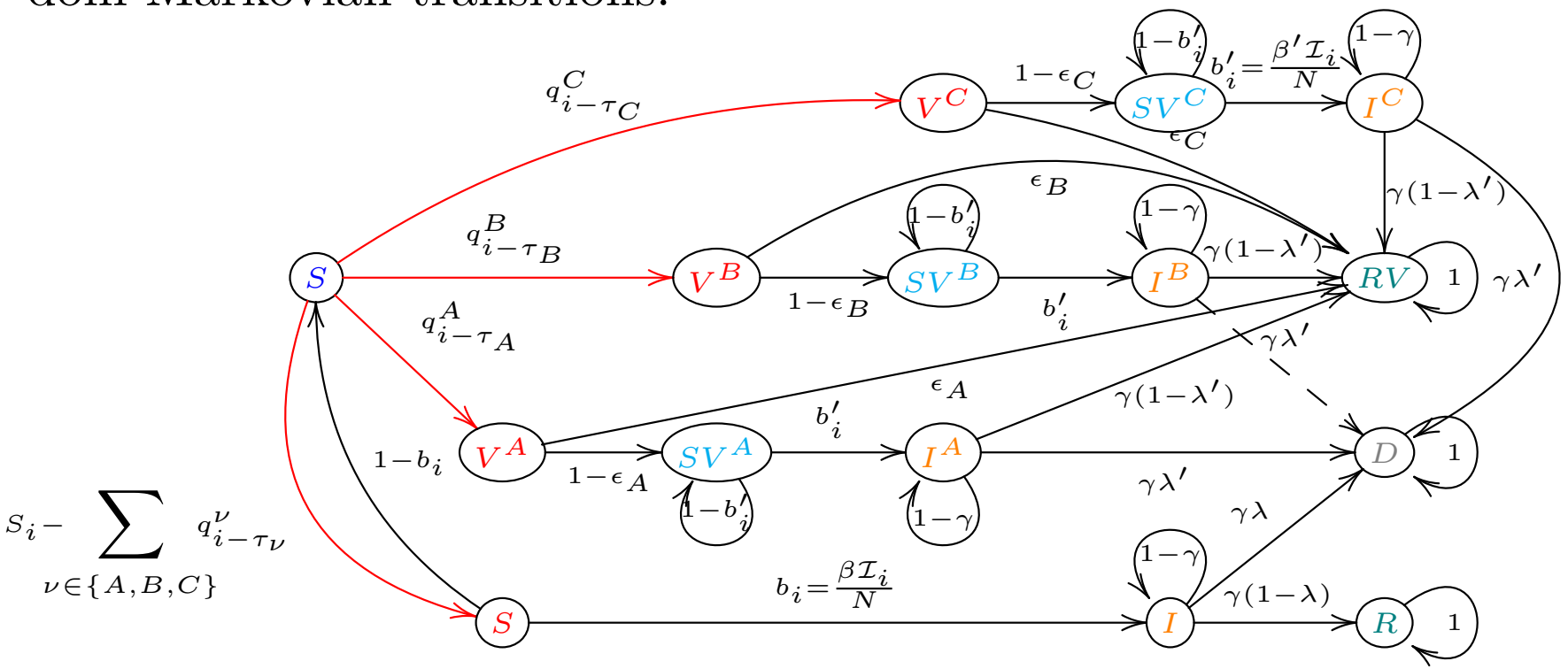
RV- Immunized people, for having been vaccinated.

I - Nonvaccinated people with active infection.

R - Recovered people.

D - Deceased people.

The model adds a few (red) deterministic movements to the random Markovian transitions:



The vaccines are A: Sinovac, B: Pfizer-BioNTech and C: AstraZeneca and their efficacies are assumed to be $\epsilon_A = 54\%, \epsilon_B = 95\%$ and $\epsilon_C = 80\%$. The delays from the inoculation up to the effectiveness of the vaccine are $\tau_A = \tau_C = 42$ days, and τ_B is also 42 for vaccines given until April 9 and 63 days from April 10 onwards, because the time between doses was enlarged to expedite the administration of the first doses to a broader part of the population.

The main diagram in Page 4 contains either the deterministic graphs for contagious rate β estimated from the COVID data available at the initial day, or a number (chosen optionally) of random paths for the same estimated β:

$$\mathcal{I} = (\mathcal{I}_i := I_i + \sum_{\nu \in \{A,B,C\}} I_i^{\nu})_{t_{ini} \le i \le t_{end}} \text{ (orange: active infections)},$$

$$J = (J_i := \mathcal{I}_i + R_i + D_i)_{t_{ini} \le i \le t_{end}} \text{ (brown: cumulative infections)},$$

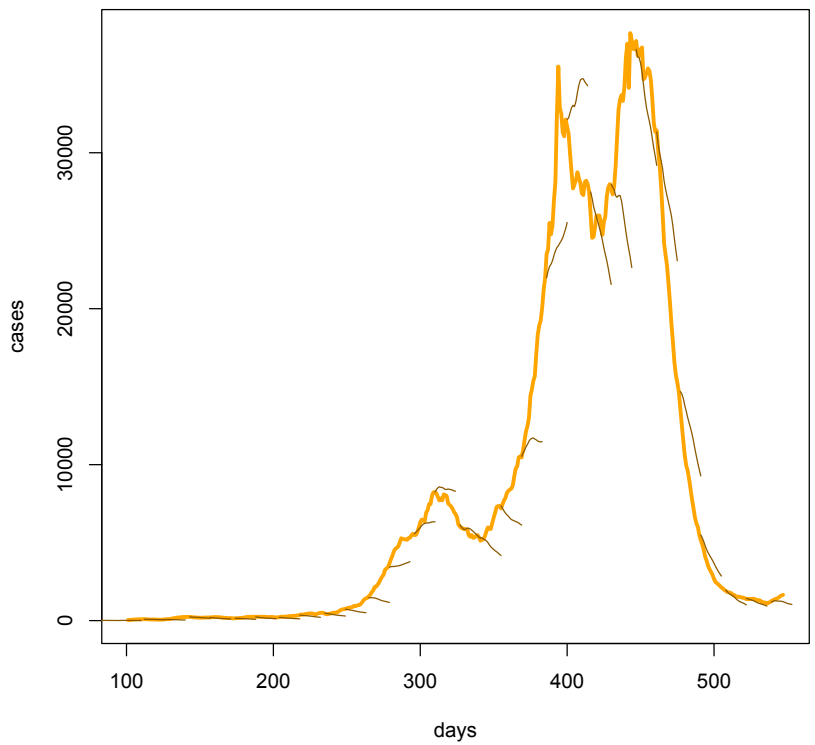
$$K = (K_i := R_i + RV_i))_{t_{ini} \le i \le t_{end}} \text{ (green: immunized people) and}$$

$$D = (D_i)_{t_{ini} \le i \le t_{end}} \text{ (gray: amount of deceased people)}.$$

The initial day is t_{ini}= *yesterday* by default, but the user can choose a previous date by means of the selector *"Time origin in days before yesterday"*. The final day t_{end} is determined by the selector *"Length in days of the timescale"*.

The Covid epidemic and the process described by the model are highly unstable, and therefore it should not be expected that predictions obtained from the model, even when it is calibrated according to the observed historical sizes of the classes, would provide a trustworthy anticipation of the future evolution of the epidemic.

In fact, the model deterministic 15-days forecasts of active infections computed the first and fifteenth day of each month are very different from the ones that actually occurred.



The forecasts in previous page are computed with the contagions rate β predicted by an AR model adjusting the past estimations from available data. It can be expected that a model with a better description of the future evolution of β, if it could be foreseen, would improve the quality of the prediction.

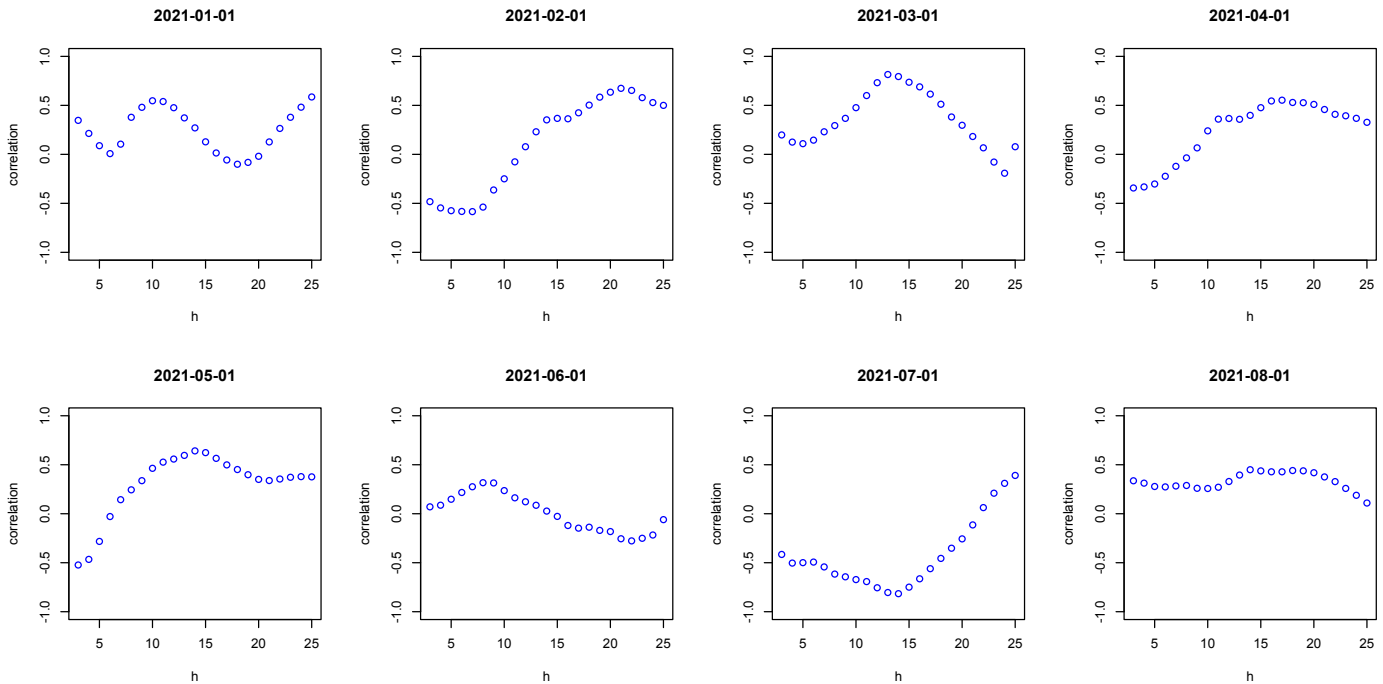
The rate β is roughly proportional to the average number of social contacts of susceptible people. Measures of such contacts are not available, but Google publishes reports showing movement trends over time sorted by geographical areas and classified into various categories of places[4].

[4] Google, *COVID-19 Community Mobility Reports*, https://www.gstatic.com/covid19/mobility/Region_Mobility_Report_CSVs.zip (2020, daily updated)

We adopt the series m_i (i ranges from February 15, 2020 to a day near *yesterday*) obtained by subtracting the mobility records of *retail and recreation* and *workplaces* from the ones of *residential mobility* as a measure of mobility. Next we explore the correlation between the contagious rate β and the mobility m.

The probability that a susceptible person be infected at day i is $b_i = \beta I_i/N$.
The expected number of contagions at day i is $b_i S_i$, estimated by the average $\overline{SI}_i = \frac{1}{7}\sum_{j=i-6}^{i} SI_j$.
This provides the estimate $\tilde{\beta}_i = N\overline{SI}_i/(I_i S_i)$ for β.

Since the contagions due to a contact that occurs one day are registered some days later due to the time of incubation and the delay until the tests of infection are applied and have a positive outcome, we explore the correlation between $\tilde{\beta}_i$ and m_{i-h} for $3 \leq h \leq 25$, for $i \in I := [t_{ini} - \Delta t, t_{ini}]$ $(\Delta t = 28)$. The results for the values of t_{ini} indicated next to each diagram are shown in next figure.



When the control button *Use Google Mobility Report* is off, Rvac adjusts an autoregressive (AR) model to the series $(\tilde{\beta}_i)_{i \leq t_{ini}}$ and substitutes the prediction $(\tilde{\beta}_i)_{i > t_{ini}}$ for $(\beta_i)_{t_{ini} < i \leq t_{end}}$.

If instead that button is on, the series m_i is extrapolated by adjusting an AR model.
Then, with $H = \arg\max_h \mathrm{correl}((\tilde{\beta}_i)_{i \in I}, (m_{i-h})_{i \in I})$, Rvac adjusts recursively a VAR to the series $(\tilde{\beta}_i, m_{i-H})_{i \leq t_{ini}+k}$, $k = 0, 1, 2, \ldots$ and uses it to predict $\tilde{\beta}_{t_{ini}+k+1}$.

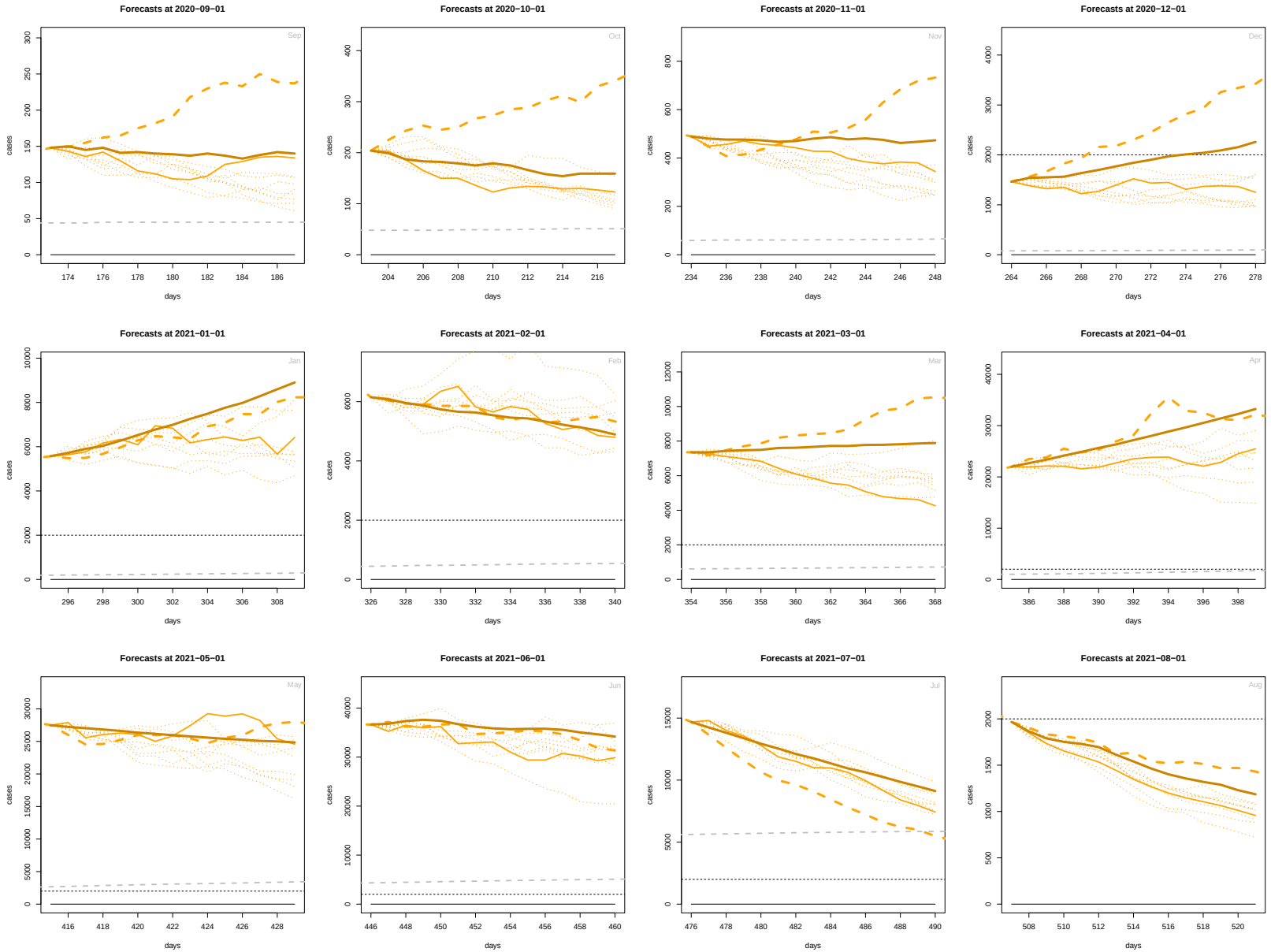
The complexities of AR models are chosen by using Akaike Information Criterion (AIC).

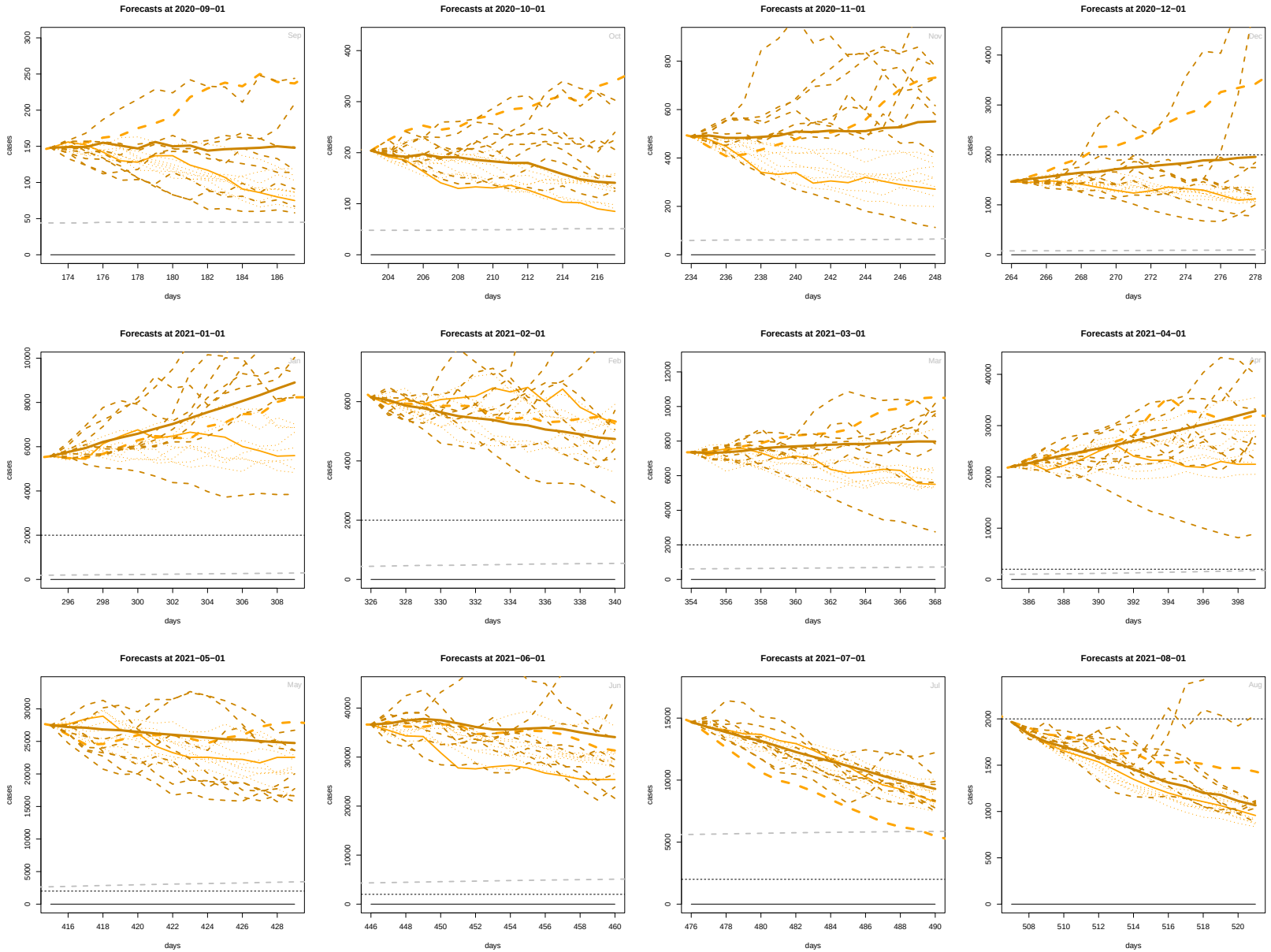
This procedure does not substantially improve the results, but provides an estimation of the variability of the parameters governing the individual Markov chains followed by each person.

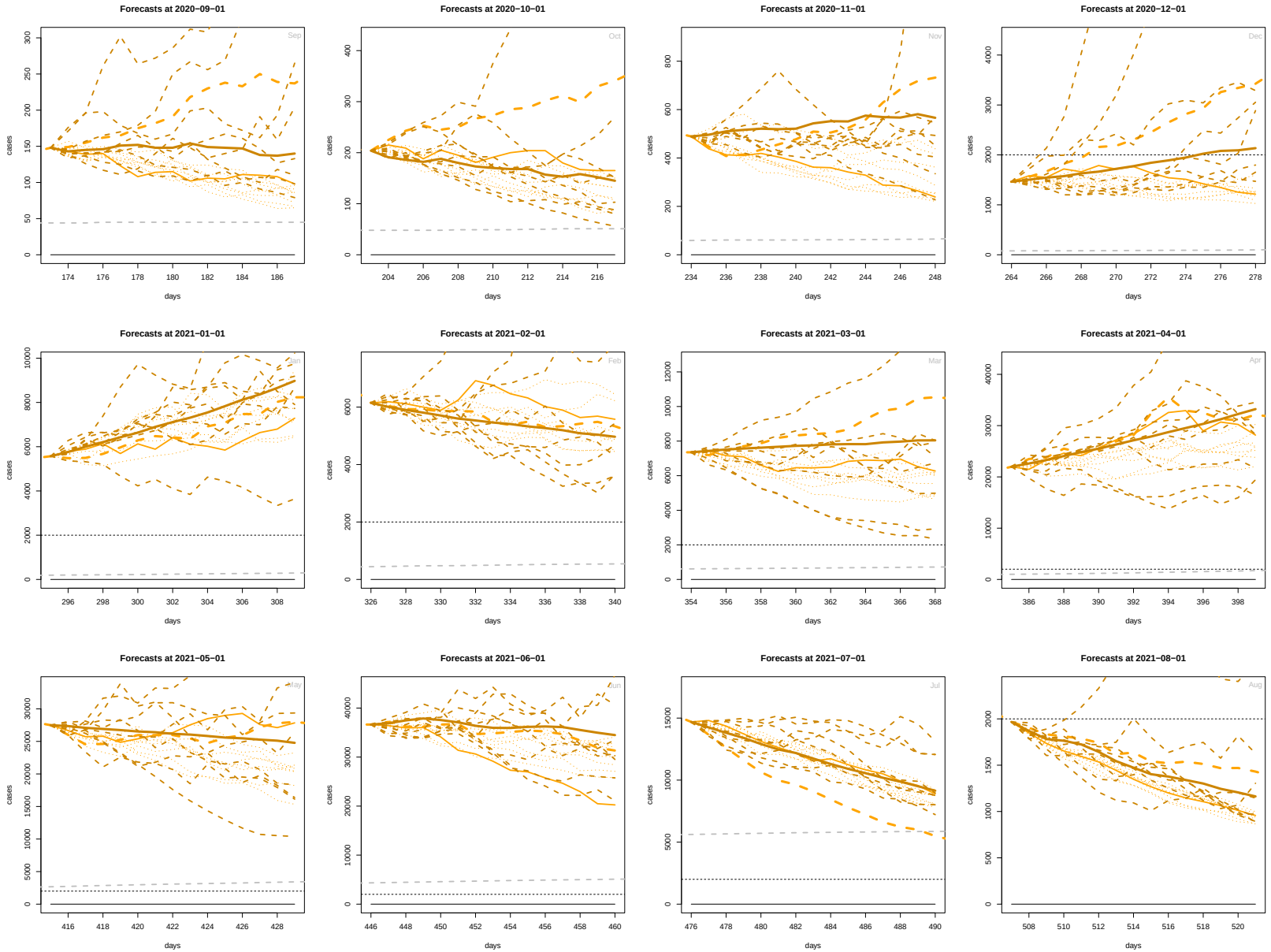
The new random paths exhibit a larger variability than the ones computed by just extrapolating $\tilde{\beta}$.

While showing greater irregularity, they confirm the impossibility of obtaining reliable predictions in a process with such an enormous degree of instability as the Covid-19 pandemic.

Next diagrams show eight random paths computed by Rvac2021 with the data available the first day of each month from September 2020 to August 2021 (Orange with mobility off, and dark orange with mobility on) besides the deterministic ones in thicker lines.

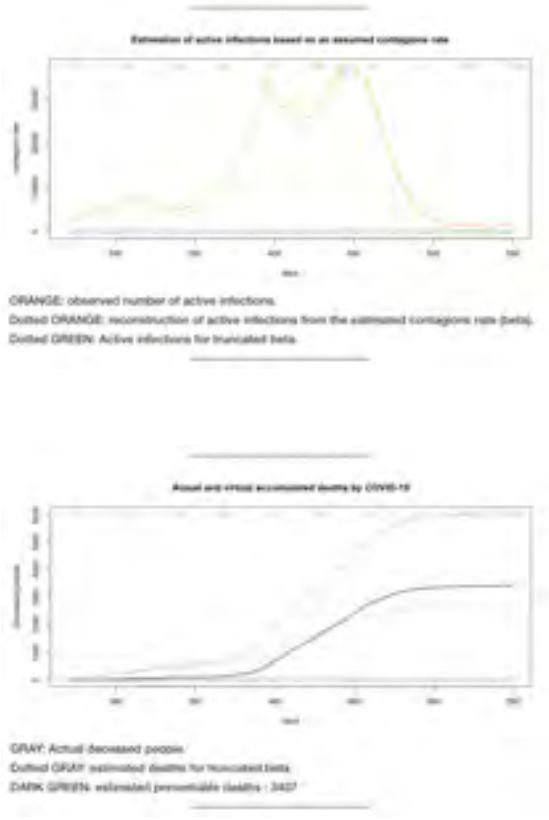

Forecasts at 2020−09−01
Sep
Forecasts at 2020−10−01
Oct
Forecasts at 2020−11−01
Nov
Forecasts at 2020−12−01
Dec
Forecasts at 2021−01−01
Jan
Forecasts at 2021−02−01
Feb
Forecasts at 2021−03−01
Mar
Forecasts at 2021−04−01
Apr
Forecasts at 2021−05−01
May
Forecasts at 2021−06−01
Jun
Forecasts at 2021−07−01
Jul
Forecasts at 2021−08−01
Aug
cases
days


Forecasts at 2020−09−01
Sep
Forecasts at 2020−10−01
Oct
Forecasts at 2020−11−01
Nov
Forecasts at 2020−12−01
Dec
Forecasts at 2021−01−01
Jan
Forecasts at 2021−02−01
Feb
Forecasts at 2021−03−01
Mar
Forecasts at 2021−04−01
Apr
Forecasts at 2021−05−01
May
Forecasts at 2021−06−01
Jun
Forecasts at 2021−07−01
Jul
Forecasts at 2021−08−01
Aug
cases
days


Forecasts at 2020−09−01
Sep
Forecasts at 2020−10−01
Oct
Forecasts at 2020−11−01
Nov
Forecasts at 2020−12−01
Dec
Forecasts at 2021−01−01
Jan
Forecasts at 2021−02−01
Feb
Forecasts at 2021−03−01
Mar
Forecasts at 2021−04−01
Apr
Forecasts at 2021−05−01
May
Forecasts at 2021−06−01
Jun
Forecasts at 2021−07−01
Jul
Forecasts at 2021−08−01
Aug
cases
days

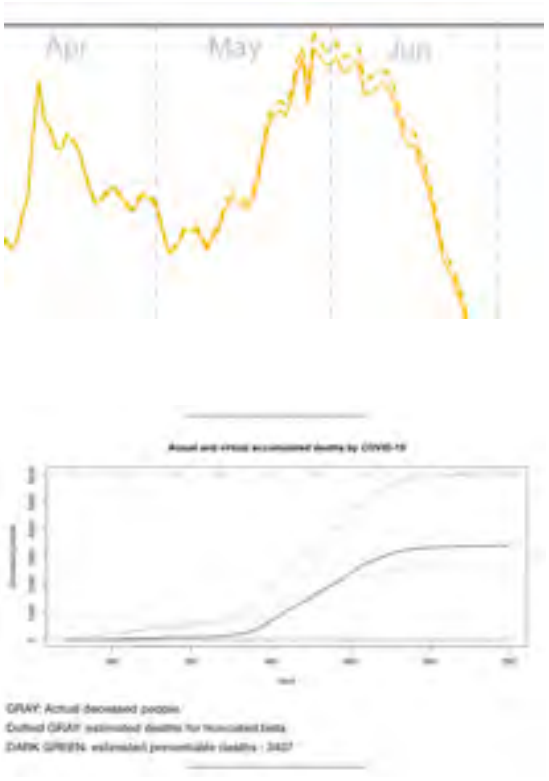
The last page of Rvac2021 is an attempt to quantify the number of deaths that could have been prevented if more severe measures to contain mobility had been taken.

The operation that leads to compute the estimate $\tilde{\beta}_i$ is inverted to reconstruct $\mathcal{I}_i$ from $\tilde{\beta}_i$, virtual paths of $\mathcal{I}_i$ are obtained by truncation of β at the level attained around December 1, 2020 and February 2021, when the government ignored warnings from its advisers, and the number of deceased people is estimated by assuming the same rate of lethality as actually occurred.



Number of deceased people:
Gray - Actual
Dotted gray - Virtual
Dark green - Preventable.

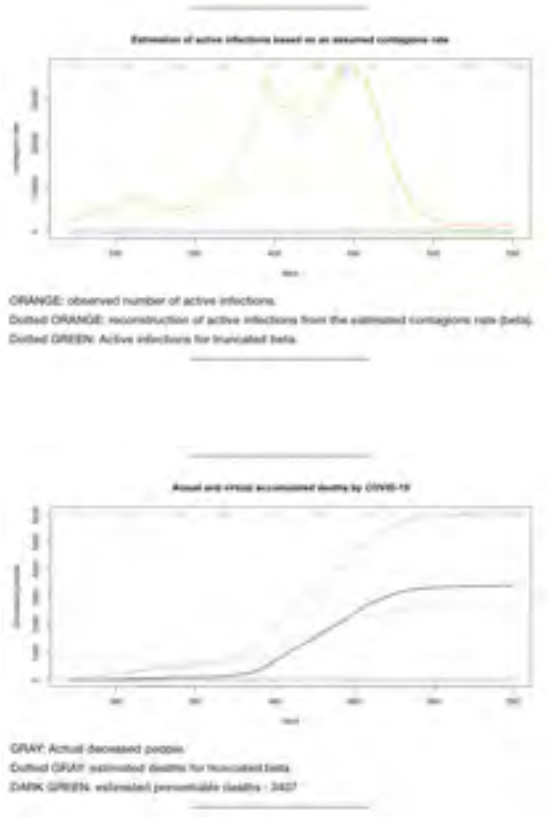
The last page of Rvac2021 is an attempt to quantify the number of deaths that could have been prevented if more severe measures to contain mobility had been taken.
The operation that leads to compute the estimate $\tilde{\beta}_i$ is inverted to reconstruct $\mathcal{I}_i$ from $\tilde{\beta}_i$, virtual paths of $\mathcal{I}_i$ are obtained by truncation of β at the level attained around December 1, 2020 and February 2021, when the government ignored warnings from its advisers, and the number of deceased people is estimated by assuming the same rate of lethality as actually occurred.



Number of deceased people:
Gray - Actual
Dotted gray - Virtual
Dark green - Preventable.

The last page of Rvac2021 is an attempt to quantify the number of deaths that could have been prevented if more severe measures to contain mobility had been taken.

The operation that leads to compute the estimate $\tilde{\beta}_i$ is inverted to reconstruct $\mathcal{I}_i$ from $\tilde{\beta}_i$, virtual paths of $\mathcal{I}_i$ are obtained by truncation of β at the level attained around December 1, 2020 and February 2021, when the government ignored warnings from its advisers, and the number of deceased people is estimated by assuming the same rate of lethality as actually occurred.



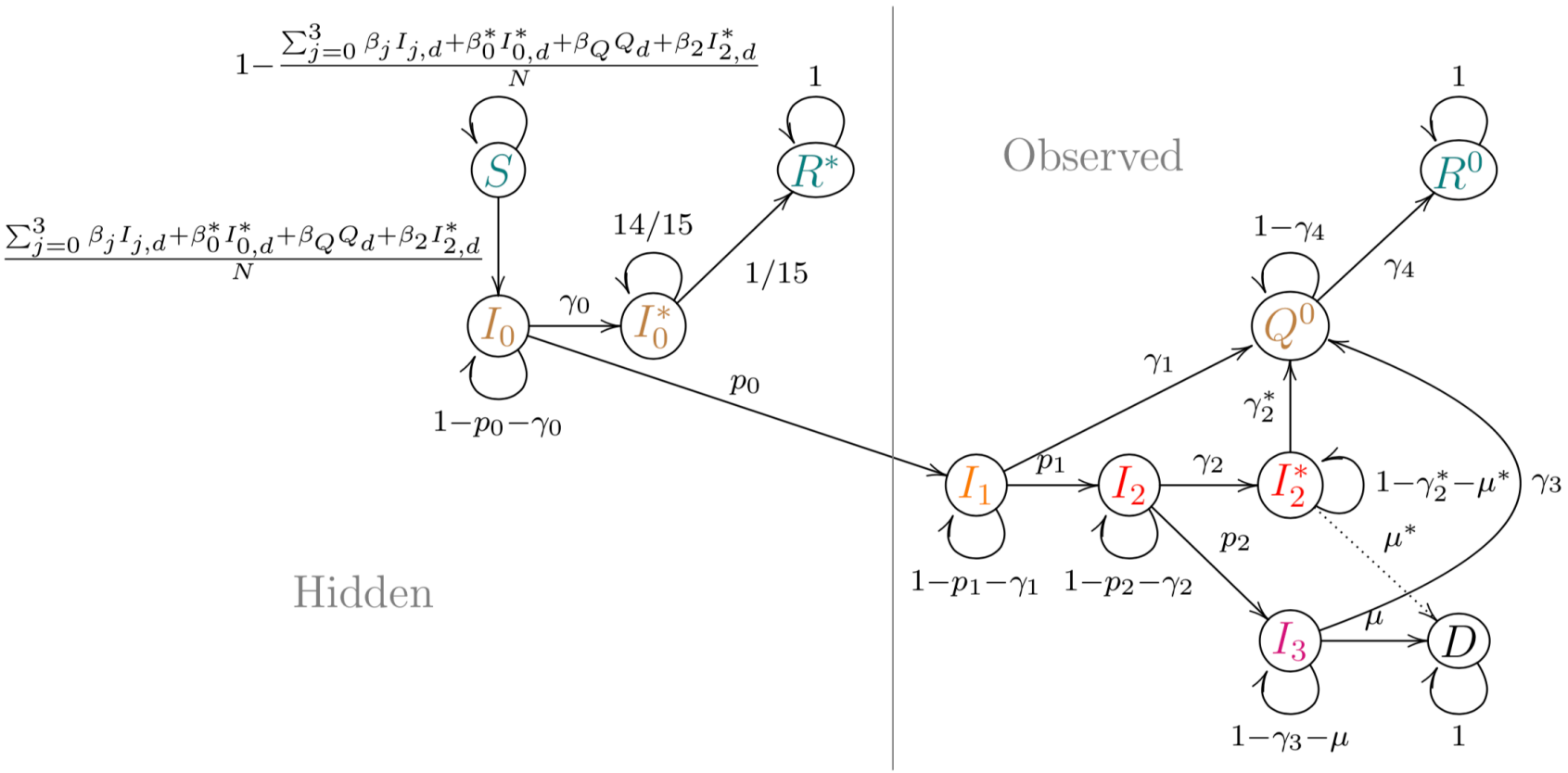
Number of deceased people:
Gray - Actual
Dotted gray - Virtual
Dark green - Preventable.

A few days after the confirmation of the first SARS-CoV-2 infections in Uruguay, the National System of Emergencies (SINAE) invited a group of scientists to advise it on an honorary basis about the Covid-19 pandemic. The group had access to classified information on the medical history of each individual infected with SARS-CoV-2, which included the dates of:

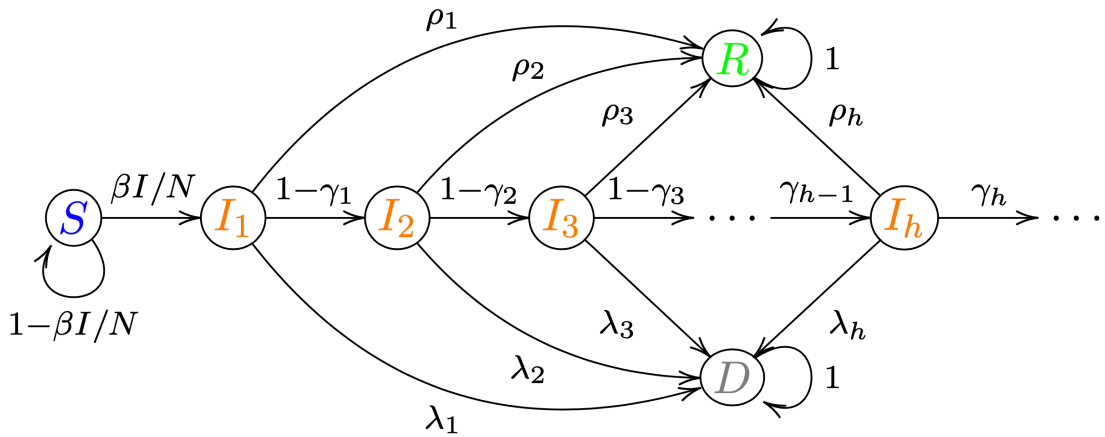
- infection,
- first symptoms,
- hospital confinement,
- transfer to intensive care unit,
- discharge from hospital,
- complete recovery or
- death,

whatever be appliable.

Having such amount of information, we proposed a model where each person describes the following Markov chain:



The infected (and infectious) people have been splitted into classes I_0: undetected, I_1: ambulatory, I_2: confined in hospital and I_3: in intensive care units.
The transition times from each state has the Geometric distribution with parameter equal to the exit probability. Starred classes have been introduced to allow the arbitrary selection of probabilities and expected times of transitions to different destinations. A detailed assignment of the exit time law can be obtained with the addition of one state for each day of stay:

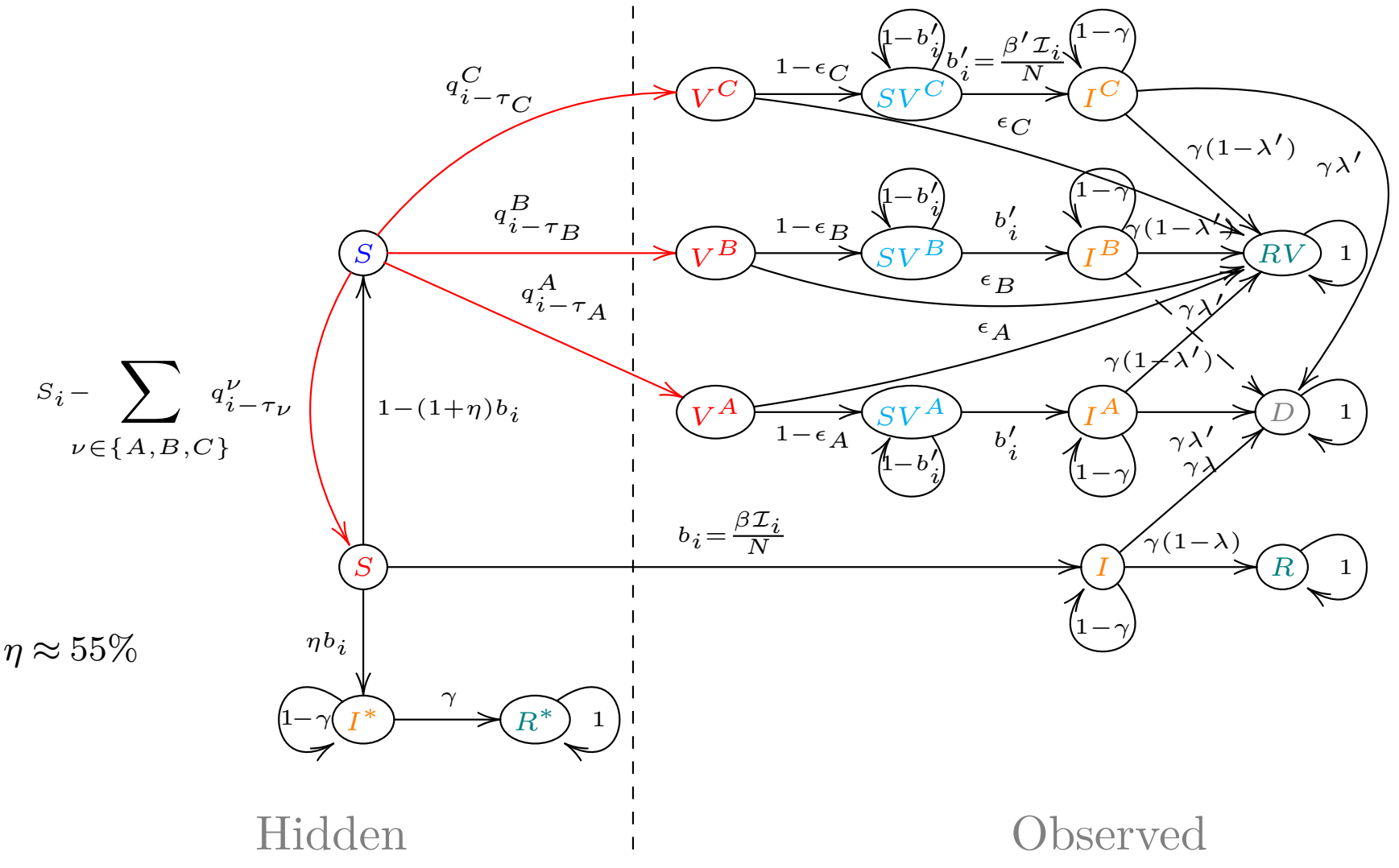


$$\rho_h + \lambda_h = \gamma_h, p_h^R = \rho_h \prod_{j<h}(1-\gamma_j), p_h^D = \lambda_h \prod_{j<h}(1-\gamma_j)$$

The chain diagram emphasizes the existence of hidden states. The available information allows to estimate separately the parameters concerning each visible state by maximum likelihood. Then the simplest way to estimate the remaining parameters, in particular, the amount of underreport, is to adjust the observed data by least squares. More accurate is the use of Viterbi algorithm, but it leads to more demanding computation because the number of states of the chain is very large.

We have discontinued the use of such detailed models because after receiving a report of our group alerting about the risk of an exponential increase of infections due to the high instability of the epidemic, the President of the Republic ordered its dissolution.

However the addition of a hidden infectious state to Rvac allows us to estimate the underreport by optimizing the quadratic error of one- and two-weeks predictions: the underreported infections are estimated in 55% of the documented cases.


$S_i - \sum_{\nu\in\{A,B,C\}} q^{\nu}_{i-\tau_\nu}$
$\eta \approx 55\%$
$q^C_{i-\tau_C}$
$q^B_{i-\tau_B}$
$q^A_{i-\tau_A}$
$1-(1+\eta)b_i$
ηb_i
$b_i = \frac{\beta \mathcal{I}_i}{N}$
$b'_i = \frac{\beta' \mathcal{I}_i}{N}$
$1-\epsilon_C$
$1-\epsilon_B$
$1-\epsilon_A$
ϵ_C
ϵ_B
ϵ_A
$1-b'_i$
b'_i
$1-\gamma$
γ
$\gamma(1-\lambda')$
$\gamma\lambda'$
$\gamma(1-\lambda)$
$\gamma\lambda$
1
S
V^C
V^B
V^A
SV^C
SV^B
SV^A
I^C
I^B
I^A
RV
D
I
R
I^*
R^*
Hidden
Observed