Supplementary Material: Compartmental epidemic model to assess undocumented infections: applications to SARS-CoV-2 epidemics in Brazil

Guilherme S. Costa [1], Wesley Cota [1], and Silvio C. Ferreira [1,2]

[1]*Departamento de Física, Universidade Federal de Viçosa, 36570-900 Viçosa, Minas Gerais, Brazil*
[2]*National Institute of Science and Technology for Complex Systems, 22290-180, Rio de Janeiro, Brazil*
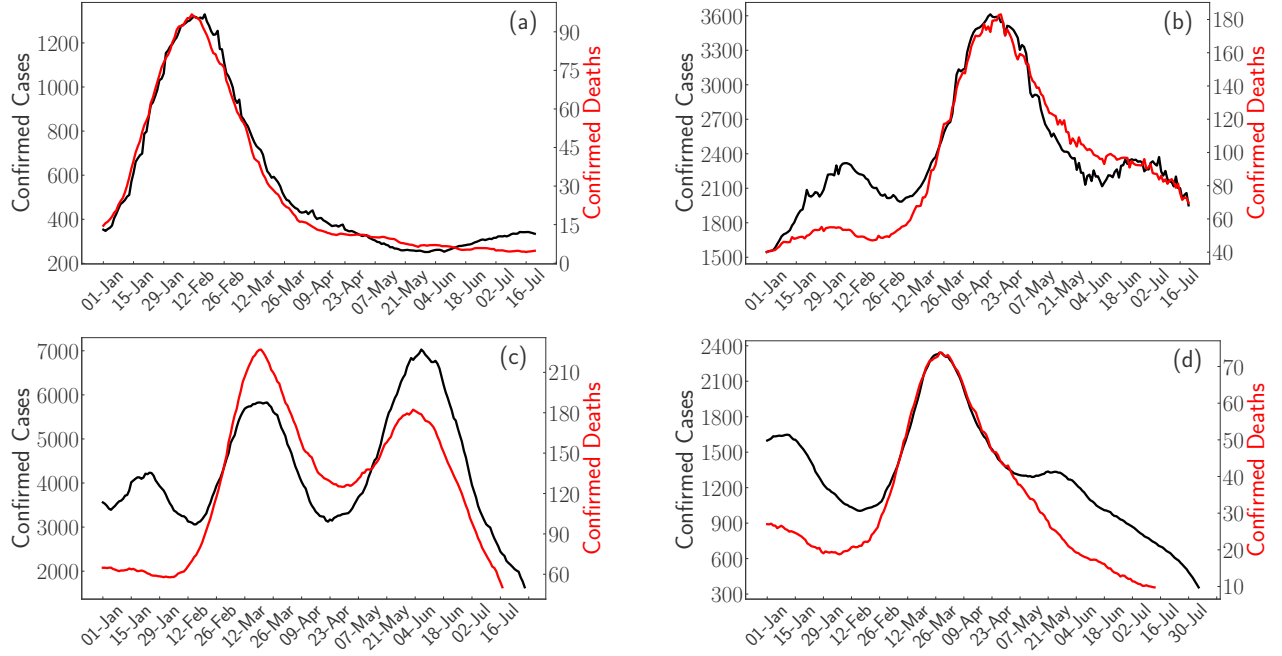
I. SUPPLEMENTARY FIGURES

FIG. S1. Time series of confirmed cases (left axis, black curves) and deaths (right axis, red curves) for (a) Manaus/AM, (b) São Paulo/SP, (c) PR state and (d) ES state. Time series of death counts are shifted in (a) 7, (b) 9, (c) 10, and (d) 20 days, respectively.
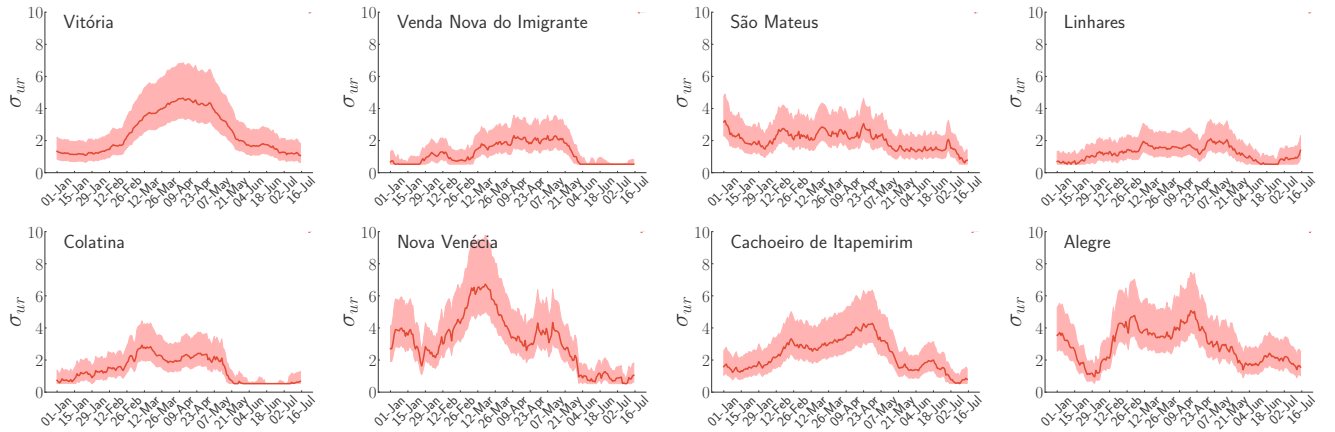

FIG. S2. Evolution of the under-reporting coefficient for immediate regions of ES state using $p_A/p_I = 0.1$ and time windows of 3 weeks. The regions are indicated in the upper-left corners for each panel.
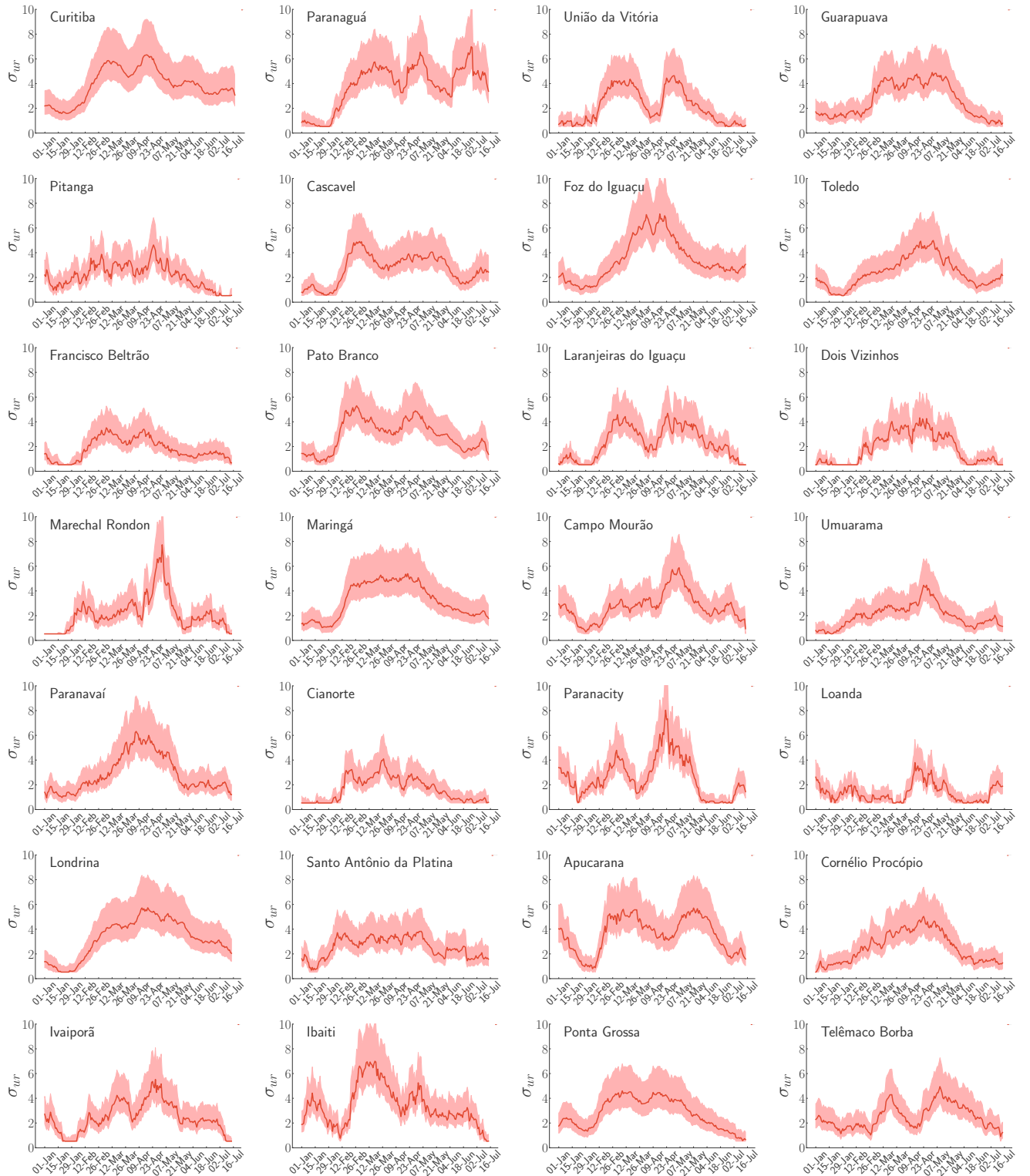
FIG. S3. Evolution of the under-reporting coefficient for immediate regions of PR state using $p_A/p_I = 0.1$ and time windows of 3 weeks. The regions are indicated in the upper-left corners for each panel.
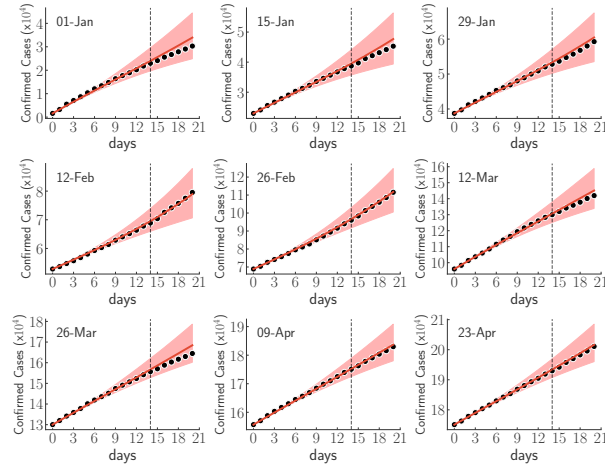
FIG. S4. Calibration curves for ES state in different time windows of 14 days indicated by the vertical lines. The initial day is indicated in the top of each panel. One week of forecasting is also shown. Symbols are the cumulative diagnosed cases while lines with shaded regions represent the calibrated curves and the corresponding confidence interval.
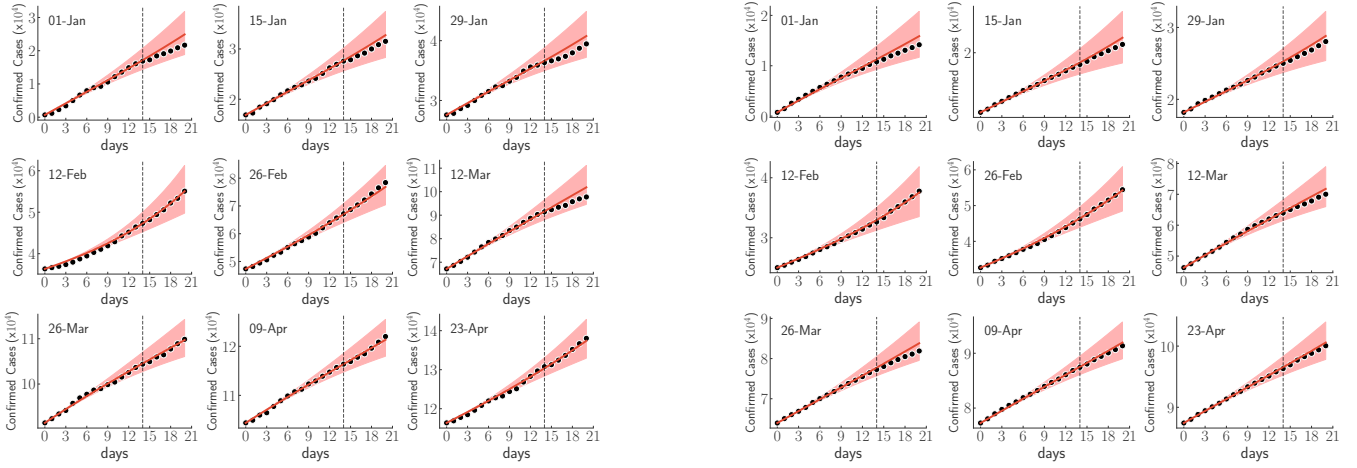

FIG. S5. Calibration curves for the immediate regions of (a-i) Curitiba and (j-r) Vitória in different time windows of 14 days indicated by the vertical lines. The initial day is indicated in the top of each panel. One week of forecasting is also shown. Symbols are the cumulative diagnosed cases while lines with shaded regions represent the calibrated curves and the corresponding confidence interval.
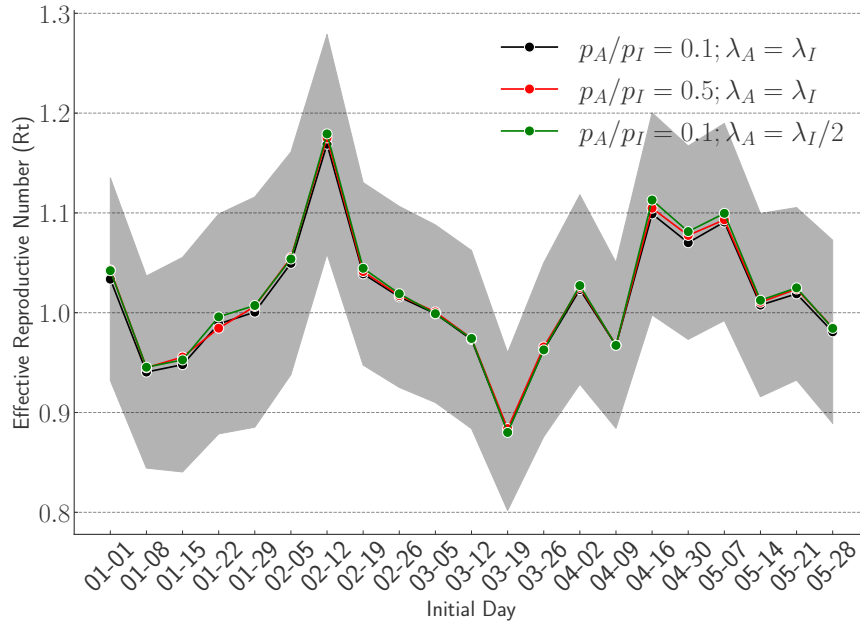
FIG. S6. Evolution of effective reproductive number computed for the PR state considering some epidemic and testing parameters. The interval of confidence of 95% is shown in the shaded region for the black curve.

II. DATA DESCRIPTION AND CODES

Dictionary of municipalities: The files (a) `dictES.csv` and (b) `dictPR.csv` of this supplemental material yield some information about municipalities of (a) ES (B) PR states. These files have six columns:

- `ID`: numeric key regarding calibration of confirmed cases time series

- `ibgeID`: official code to identify the city

- `name`: name of the city

- `intermID`: official code of intermediate region to which the city belongs

- `imedID`: official code of immediate region to which the city belongs

- `totPop2019`: population of the city estimated in 2019

Dictionaries of immediate and intermediate regions: The files (a) `dictImed.csv` and (b) `dictInterm.csv` of this supplemental material yield some information about (a) Immediate and (b) Intermediate regions. They were aggregated from `dictPR.csv` and `dictES.csv` files. These files have five columns:

- `ID`: numeric key regarding calibration of confirmed cases time series

- `imedID` or `intermID`: official code to identify the region

- `name`: name of the region

- `state`: state to which the region belongs

- `totPop2019`: population of the region estimated in 2019

Dictionary of states: The file `dictUF.csv` of this supplemental material yield some information about PR and ES states. They were aggregated from `dictPR.csv` and `dictES.csv` files. These files have five columns:

- `ID`: numeric key regarding calibration of confirmed cases time series

- `ibgeID`: official code to identify the state

- `name`: name of the state

- `uf`: abbreviation of the state's name

- `totPop2019`: population of the state estimated in 2019

Time series of cases and deaths: The files (a) `PR.csv`, (b) `ES.csv`, (c) `saopaulo.csv` and (d) `manaus.csv` of this supplemental material yield the time series of confirmed cases and deaths since April 1st, 2020 for (a) All cities of PR state, (b) All cities of ES state, (c) São Paulo city and (d) Manaus city. These files have seven columns:

- `date`: date

- `ibgeID`: official code to identify the city

- `newCases`: new confirmed cases on that day

- `newDeaths`: new confirmed deaths on that day

- `city`: name of the city

- `totalCases`: accumulated cases

- `totalDeaths`: accumulated deaths

Time series for calibration: Within files (a) `imed.zip` and (b) `state.zip` we have the time series of accumulated cases and fatality ratio, aggregated for different geographical levels. In this, we have two types of files: `casesXX.dat` (XX refers to the calibrating IDs mentioned before) are accumulated cases while `lethXX.dat` are the daily fatalities).

Main code for calibration : The file `calibra.f90` is a program written in Fortran that executes the calibration algorithm described on Methods section of the main paper 1000 times with different epidemiological parameters. This program has four inputs: the time series of accumulated cases and fatality, the initial date for calibration and the population of the region (state, city, etc). Besides that, this program has two output files: `epiQuantities.dat` and `hiddenCompart.dat`. The first has seven columns: days from the initial time, calibrated confirmed cases, reference cases, R_t, fraction of susceptible population, σ_{ur} and sample. On `hiddenCompart.dat`, we have time series for some compartments in the model: from left to right S, E, A, I, $C_A + C_I$, $R + R_I + R_A + D$ and sample number.

Calculation of underreporting coefficient σ_{ur}: the file `underreporting.ipynb` is a I-python script that calculates σ_{ur} starting from a time series of confirmed cases and deaths. At the end, it exhibits a graphic showing the evolution of this coefficient.

Template for figures: The majority of figures in this work were generated with `matplotlib` and `seaborn` packages of Python 3.7. File `format_covid19br.mplstyle` contains the template (font family and sizes) for generating those figures and graphics.