Supplementary Material: Evolution of transmission rate during the course of SARS-COV-2: patterns and determinants

Jie Zhu and Blanca Gallego

Methods

Our analysis is based on the publicly available data of the new confirmed daily cases reported by Johns Hopkins University (JHU) from 22ed of January until the 15th of May. Based on the released data, we attempted to estimate the mean values of the epidemiological parameters (i.e., the daily reproduction number (R_t) and the mortality rate (ξ)). We used calibrated parameters of the SIERD model to predict the reported incidence beyond the fitting period for 101 selected countries and regions.

SEIRD model calibration, fitting and prediction

We used an extension of the traditional SEIRD model that allows for variation in delays from onset to report and the detection rate of testing in each country. In this way, if S_k, E_k, I_k, R_k, and D_k represent respectively the susceptible, exposed, infectious, recovered and dead for each country or region k, their change in time is modelled as

$$dS_k(t) = -\frac{\beta_k(t)}{N_k(t)} S_k(t) I_k(t)\, dt \tag{1}$$

$$dE_k(t) = \frac{\beta_k(t)}{N_k(t)} S_k(t) I_k(t)\, dt - \sigma E_k(t)\, dt \tag{2}$$

$$dI_k(t) = \sigma E_k(t)\, dt - \gamma I_k(t)\, dt + \xi_k(t) I_k(t)\, dt \tag{3}$$

$$dR_k(t) = \gamma I_k(t)\, dt \tag{4}$$

$$dD_k(t) = \xi_k(t) I_k(t)\, dt \tag{5}$$

Where $N_k(t) = S_k(t) + E_k(t) + I_k(t) + R_k(t) + D_k(t)$ is the total population after the travel ban if we ignore births and deaths unrelated to COVID-19. The effective transmission rate, $\beta_k(t)$, *and* the mean fatality rate, $\xi_k(t)$, for each country are assumed to follow geometric Brownian motions:

$$d\log\beta_k(t) = \beta_k(t)\, dB(t) \tag{6}$$

$$d\log\xi_k(t) = \xi_k(t)\, dB(t) \tag{7}$$

Meanwhile, the clinical parameters, σ, the mean rate of becoming symptomatic (that is 1/incubation period), and γ, the mean rate of recovery (that is 1/duration of illness) are considered constant and obtained from the literature: $\sigma = (5.2\ days)^{-1}$ and $\gamma = (15.5\ days)^{-1}$. We therefore implicitly assume individuals become infectious when they are symptomatic.

The unknown parameters of interest, $\beta_k(t)$ and $\xi_k(t)$, are estimated by jointly fitting the daily observed cumulative number of confirmed cases ($\widehat{C_k}$), confirmed recoveries ($\widehat{R_k}$) and deaths ($\widehat{D_k}$) using

$$d\hat{C}(t) = \eta_k(t)\rho_k(\sigma E_k(t)\, dt - \gamma I_k(t)\, dt + \xi_k(t) I_k(t)\, dt), \tag{8}$$

$$d\hat{R}(t) = \eta_k(t)\rho_k\gamma I_k(t)\, dt, \text{and} \tag{9}$$

$$d\hat{D}(t) = \eta_k(t)\rho_k\xi_k(t)I_k(t)\,dt, \tag{10}$$

where ρ_k represents a delay from the actual incidence, recovery or death date to their reported date and $\eta_k(t)$, is the time-varying case detection rate in each country.

We estimated the time varying transmission rate,$\beta_k(t)$ and mortality rate, $\xi_k(t)$ using sequential Monte Carlo (SMC) by jointly fitting to three data sets:

1. Daily number of new cases reported in each country by date of onset. We only consider onsets since the first incidence in each country and up to 15th May 2020.
2. Daily number of recoveries reported in each country up to 15th May 2020
3. Daily number of deaths reported in each country up to 15th May 2020

The SMC process proposes trajectories of the unobserved state process according to the probabilistic rule specified in equations (1) to (10). For each country, we generate 300 trajectories $x_{n,\,j}$ with 5000 particles ($n = 1, 2, \ldots, 5000, \quad j = 1, 2, \ldots, 300$) for $x = \beta_k(t)$ and $\xi_k(t)$ respectively. We first used the model outputs to calculate expected trajectories for each of the datasets we were fitting to, then we used a negative binomial observation model to compute the likelihood of fitting to datasets 1–3, using the expected values from model outputs. The time varying basic reproduction number can be calculated as $R_t = \beta(t)/\gamma$. The value of assumed delay from onset to confirmation ρ_k, the detection rate at each time point $\eta_k(t)$, and the volatility of the random walks were found by grid search (see Figures 1 and 2). The assumed initial expected values of $\beta_k(0)$ and $\xi_k(0)$ are 0.

Profile likelihoods (see Figure 3) for each parameter were constructed based on the joint likelihood distribution. We assumed the outbreak started with 1 infectious individual in each country and the country's population was initially fully susceptible. The accuracy of the fitting is presented in Figure 4.

We ran our SEIRD model by fixing its parameters to those fitted on 15th of May 2020 and predicted the next 30-day reported cases in each country and compared with predictions using fitted paramters on 10th and 5th of May 2020. The bias of the prediction is presented in Figures 5, 6 and 7.

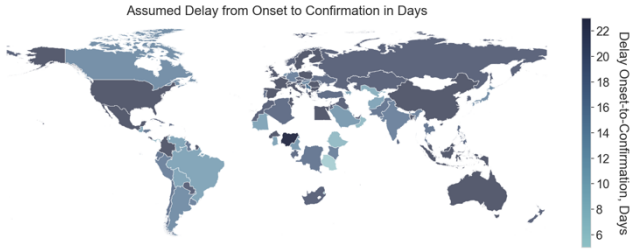

Figure 1: Global variation in assumed delay from onset to case confirmation.

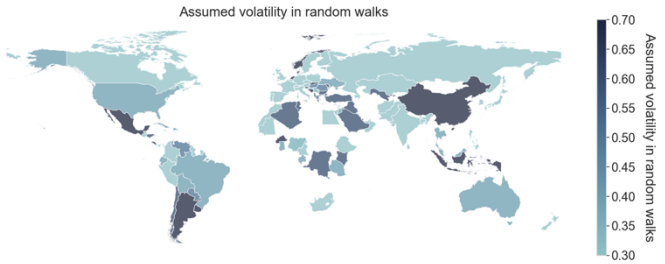

FIGURE 2: Global variation in assumed volatility of transmission rate.

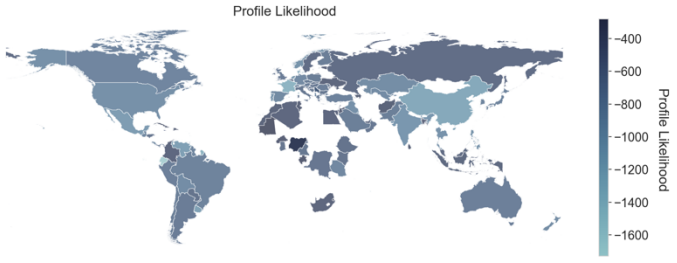

FIGURE 3: The likelihood profile of the fitting.

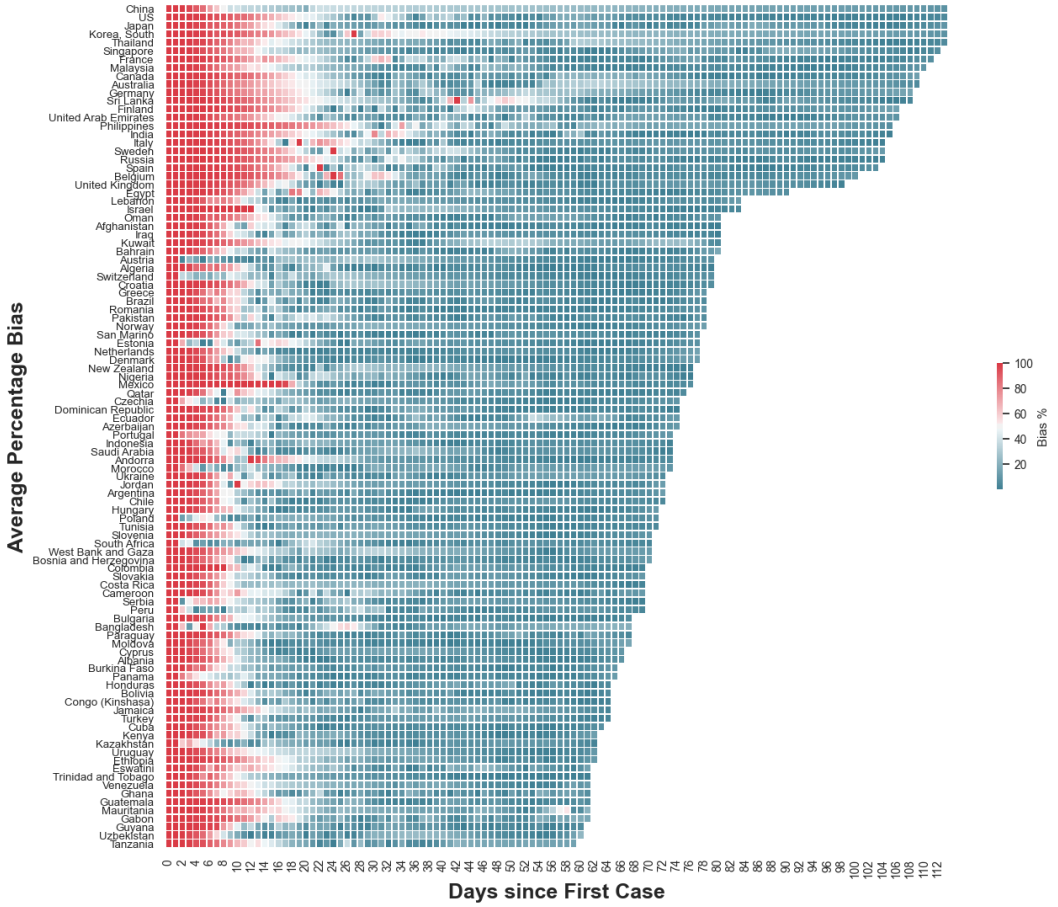

FIGURE 4: Percentage bias for fitted reported cases since outbreak onset in each country to 15th of May 2020.

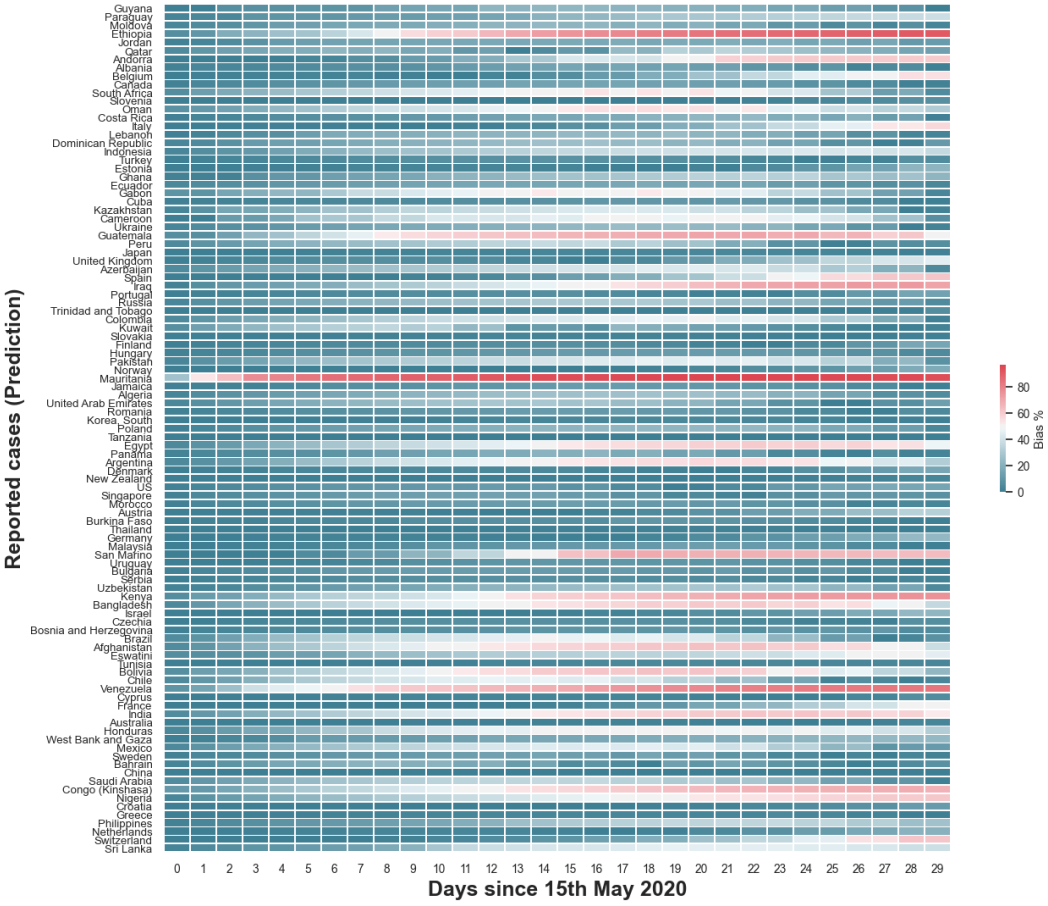

FIGURE 5: Percentage bias for predicted reported cases between 15th of May to 15th of June 2020 using parameters fitted on 15th of May 2020.

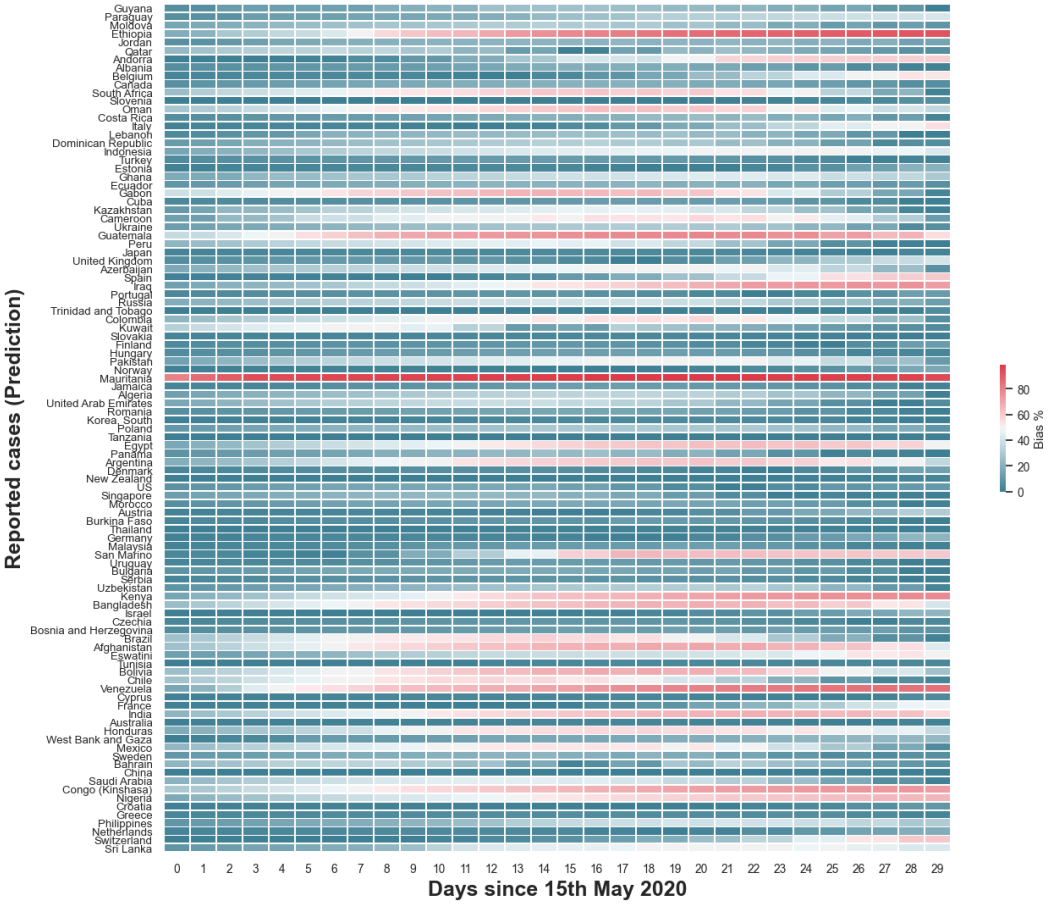

FIGURE 6: Percentage bias for predicted reported cases between 15th of May to 15th of June 2020 using parameters fitted on 10th of May 2020.

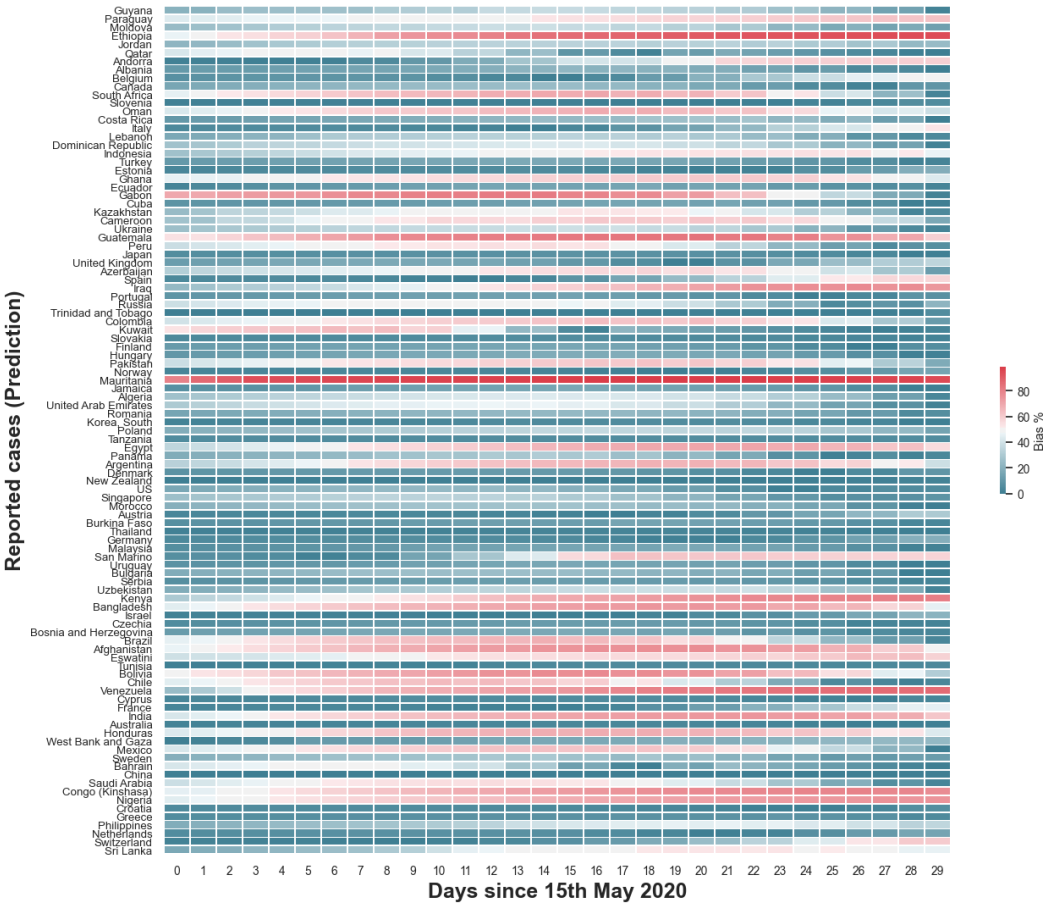

FIGURE 7: Percentage bias for predicted reported cases between 15th of May to 15th of June 2020 using parameters fitted on 5th of May 2020.