DECLINE IN GLOBAL TRANSMISSION RATES OF COVID-19 THROUGH MAY 6 2020

A PREPRINT

Ethan Romero-Severson
Theoretical biology and biophysics
Los Alamos National Laboratory
Los Alamos, NM
eoromero@lanl.gov

Nick Hengartner
Theoretical biology and biophysics
Los Alamos National Laboratory
Los Alamos, NM
nickh@lanl.gov

Grant David Meadors
Computational physics
Los Alamos National Laboratory
Los Alamos, NM
gdmeadors@lanl.gov

Ruian Ke
Theoretical biology and biophysics
Los Alamos National Laboratory
Los Alamos, NM
rke@lanl.gov

May 21, 2020

ABSTRACT

We analyzed COVID-19 data through May 6th, 2020 using a partially observed Markov process. Our method uses a hybrid deterministic and stochastic formalism that allows for time variable transmission rates and detection probabilities. The model was fit using iterated particle filtering to case count and death count time series from 55 countries. We found evidence for a shrinking epidemic in 30 of the 55 examined countries. Of those 30 countries, 27 have significant evidence for subcritical transmission rates, although the decline in new cases is relatively slow compared to the initial growth rates. Generally, the transmission rates in Europe were lower than in the Americas and Asia. This suggests that global scale social distancing efforts to slow the spread of COVID-19 are effective although they need to be strengthened in many regions and maintained in others to avoid further resurgence of COVID-19. The slow decline also suggests alternative strategies to control the virus are needed before social distancing efforts are partially relaxed.

***K*eywords** COVID-19 · Partially observed Markov process · Iterated particle filtering

1 Introduction

Since its initial outbreak in Wuhan, China in late 2019 and early 2020 (1), COVID-19 caused repeated rapid outbreaks across the global from February through April 2020. The extremely rapid spread of COVID-19 in China (2) does not appear to be an anomaly: the disease has shown a short doubling time (2.4-3.6 days) outside of China as well (3). As of May 21, 2020, the virus caused 5,034,458 reported infections, and 328,730 deaths globally (4). In response, most affected countries/regions have implemented strong social distancing efforts, such as school closures, working-from-home, shelter-in-place orders. As a result, the spread of COVID-19 slowed down substantially in some countries (5), leading to a flattening of the epidemic curve. As social distancing induces high costs to both society and individuals, plans to relax social distancing are discussed. However, changes in both the transmission rates and detection probabilities over time coupled with stochasticity due to reporting delays makes differentiating between truly subcritical dynamics and simply reduced transmission difficult.

In this report, we developed a deterministic-stochastic hybrid model and fitted the model to case incidence and death incidence time series data from 55 countries. Following the approach suggested by King et al. (6), we use a (partially) stochastic model and base our estimates on incidence rather than cumulative incidence data . Using both case count and

death count of data allowed us to disentangle changes in surveillance intensity from changes in transmission (3). We found evidence for large decreases in the country-level transmission rates, in several of the worst-affected countries. Importantly, using data up to May 6, 2020, we computed 99% confidence intervals to test whether or not the data were consistent with subcritical dynamics (i.e. the reproductive number R was below 1 on May 6, 2020). Most countries showed large decreases in transmission rates over time, and more than half of studied countries have transmission rates below the epidemic threshold. On the other hand, many countries still appear to be showing rapid exponential growth. Given its highly contagious nature, COVID-19 can spread rapidly when strong social distancing measures are lifted, even partially (3). Alternative strategies that can effective control the virus are needed when social distancing measures are relaxed.

2 Materials and Methods

2.1 Data

Case count and death data were downloaded from the Johns Hopkins GitHub repository (`https://github.com/CSSEGISandData/COVID-19`) through May 6, 2020. Any country in the data that had more than 2000 cumulative cases and 100 deaths by May 6, 2020 was modeled. To minimize the effect of repatriated cases we started each time series on the first day when the cumulative number of cases exceeded 100. All data processing and model fitting were otherwise done on the incidence scale. To address obvious bulk-reporting issues in the data (e.g. sudden zeros in the data followed by very large numbers), we smoothed the data using Tukey's 3-median method. Because several countries had days with a single death surrounded by days with no deaths, which the smoothing method would set to zero, days with a single death were not replaced with smoothed values. The original data and the smoothed data used for estimation are shown in Supplemental figure S1.

2.2 Model

We model the spread of COVID-19 as a partially observed Markov process with real-valued states S (susceptible), E (exposed), I (infected), and R (recovered) to describe the latent population dynamics, and integer-valued states C_0 (to be counted), Y_1 (counted cases), $D_{0:3}$ (dying), and Y_2 (counted deaths) to model sampling into the data. The model and all of its parameters have time units of days. The latent population model is governed by the following ordinary differential equations,

$$\begin{aligned}
\frac{dS}{dt} &= -\chi(t) I \frac{S}{N}, \\
\frac{dE}{dt} &= \chi(t) I \frac{S}{N} - \lambda_{EI} E, \\
\frac{dI}{dt} &= \lambda_{EI} E - \lambda_{IR} E, \\
\frac{dR}{dt} &= \lambda_{IR} E,
\end{aligned}$$

where $\chi(t)$ is the time-variable transmission rate and $N = S + E + I + R$ is assumed to be fixed over the run of the model. λ_{EI} is the rate at which exposed persons become infectious and λ_{IR} is the rate at which cases recover (i.e. are either no longer infectious or die). At every time interval, we sample persons moving from the E to I states into a stochastic arm of the model that is used to calculate the likelihood of the data.

To relate the latent population model to data, we randomly sample individuals from the unobserved population into a stochastic process that models the random movement from infection to being either counted as an observed case or counted as an observed death. The number of persons sampled into the observation arm of the model over a time interval dt is given by $X = (X_1, X_2, X_3, X_4) \sim \text{Multinomial}\big(g(E\lambda_{EI}dt), \{(\rho^c(t)\omega^c, \rho^c(t)\omega, \rho(t)\omega^c, \rho(t)\omega\}\big)$ where $\rho(t)$ and ω are the probabilities that an infected person will be counted or die respectively, $\rho^c(t)$ and ω^c are the probabilities of being not sampled or not dying respectively, and $g(u)$ is a stochastic function that maps from real-value u to integer $g(u)$; it takes value $\lfloor u \rfloor$ with probability $u \mod 1$ and value $\lceil u \rceil$ with probability $1 - (u \mod 1)$. In plain English, the model tracks the random fate of each newly infected person as they move from the exposed to infected state with respect to eventually being observed as a case and/or a death in data. At each time step of length dt, the change in state

space is given by,

$$
\begin{aligned}
C_0(t+dt) &= C_0(t) + X_3 + X_4 - F_t, \\
Y_1(t+dt) &= Y_1(t) + F_t, \\
D_0(t+dt) &= D_0(t) + X_2 + X_4 - H_{1t}, \\
D_1(t+dt) &= D_1(t) + H_{1t} - H_{2t}, \\
D_2(t+dt) &= D_2(t) + H_{2t} - H_{3t}, \\
D_3(t+dt) &= D_3(t) + H_{3t} - H_{4t}, \\
Y_2(t+dt) &= Y_2(t) + H_{4t},
\end{aligned}
$$

where $F_t|C_0$ is a random variate from $\text{Bin}(C_0, 1 - \exp(-\lambda_{Y_1} dt))$, and H_{it} is a random variate from $\text{Bin}\left(D_{i-1}, 1 - \exp\left(-\frac{4}{21}dt\right)\right)$. X_i indicates the ith element of Multinomial random variate defined above. The rate λ_{Y1} determines the rate at which persons who will be counted are counted (i.e. lower values of λ_{Y1} mean a longer delay in cases showing up in the data). The values of $Y_{1:2}$ are set to zero at the beginning of every day such that they accumulate the simulated number of cases and deaths that occur each day. Both the transmission rate and detection probability are allowed to vary with time in the following way:

$$
\chi(t) = \begin{cases} \chi_0 & \text{if } t \leq t_\chi^{(1)}, \\ \chi_0 + \frac{\chi_f - \chi_0}{t_f - t_\chi}(t_\chi - t) & \text{if } t_\chi^{(1)} < t \leq t_\chi^{(2)}, \\ \chi_f & \text{if } t > t_\chi^{(2)}, \end{cases}
$$

$$
\rho(t) = \begin{cases} \rho_0 & \text{if } t \leq t_\rho, \\ \rho_0 + \frac{\rho_f - \rho_0}{t_f - t_\rho}(t_\rho - t) & \text{if } t_\rho < t \leq (t_f - 20), \\ \rho_f & \text{if } t > (t_f - 20), \end{cases}
$$

where t_f is the final time in the given dataset. In plain English, the transmission rate is constant up to some time, $t_\chi^{(1)}$, where it linearly increases or decreases to the value χ_f by time $t_\chi^{(2)}$. The model is constrained such that $t_\chi^{(2)} < t_f$, that is, the transmission rate must be constant for at least 20 days before the end of the data collection period. This constraint is in place to avoid overfitting the final transmission rate. The detection rate likewise has a linear increase or decrease beginning at time t_ρ; however, the increase or decrease continues to a fixed point that is 20 days before the final datum. Variation in the testing rate over the course of the epidemic can strongly bias estimates of the population growth rate (derivation for the exponential growth case in Supplemental Material); not allowing the detection probability to change over time could lead to discordance in the case count and death count time series.

2.3 Parameter estimation

We assume that the data are Negative Binomial-distributed, conditional on the simulated number of cases and deaths that occur in a given time interval, i.e. the number of cases in the ith observation period has density $\text{NegBin}(Y_1, \epsilon_1)$ and the number of deaths in the ith observation period has density $\text{NegBin}(Y_2, \epsilon_2)$. The Negative Binomial is parameterized such that the first argument is the expectation and the second is an inverse overdispersion parameter that controls the variance of the data about the expectation; as ϵ_i becomes large, the data model approaches a Poisson with parameter Y_i. Both ϵ_1 and ϵ_2 were estimated from the data.

Parameter estimation had two distinct steps: model selection and computation of confidence intervals. In the model selection phase, the model was fit to the data using an iterated particle-filtering method, implemented in the `pomp` R library (7). To optimize the likelihood of the data, we used 1500 particles in 125 iterations for each country. The reported likelihoods were measured using 4500 particles at the optimized Maximum Likelihood Estimates (MLEs). The Monte Carlo error in the likelihood measurement was less than 0.1 log units.

For all fits, the initial state at time zero is computed by assuming there were $I(0)$ infected persons, 21 days before the first reported death (by definition time one). The number of initial number of susceptibilities was assumed to be the predicted 2020 population size of the given country (8). The model is then simulated forward for 21 days, assuming exponential growth with transmission rate χ_0, which is taken at the initial state of the model at time zero. Parameters have constraints $I(0) \in [1, 1000]$, $\rho_0 \in [0, 1]$, $\rho_f \in [0, 1]$, $\lambda_{Y1}^{-1} \in [5, 21]$ t_χ and $t_\rho \in [0, t_f - 10]$; all other parameters are constrained to be positive.

Because the data were highly variable in the complexity of the patterns they showed, we considered three nested models of increasing complexity for each country. The first model (model 1) assumed simple exponential growth with a constant sampling probability, i.e. $\rho_0 = \rho_f$ and $t_\chi^{(1)} = t_\chi^{(1)} = 0$. This amounts to a period of exponential growth with

transmission rate χ_0 in the pre-data period and then a constant transmission rate χ_f over the observation period. The second model (model 2) allowed the transmission rate to vary but kept the detection probability constant, i.e. $\rho_0 = \rho_f$. The third model (model 3) allowed both the transmission rate and detection rate to vary, i.e. all parameters estimated. We determined the best model for each country by a sequence of likelihood ratio tests first comparing model 1 and model 2 and then model 2 to model 3.

Because we are using an optimization method, we do not have direct access to samples from the likelihood surface directly. To obtain estimates of the parametric uncertainly in the final transmission rate we computed 99% confidence intervals using the profile likelihood method (9). We used the 99% confidence level to minimize the effect of numerical noise coming from both the optimization procedure and the Monte Carlo error in the measurement of the likelihood function.

3 Results

We fit our model to data collected from 55 countries. Model fits are shown in Figure 1, and parameter values are given in Table 5. The model can capture the data well, with a few exceptions. The model was not able to find a robust fit to the data from Bangladesh; in general, the upside down 'V' shape to the deaths could not be captured well in such a short time series. Algeria had a similarly odd pattern in the deaths time series that the model could not capture in detail, although the overall trend in deaths was recovered. In previous versions of this paper, the model had a hard time fitting data from both Italy and Spain. However, given the longer time series and modifications to the model form, we now find that both Italy and Spain are well-captured by the model.

Overall, the model slightly overestimates the number of deaths around the time where deaths begin to decrease. However, this was generally corrected if the time series was long enough. Including temporal heterogeneity in the time from infection to detection of a COVID-19 death would likely correct this; however, it is not clear that this is advisable, as death counts are likely under-reported.

All countries except Japan and Saudi Arabia were found to have lower transmission rates on May 6, 2020 than at the beginning of the observation period, suggesting a global decline in the transmission of COVID-19 though May 6, 2020. However, the initial transmission rate should be interpreted cautiously, as we allowed a wide range of infected persons to exist 21 days before the first observation. That is, the initial transmission rate parameter is rather a convolution of the unknown number of infected persons and initial growth rate consistent with the data.

We found significant evidence for subcritical dynamics in 27 countries (3 countries had subcritical point estimates but their CIs contained the epidemic threshold). Figure 2 shows the point estimates and CIs of the final transmission rate on May 6, 2020 for all countries, stratified by continent (10). European countries had the highest probability of being subcritical (21 of 25) with Asia (7 of 15) and the Americas (2 of 11) having fewer subcritical countries. None of the countries in Africa that met the inclusion criteria had subcritical dynamics.

Figure 3 shows the predicted number of deaths projected out to July 5, 2020 assuming that all parameter values are constant over the period May 6 through July 5 2020. The average duration of the country-level epidemic in European countries is longer than in Asia, leading to a higher level of death, despite Asian countries having on-average higher growth rates. However, in the Americas, the predicted deaths are higher with 8 of 11 countries having total predicted deaths greater than 0.1% of the total population by July 5, 2020. The model predicts 1,320,170 total COVID-19 deaths in all 55 countries on July 5, 2020; of those deaths, 21% are predicted to occur in the US. The deaths due to COVID-19 in Europe are lower than the average number of reported deaths in a period of the same length for all countries in the data set that also had all-cause death counts from previous years. However, in the Americas, the COVID-19 death counts are approaching the all-cause death levels in several countries, suggesting that COVID-19 deaths are approaching a doubling of all-cause deaths in those countries.

4 Discussion

Our model found evidence for reductions in transmission rates of COVID-19 in 53 of 55 examined countries. Encouragingly, of those countries, we found statistical evidence that the size of the epidemic is decreasing in 27 countries, i.e. the effective reproductive number is less than 1, using data up to May 6, 2020. This suggests that, despite the highly heterogeneous populations represented by these countries, the growth of COVID-19 outbreak can be reverted. Although our model cannot attribute exact causes to the global decline in transmissions rates, most countries implemented sustained, population-level social distancing efforts over a period of weeks to months. These efforts are highly likely to play a major role in reducing the transmission of COVID-19 (5).

We estimated that in countries with decreasing transmission, the rate of decrease is in general less than 0.1/day (average -0.04/day). Based on data from 8 European countries, the US, and China, we previously estimated that in the absence of intervention efforts, the epidemic can grow at rates between 0.19-0.29/day (2). This means that the outbreak can grow rapidly and quickly wipe gains made though public heath efforts made if social distancing measures are completely relaxed. For example, if the rate of decrease under strong public health interventions is 0.1/day and the growth rate in the absence of public heath interventions is 0.2/day, then, the number of cases averted in two weeks of intervention will be regained in only one week. Social distancing measures have their own social costs. Our results suggests alternative strategies to control the virus are needed in place when social distancing efforts are relaxed. Due to the uncertainties in the impact of each specific control measures, changes to policies should be made slowly because the signal of changing transmission can take weeks to fully propagate into current data streams as a result of the long lag between infection to case confirmation (as we estimated to be on average approximately 2 weeks).

Overall, our results suggest that COVID-19 is controllable in diverse settings using a full range of strong and comprehensive non-pharmaceutical measures, and that future deaths from the disease are avoidable.

5 Tables

Symbol	Description	Value
λ_{EI}	Latency rate	$\frac{1}{3}$
λ_{IR}	Recovery rate	$\frac{1}{10}$
ω	Case fatality ratio	1%
χ_0	Initial transmission rate	fit
$t_\chi^{(1)}$	Time of beginning of trans. rate change	fit
$t_\chi^{(2)}$	Time of ending of trans. rate change	fit
χ_f	Final transmission rate	fit
ρ_0	Initial probability of being counted	fit
t_ρ	Time of counting prob. change	fit
ρ_f	Final counting prob.	fit
$I(0)$	Initial number infected	fit

Country	Model	χ_f	CI χ_f	r_f	CI r_f	ρ_0	ρ_f
Algeria	2	0.111	0.109, 0.117	0.008	0.007, 0.013	0.08	0.08
Argentina	1	0.111	0.11, 0.123	0.008	0.008, 0.017	0.16	0.16
Austria	2	0.024	0.022, 0.025	-0.07	-0.072, -0.069	0.25	0.25
Bangladesh	3	0.197	0.195, 0.199	0.065	0.064, 0.066	0.1	0.1
Belarus	1	0.163	0.119, 0.223	0.044	0.014, 0.08	0.79	0.79
Belgium	3	0.092	0.088, 0.095	-0.006	-0.009, -0.004	0.05	<0.01
Brazil	2	0.185	0.183, 0.186	0.058	0.057, 0.059	0.05	0.05
Canada	3	0.156	0.154, 0.157	0.039	0.038, 0.04	0.11	<0.01
Chile	1	0.154	0.153, 0.156	0.038	0.037, 0.039	0.55	0.55
China	1	0.025	0.023, 0.027	-0.069	-0.071, -0.067	0.24	0.24
Colombia	1	0.15	0.137, 0.154	0.035	0.027, 0.038	0.18	0.18
Czechia	2	0.027	0.009, 0.039	-0.066	-0.088, -0.054	0.28	0.28
Denmark	2	0.05	0.044, 0.061	-0.042	-0.048, -0.032	0.19	0.19
Dominican Republic	1	0.122	0.12, 0.136	0.016	0.015, 0.026	0.24	0.24
Ecuador	1	0.163	0.134, 0.194	0.044	0.025, 0.063	0.13	0.13
Egypt	2	0.165	0.163, 0.165	0.045	0.044, 0.046	0.12	0.12
Finland	3	0.131	0.128, 0.133	0.023	0.021, 0.024	0.23	0.03
France	2	0.002	<0.001, 0.004	-0.097	-0.1, -0.095	0.07	0.07
Germany	3	0.023	0.019, 0.057	-0.071	-0.076, -0.036	0.38	0.02
Greece	2	0.024	<0.001, 0.048	-0.07	-0.1, -0.044	0.18	0.18
Hungary	2	0.087	0.077, 0.093	-0.011	-0.019, -0.005	0.08	0.08
India	3	0.158	0.156, 0.16	0.041	0.04, 0.042	0.19	0.05
Indonesia	2	0.122	0.116, 0.123	0.017	0.012, 0.017	0.07	0.07
Iran	1	0.092	0.09, 0.093	-0.007	-0.008, -0.005	0.15	0.15
Iraq	1	0.112	0.105, 0.118	0.009	0.004, 0.014	0.21	0.21
Ireland	2	0.069	0.063, 0.081	-0.025	-0.03, -0.015	0.12	0.12
Israel	2	0.012	0.002, 0.026	-0.085	-0.097, -0.067	0.66	0.66
Italy	2	0.051	0.049, 0.054	-0.042	-0.043, -0.039	0.06	0.06
Japan	1	0.18	0.177, 0.183	0.054	0.053, 0.056	0.18	0.18
Malaysia	2	0.043	0.037, 0.045	-0.049	-0.056, -0.047	0.54	0.54
Mexico	2	0.175	0.174, 0.192	0.052	0.051, 0.062	0.05	0.05
Moldova	3	0.085	0.079, 0.103	-0.012	-0.017, 0.003	0.3	<0.01
Morocco	1	0.121	0.12, 0.125	0.016	0.015, 0.019	0.23	0.23
Netherlands	3	0.046	0.043, 0.048	-0.046	-0.049, -0.045	0.08	0.01
Norway	2	0.063	0.058, 0.07	-0.03	-0.036, -0.024	0.35	0.35
Pakistan	1	0.183	0.176, 0.194	0.056	0.052, 0.063	0.34	0.34
Panama	2	0.061	0.059, 0.084	-0.032	-0.034, -0.012	0.34	0.34
Peru	1	0.255	0.255, 0.271	0.097	0.097, 0.106	0.1	0.1
Philippines	2	0.082	0.081, 0.084	-0.015	-0.015, -0.013	0.16	0.16
Poland	2	0.071	0.068, 0.072	-0.024	-0.027, -0.023	0.19	0.19
Portugal	2	0.046	0.043, 0.049	-0.046	-0.05, -0.044	0.22	0.22
Romania	2	0.084	0.076, 0.085	-0.013	-0.019, -0.012	0.15	0.15
Russia	2	0.148	0.146, 0.15	0.034	0.033, 0.036	0.43	0.43
Saudi Arabia	1	0.222	0.218, 0.227	0.079	0.077, 0.082	0.96	0.96
Serbia	2	0.001	<0.001, 0.005	-0.099	-0.1, -0.094	0.41	0.41
South Africa	1	0.148	0.133, 0.151	0.034	0.024, 0.036	0.5	0.5
South Korea	3	0.059	0.059, 0.061	-0.034	-0.035, -0.033	0.45	0.01
Spain	2	0.014	0.005, 0.027	-0.082	-0.093, -0.067	0.08	0.08
Sweden	3	0.169	0.093, 0.182	0.048	-0.006, 0.056	0.08	0.07
Switzerland	2	0.024	0.023, 0.026	-0.069	-0.071, -0.067	0.17	0.17
Turkey	2	0.059	0.058, 0.061	-0.034	-0.035, -0.032	0.22	0.22
Ukraine	2	0.085	0.082, 0.101	-0.012	-0.014, 0.001	0.33	0.33
United Arab Emirates	1	0.193	0.19, 0.23	0.063	0.061, 0.084	0.99	0.99
United Kingdom	3	0.095	0.094, 0.097	-0.004	-0.004, -0.002	0.05	<0.01
US	3	0.1	0.098, 0.101	0	-0.001, 0.001	0.09	0.07

Table 1: Table of key parameter values. The 'Model' column referrers to the best model defined in the model selection procedure. The values r_0 and r_f are the initial and final exponential growth rates implied by the other model parameters. The growth rate r and transmission rate χ are derived according to (11). χ_0, χ_f, r_0, r_f, have day^{-1} units; ρ_0 and ρ_f are probabilities.

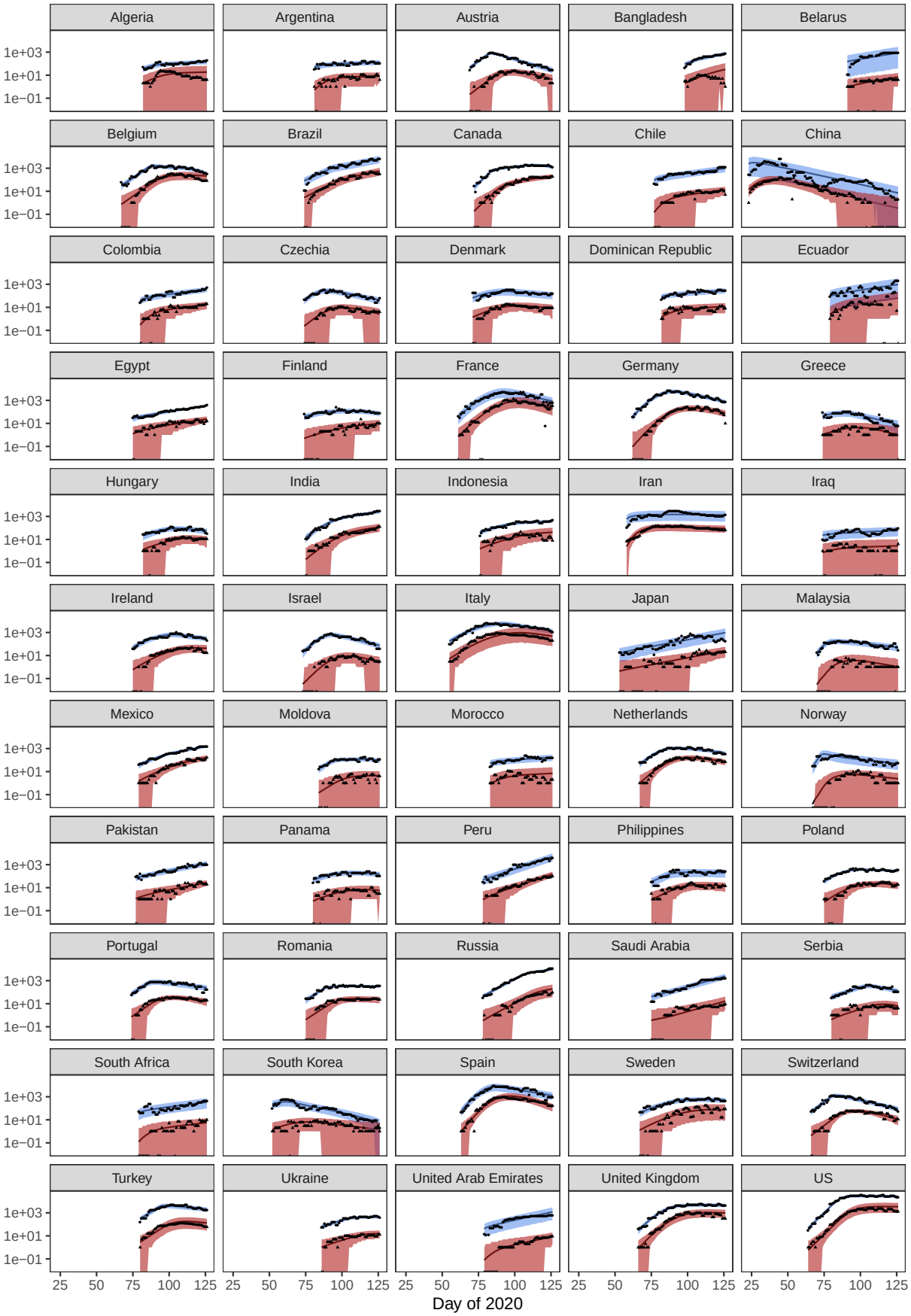

Figure 1: Model fits. Smoothed data points are shown in black and the expectations from the model are shown as solid lines. The envelope shows the 95% CIs of the distribution of the data conditional on the maximum likelihood model fit for the best model for each country. The x-axis is measured in sequential days of the year 2020.

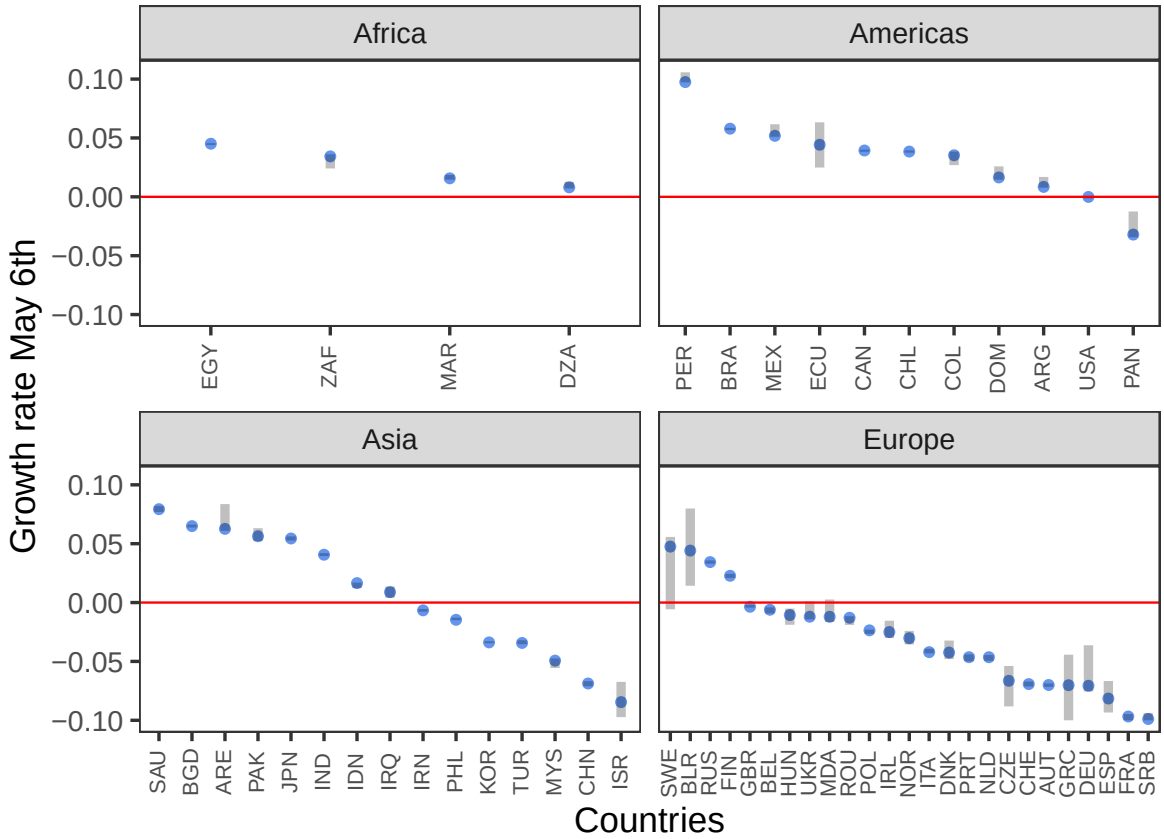

Figure 2: Values of the final exponential growth rates as of May 6 2020. Blue dots indicate point estimates and grey bars indicate the 99% confidence intervals. The red line delineates exponential growth from exponential decay. Countries below this line have a shrinking epidemic. Country names are indicated by ISO Alpha-3 codes.

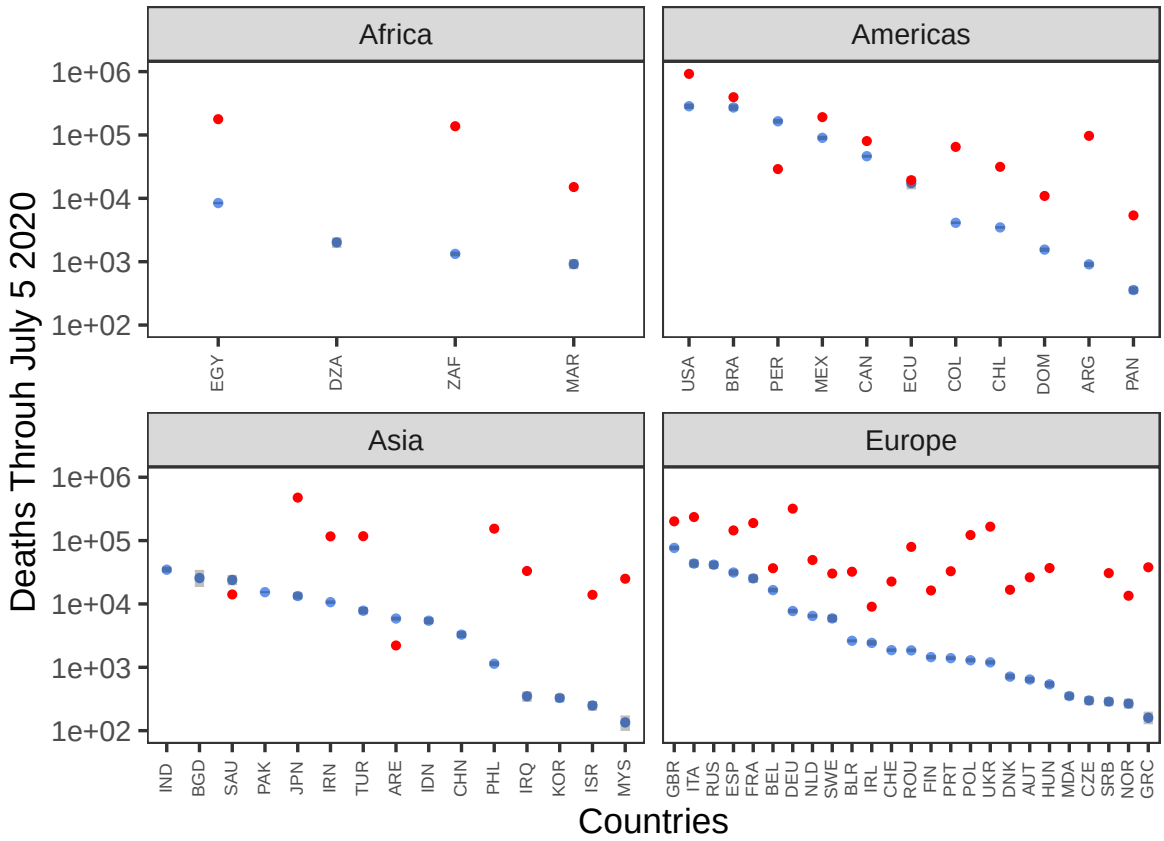

Figure 3: Predicted number of cumulative deaths through July 5 2020. The distribution of the total number of predicted deaths in the observation period plus 60 days assuming final model maximum likelihood values on May 6 2020 are constant though July 5 2020. Blue dots indicate median values and grey bars show the 95% confidence interval. Red dots show the average number of deaths in an average period of the same length in the World Health Organization death data from the most recent year available. The y-axis is on the log10 scale. Country names are indicated by ISO Alpha-3 codes

6 Supplemental

6.1 Estimating the infection and counting rates with exponential growth

Consider an exponential growth model early in the outbreak where infected (I) persons transmit at rate β, are counted (C) at rate $\alpha(t)$, recover (R) at rate γ_1, and die (D) at rate γ_2. When individuals are counted, they are not removed from I and transmit at the same rate as uncounted persons. The differential equations for I and C cases in time t, where (I, C, R, D) are dimensionless numbers, but $(\alpha, \beta, \gamma_1, \gamma_2)$ have units of $(\text{time})^{-1}$, state,

$$\frac{dI(t)}{dt} = \beta I(t) - \gamma_1 I(t) - \gamma_2 I(t), \tag{1}$$

$$\frac{dC(t)}{dt} = \alpha(t) I(t), \tag{2}$$

$$\frac{dR(t)}{dt} = \gamma_1 I(t), \tag{3}$$

$$\frac{dD(t)}{dt} = \gamma_2 I(t). \tag{4}$$

Let us also tie C to R. We have that,

$$\frac{C(t)}{dt} = \alpha(t) I(t), \tag{5}$$

$$= \frac{\alpha(t)}{\gamma_2} \frac{dD(t)}{dt}. \tag{6}$$

The problem is that we only observe $dC(t)/dt$ and $dD(t)/dt$. So algebraically,

$$\alpha(t) = \gamma_2 \frac{dC(t)/dt}{dD(t)/dt}. \tag{7}$$

Given the model specified, it is true that $\alpha = \gamma_2 (dC/dt)/(dD/dt)$. Since they are total derivatives, it is also true that $\alpha = \gamma_2 dC/dD$.

6.1.1 Constant counting rate

It is a first-order problem to show that,

$$I = I_0 \exp\left[(\beta - \gamma_1 - \gamma_2)t\right], \tag{8}$$

$$S = -\beta(\beta - \gamma_1 - \gamma_2)^{-1} I, \tag{9}$$

$$R = \gamma_1(\beta - \gamma_1 - \gamma_2)^{-1} I, \tag{10}$$

$$D = \gamma_2(\beta - \gamma_1 - \gamma_2)^{-1} I, \tag{11}$$

as well as, when $d\alpha/dt = 0$,

$$C = \alpha(\beta - \gamma_1 - \gamma_2)^{-1} I. \tag{12}$$

An interpretation is that $I(t) = C(t + \tau)$, with time lag of τ applied to t, where,

$$\tau = \frac{1}{\beta - \gamma_1 - \gamma_2} \log\left(\frac{\alpha}{\beta - \gamma_1 - \gamma_2}\right). \tag{13}$$

This simply shows that if the counting rate is constant, then the value of C is simply equal to I shifted in time by τ.

6.1.2 Time-varying counting rate

For time-varying α, at time t with initial count C_0 at time t_0, in terms of (the only quantities observable) C & D,

$$C = C_0 + (\gamma_2)^{-1} \int_{t_0}^{t} \alpha dD, \tag{14}$$

$$= C_0 + (\gamma_2)^{-1} \left[\alpha(t)D(t) - \alpha(t_0)D(t_0) - \int_{t_0}^{t} D(t')(d\alpha/dt')dt' \right]. \tag{15}$$

Stipulating that $C_0 = (\gamma_2)\alpha(t_0)D(t_0)$, up to an additive constant,

$$C = (\gamma_2)^{-1} \left[\alpha(t)D(t) - \int_{t_0}^{t} D(t')(d\alpha/dt')dt' \right], \tag{16}$$

which reduces back, if $d\alpha/dt = 0$, to the time-integral of Equation 7:

$$C = (\gamma_2)^{-1} \alpha D(t). \tag{17}$$

6.1.3 Exponential-growth difference between counts and infections

In the case where $d\alpha/dt > 0$, we find the difference $(d \log C)/dt - (d \log I)/dt$. It is generally true that,

$$\frac{dI}{dt} = \alpha^{-1}(\beta - \gamma_1 - \gamma_2)\frac{dC}{dt}. \tag{18}$$

Using $dx = x d \log x$ to transform C, I,

$$\frac{d \log I}{dt} = \frac{C}{I}\alpha^{-1}(\beta - \gamma_1 - \gamma_2)\frac{d \log C}{dt}. \tag{19}$$

Yet integration by parts shows that, for general $d\alpha/dt$,

$$C = (\beta - \gamma_1 - \gamma_2)^{-1} \left[\alpha(t)I(t) - \int_{t_0}^{t} I(t')(d\alpha/dt')dt' \right], \tag{20}$$

substitution of which yields,

$$\frac{d \log I}{dt} = \left(1 - \frac{1}{\alpha(t)I(t)} \int_{t_0}^{t} I(t')(d\alpha/dt')dt \right) \frac{d \log C}{dt}, \tag{21}$$

Whenever $d\alpha/dt$ is positive and, as almost always the case, $\alpha, I, dI/dt > 0$, then Equation 21 implies that $\log C$ grows faster than $\log I$. In the approximation that α changes slowly,

$$\int I(d\alpha/dt')dt' \ll \alpha(t)I(t),$$

then Taylor expanding and rearranging terms yields a simpler expression,

$$\frac{d \log C}{dt} - \frac{d \log I}{dt} \approx \left(\frac{1}{\alpha(t)I(t)} \int_{t_0}^{t} I(t') \frac{d\alpha(t')}{dt'} dt' \right) \frac{d \log I}{dt}, \tag{22}$$

$$= \frac{1}{\alpha(t)e^{(\beta-\gamma_1-\gamma_2)t}} \int_{t_0}^{t} e^{(\beta-\gamma_1-\gamma_2)t'} \frac{d\alpha(t')}{dt'} dt'. \tag{23}$$

For all positive $\alpha, d\alpha/dt$, the right-hand side is positive. Evidently, the growth rate in logarithmic C exceeds that in logarithmic I. The growth rate of logarithmic C can be readily inferred from a log-linear plot, but I is generally unknown. When the testing rate increases during an epidemic's exponential growth phase, the number of counts C increases faster than the number of infections I.

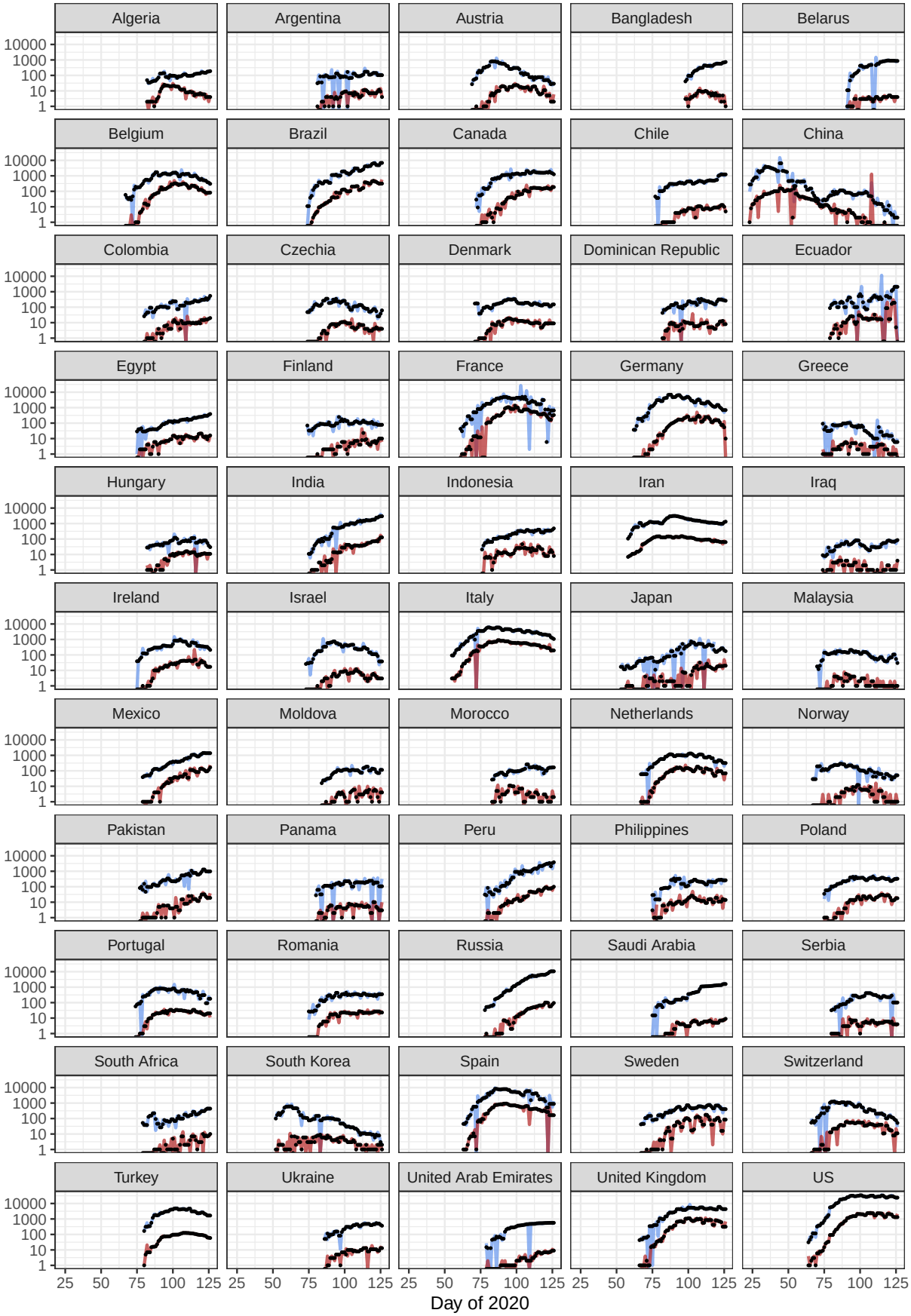

Figure S1: Raw and smoothed data. Raw data for cases (blue lines) and deaths (red lines) are shown with the smoothed time series (black) used for model fitting. The x-axis is measured in sequential days of 2020.

References

[1] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. Leung, E. H. Lau, J. Y. Wong, X. Xing, N. Xiang, Y. Wu, C. Li, Q. Chen, D. Li, T. Liu, J. Zhao, M. Liu, W. Tu, C. Chen, L. Jin, R. Yang, Q. Wang, S. Zhou, R. Wang, H. Liu, Y. Luo, Y. Liu, G. Shao, H. Li, Z. Tao, Y. Yang, Z. Deng, B. Liu, Z. Ma, Y. Zhang, G. Shi, T. T. Lam, J. T. Wu, G. F. Gao, B. J. Cowling, B. Yang, G. M. Leung, and Z. Feng, "Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia," *New England Journal of Medicine*, vol. 382, no. 13, pp. 1199–1207, Mar. 2020, publisher: Massachusetts Medical Society _eprint: https://doi.org/10.1056/NEJMoa2001316. [Online]. Available: https://doi.org/10.1056/NEJMoa2001316

[2] S. Sanche, Y. T. Lin, C. Xu, E. Romero-Severson, N. Hengartner, and R. Ke, "Early Release - High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2 - Volume 26, Number 7—July 2020 - Emerging Infectious Diseases journal - CDC." [Online]. Available: https://wwwnc.cdc.gov/eid/article/26/7/20-0282_article

[3] R. Ke, S. Sanche, E. Romero-Severson, and N. Hengartner, "Fast spread of COVID-19 in Europe and the US and its implications: even modest public health goals require comprehensive intervention," *medRxiv*, p. 2020.04.04.20050427, Apr. 2020, publisher: Cold Spring Harbor Laboratory Press. [Online]. Available: https://www.medrxiv.org/content/10.1101/2020.04.04.20050427v1

[4] E. Dong, H. Du, and L. Gardner, "An interactive web-based dashboard to track COVID-19 in real time," *The Lancet Infectious Diseases*, vol. 0, no. 0, Feb. 2020, publisher: Elsevier. [Online]. Available: https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)30120-1/abstract

[5] S. Flaxman, S. Mishra, and A. Gandy, "2020-03-30-COVID19-Report-13.pdf," Imperial College London, Tech. Rep., Mar. 2020. [Online]. Available: https://www.imperial.ac.uk/media/imperial-college/medicine/mrc-gida/2020-03-30-COVID19-Report-13.pdf

[6] A. A. King, M. D. d. Cellès, F. M. G. Magpantay, and P. Rohani, "Avoidable errors in the modelling of outbreaks of emerging pathogens, with special reference to Ebola," *Proc. R. Soc. B*, vol. 282, no. 1806, p. 20150347, May 2015. [Online]. Available: http://rspb.royalsocietypublishing.org/content/282/1806/20150347

[7] A. King, E. Ionides, C. Bretó, S. Ellner, B. Kendall, H. Wearing, M. Ferrari, M. Lavine, and D. Reuman, *pomp: Statistical inference for partially observed Markov processes (R package)*, 2009.

[8] C. I. A. Washington DC, "The World Factbook 2020." [Online]. Available: https://www.cia.gov/library/publications/resources/the-world-factbook/index.html

[9] E. L. Ionides, "Maximum smoothed likelihood estimation," *Statistica Sinica*, pp. 1003–1014, 2005. [Online]. Available: http://www.jstor.org/stable/24307347

[10] V. Arel-Bundock, N. Enevoldsen, and C. Yetman, "countrycode: An R package to convert country names and country codes," *Journal of Open Source Software*, vol. 3, no. 28, p. 848, Aug. 2018. [Online]. Available: https://joss.theoj.org/papers/10.21105/joss.00848

[11] J. Ma, "Estimating epidemic exponential growth rate and basic reproduction number," *Infectious Disease Modelling*, vol. 5, pp. 129–141, Jan. 2020. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6962332/