



# The Quixotic Task of Forecasting Peaks of COVID-19: Rather Focus on Forward and Backward Projections

Ruy Freitas Reis<sup>1</sup>, Rafael Sachetto Oliveira<sup>2</sup>, Bárbara de Melo Quintela<sup>1</sup>, Joventino de Oliveira Campos<sup>3</sup>, Johnny Moreira Gomes<sup>4</sup>, Bernardo Martins Rocha<sup>1,4</sup>, Marcelo Lobosco<sup>1,4</sup> and Rodrigo Weber dos Santos<sup>1,4\*</sup>

<sup>1</sup> Departamento de Ciência da Computação, Universidade Federal de Juiz de Fora, Juiz de Fora, Brazil, <sup>2</sup> Departamento de Ciência da Computação, Universidade Federal de São João del-Rei, São João del-Rei, Brazil, <sup>3</sup> Centro Federal de Educação Tecnológica de Minas Gerais, Leopoldina, Brazil, <sup>4</sup> Pós-Graduação em Modelagem Computacional, Universidade Federal de Juiz de Fora, Juiz de Fora, Brazil

## OPEN ACCESS

### Edited by:

Marc Jean Struelens,  
Université libre de Bruxelles, Belgium

### Reviewed by:

Julio Elias Normey-Rico,  
Federal University of Santa Catarina,  
Brazil

Theodore Gyle Lewis,  
Naval Postgraduate School,  
United States

### \*Correspondence:

Rodrigo Weber dos Santos  
rodrigo.weber@uffj.edu.br

### Specialty section:

This article was submitted to  
Infectious Diseases - Surveillance,  
Prevention and Treatment,  
a section of the journal  
Frontiers in Public Health

**Received:** 30 October 2020

**Accepted:** 10 February 2021

**Published:** 16 March 2021

### Citation:

Reis RF, Oliveira RS, Quintela BdM, Campos JdO, Gomes JM, Rocha BM, Lobosco M and dos Santos RW (2021) The Quixotic Task of Forecasting Peaks of COVID-19: Rather Focus on Forward and Backward Projections. *Front. Public Health* 9:623521. doi: 10.3389/fpubh.2021.623521

Over the last months, mathematical models have been extensively used to help control the COVID-19 pandemic worldwide. Although extremely useful in many tasks, most models have performed poorly in forecasting the pandemic peaks. We investigate this common pitfall by forecasting four countries' pandemic peak: Austria, Germany, Italy, and South Korea. Far from the peaks, our models can forecast the pandemic dynamics 20 days ahead. Nevertheless, when calibrating our models close to the day of the pandemic peak, all forecasts fail. Uncertainty quantification and sensitivity analysis revealed the main obstacle: the misestimation of the transmission rate. Inverse uncertainty quantification has shown that significant changes in transmission rate commonly precede a peak. These changes are a key factor in forecasting the pandemic peak. Long forecasts of the pandemic peak are therefore undermined by the lack of models that can forecast changes in the transmission rate, i.e., how a particular society behaves, changes of mitigation policies, or how society chooses to respond to them. In addition, our studies revealed that even short forecasts of the pandemic peak are challenging. Backward projections have shown us that the correct estimation of any temporal change in the transmission rate is only possible many days ahead. Our results suggest that the distance between a change in the transmission rate and its correct identification in the curve of active infected cases can be as long as 15 days. This is intrinsic to the phenomenon and how it affects epidemic data: a new case is usually only reported after an incubation period followed by a delay associated with the test. In summary, our results suggest the phenomenon itself challenges the task of forecasting the peak of the COVID-19 pandemic when only epidemic data is available. Nevertheless, we show that exciting results can be obtained when using the same models to project different scenarios of reduced transmission rates. Therefore, our results highlight that mathematical modeling can help control COVID-19 pandemic by backward projections that characterize the phenomena's essential features and forward projections when different scenarios and strategies can be tested and used for decision-making.

**Keywords:** COVID-19, epidemiology, mathematical modeling, uncertainty quantification, projection, forecasting

## 1. INTRODUCTION

Epidemiology is defined in the International Epidemiological Association's dictionary as "the study of the occurrence and distribution of health-related events, states, and processes in specified populations, including the study of the determinants influencing such processes, and the application of this knowledge to control relevant health problems" (1). One of its main objectives, as stated in the definition, is to provide data so governments can plan and execute actions to prevent and control diseases. The current COVID-19 pandemic has put epidemiology at the center of the debate as, to date, there are no antivirals with proven efficacy against the disease (2–6). The first vaccines have just become available, but it is unknown how long does the immunity last after vaccination. Due to the lack of pharmaceutical treatments, non-pharmaceutical interventions suggested by epidemiologists have been used by many countries to deal with the pandemic, more specifically to reduce transmission and the impact on healthcare systems (7–12).

Mathematical and computational tools can be used by epidemiological studies to describe and predict the dynamics of the spread of a disease over time and space (13, 14). In addition, these tools can be used to assess the impact of non-pharmaceutical interventions, such as isolation (15–18). Several models have been proposed to describe the spread of diseases (14). Over the last months, many more have been developed to represent the dynamics of populations and their interactions, as well as to forecast the dynamics of the COVID-19 pandemic. Most describe the spread of COVID-19 based on ODEs (Ordinary Differential Equations) (17, 19–28), but statistical (29, 30), chaotic (27), and stochastic/probabilistic models (16, 18, 31, 32) have also been used.

Determining the pandemic's peak is a piece of valuable information for planning the health resource needed to cope with the disease. In the case of COVID-19, it is also relevant for economic reasons since many countries adopted lockdowns to reduce the spread of the disease, impairing their gross domestic product and, consequently, their budget while increasing their health and social protection costs systems. Although most of the models found in the literature can have their parameters adjusted to COVID-19 data, i.e., to describe the behavior of its spread in different cities, regions, or countries, they usually fail to forecast the peak of the pandemic accurately (26–28). One could ask if one particular modeling technique would be more appropriate to forecast the dynamics of COVID-19 than others. However, a review of the literature does not clearly show that this hypothesis holds. In fact, in the literature we can find examples of forecasts that failed using SEIR (26), SIRD (25) extensions of SEIR with more compartments (28), statistical (33, 34), agent-based (35), machine-learning (36), and chaos-based theory models (27).

Another possible explanation for the failed forecasts could be related to the classical problem of overfitting (37), where the model can replicate the data it is adjusted to but fails on any attempt of extrapolation or forecasting. In this paper, we took several precautions to prevent the issue of overfitting. A simple mathematical model, based on the classical SIRD model, was adopted with a reduced number of parameters. We decided

to keep the model as simple as possible since adding more compartments increases the number of unknown parameters to be estimated, which hinders the accurate calibration of the model. We also used the methods of forward and inverse uncertainty quantification (UQ). The parameters of the models were treated as probability density functions (PDFs) during the task of model-to-data fitting (via inverse UQ) and during the tasks of forecasting and projections (via forward UQ). In addition, during the fitting phase, we also considered a possible discrepancy between model and reality (38). Nevertheless, in this paper, we show that the above precautions to avoid overfitting did not solve the problem of mispredicting the peak of COVID-19.

This work shows that this common pitfall is likely due to fast and unpredictable changes in the disease's transmission rate. The models are useful for predictions in a more controlled environment. It is like trying to predict the trajectory of a paper aeroplane on a windless day. It is much easier than during a storm. Likewise, it is challenging to predict significant changes in how a particular society behaves, mitigation policies, or how society responds to them during a pandemic. However, these all have a direct impact on the transmission rate, which in turn significantly affects the dynamics of the pandemic, as shown in previous studies (20, 39, 40). Nevertheless, differently from the weather, mitigation policies can be planned and controlled to some extent. This fact brings up the importance of projections of different scenarios during this pandemic. Different from forecasts, projections aim to study one or more hypothetical scenarios. In contrast, forecasts use the available data and try to predict future trends (41).

The models and techniques used in this work were first described in previous work (20). The model consists of a non-linear system of ordinary differential equations subject to uncertainty in some of its parameters and initial conditions. Probability density functions (PDFs) were used to describe the uncertainties associated with these parameters, so they are not scalars. Some of them, such as the transmission rate, are additionally time-dependent. After we adjusted the model's PDFs to a particular data (via inverse Uncertainty Quantification), the model was able to provide useful insights in terms of characterization of the pandemic dynamics in a particular country.

The current study uses four countries as examples, with distinct population sizes and demographics: Austria, Germany, Italy, and South Korea. For each country, we first show that the proposed model and methods correctly described the dynamics of total reported cases, active infected, and deaths when fitting model to data, i.e., our models can reproduce the different dynamics and peaks. Next, we show that our models can forecast the pandemic dynamics 20–30 days ahead when far from the peaks. However, all pandemic peaks' forecasts fail, even when adjusting the model to the data up to 5 days before the peak in each country. We further investigate this weakness, which is shared by many distinct models presented in the literature, by analyzing which parameter was misestimated via backward projection or inverse UQ. The analysis pointed to the misestimation of changes in the transmission rate near the peak as the primary source of error. In addition, the correct estimation

of any temporal change in the transmission rate was only possible many days ahead.

Finally, we performed projections, adjusted the model to the data up to 10 days before the peak, and focused on different scenarios that considered changes in the transmission rate. The projections that simulated significant reductions in the transmission rate were the ones where the pandemic peaks were closest to the real observed ones.

Therefore, our results highlight how mathematical models can help the fight against the COVID-19 pandemic: by characterizing important parameters that dictate the pandemic dynamics, as performed before in our previous work (20); and via projections, when different scenarios and strategies can be tested and used for decision-making. In addition, our analysis suggests that the misestimation of changes in the transmission rate near the peaks is the main source of error during the task of forecasting the peaks of COVID-19 pandemic.

## 2. MATERIALS AND METHODS

To demonstrate the impact of the transmission rate in forecasting the peak of COVID-19 pandemic, the parameters of our model (20) were calibrated according to total and active COVID-19 cases and deaths in three countries that have already achieved the pandemic peak: Austria, Germany, South Korea, and Italy. Peak predictions are performed considering a distinct number of available days.

The characterization of COVID-19 in these four countries is performed using inverse UQ techniques. Therefore, during the calibration of the model, the coefficients are treated as unknown probability density functions. Once estimated, the PDFs of the coefficients, their means, standard deviations (SD), and shape provide important information on model parameters that are essential in the characterization of the COVID-19 pandemic. The model and how it is adjusted are briefly described in this section to facilitate the understanding of the results. More details about the model, calibration of the parameters, and uncertainties can be found in our previous work (20).

### 2.1. Mathematical Model

The model used in this work (20) is based on the classic compartmental SIRD model (13, 14, 42–44), and was kept as simple as possible to reduce the number of unknown parameters to be estimated.

The model is described by the following set of equations:

$$\begin{cases} \frac{dS}{dt} = -\frac{\alpha(t)}{N}SI, \\ \frac{dI}{dt} = \frac{\alpha(t)}{N}SI - \beta I - \gamma I, \\ \frac{dR}{dt} = \gamma I, \\ \frac{dD}{dt} = \beta I, \\ I_r = \theta I, \\ R_r = \theta R, \\ C = I_r + R_r + D, \end{cases} \quad (1)$$

where  $S$ ,  $I$ ,  $R$ ,  $D$ ,  $I_r$ ,  $R_r$ , and  $C$  are the variables that represent the number of individuals within a population of size  $N$  that

are susceptible, infected, recovered, dead, reported as infected, reported as recovered, and total confirmed cases, respectively. The term  $\alpha(t) = a(t)b$  denotes the rate at which a susceptible individual becomes infected; where  $a(t)$  denotes the probability of contact and  $b$  the rate of infection. The function  $a(t)$  models temporal changes in the transmission rate:

$$a(t) = \begin{cases} 1, & \text{if } t < t_i, \\ \frac{r-1}{\Delta}(t - t_i) + 1, & \text{if } t_i \leq t \leq t_i + \Delta, \\ r, & \text{otherwise.} \end{cases} \quad (2)$$

Each transmission change starts at  $t_i$ , and is changed by a factor  $r$  at the final time  $t_i + \Delta$ . The mortality rate of infected individuals is modeled by the constant  $\beta = m(1/\tau_o)$ , where  $m$  is the probability of death and as the percentage of death among the reported cases of positive infection ( $I_r$ ). The number of days from infection until death is represented by  $\tau_o = \tau_1 + \tau_2$ , where  $\tau_1$  is the incubation time of the virus and  $\tau_2$  is the time between the first symptoms until death. Similarly, represented by  $\tau_r = \tau_1 + \tau_3$ , where  $\tau_r$  is the number of days from infection until recovery and  $\tau_3$  is the time between the first symptoms until recovery. The rate at which infected individuals recover from the virus is given by constant  $\gamma = (1 - m)(1/\tau_r)$ . Lastly, the percentage of confirmed infected individuals that are notified or reported is represented by  $\theta$ .

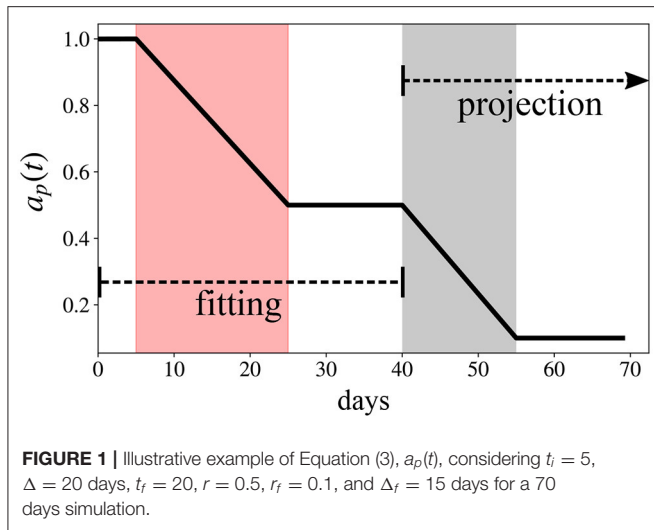
For making projections using the model beyond the last day used during the fitting, we consider  $\alpha(t) = a_p(t)b$ :

$$a_p(t) = \begin{cases} 1, & \text{if } t < t_i, \\ \frac{r-1}{\Delta}(t - t_i) + 1, & \text{if } t_i \leq t \leq t_i + \Delta, \\ r, & \text{if } t_i + \Delta < t \leq t_f, \\ \frac{r_f-r}{\Delta_f}(t - t_f) + r, & \text{if } t_f < t \leq t_f + \Delta_f, \\ r_f, & \text{otherwise.} \end{cases} \quad (3)$$

The function  $a_p(t)$  is similar to  $a(t)$  but it adds different scenarios in terms of how the transmission rate evolves after the last day of data used for model fitting. The constant  $t_f$  is the last day used during the calibration of the model,  $r_f$  is the final value of  $a_p(t)$  during the projection phase, and  $\Delta_f$  is the time interval for  $a_p(t)$  to change from  $r$  to  $r_f$ . **Figure 1** is an illustrative example for Equation (3).

### 2.2. Parameter Estimation and Uncertainty Quantification

Model parameters were adjusted using the differential evolution (DE) optimization method (45) implemented in the C programming language. The DE was used to estimate each of the parameters of the proposed mathematical model, respecting the limits established for each one of them [see (20)]. The parameter values were estimated based on official data from the epidemic reported in each country (Austria, Germany, S. Korea, and Italy) and made available by Dong et al. (46). Here, we are using  $\hat{I}(t)$  as the reported numbers of active cases,  $\hat{D}(t)$  the number of deaths, and  $\hat{C}(t)$  the total confirmed cases. The



following objective function, Equation (5), was used to minimize the relative error ( $R_E(\lambda, \hat{\lambda})$ ) between the data and the model:

$$R_E(\lambda, \hat{\lambda}) = \frac{\|\lambda(t, p) - \hat{\lambda}(t)\|_1}{\|\hat{\lambda}(t)\|_1}, \quad (4)$$

$$\min_p \left( \omega_1 R_E(I, \hat{I}) + \omega_2 R_E(D, \hat{D}) + \omega_3 R_E(C, \hat{C}) \right), \quad (5)$$

where  $p$  is the set of parameters to be estimated and  $\omega_n$  is a weight. For this work, we used  $\omega_1 = \omega_2 = \omega_3 = 1.0$ .

Some input parameters of the model are subject to uncertainties and variations, due to measurement errors, technical limitations, and resource availability. Parameters such as the incubation period, mortality, period from symptoms to death, period from symptoms to recovery, and the effectiveness of contact reduction are hampered by uncertainties, as suggested in data from literature. Therefore, when evaluating models such as the one studied in this work, it is important to evaluate the uncertainties from estimated input parameters, providing a confidence interval for the predictions.

Although the parameter intervals are reported in the literature, we still do not know how their probability densities functions behave. So, an inverse UQ technique was used to estimate the PDFs and corresponding uncertainties of the input parameters or coefficients of the model during model calibration. For each model parameter, we determined its PDF from the fitting procedure using the DE method. Among the offsprings generated by the DE, during the fitting process, we selected individuals with  $o(p) \leq 25\%$ , where  $o(p)$  is defined in Equation (6).

$$o(p) = \max_p \left( \omega_n R_E(I, \hat{I}), \omega_n R_E(D, \hat{D}), \omega_n R_E(C, \hat{C}) \right), \quad (6)$$

in which  $p$  is the set of parameters to be estimated and  $\omega_n$  are the same weights defined in Equation (5). Using this procedure,

we acknowledge a possible discrepancy of up to 25% between the model and reality.

Thus, from these samples, we estimate the covariance matrix and mean of all parameters. We use these data to generate a multivariate normal distribution to perform a forward UQ analysis via the Monte Carlo method with a total of 10,000 samples using the ChaosPy library (47).

Finally, we evaluate how the uncertainties in the input parameters of the model impact its outputs using forward UQ techniques. The forward UQ technique propagates the uncertainty of the input parameters to the outputs. Among the consolidated methods from the literature, Monte Carlo is one of the most used to perform uncertainty propagation (48–50). Briefly, this method draws samples of the input parameters and evaluates the model using them to provide statistical properties for the quantities of interest.

### 2.3. Data Sources

The model was calibrated using the data reported by the Center for Systems Science and Engineering at Johns Hopkins University (46), between 01/22/2020 and 12/20/2020. The bounds used for the parameters are described in (20).

### 2.4. Characterization

Tables 1–4 show the mean and standard deviation of all offspring solution with  $o(p) \leq 25\%$ . The value of  $N$  is  $9.00 \times 10^6$  for Austria,  $83.02 \times 10^6$  for Germany,  $60.42 \times 10^6$  for Italy and  $51.47 \times 10^6$  for South Korea.

## 3. RESULTS

### 3.1. The Calibrated Model Captures the Peak of the COVID-19 Pandemic

First, to check if the proposed model is able to fit the available data of countries during the peaks. A summary of the inverse UQ analysis results is presented in Table 1, which presents the mean and standard deviation (SD) of the estimated PDFs of the parameters for the three countries. Figure 2 compares the results of the fitted models to the original data for each country. For each time instant  $t$ ,  $I(t)$ ,  $C(t)$  and  $D(t)$  are PDFs, in response to the process of forward uncertainty quantification. It should be noted that the same model, with different parameters, was able to reproduce the distinct scenarios and peaks of the COVID-19 pandemic in Germany, Austria, Italy, and S. Korea. For the case of Italy, we fitted the recent second peak that has just been reached.

### 3.2. The Models Correctly Forecast the Dynamics of COVID-19 Away From the Peaks

Figure 3 shows that the models can correctly forecast the dynamics of COVID-19 for the four countries when away from the peaks. The match between forecasts and real data is observed to last between 20 and 30 days after the fitting phase, which used 30 days for all countries. We have chosen different but all recent phases for these forecasts. The curves' shapes are very distinct, but as long as the forecast is away from the peak, the

**TABLE 1** | Values of parameters used to fit model to data.

Symbol	Austria		Germany		S. Korea		Italy	
	Mean	STD	Mean	STD	Mean	STD	Mean	STD
$b$	$3.01 \times 10^{-1}$	$5.98 \times 10^{-3}$	$3.35 \times 10^{-1}$	$4.45 \times 10^{-3}$	$5.01 \times 10^{-1}$	$8.19 \times 10^{-3}$	$6.40 \times 10^{-2}$	$5.39 \times 10^{-4}$
$r$	$1.05 \times 10^{-2}$	$2.78 \times 10^{-3}$	$8.19 \times 10^{-2}$	$3.73 \times 10^{-3}$	$1.01 \times 10^{-2}$	$5.96 \times 10^{-4}$	$2.75 \times 10^{-1}$	$2.81 \times 10^{-2}$
$t_i$	$1.04 \times 10^1$	$1.06 \times 10^0$	$9.16 \times 10^0$	$9.47 \times 10^{-1}$	$5.53 \times 10^{-1}$	$3.63 \times 10^{-1}$	$9.47 \times 10^1$	$2.21 \times 10^0$
$\Delta$	$1.85 \times 10^1$	$1.47 \times 10^0$	$3.20 \times 10^1$	$1.23 \times 10^0$	$1.92 \times 10^1$	$4.66 \times 10^{-1}$	$5.48 \times 10^0$	$3.90 \times 10^0$
$\theta$	$8.95 \times 10^{-1}$	$1.80 \times 10^{-1}$	$6.38 \times 10^{-1}$	$2.72 \times 10^{-2}$	$9.71 \times 10^{-1}$	$6.21 \times 10^{-2}$	$9.75 \times 10^{-1}$	$9.76 \times 10^{-2}$
$\tau_1$	$3.20 \times 10^0$	$9.81 \times 10^{-1}$	$9.33 \times 10^0$	$2.18 \times 10^{-1}$	$1.31 \times 10^1$	$3.77 \times 10^{-1}$	$1.60 \times 10^1$	$3.71 \times 10^0$
$\tau_2$	$1.17 \times 10^1$	$2.05 \times 10^0$	$6.26 \times 10^0$	$4.71 \times 10^{-1}$	$1.74 \times 10^1$	$2.04 \times 10^0$	$2.47 \times 10^1$	$4.27 \times 10^0$
$\tau_3$	$1.56 \times 10^1$	$9.20 \times 10^{-1}$	$7.15 \times 10^0$	$1.11 \times 10^{-1}$	$1.64 \times 10^1$	$3.47 \times 10^{-1}$	$2.85 \times 10^1$	$3.44 \times 10^0$
$m$	$3.09 \times 10^{-2}$	$4.45 \times 10^{-3}$	$3.13 \times 10^{-2}$	$9.88 \times 10^{-4}$	$2.75 \times 10^{-2}$	$2.49 \times 10^{-3}$	$3.01 \times 10^{-2}$	$2.85 \times 10^{-3}$

$b$ , COVID-19 transmission rate;  $m$ , death probability;  $r$ , contact reduction;  $t_i$ , start of intervention policy;  $\Delta$ , duration of intervention policy;  $\tau_1$ , incubation period;  $\tau_2$ , period from symptoms to death;  $\tau_3$ , period from symptoms to recovery;  $\theta$ , fraction of notified cases. The model parameters were calibrated using data from the first day with more than 100 cases to 05/11/2020 for each country for Austria, Germany, and S. Korea. For Italy the model parameters were calibrated from 08/17/2020 to 12/14/2020.

**TABLE 2** | Value of parameters used to make short forecasts.

Symbol	Austria		Germany		S. Korea		Italy	
	Mean	STD	Mean	STD	Mean	STD	Mean	STD
$b$	$7.59 \times 10^{-2}$	$3.57 \times 10^{-3}$	$5.02 \times 10^{-2}$	$2.73 \times 10^{-3}$	$6.27 \times 10^{-2}$	$1.10 \times 10^{-3}$	$1.90 \times 10^{-2}$	$1.59 \times 10^{-3}$
$r$	$1.96 \times 10^0$	$9.81 \times 10^{-2}$	$2.07 \times 10^0$	$1.15 \times 10^{-1}$	$1.83 \times 10^0$	$1.62 \times 10^{-1}$	$4.10 \times 10^{-1}$	$1.43 \times 10^{-1}$
$t_i$	$7.51 \times 10^{-1}$	$2.97 \times 10^0$	$1.50 \times 10^1$	$1.14 \times 10^0$	$2.63 \times 10^1$	$1.45 \times 10^0$	$5.48 \times 10^{-1}$	$1.75 \times 10^0$
$\Delta$	$2.85 \times 10^1$	$3.43 \times 10^0$	$1.26 \times 10^1$	$1.33 \times 10^0$	$1.83 \times 10^0$	$1.34 \times 10^0$	$1.19 \times 10^1$	$2.06 \times 10^0$
$\theta$	$3.90 \times 10^{-1}$	$3.21 \times 10^{-2}$	$2.77 \times 10^{-1}$	$4.00 \times 10^{-2}$	$3.85 \times 10^{-1}$	$3.45 \times 10^{-2}$	$6.49 \times 10^{-1}$	$5.76 \times 10^{-2}$
$\tau_1$	$4.33 \times 10^0$	$1.23 \times 10^0$	$5.20 \times 10^0$	$1.70 \times 10^0$	$8.99 \times 10^0$	$2.61 \times 10^0$	$7.95 \times 10^0$	$3.05 \times 10^0$
$\tau_2$	$1.63 \times 10^1$	$3.00 \times 10^0$	$1.75 \times 10^1$	$2.21 \times 10^0$	$1.67 \times 10^1$	$2.86 \times 10^0$	$1.74 \times 10^1$	$3.92 \times 10^0$
$\tau_3$	$9.28 \times 10^0$	$1.22 \times 10^0$	$9.91 \times 10^0$	$1.63 \times 10^0$	$1.18 \times 10^1$	$2.52 \times 10^0$	$1.97 \times 10^1$	$3.61 \times 10^0$
$m$	$4.33 \times 10^{-3}$	$7.07 \times 10^{-4}$	$5.62 \times 10^{-3}$	$1.65 \times 10^{-3}$	$1.00 \times 10^{-2}$	$1.21 \times 10^{-3}$	$4.27 \times 10^{-2}$	$3.61 \times 10^{-3}$

$b$ , COVID-19 transmission rate;  $m$ , death probability;  $r$ , contact reduction;  $t_i$ , start of intervention policy;  $\Delta$ , duration of intervention policy;  $\tau_1$ , incubation period;  $\tau_2$ , period from symptoms to death;  $\tau_3$ , period from symptoms to recovery;  $\theta$ , fraction of notified cases. The model parameters were calibrated using 30 days of data. The model was then used to forecast the following 30 days for Austria, Germany, S. Korea, and Italy.

**TABLE 3** | Value of parameters used to predict and to project the pandemic peak considering 5 days before the peak.

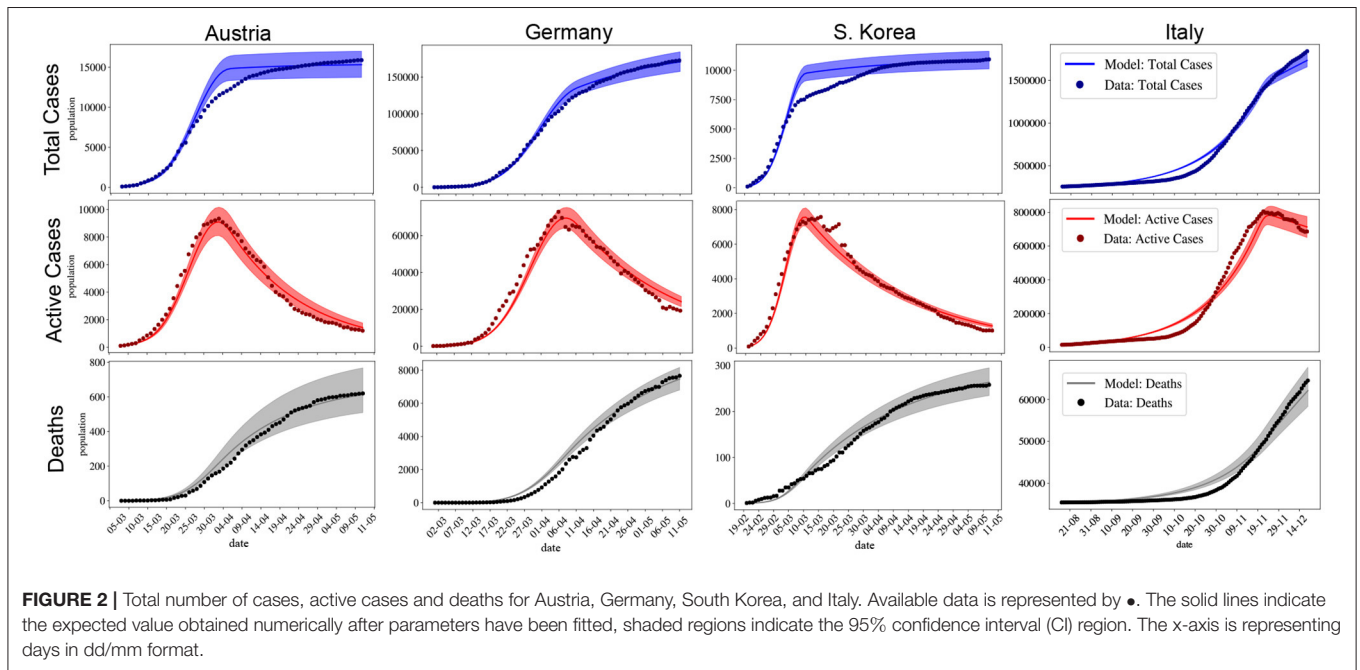
Symbol	Austria		Germany		S. Korea		Italy	
	Mean	STD	Mean	STD	Mean	STD	Mean	STD
$b$	$2.94 \times 10^{-1}$	$2.10 \times 10^{-3}$	$3.02 \times 10^{-1}$	$1.36 \times 10^{-3}$	$5.88 \times 10^{-1}$	$5.69 \times 10^{-3}$	$4.61 \times 10^{-2}$	$1.97 \times 10^{-4}$
$r$	$3.95 \times 10^{-1}$	$1.74 \times 10^{-2}$	$2.10 \times 10^{-1}$	$9.83 \times 10^{-3}$	$1.58 \times 10^{-1}$	$6.25 \times 10^{-3}$	$1.77 \times 10^0$	$1.48 \times 10^{-2}$
$t_i$	$9.07 \times 10^0$	$5.62 \times 10^{-1}$	$1.32 \times 10^1$	$4.24 \times 10^{-1}$	$5.53 \times 10^{-2}$	$1.96 \times 10^{-1}$	$4.87 \times 10^1$	$1.00 \times 10^0$
$\Delta$	$1.19 \times 10^1$	$5.88 \times 10^{-1}$	$1.78 \times 10^1$	$4.44 \times 10^{-1}$	$1.34 \times 10^1$	$3.13 \times 10^{-1}$	$2.52 \times 10^0$	$1.70 \times 10^0$
$\theta$	$8.52 \times 10^{-1}$	$3.42 \times 10^{-2}$	$6.81 \times 10^{-1}$	$2.35 \times 10^{-2}$	$9.25 \times 10^{-1}$	$4.22 \times 10^{-2}$	$1.52 \times 10^{-1}$	$2.37 \times 10^{-2}$
$\tau_1$	$1.40 \times 10^1$	$1.66 \times 10^{-1}$	$1.40 \times 10^1$	$2.16 \times 10^{-1}$	$1.40 \times 10^1$	$1.91 \times 10^{-1}$	$3.96 \times 10^1$	$1.36 \times 10^0$
$\tau_2$	$1.77 \times 10^1$	$1.33 \times 10^0$	$2.14 \times 10^1$	$1.10 \times 10^0$	$1.97 \times 10^1$	$1.18 \times 10^0$	$5.94 \times 10^1$	$3.02 \times 10^0$
$\tau_3$	$1.70 \times 10^1$	$1.92 \times 10^{-1}$	$1.69 \times 10^1$	$2.86 \times 10^{-1}$	$1.70 \times 10^1$	$1.22 \times 10^{-1}$	$2.09 \times 10^1$	$1.82 \times 10^0$
$m$	$3.39 \times 10^{-2}$	$6.41 \times 10^{-4}$	$3.38 \times 10^{-2}$	$5.69 \times 10^{-4}$	$3.35 \times 10^{-2}$	$1.87 \times 10^{-3}$	$1.01 \times 10^{-2}$	$7.36 \times 10^{-4}$

$b$ , COVID-19 transmission rate;  $m$ , death probability;  $r$ , contact reduction;  $t_i$ , start of intervention policy;  $\Delta$ , duration of intervention policy;  $\tau_1$ , incubation period;  $\tau_2$ , period from symptoms to death;  $\tau_3$ , period from symptoms to recovery;  $\theta$ , fraction of notified cases. The model parameters were calibrated using data from the first day with more than 100 cases to 5 days before the pandemic peak for Austria, Germany, and S. Korea. For Italy the model parameters were calibrated from 08/17/2020 to 5 days before the second peak.

**TABLE 4** | Value of parameters used to predict and to project the pandemic peak considering 10 days before the peak.

Symbol	Austria		Germany		S. Korea		Italy	
	Mean	STD	Mean	STD	Mean	STD	Mean	STD
$b$	$3.15 \times 10^{-1}$	$1.97 \times 10^{-3}$	$3.04 \times 10^{-1}$	$1.33 \times 10^{-3}$	$5.87 \times 10^{-1}$	$4.20 \times 10^{-3}$	$4.56 \times 10^{-2}$	$9.43 \times 10^{-5}$
$r$	$5.97 \times 10^{-1}$	$1.72 \times 10^{-2}$	$2.90 \times 10^{-1}$	$2.38 \times 10^{-2}$	$1.29 \times 10^{-1}$	$5.29 \times 10^{-3}$	$1.92 \times 10^0$	$7.42 \times 10^{-3}$
$t_i$	$4.73 \times 10^0$	$5.58 \times 10^{-1}$	$1.46 \times 10^1$	$4.06 \times 10^{-1}$	$2.76 \times 10^{-2}$	$1.38 \times 10^{-1}$	$5.12 \times 10^1$	$4.55 \times 10^{-1}$
$\Delta$	$1.13 \times 10^1$	$5.62 \times 10^{-1}$	$1.14 \times 10^1$	$4.50 \times 10^{-1}$	$1.40 \times 10^1$	$1.42 \times 10^{-1}$	$3.07 \times 10^0$	$8.16 \times 10^{-1}$
$\theta$	$9.96 \times 10^{-1}$	$1.92 \times 10^{-2}$	$9.76 \times 10^{-1}$	$2.83 \times 10^{-2}$	$8.38 \times 10^{-1}$	$3.57 \times 10^{-2}$	$1.74 \times 10^{-1}$	$3.23 \times 10^{-2}$
$\tau_1$	$1.40 \times 10^1$	$3.00 \times 10^{-1}$	$1.40 \times 10^1$	$2.65 \times 10^{-1}$	$1.40 \times 10^1$	$1.91 \times 10^{-1}$	$3.93 \times 10^1$	$1.53 \times 10^0$
$\tau_2$	$2.18 \times 10^1$	$7.73 \times 10^{-1}$	$2.16 \times 10^1$	$9.60 \times 10^{-1}$	$1.67 \times 10^1$	$1.26 \times 10^0$	$5.90 \times 10^1$	$3.10 \times 10^0$
$\tau_3$	$1.70 \times 10^1$	$2.44 \times 10^{-1}$	$1.69 \times 10^1$	$3.97 \times 10^{-1}$	$1.70 \times 10^1$	$2.84 \times 10^{-1}$	$2.21 \times 10^1$	$1.53 \times 10^0$
$m$	$2.97 \times 10^{-2}$	$1.01 \times 10^{-3}$	$3.36 \times 10^{-2}$	$1.57 \times 10^{-3}$	$3.37 \times 10^{-2}$	$1.20 \times 10^{-3}$	$1.02 \times 10^{-2}$	$1.03 \times 10^{-3}$

$b$ , COVID-19 transmission rate;  $m$ , death probability;  $r$ , contact reduction;  $t_i$ , start of intervention policy;  $\Delta$ , duration of intervention policy;  $\tau_1$ , incubation period;  $\tau_2$ , period from symptoms to death;  $\tau_3$ , period from symptoms to recovery;  $\theta$ , fraction of notified cases. The model parameters were calibrated using data from the first day with more than 100 cases to 10 days before the pandemic peak for Austria, Germany, and S. Korea. For Italy the model parameters were calibrated from 08/17/2020 to 10 days before the second peak.



**FIGURE 2** | Total number of cases, active cases and deaths for Austria, Germany, South Korea, and Italy. Available data is represented by •. The solid lines indicate the expected value obtained numerically after parameters have been fitted, shaded regions indicate the 95% confidence interval (CI) region. The x-axis is representing days in dd/mm format.

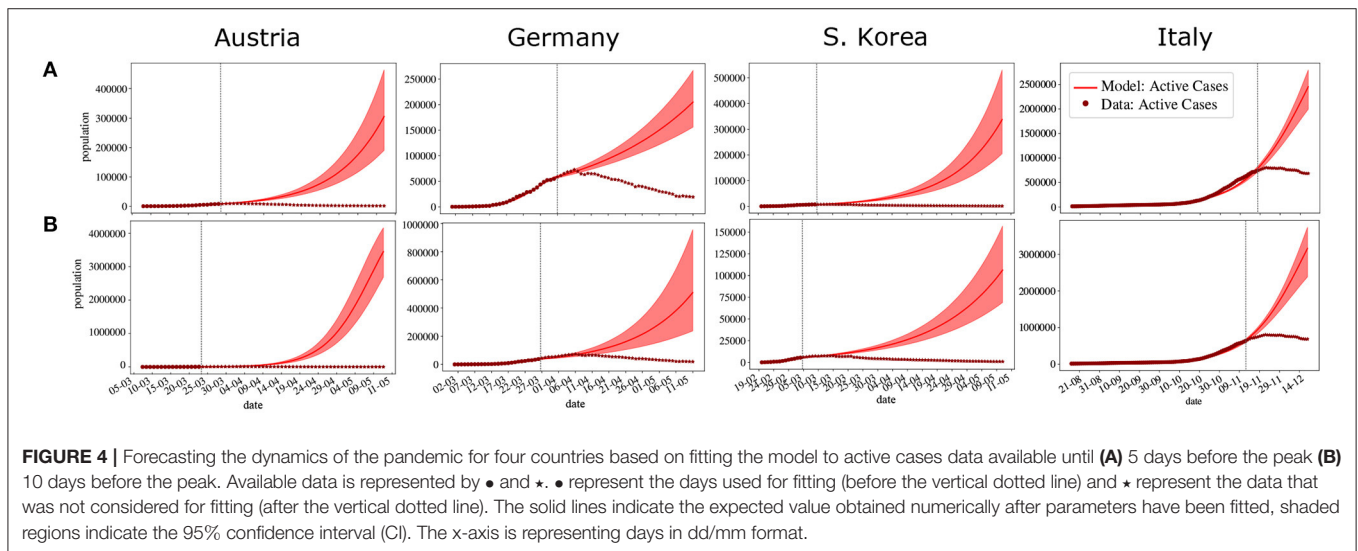
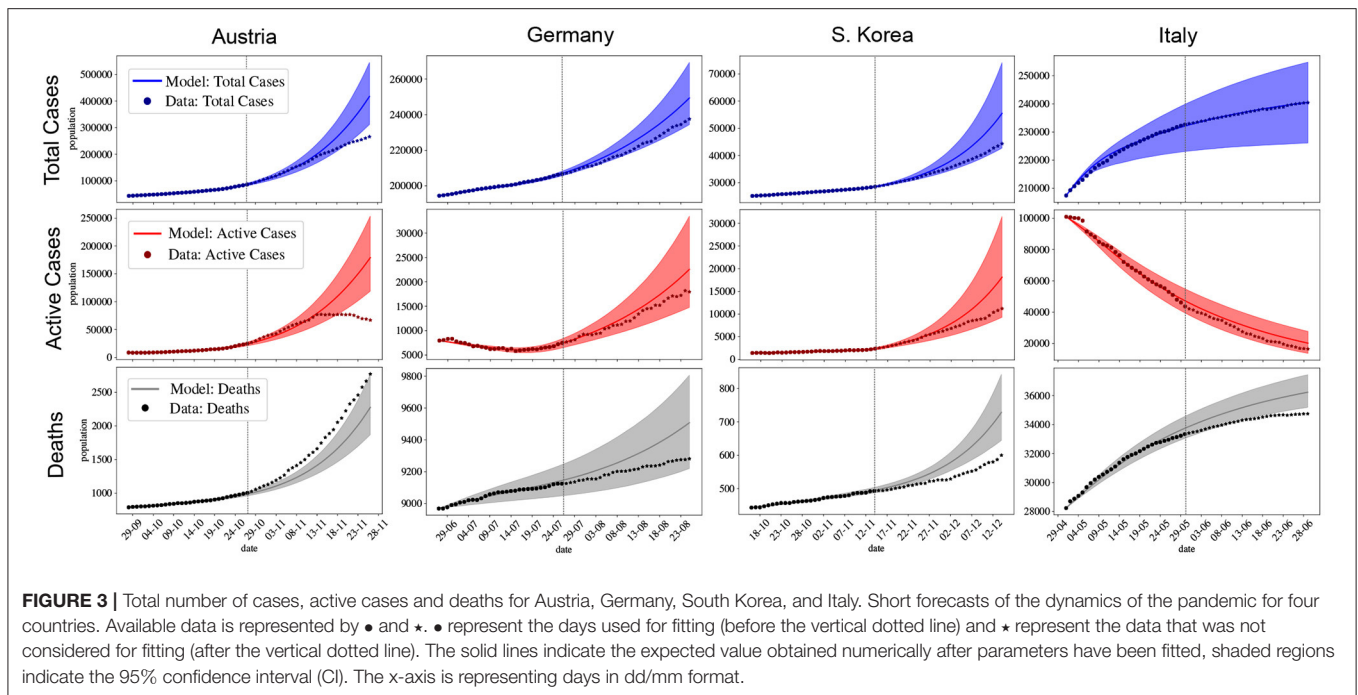
prediction is within the calculated interval of confidence. Note the particular valley shape captured by the forecast for Germany. Table 2 presents the results of the calibration process.

### 3.3. A Shared Weakness: Forecasting the Peak of COVID-19

As mentioned in the introduction, a literature review shows that many models fail to forecast the peak of the pandemic (26–28), regardless if they are based on SEIR models (26), SIRD (25) extensions of SEIR with more compartments (28), statistical (33, 34), agent-based (35), machine-learning (36), or chaos-based theory models (27).

The same happens with our model. The following experiment was performed: the parameters were adjusted again using

data available for active cases until 10 days before achieving the pandemic peak. Then, we tried to predict the number of active cases in the next days. The same experiment was then performed adjusting the data available until 5 days before achieving the pandemic peak. We chose to calibrate the models up to 5 or 10 days before the peak to show that even when it is very close to occurring, forecasts may fail. A summary of the results of the inverse UQ analysis is presented in Tables 3, 4, presenting the mean and standard deviation of the estimated PDFs of the parameters adjusted for the three countries using data until 5 and 10 days, respectively, before the pandemic peak. Figure 4 presents the forecasts. As one can observe, even in the case where the model was adjusted using data available 5 days before the



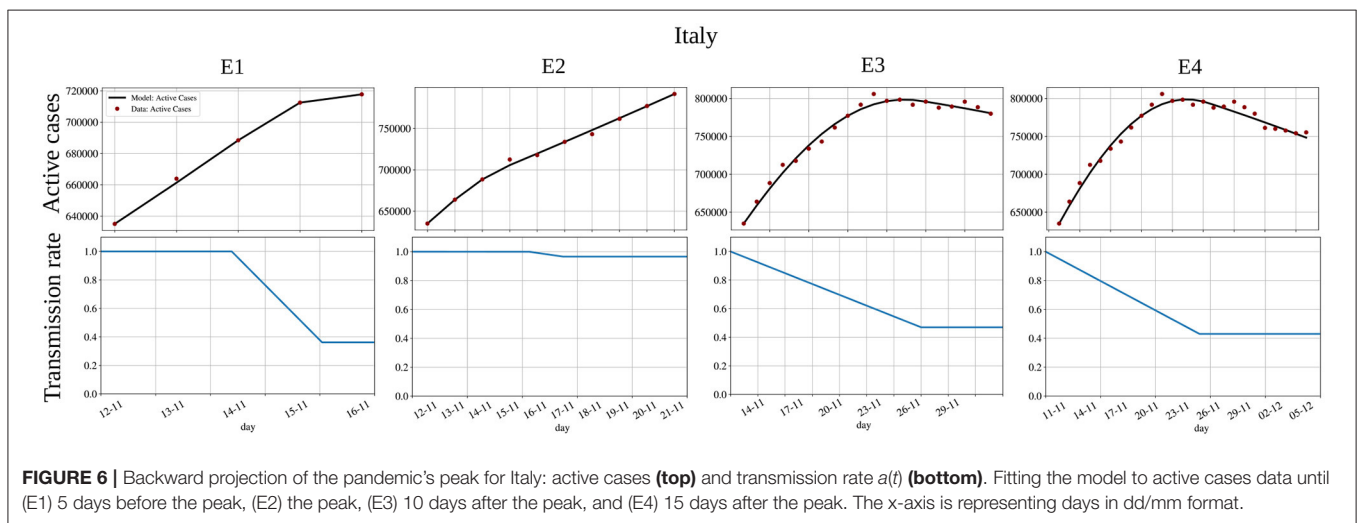
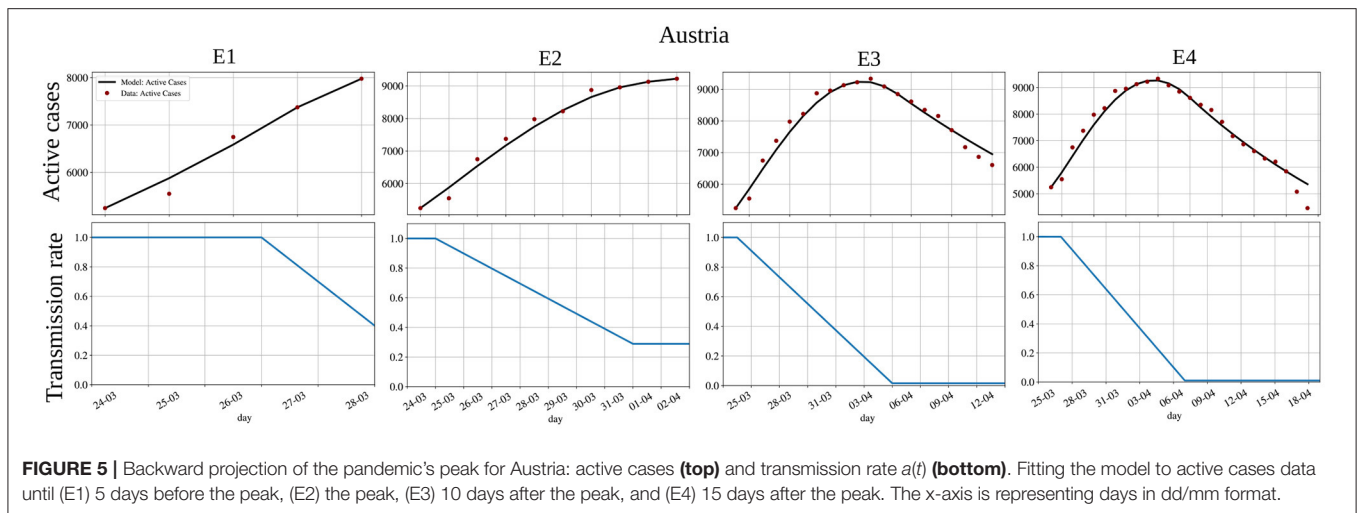
peak, the calibrated model was not able to correctly predict the peaks.

### 3.4. The Main Source of Error: Misestimation of Transmission Rates Near the Peak

In this section we investigate the possible sources of errors when forecasting the peaks of COVID-19. First, by comparing the inverse uncertainty quantification results presented in **Tables 1, 3, 4** we can observe that the main difference between the simulations that capture the peaks (**Figure 2**) and those that do not (**Figure 4**) lies on the estimation of the parameters that

describe the time-varying transmission rate ( $a(t)$ ):  $t_i$ ,  $r$ , and  $\Delta$ . When the peak is well capture by the model,  $a(t)$  reduces significantly near the peak.

We continue this investigation by performing a sequence of four backward projection experiments, named E1 to E4. To describe these experiments, let  $P$  denotes the day of the peak in the corresponding country considered in this study. The first experiment (E1) calibrates the model parameters up to 5 days before the peak, as performed before, but focusing near the peak, from  $P - 10$  to  $P - 5$  days. For the second experiment (E2), we expand the model calibration toward the peak, from  $P - 10$  to  $P$ . We continue expanding the window surrounding the peak in the third experiment (E3), performing the adjustment from



$P - 10$  to  $P + 10$  days. Finally, in the last experiment (E4), the last calibration is performed from  $P - 10$  to  $P + 15$  days. Then we compare all the parameters to check those that vary most between the different calibrations, which include the one we used before that failed forecasting the peak (E1,  $P - 10$  to  $P - 5$ ) and the one that captures the peak (E4). Once again, the main difference between the different experiments' estimated parameters was on those that describe the time-varying transmission rate ( $a(t)$ ):  $t_i$ ,  $r$ , and  $\Delta$ .

**Figure 5** presents the experiments E1-E4 for Austria and each corresponding estimated  $a(t)$ . In this case, we note that from the adjusted and experimental data of E1 (active cases), there is little to no indication that we are slowing down and reaching the peak. Also, by comparing the shapes of the estimated  $a(t)$ , we observe that only 10 days after the peak (E3), the crucial information on how the transmission rate evolves before the peak converges [the shapes of  $a(t)$  obtained in E3 and E4 are nearly the same].

**Figure 6** presents the experiments E1-E4 for Italy and each corresponding estimated  $a(t)$ . In this case, we note that even

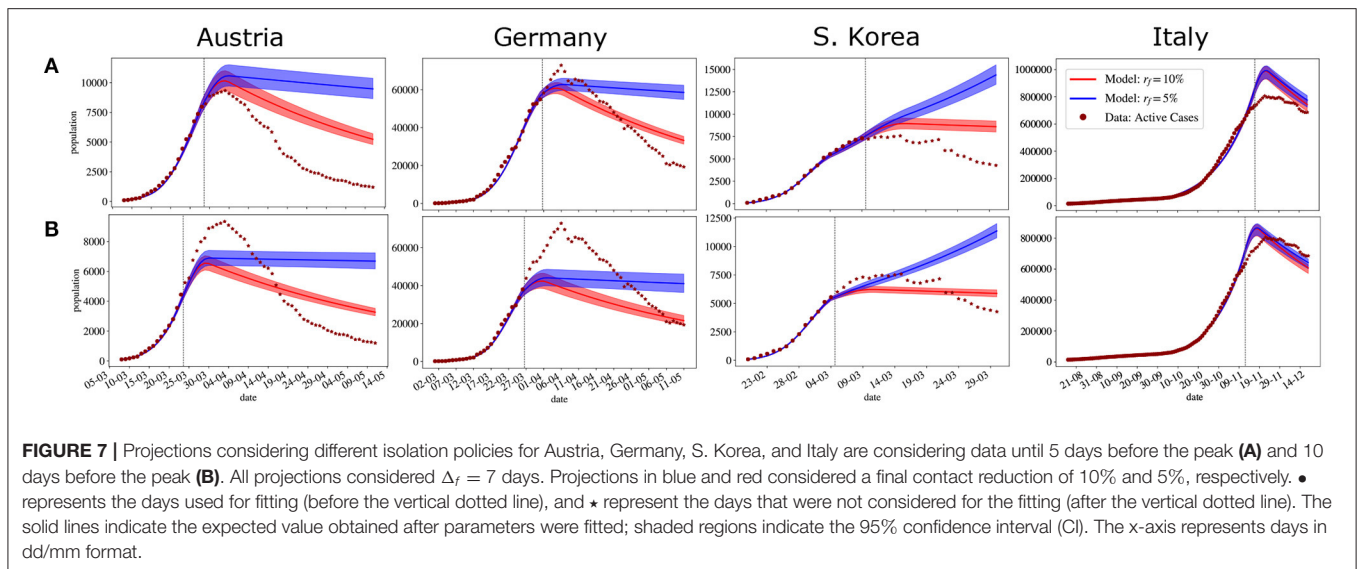
from the adjusted and experimental data of E2 (active cases from  $P - 10$  to  $P$ ), there is little to no indication that we are slowing down and reaching the peak. Again, by comparing the shapes of the estimated  $a(t)$ , we observe that only 15 days after the peak (E4), we can correctly estimate the shape of  $a(t)$  before the peak.

### 3.5. Projections Considering Different Scenarios of Transmission Rate Reduction

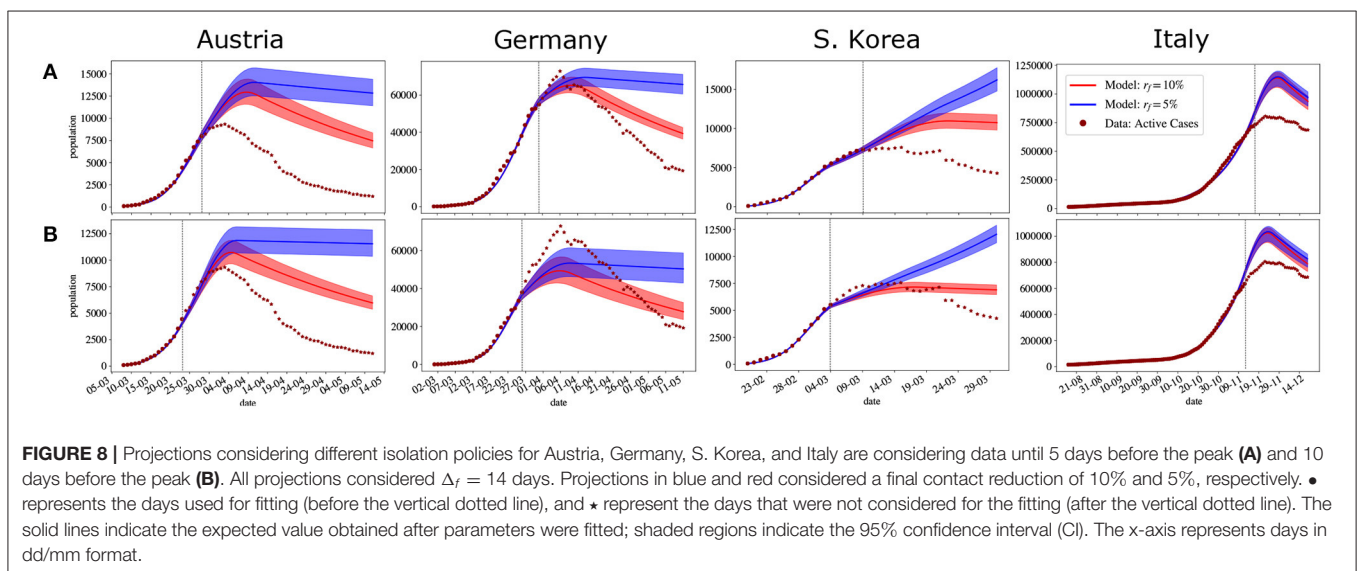
The last experiment makes projections considering different scenarios of transmission rate reduction. Again, the parameters were adjusted using data available for active cases until 5 and 10 days before achieving the peak (in the active cases) for the four countries. The idea here is to evaluate the impact of different shapes of  $a_p(t)$  after the calibration, i.e., different scenarios of transmission rate reduction.

Since the projection is done considering the same dates used for the forecasts, the values in **Tables 3, 4** are the same for both experiments. In the projections we consider





**FIGURE 7 |** Projections considering different isolation policies for Austria, Germany, S. Korea, and Italy are considering data until 5 days before the peak (A) and 10 days before the peak (B). All projections considered  $\Delta_f = 7$  days. Projections in blue and red considered a final contact reduction of 10% and 5%, respectively. ● represents the days used for fitting (before the vertical dotted line), and ★ represent the days that were not considered for the fitting (after the vertical dotted line). The solid lines indicate the expected value obtained after parameters were fitted; shaded regions indicate the 95% confidence interval (CI). The x-axis represents days in dd/mm format.



**FIGURE 8 |** Projections considering different isolation policies for Austria, Germany, S. Korea, and Italy are considering data until 5 days before the peak (A) and 10 days before the peak (B). All projections considered  $\Delta_f = 14$  days. Projections in blue and red considered a final contact reduction of 10% and 5%, respectively. ● represents the days used for fitting (before the vertical dotted line), and ★ represent the days that were not considered for the fitting (after the vertical dotted line). The solid lines indicate the expected value obtained after parameters were fitted; shaded regions indicate the 95% confidence interval (CI). The x-axis represents days in dd/mm format.

two distinct values for  $r_f$ , 0.10 and 0.05, i.e., two different final values for  $a_p(t)$ , and two distinct values for  $\Delta_f$ , 7 and 14. All projections considered  $t_i = 0$ . Therefore, we are considering different scenarios where transmission rate reduces after the calibration  $a_p(t)$ , decreasing to 0.10 or 0.05, after 1 or 2 weeks.

Figure 8 presents the projections for the four countries with  $\Delta_f = 7$ . The peaks become visible when projecting a more significant reduction in the transmission rates, i.e., with  $r = 0.05$ . Figure 7 presents the projections with  $\Delta_f = 14$ . The results are similar to the previous projections. The main difference is that the peaks are higher and occur further ahead. These results support that the fastest way to control the pandemic is with strict mitigation policies that can significantly reduce the transmission rate in a short period.

### 4. DISCUSSION

First of all, one should observe that the model used in this work can reproduce the dynamics of COVID-19 for distinct countries. The pandemic peak for all countries considered here was utterly determined since the adjusted model captures both the day in which the peak occurs as well as its maximum value, as shown in Figure 2. Figure 3 also shows that the model can be very useful in forecasting the dynamics of COVID-19.

There is, however, one main weakness of this model: forecasts near the pandemic peak usually fail. In Figure 4, the same method was applied to adjust the parameters of the model, but this time we did not use the entire dataset. The model was adjusted using data until 5 and 10 days before the peak of the pandemic. As can be observed in Figure 4, the forecasts

overestimate the number of active cases and mispredict the peaks by more than a month.

In this work, we have used many sophisticated tools based on forward and inverse UQ to identify the source of this problem. First, we compared the inverse UQ results presented in **Tables 1, 3, 4** and observed that the main difference between the simulations that capture the peaks (**Figure 2**) and those that do not (**Figure 4**) lies on the estimation of the parameters that describe the time-varying transmission rate ( $a(t)$ ):  $t_i$ ,  $r$ , and  $\Delta$ . When the peak is well-captured by the model,  $a(t)$  reduces significantly near the peak. This reduction in the transmission rate is key in forecasting the pandemic peak. Long forecasts of the pandemic peak are therefore undermined by the lack of models that can forecast changes in the transmission rate, i.e., how a particular society behaves, changes of mitigation policies, or how society chooses to respond to them.

In addition, we performed a sequence of four backward projection experiments, named (E1) to (E4). The first experiment (E1) calibrates the model parameters up to 5 days before the peak. The second one (E2) calibrates it up to the peak, (E3) up to the peak plus 10 days, and (E4) up to the peak plus 15 days. Then we compared all the parameters to check those that vary most between the different calibrations. Once again, the main difference between the different experiments' estimated parameters was on those that describe the time-varying transmission rate,  $a(t)$ . **Figures 5, 6** show also that the shape of the estimated  $a(t)$  only converges to the correct one when using epidemic data that includes many days after the peak. These results clearly show a delay between changes in the transmission rate and their impact on the curve of active cases, which is about 15 days. This is intrinsic to the phenomenon and how it affects epidemic data: a new case is usually only reported after an incubation period followed by a delay associated with the test.

The presence of this delay suggests that the challenging task of forecasting the pandemic peak might require additional data and constant monitoring to capture the transmission rate better.

The aforementioned results clearly show how inverse UQ and backward projections can provide important information on the dynamics of the COVID-19 pandemic. Finally, we have also performed forward projections to assess different scenarios of transmission rate reduction. **Figures 7, 8** show how significant

changes in the transmission rate impacts the dynamics and influences the location of the peaks. In addition, these results support that the fastest way to control the pandemic is with strict mitigation policies that can significantly reduce the transmission rate in a short period.

In summary, our results highlight how mathematical models can help the fight against the COVID-19 pandemic: by characterizing important parameters that dictate the dynamics of the pandemic, as performed before in our previous work (20); and via projections, when different scenarios can be tested and used for decision-making. In addition, they suggest that forecasting the peaks of COVID-19 can be quixotic due to the challenges that involve a precise estimation of how the transmission rate evolves.

## DATA AVAILABILITY STATEMENT

The dataset used for this study can be found in the CSSEGIS and Data at: <http://github.com/CSSEGISandData/COVID-19>.

## AUTHOR CONTRIBUTIONS

RF, BM, and RS: software, methodology, and formal analysis. BdM and JM: formal analysis. JdO: software and formal analysis. ML: methodology and formal analysis. RW: conceptualization, software, methodology, and formal analysis. All authors: writing the original draft, review and editing, and have read and approved the final manuscript.

## FUNDING

This work was partially supported by Fapemig, CNPq, CAPES, UFJF, UFSJ, and CEFET-MG.

## ACKNOWLEDGMENTS

The authors would like to thank the researchers Guilherme Côrtes Fernandes, Thaiz Ruberti Schmal and Luis Paulo da Silva Barra for the motivating discussions that improve the quality of this work.

## REFERENCES

- Porta M, Greenland S, Hernán M, dos Santos Silva I, Last JM. *A Dictionary of Epidemiology*. New York, NY: Oxford University Press (2014).
- Le TT, Andreadakis Z, Kumar A, Roman RG, Tollefsen S, Saville M, et al. The COVID-19 vaccine development landscape. *Nat Rev Drug Discov*. (2020). 19:305–6. doi: 10.1038/d41573-020-00073-5
- Zhou Y, Hou Y, Shen J, Huang Y, Martin W, Cheng F. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discov*. (2020) 6:14. doi: 10.1038/s41421-020-0153-3
- Dong L, Hu S, Gao J. Discovering drugs to treat coronavirus disease 2019. (COVID-19). *Drug Discov Therapeut*. (2020) 14:58–60. doi: 10.5582/ddt.2020.01012
- Baden LR, Rubin EJ. COVID-19—The search for effective therapy. *N Engl J Med*. (2020) 382:1851–2. doi: 10.1056/NEJMe2005477
- Duan K, Liu B, Li C, Zhang H, Yu T, Qu J, et al. Effectiveness of convalescent plasma therapy in severe COVID-19 patients. *Proc Natl Acad Sci USA*. (2020) 117:9490–6. doi: 10.1073/pnas.2007408117
- Ferguson NM, Laydon D, Nedjati-Gilani G, Imai N, Ainslie K, Baguelin M, et al. Report 9: impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. *Imperial Coll Lond*. (2020). doi: 10.25561/77482
- Wilder-Smith A, Freedman DO. Isolation, quarantine, social distancing and community containment: pivotal role for old-style public health measures in the novel coronavirus (2019-nCoV) outbreak. *J Travel Med*. (2020) 27. doi: 10.1093/jtm/taaa020

9. Wilder-Smith A, Chiew CJ, Lee VJ. Can we contain the COVID-19 outbreak with the same measures as for SARS? *Lancet Infect Dis.* (2020). 20:e102–7. doi: 10.1016/S1473-3099(20)30129-8
10. Lee VJ, Chiew CJ, Khong WX. Interrupting transmission of COVID-19: lessons from containment efforts in Singapore. *J Travel Med.* (2020) 27:taaa039. doi: 10.1093/jtm/taaa039
11. Heymann DL, Shindo N. COVID-19: what is next for public health? *Lancet.* (2020) 395:542–5. doi: 10.1016/S0140-6736(20)30374-3
12. Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science.* (2020) 368:395–400. doi: 10.1126/science.aba9757
13. Diekmann O, Heesterbeek JAP. Mathematical epidemiology of infectious diseases: model building, analysis and interpretation. *Wiley Ser Math Comput Biol.* (2000) 30:186. doi: 10.1093/ije/30.1.186
14. Hethcote HW. The mathematics of infectious diseases. *SIAM Rev.* (2000) 42:599–653. doi: 10.1137/S0036144500371907
15. Kantner M, Koprucki T. Beyond just “flattening the curve”: Optimal control of epidemics with purely non-pharmaceutical interventions. *J Math Industry.* (2020) 10:1–23. doi: 10.1186/s13362-020-00091-3
16. Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *Lancet Glob Health.* (2020) 8:e488–96. doi: 10.1016/S2214-109X(20)30074-7
17. Giordano G, Blanchini F, Bruno R, Colaneri P, Di Filippo A, Di Matteo A, et al. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nat Med.* (2020). 26:855–60. doi: 10.1038/s41591-020-0883-7
18. Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect Dis.* (2020) 20:553–8. doi: 10.1016/S1473-3099(20)30144-4
19. Morato MM, Bastos SB, Cajueiro DO, Normey-Rico JE. An optimal predictive control strategy for COVID-19 (SARS-CoV-2) social distancing policies in Brazil. *Annu Rev Control.* (2020) 50:417–31. doi: 10.1016/j.arcontrol.2020.07.001
20. Reis RF, de Melo Quintela B, de Oliveira Campos J, Gomes JM, Rocha BM, Lobosco M, et al. Characterization of the COVID-19 pandemic and the impact of uncertainties, mitigation strategies, and underreporting of cases in South Korea, Italy, and Brazil. *Chaos Solitons Fractals.* (2020) 136:109888. doi: 10.1016/j.chaos.2020.109888
21. Morato MM, Pataro IML, Americano da Costa MV, Normey-Rico JE. A parametrized nonlinear predictive control strategy for relaxing COVID-19 social distancing measures in Brazil. *ISA Trans.* (in press). doi: 10.1016/j.isatra.2020.12.012
22. Lin Q, Zhao S, Gao D, Lou Y, Yang S, Musa SS, et al. A conceptual model for the coronavirus disease (2019). (COVID-19) outbreak in Wuhan, China with individual reaction governmental action. *Int J Infect Dis.* (2020). 93:211–6. doi: 10.1016/j.ijid.2020.02.058
23. Anastassopoulou C, Russo L, Tsakris A, Siettos C. Data-based analysis, modelling and forecasting of the COVID-19 outbreak. *PLoS ONE.* (2020) 15:e0230405. doi: 10.1371/journal.pone.0230405
24. Wu JT, Leung K, Bushman M, Kishore N, Niehus R, de Salazar PM, et al. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nat Med.* (2020) 26:506–10. doi: 10.1038/s41591-020-0822-7
25. Fanelli D, Piazza F. Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos Solitons Fractals.* (2020) 134:109761. doi: 10.1016/j.chaos.2020.109761
26. Kuniya T. Prediction of the epidemic peak of Coronavirus Disease in Japan, 2020. *J Clin Med.* (2020) 9:789. doi: 10.3390/jcm9030789
27. Mangiarotti S, Peyre M, Zhang Y, Huc M, Roger F, Kerr Y. Chaos theory applied to the outbreak of COVID-19: an ancillary approach to decision making in pandemic context. *Epidemiol Infect.* (2020) 148:e95. doi: 10.1017/S0950268820000990
28. Liu Z, Magal P, Webb G. Predicting the number of reported and unreported cases for the COVID-19 epidemics in China, South Korea, Italy, France, Germany and United Kingdom. *J Theor Biol.* (2021) 509:110501. doi: 10.1016/j.jtbi.2020.110501
29. Roosa K, Lee Y, Luo R, Kirpich A, Rothenberg R, Hyman JM, et al. Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. *Infect Dis Modell.* (2020) 5:256–63. doi: 10.1016/j.idm.2020.02.002
30. Roosa K, Lee Y, Luo R, Kirpich A, Rothenberg R, Hyman JM, et al. Short-term forecasts of the COVID-19 epidemic in Guangdong and Zhejiang, China: February 13–23 2020. *J Clin Med.* (2020) 9:596. doi: 10.3390/jcm9020596
31. Coelho FC, Lana RM, Cruz OG, Villela D, Bastos LS, Piontti A, et al. Assessing the potential impact of COVID-19 in Brazil: mobility, morbidity and social vulnerability. *PLoS ONE.* (2020) 15:e0238214. doi: 10.1371/journal.pone.0238214
32. Costa JA Jr, Martinez AC, Geromel JC. On an alternative susceptible-infected-removed epidemic model in discrete-time. *Anais da Sociedade Brasileira de Automática.* (2020) 2. doi: 10.48011/asba.v2i1.995
33. Li L, Yang Z, Dang Z, Meng C, Huang J, Meng H, et al. Propagation analysis and prediction of the COVID-19. *Infect Dis Modell.* (2020) 5:282–92. doi: 10.1016/j.idm.2020.03.002
34. Jewell NP, Lewnard JA, Jewell BL. Caution warranted: using the institute for health metrics and evaluation model for predicting the course of the COVID-19 pandemic. *Ann Internal Med.* (2020). 173:226–7. doi: 10.7326/M20-1565
35. Chang SL, Harding N, Zachreson C, Cliff OM, Prokopenko M. Modelling transmission and control of the COVID-19 pandemic in Australia. *arXiv preprint arXiv:200310218.* (2020). doi: 10.1038/s41467-020-19393-6
36. Alzahrani SI, Aljamaan IA, Al-Fakih EA. Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using ARIMA prediction model under current public health interventions. *J Infect Public Health.* (2020). 13:914–9. doi: 10.1016/j.jiph.2020.06.001
37. Hawkins DM. The problem of overfitting. *J Chem Inform Comput Sci.* (2004) 44:1–12. doi: 10.1021/ci0342472
38. Lei CL, Ghosh S, Whittaker DG, Aboelkassem Y, Beattie KA, Cantwell CD, et al. Considering discrepancy when calibrating a mechanistic electrophysiology model. *Philos Trans R Soc A Math Phys Eng Sci.* (2020) 378:20190349. doi: 10.1098/rsta.2019.0349
39. Maier BF, Brockmann D. Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. *Science.* (2020) 368:742–6. doi: 10.1126/science.ab4557
40. Zhao S, Stone L, Gao D, Musa SS, Chong MKC, He D, et al. Imitation dynamics in the mitigation of the novel coronavirus disease (COVID-19) outbreak in Wuhan, China from 2019 to 2020. *Ann Transl Med.* (2020) 8. doi: 10.21037/atm.2020.03.168
41. Brown RG. *Smoothing, Forecasting and Prediction of Discrete Time Series.* New York, NY: Courier Corporation (2004).
42. Kermack WO, McKendrick AG, Walker GT. A contribution to the mathematical theory of epidemics. *Proc R Soc Lond Ser A.* (1927) 115:700–21. doi: 10.1098/rspa.1927.0118
43. Kermack WO, McKendrick AG. Contributions to the mathematical theory of epidemics–I. *Bull Math Biol.* (1991) 53:33–55. doi: 10.1016/S0092-8240(05)80040-0
44. Keeling MJ, Rohani P. *Modeling Infectious Diseases in Humans and Animals.* Princeton, NJ: Princeton University Press (2011). doi: 10.2307/j.ctvcwm4gk0
45. Storn R, Price K. Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. *J Global Optimizat.* (1997) 11:341–59. doi: 10.1023/A:1008202821328
46. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis.* (2020). 20:533–4. doi: 10.1016/S1473-3099(20)30120-1
47. Feinberg J, Langtangen HP. Chaospy: an open source tool for designing methods of uncertainty quantification. *J Comput Sci.* (2015) 11:46–57. doi: 10.1016/j.jocs.2015.08.008
48. Papadopoulos CE, Yeung H. Uncertainty estimation and Monte Carlo simulation method. *Flow Meas Instrument.* (2001) 12:291–8. doi: 10.1016/S0955-5986(01)00015-2
49. Gilks WR. Markov Chain Monte Carlo. *Encyclopedia Biostat.* (2005) 4. doi: 10.1002/0470011815.b2a14021

50. Coleman HW, Steele WG. Engineering application of experimental uncertainty analysis. *AIAA J.* (1995) 33:1888–96. doi: 10.2514/3.12742

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2021 Reis, Oliveira, Quintela, Campos, Gomes, Rocha, Lobosco and dos Santos. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*