RESEARCH ARTICLE - APPLICATION

WILEY

A mathematical dashboard for the analysis of Italian COVID-19 epidemic data

Nicola Parolini[1] | **Giovanni Ardenghi[1]** | **Luca Dede'[1]** | **Alfio Quarteroni[1,2]**

[1]MOX, Dipartimento di Matematica, Politecnico di Milano, Milan, Italy

[2]Mathematics Institute, EPFL (Professor Emeritus), Lausanne, Switzerland

Correspondence
Luca Dede', MOX, Dipartimento di Matematica, Politecnico di Milano, Milan, Italy.
Email: luca.dede@polimi.it

Abstract

An analysis of the COVID-19 epidemic is proposed on the basis of the *epiMOX* dashboard (publicly accessible at https://www.epimox.polimi.it) that deals with data of the epidemic trends and outbreaks in Italy from late February 2020. Our analysis provides an immediate appreciation of the past epidemic development, together with its current trends by fostering a deeper interpretation of available data through several critical epidemic indicators. In addition, we complement the *epiMOX* dashboard with a predictive tool based on an epidemiological compartmental model, named SUIHTER, for the forecast on the near future epidemic evolution.

1 | INTRODUCTION

For the COVID-19 pandemic, the development of reliable mathematical models, supported by the availability and analysis of complete and accurate data, is a fundamental tool for the interpretation and understanding of the epidemic, as well as for providing support to digital health.[1] The availability of accurate and complete historical data together with a suitable epidemiological model for short-term forecasting are essential to provide institutional bodies and authorities with factual quantitative information prior to the adoption of non-pharmaceutical interventions (NPIs).

The first goal of this paper is to present a newly developed mathematical dashboard named *epiMOX* (accessible online at the address https://www.epimox.polimi.it), which gathers the time histories of recorded epidemiological compartments and their first and second rates of variation. We believe that this analysis, which focuses on the situation in Italy at both national and regional scales—for 19 Italian Regions and the two autonomous provinces of Trento and Bolzano—will enhance the interpretation and transparency of available data thanks to a thorough understanding of the past epidemic development, and of its current trends. As several other countries, Italy has been hit since February 2020 by a sequence of epidemic waves: the first one initiated on February 21, 2020 and exhausted in early June 2020, while the second wave raised in early October 2020 and peaked in late November 2020, while the outbreak of a third wave is being observed at the time of writing (March 7, 2021).

The second goal of the paper is to enable a comparison between the first and the second epidemic waves in Italy that feature indeed several analogies, as well as significant differences that we aim to highlight and analyze.

Finally, a preview on the expected trend of the epidemic for the near future—late March 2021—is discussed, based on a compartmental epidemiological model, named SUIHTER, which was purposely designed for the Italian COVID-19 epidemic.[2]

Our analysis highlights several important features of the Italian COVID-19 epidemic. The most relevant are the following:

1. In the early phase of the exponential outbreak, the timeline for the implementation of NPIs is crucial. It is indeed far more efficient to take more stringent restrictions for a short time span in the early phase of the outbreak rather than implementing less severe (or even the same) restrictions for a longer interval later.

2. The second epidemic wave showed a slower pace, but a widespread diffusion in the Country that eventually yielded far worse figures (in terms of fatalities, patients hosted in intensive care units, etc.) than the first wave.

3. The epidemiological mathematical model SUIHTER, which is here proposed in an improved version accounting for the effect of the UK variant, allows the investigation of various scenarios that conform to the different restrictive NPIs devised by the Italian government. We identify those that are potentially more effective to contrast the near future epidemic development.

4. A retrospective (a posteriori) analysis allows the validation of the SUIHTER epidemiological model on different phases of the epidemic.

5. Thanks to the SUIHTER model, we can carry out a *what-if* analysis aimed at simulating different epidemic trends that would have occurred in case different NPIs had been implemented by the Italian authorities at the outbreak of the second wave.

In this paper, epidemiological data available up to March 7, 2021 are used to validate the suitability of the *epiMOX* dashboard to provide fast and in-depth analyses of the past trends of the Italian outbreak and predict its evolution through the SUIHTER epidemiological model. Needless to say, the *epiMOX* dashboard will continue to monitor the evolution of the epidemic providing updated predictions based on data made available on a daily basis.

An outline of this paper is as follows: in section 2, a description of the COVID-19 epidemic data time series is supplied together with a description of the data processing tools (filtering, derivative, scaling, normalization) used to straightforwardly identify the main characteristics of each time series. The first and second epidemic waves are analyzed and the results are compared at both the national (section 2.1) and regional (section 2.2) scales. Some critical indicators for the epidemic, that can be displayed by the dashboard, are analyzed in section 3. In section 4, we first recall the SUIHTER model and analyze its interplay with the *epiMOX* dashboard. In particular, through the dashboard, we validate the model in section 4.1 against the data of the second wave at the national level. Then, we propose what-if-scenarios for the analysis of alternative NPIs in section 4.2, we analyze the impact of the UK variant in section 4.3, and we compare different short term future scenarios in section 4.4. Conclusions follow in section 5.

2 | DATA ACQUISITION, PROCESSING, AND ANALYSIS

The data used in this paper are those made available by the Italian *Dipartimento della Protezione Civile* (DPC) through the open data repository https://github.com/pcm-dpc/COVID-19. Data are communicated on a daily basis and include the number of individuals who are currently *positive*, *isolated at home*, *hospitalized*, *hosted in ICUs* (Intensive Care Units), the *daily positive* cases, the cumulative number of *deaths* and *recovered* since the beginning of the pandemic, and the number of *swabs* performed. All these data are supplied at both the regional and national levels, while the available data on a finer scale (provinces) are unfortunately limited to the count of *positive* cases since the beginning of the epidemic.

From now on, we will refer to these data as *raw* data. These raw data can then be smoothed by resorting to a local polynomial regression based on the Savitzky–Golay convolution filter[3]: at day n we attribute the value of the polynomial of degree r that approximates, in the least squares sense, the $2q + 1$ values of the raw data centered on day n, that is, in the range $[n - q, n + q]$, where $r \leq 2q$. Here, we consider a cubic least squares polynomial and a window size of 21 days (i.e., $r = 3$ and $q = 10$). A standard approach based on a *weekly moving average* could be obtained by taking $r = 0$ and $q = 3$.

From now on, if not otherwise specified, the time series that will be presented will be smooth curves obtained using the Savitzky–Golay convolution filter described above.

The second step is to calculate the first and second rates of variation of the different compartments. First rate of variation describes how fast a trend is either increasing or decreasing: change of sign from positive to negative for the first rate of variation indicates switching from increasing to decreasing in the corresponding curves, whereas a change of sign on the second rate of variations denotes a change of convexity (a point of inflection).

2.1 | The epidemic at the national scale

The time series of the filtered data (solid lines) for some relevant epidemiological compartments at the national Italian level, namely the *daily positive* cases, the *daily deceased*, the number of individuals that are *hospitalized* with symptoms and those *hosted in ICUs*, are reported in Figure 1. On top of each figure, we display a color bar. The color code used therein is corresponding to the severity of NPIs, as stated in DPCM (the decree issued by the Italian Government) on November 4, 2020.*

The color codes apply at regional level and are defined as follows, according to an algorithm based on different risk indicators:

- *yellow*: moderate risk zones with basic NPIs in place (mandatory use of masks, curfew from 10 p.m. to 5 a.m., limitations on the activity of shops, bars and restaurants, strong limitations of sport and leisure activities, [partial] distance learning for secondary schools);
- *orange*: elevate risk zones with stricter NPIs (municipality confinement, food service activities suspended);
- *red*: maximum risk zones with the even stricter NPIs (home confinement, distance learning from grade 8, non-essential commercial activities suspended, all sport and leisure activities suspended).

Since November 6, 2020, the color of each region has been updated weekly. At the national level (e.g., in Figure 1), the color code is obtained as a suitable averaging of those in place in the regions, weighted by their population. The strict lockdown imposed at the national level in Spring 2020 is conventionally identified with the black color.

By analyzing the time series of the *positive* cases using a log scale (see Figure 2) along with the full evolution of the epidemic in Italy, we can observe, for the first epidemic wave, an exponential growth $f(t) = Ce^{\lambda_1 t}$ (which is linear in log scale with slope λ_1) for the first 2 weeks of March with a doubling time of approximately 3 days.

During the second half of August, very likely because of the relaxation of social distances during the summer holidays and the newly imported cases from abroad, a few days of exponential growth can be observed, although featuring a much lower growth factor λ_2, yielding a doubling time of approximately 15 days. Finally, what we will refer to as the second wave had its exponential growth last October, with a growth factor λ_3 and a doubling time of about 8 days. This was very likely connected to the increased number of contacts associated with the school opening in September and related commuting, the restart of recreational activities in closed ambiences, seasonality, and the full recovery of the working activities that were dramatically reduced during the spring lockdown, as well as (although at a minor extent) during the summer period. Finally, we observe an exponential growth starting by the end of February 2021 and due to new virus variants featuring a higher transmission rate (see the discussion in section 4.3). Although, in this case, the growth rate λ_4 appears to be lower than previous ones with a corresponding doubling time of approximately 36 days,
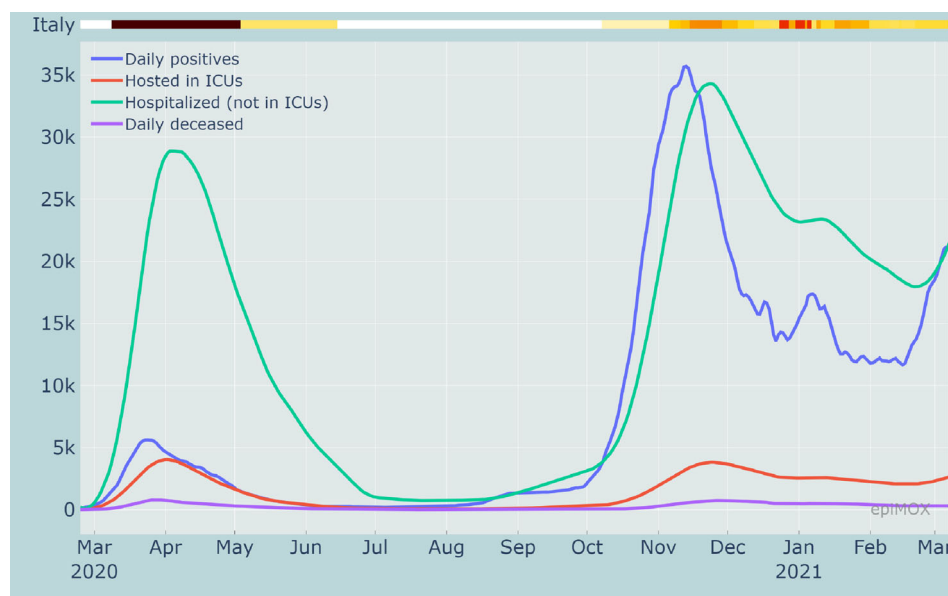

FIGURE 1 Time series of different compartments during the whole epidemic history in Italy

FIGURE 2 Identification of four exponential growth phases in the time series of the *positive* cases in logarithmic scale
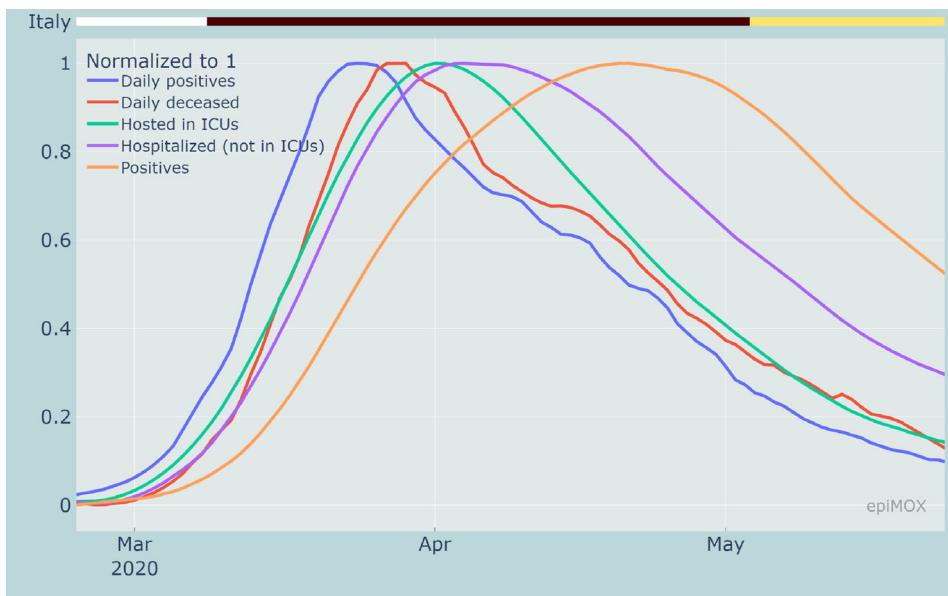

FIGURE 3 Normalized time series of different compartments highlighting the time-shifts between the different peaks during the first epidemic wave

the fact that this new outbreak is starting from a much larger number (more than half a million) of positive individuals makes the situation at present (early March 2021) extremely critical.

2.1.1 | The first epidemic wave

We first focus on the first epidemic wave that occurred in Italy last Spring 2020. In our dashboard, a specific time range can be selected and each time series can be normalized with respect to its maximum attained in the prescribed time
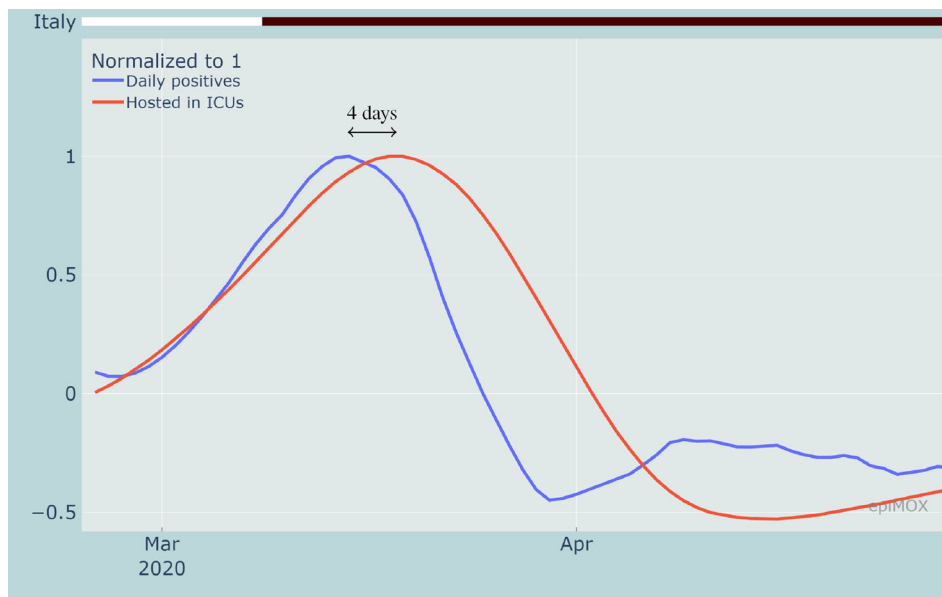
FIGURE 4 Rate of change of *daily positive* cases and *hosted in ICUs* highlighting the time shifts between the inflection points during the first epidemic wave

range. This allows for an immediate identification of the day the different peaks have occurred and of their relative positions (relative delays). As displayed in Figure 3, it can be noticed that the first compartment that reaches a peak is that of the number of *daily positive* cases (on March 24, 2020), followed 5 days later on March 29, 2020 by the peak of *daily deceased*. The peaks for the number of individuals *hosted in ICUs* and those *hospitalized* occur on April 1 and 3, respectively, that is 7 and 9 days past the peak of *daily positive* cases.

Even if, at a first glance, the fact that the peak of *daily deceased* occurs before those of the hospitalized and *hosted in ICUs* may appear surprising, this is instead reasonable since the latter data do not refer to new daily entries but to the total number of individuals who are hosted in hospitals or ICUs at a specific date. Unfortunately, information (raw data) on the daily admissions in hospitals and ICUs are not available, the latter being supplied starting only on December 3, 2020. Finally, the peak of *positive* cases was reached on April 21, 2020, that is almost 1 month past the peak of *daily positive* cases.

A similar analysis carried out on the rate of change (first derivative) of the different indicators may be used to determine when the initial exponential phase for each indicator is over. In Figure 4, the normalized first derivatives of the same indicators discussed above are displayed. In particular, we notice that the growth rate of the *daily positive* cases reached its maximum on March 15, 2020, while the growth rate of the number of *hosted in ICUs* is attained just 4 days later (March 19, 2020). This corresponds to an inflection point in the time series and indicates the time at which the growth rate starts decreasing.

2.1.2 | The second epidemic wave

The second wave of the COVID-19 epidemic in Italy started on the first half of October 2020. The peak of *daily positive* cases was reached on November 14, 2020 (see Figure 5).

Other relevant compartments reached their peaks with a time shift that is consistent with the epidemic evolution. In particular, the maximum number of *hospitalized*, *hosted in ICUs* and *daily deceased* were reached on November 23, 24 and 29, respectively (see the normalized time series in Figure 6).

As already observed for the first wave, monitoring the growing rate (first derivative) of the time series can provide a preliminary indication on when the peaks should be expected. For instance, as displayed in Figure 7, the peak on the growth rate of the *daily positive* (on October 24, 2020) anticipates the peak of the *hosted in ICUs* by 12 days (November 5, 2020).
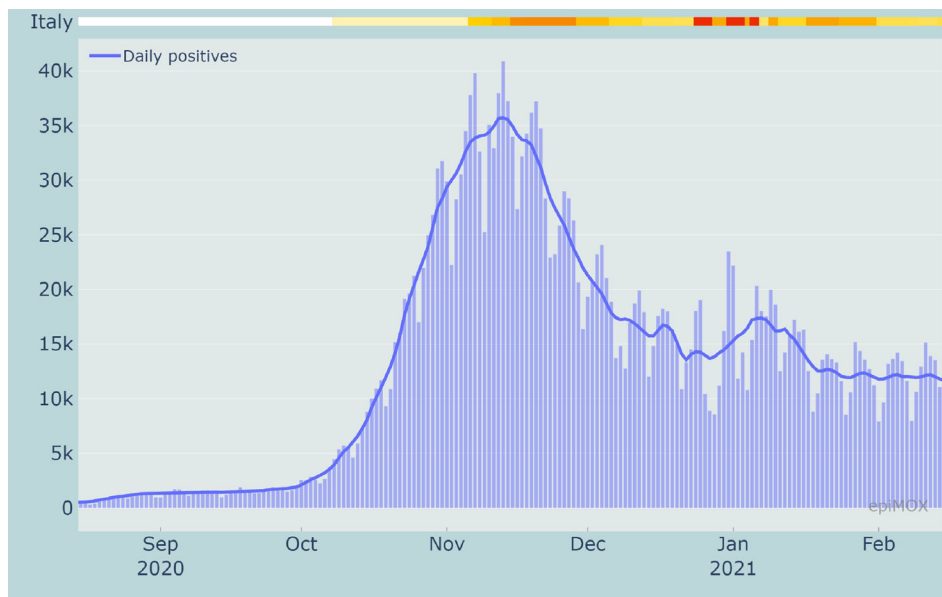
FIGURE 5 Raw (bars) and smoothed (solid line) time series of daily new cases during the second wave
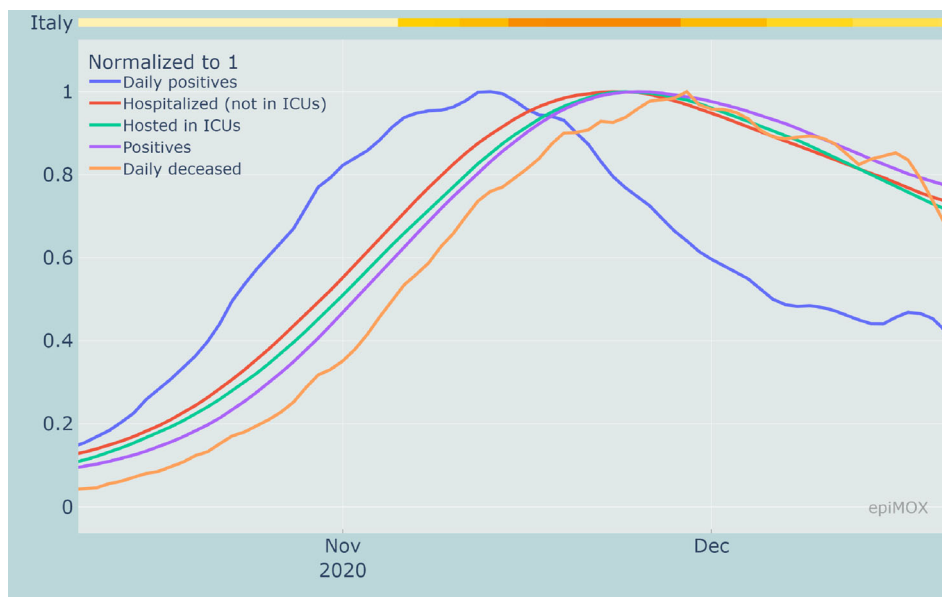

FIGURE 6 Normalized time series of different compartments highlighting the time-shifts between the different peaks during the second epidemic wave

For the second wave, a consistent time lag between the peaks of the first derivative and the peak of the corresponding time series can be noticed. As shown in Table 1 for most of the compartments (*Positives*, *Daily positives Hospitalized* and *Hosted in ICUs*) there is a 20-days delay between inflection point and the maximum, while this delay is slightly longer (24 days) for the *Daily deceased* time series. A consistent time lag between inflection point and maximum point (the one corresponding to the peak) can be used as a preliminary indicator in future epidemic waves to predict the day at which the peak will possibly occur.

A similar correlation between the peak of the derivative and that of the corresponding time series is missing during the first wave. This is likely due to the severe under-estimation of data collected for some epidemiological compartments and then provided by the DPC.

FIGURE 7 Rate of change of *daily positive* cases and *hosted in ICUs* highlighting the time shifts between the inflection points during the second epidemic wave

TABLE 1 Peak date of the time series of several epidemiological compartments and that of the corresponding first derivative relative to the second epidemic wave in Italy

	Peak date				
Trend	**Positives**	**Daily positives**	**Hospitalized**	**ICUs**	**Daily deceased**
Time series	November 25, 2020	November 13, 2020	November 23, 2020	November 24, 2020	November 29, 2020
First derivative	November 05, 2020	October 24, 2020	November 04, 2020	November 05, 2020	November 05, 2020

2.2 | The epidemic at the regional scale

The same analyses that have been proposed at the national Italian level can be carried out at the regional scale, that is for each of the 19 Italian Regions and the two autonomous provinces of Trento and Bolzano. In Figure 8, the normalized time series of *daily positive* cases, *hospitalized*, *hosted in ICUs* and *daily deceased* are presented for 8 Italian regions, among which Lombardia, Emilia Romagna, Piemonte, Veneto that were severely hit by the first epidemic wave, and Lazio, Puglia, Campania, Sicilia that are far more evidently affected by the latter epidemic wave than by the former.

In Figure 9, we display the evolution of the pandemic in the eight largest Italian regions since the early stages of the second wave (September 2020). We compare two compartments: *positives*, and *hosted in ICUs*, in three regions in northern Italy (Lombardia, Veneto, Emilia-Romagna) that were severely affected by the first wave. The same comparison is carried out in Figure 10 concerning three southern Italian regions (Campania, Sicilia and Sardegna, the latter two being insular) that were very mildly affected by the first wave last Spring 2020. We recall that, in all figures, the upper color bars indicate the non-uniform NPIs. If we focus on the number of *Positives* per 100,000 inhabitants (Figures 9A and 10A) that were diagnosed as positive to the virus, we notice that the red regime yields a pronounced decrease of the curves. Both the orange and yellow regimes serve to mitigate the outbreak, without however significantly damp it. This behavior is less evident on the *hosted in ICUs* compartment (Figures 9B and 10B). A possible explanation is that this latter compartment is slightly more "indifferent" to the stringency of the NPIs that are being implemented. The *hosted in ICUs* individuals are at a very large extent elder people who were less directly exposed to outdoor contacts as either they were residents of healthcare facilities or they spent a substantial part of their time at home.
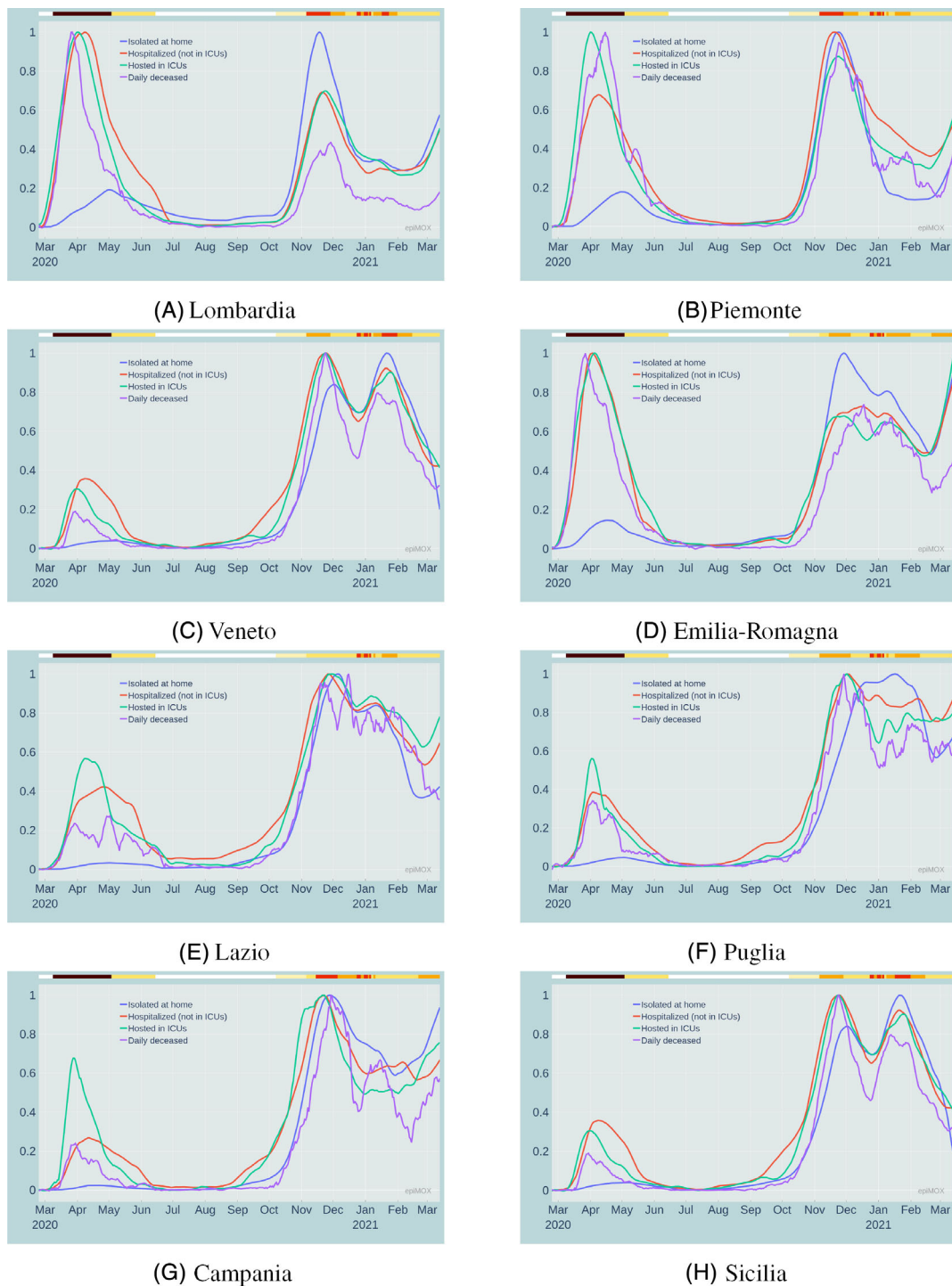
FIGURE 8 Normalized time series of different compartments during the whole epidemic history in 8 Italian regions

3 | ANALYSIS OF SOME CRITICAL INDICATORS

The *epiMOX* dashboard can display several critical indicators for monitoring the progress of the pandemic, namely:

- The daily evolution of R*, an estimate of the reproduction number Rt proposed in Reference 4.
- The daily evolution of the fatality rate, estimated through the CFR (Case Fatality Ratio) as the ratio between the number of cumulative deaths at a given date and the number of resolved cases at the same date
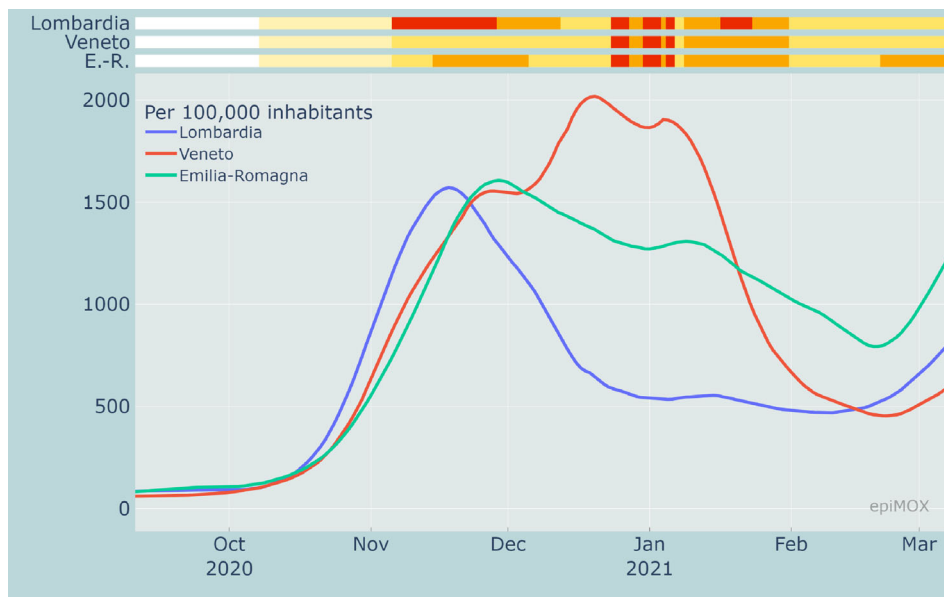
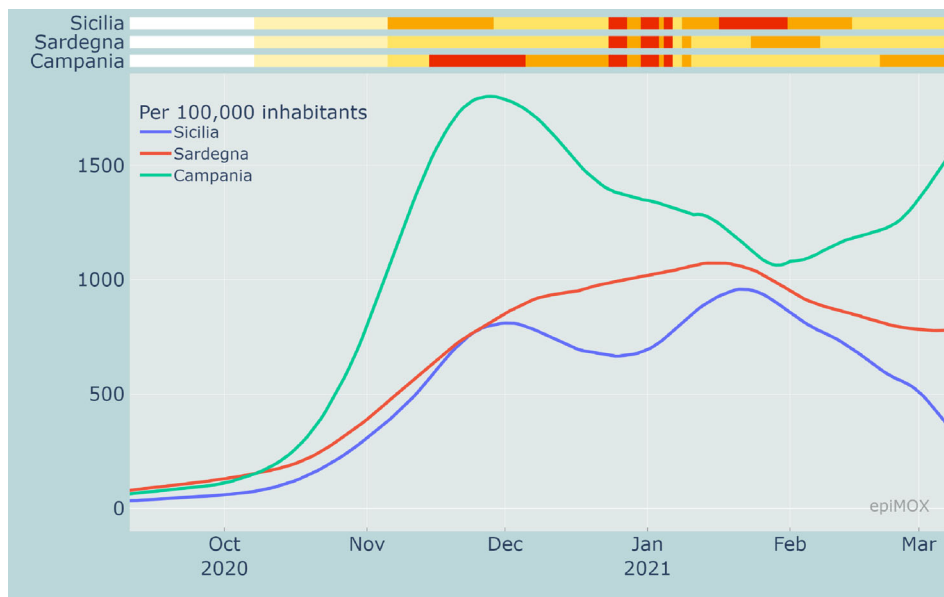**(A)** Positives
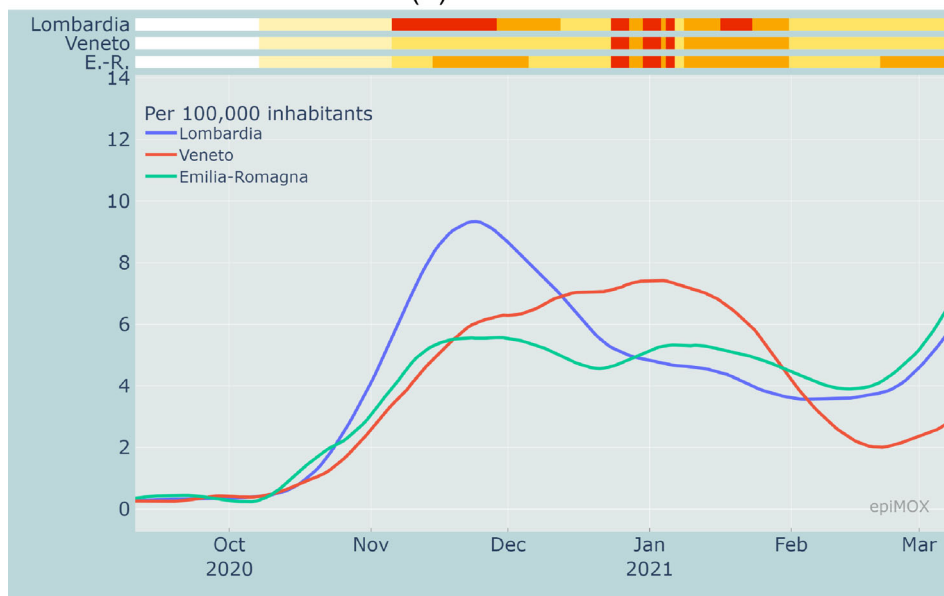
(B) ICUs

FIGURE 9 Comparison of *positive* cases and *hosted in ICUs* per 100,000 inhabitants in Emilia Romagna, Lombardia and Veneto during the second wave in relation with the measures adopted by each region

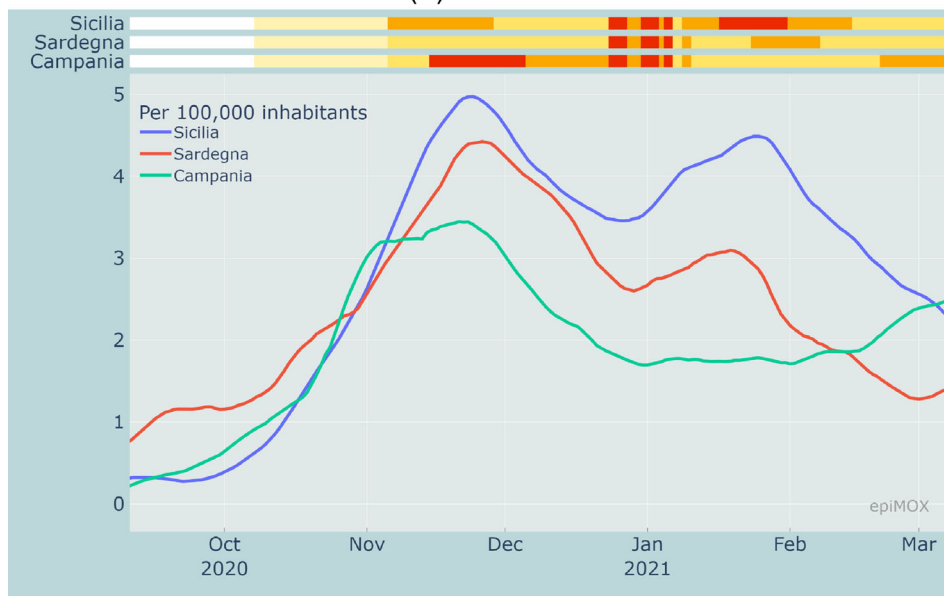$$\text{CFR} = \frac{\text{Deceased}}{\text{Healed} + \text{Deceased}}.$$

- The daily evolution of the positivity rate (PR) defined as

$$\text{PR} = \frac{\text{Daily positives}}{\text{Daily swabs performed}}.$$

- The daily evolution of the hospitalized case ratio (HCR), defined as

(A) Positives

(B) ICUs

FIGURE 10 Comparison of *positive* cases and *hosted in ICUs* per 100,000 inhabitants in Campania, Sardegna and Sicilia during the second wave in relation with the measures adopted by each region

$$HCR = \frac{\text{Total hospitalized}}{\text{Positives}} = \frac{\text{Hospitalized} + \text{Hosted in ICUs}}{\text{Positives}}.$$

- The daily evolution of the 14-day notification rate, that is, the number of new positive cases emerged in the previous 14 days per 100,000 population, an indicator used by the European Union to monitor the evolution of the epidemic in the different European regions.

From Figures 11, 12 and 13, we can notice that CFR, PR and HCR were far much higher in the first wave than in the second one because of a significant underestimation of the value of the denominators. The *PR* indicator dramatically dropped around mid-January 2021 because of the different counting strategy for cases that was adopted by authorities. Indeed, since then the denominator accounts for both molecular and genetic swabs. Figure 14 displays the value
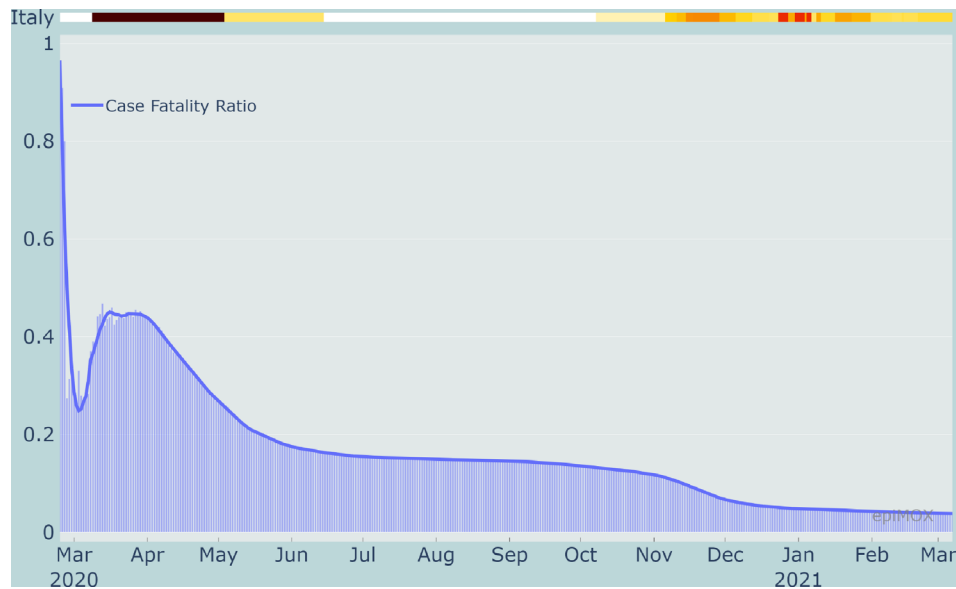
FIGURE 11 Case Fatality Ratio (CFR) in Italy since the beginning of the epidemic
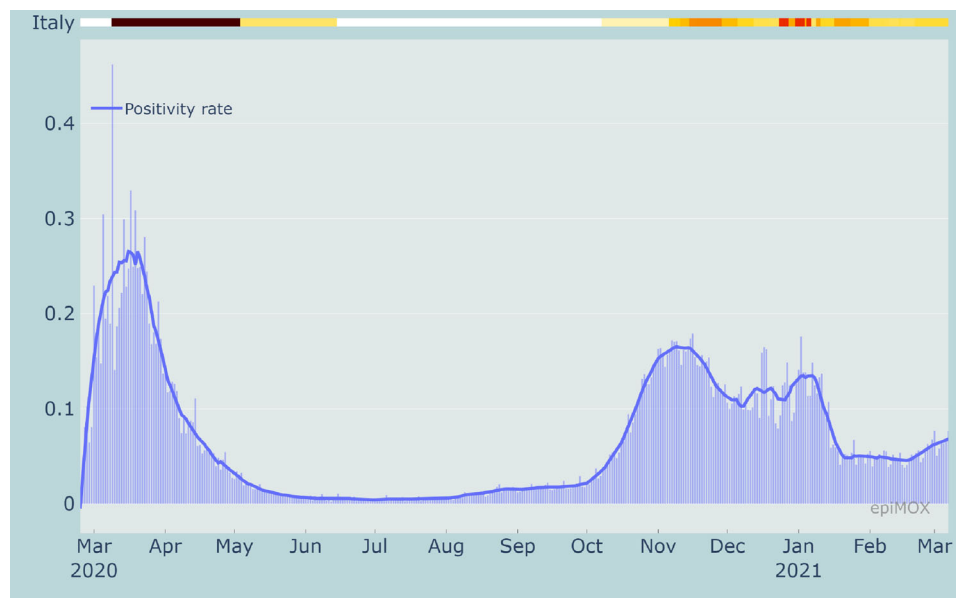

FIGURE 12 Positivity rate (PR) in Italy since the beginning of the epidemic

of the (estimated) reproduction number R^* in three different Italian Regions. We can notice that a significant increase of R^* is reported over the summer period in two touristic regions like the Sardegna island and the Trento province in the Dolomites due to a substantial increase of the population and a very likely relaxation of the restrictions (limited use of facial masks and lacking of the social distancing).

4 | EMPOWERING *epiMOX* WITH THE SUIHTER EPIDEMIOLOGICAL MODEL

Epidemic forecasting of COVID-19 is difficult because of the intrinsic variability and uncertainty of the pandemic. Incomplete, uncertain, or inaccurate data, in terms of both initial conditions and time series of the different compartments,
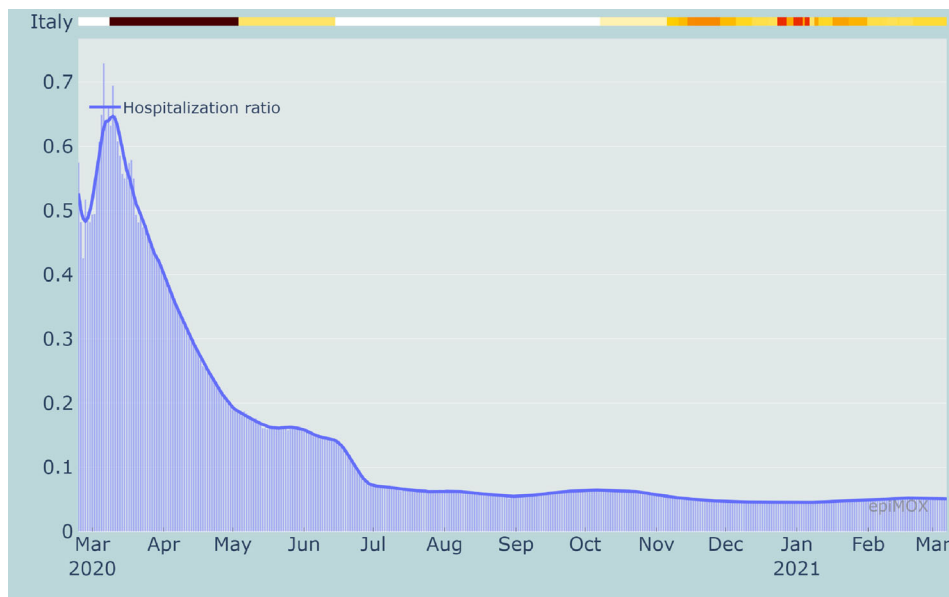
FIGURE 13   Hospitalization Case Ratio (HCR) in Italy since the beginning of the epidemic
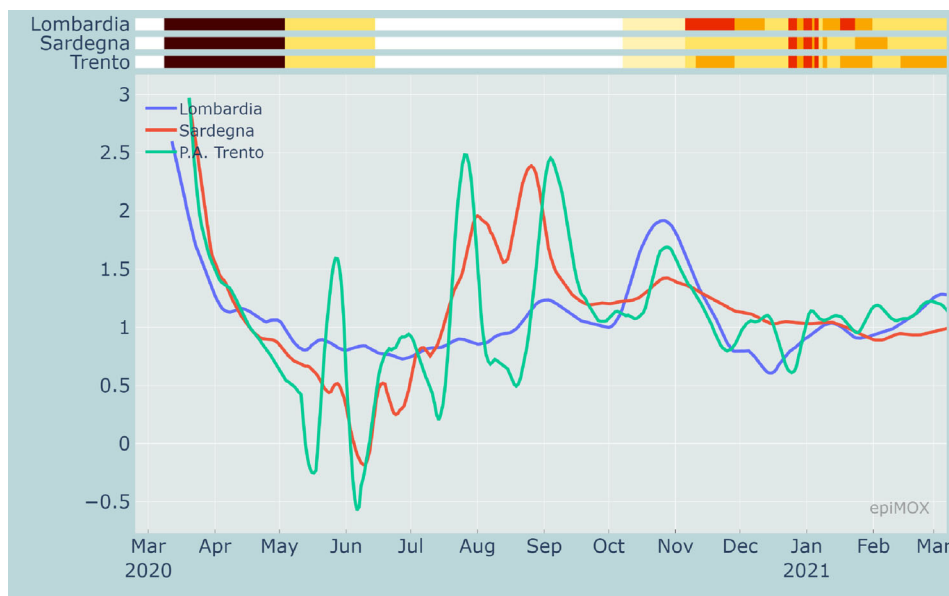

FIGURE 14   Evolution of R^* in three different regions since the beginning of the epidemic

represent a serious limitation. The partial knowledge of the behavior of the specific infecting agent and the dynamic evolution of environmental and social conditions is a further source of uncertainty.

Since the seminal work in Reference 5, where the first SIR compartmental model based on a system of nonlinear ordinary differential equations was presented, a large variety of models have been proposed—see, for instance,[6–11]—each one attempting to cope with specific aspects of the problem. If in the original SIR model the epidemic evolution is described by the number of individuals belonging to the susceptible (S), infected (I) and removed (R) compartments, several models consider increasing the number of compartments to include, for instance:

- the exposed individuals (those who have already been exposed to the infecting agent but are not yet infectious);
- possible splitting of the infected compartment into different classes according to the actual level of severity;
- a distinction of the removed compartment in recovered and dead.

These models, although typically derived under many simplifying assumptions, enable forecasting analyses that go beyond mere data extrapolation. Moreover, simulations regarding the future dynamics may allow the investigation of different scenarios corresponding to different NPIs. One of the most critical aspects in the development of complex compartmental models is indeed related to their *calibration* based on available data. On the one hand, data related to the different compartments may not be available (or they may not even be collected); on the other hand, the resulting data assimilation problem may suffer from limited identifiability of the parameters, as discussed, for example, in the paper[12] for SIR-like models of COVID-19. A new compartmental model named SUIHTER has been recently introduced in Reference 2 with the goal of facing the first of these two issues, that is, defining a model best suited for the data actually available. In the context of the COVID-19 epidemic, the time series that are daily collected and made available (which have already been described discussing the dashboard for the data analysis) led us to consider for the new model the following compartments:

- *S*: *susceptible* (uninfected) individuals;
- *U*: *undetected* (both asymptomatic and symptomatic) infected individuals;
- *I*: *isolated* (quarantined) individuals;
- *H*: *hospitalized* individuals;
- *T*: *threatened* (acutely symptomatic infected, hosted in ICUs) individuals;
- *E*: *extinct* individuals;
- *R*: *recovered* individuals,

In this paper, we use the following specific version of the SUIHTER model:

$$
\begin{aligned}
\dot{S}(t) &= -S(t)\frac{\beta_U U(t) + \beta_I I(t)}{N}, \\
\dot{U}(t) &= S(t)\frac{\beta_U U(t) + \beta_I I(t)}{N} - (\delta + \rho_U)U(t), \\
\dot{I}(t) &= \delta U(t) - (\rho_I + \omega_I + \gamma_I)I(t), \\
\dot{H}(t) &= \omega_I I(t) - (\rho_H + \omega_H)H(t), \\
\dot{T}(t) &= \omega_H H(t) - (\rho_T + \gamma_T)T(t), \\
\dot{E}(t) &= \gamma_I I(t) + \gamma_T T(t), \\
\dot{R}(t) &= \rho_U U(t) + \rho_I I(t) + \rho_H H(t) + \rho_T T(t),
\end{aligned}
\tag{1}
$$

where $N = S + U + I + H + T + E + R$ denotes the total population (assumed constant). The model is characterized by the following 11 parameters, which are chosen as time dependent piecewise constant functions:

- β_U, β_I denote the transmission rates due to contacts between a susceptible subject and an undetected infected, a quarantined, or a hospitalized subject, respectively;
- ω_I denotes the rate at which *I*-individuals develop clinically relevant symptoms, while ω_H denotes the rate at which *H*-individuals develop life-threatening symptoms;
- δ denotes the probability rate of detection, relative to undetected infected individuals;
- ρ_U, ρ_I, ρ_H and ρ_T denote the rate of recovery for four classes (*U*, *I*, *H* and *T*, respectively) of infected subjects;
- γ_I and γ_T denote the mortality rates for the individuals isolated at home, and hosted in ICUs, respectively.

Model calibration through data fitting is essential to reproduce the past history of the epidemic and to perform short-term forecasts by inferring the epidemiological characteristics of COVID-19. Here we use reported isolated, hospitalized, threatened and extinct cases data to estimate the parameters of the SUIHTER model. In particular, we perform the calibration in two steps. At first, we find a set of parameter values by using an (ordinary) least squares (LS) estimator. Then, we perform a Bayesian calibration using the delayed rejection adaptive Metropolis (DRAM) algorithm[13] implemented in the Python library `pymcmcstat`,[14] starting from a prior distribution of the parameters centered about the LS estimate.

System (1) is solved in the time interval $[t_I, t_F]$ which is subdivided into several sub-intervals, named *phases*. Typically, phases change when the NPIs are modified by the Italian Authorities. When n_{ph} phases are considered, the

calibration leads to the optimization of $n_p = 11n_{ph}$ parameters in total. Namely, for each phase of the epidemic, we have the 11 parameters given by $[\beta_U, \beta_I, \omega_I, \omega_H, \delta, \rho_U, \rho_I, \rho_H, \rho_T, \gamma_I, \gamma_T]$.

Unfortunately, so many parameters make the calibration process difficult. In what follows, we calibrate our model under the following simplifying assumptions:

- β_I is set to be a fraction of β_U, namely $\beta_I = \alpha\beta_U$ with α constant in all the time frame;
- $\delta, \rho_U, \rho_I, \rho_H, \rho_T, \gamma_I$, are constant on $[t_I, t_F]$.

With these restrictions, the total number of parameters to be calibrated is reduced to $4n_{ph} + 7$.

More details on the initialization and numerical solution of system (1) are reported in Reference 2.

4.1 | Results' assessment

Scope of this section is to provide an assessment of the forecasting capability of the SUIHTER model. More specifically, we will analyze suitability of the epidemiological model to predict the date and the height of the peaks of the curves representing the hospitalized and hosted in ICUs compartments, focusing on the second wave at the national scale. For each compartment, by proceeding retrospectively, in Figure 15 we display the predictions made by the SUIHTER model at different dates prior to the date at which the corresponding curve has attained its peak value. By comparing the predictions with the real data (see Table 2) we can draw the conclusion that the SUIHTER model can provide accurate predictions within a week (with errors lower than 1% on the peak value and 1 day on the peak date), and reasonably good predictions 2 weeks in advance.

4.2 | What-if scenarios

In this section, we aim at highlighting another feature of the dashboard, the possibility of carrying out *what-if* scenarios. An example is illustrated in Figure 16. Here we simulate (by means of the SUIHTER model) what would have happened in the case the Italian Government had anticipated by 10 days the restriction measures introduced on November 4, 2020.† The difference is indeed quite striking: the peak value for the *Isolated* during the second epidemic wave would have been 561,714 instead of 759,993, for the *Hospitalized* 25,169 instead of 34,697, for the *Hosted in ICUs* 2793 instead of 3823. More importantly, according to our estimates, a remarkable number of human lives would have likely been saved. Obviously, once again we point out that these conclusions should be taken with a grain of salt, because of the accuracy limitations of the epidemiological model. These impressive figures prove that the most effective strategy in the early phase of an outbreak is to enforce very strict confinement measures as early as possible. This strategy, even if maintained for a short period, is indeed far more effective than adopting more permissive measures (or even the very same measures) for a much longer duration at a *later* stage.

4.3 | Impact of the UK variant

The appearance of different virus variants (UK, Brazilian, South African) at the end of 2020 has introduced an additional level of complexity on the managing of the epidemic as well as on the effort of the scientific community in modeling its dynamics (see e.g., Reference 15). A recent study in Reference 16 indicates that "SARS-CoV-2 is at an advanced stage of evolution for human infection." Other than questioning the vulnerability of available vaccines and current therapeutics to virus variants, the mutations can lead to different characterizations of the epidemic, in particular of the transmission rates (see Reference 17). In this respect, in the present paper, we limit our study to the effect of virus mutations on the transmission rates that can be readily incorporated in existing compartmental epidemiological models.

In particular, in this paper, we adapt the SUIHTER model to this new challenge by splitting each of the compartments that spread the infection (in particular the *Undetected* and the *Isolated* compartments) into two sub-compartments, one carrying the baseline virus variant and the other the UK virus variant (VOC 202012/01, lineage B.1.1.7). Recent studies
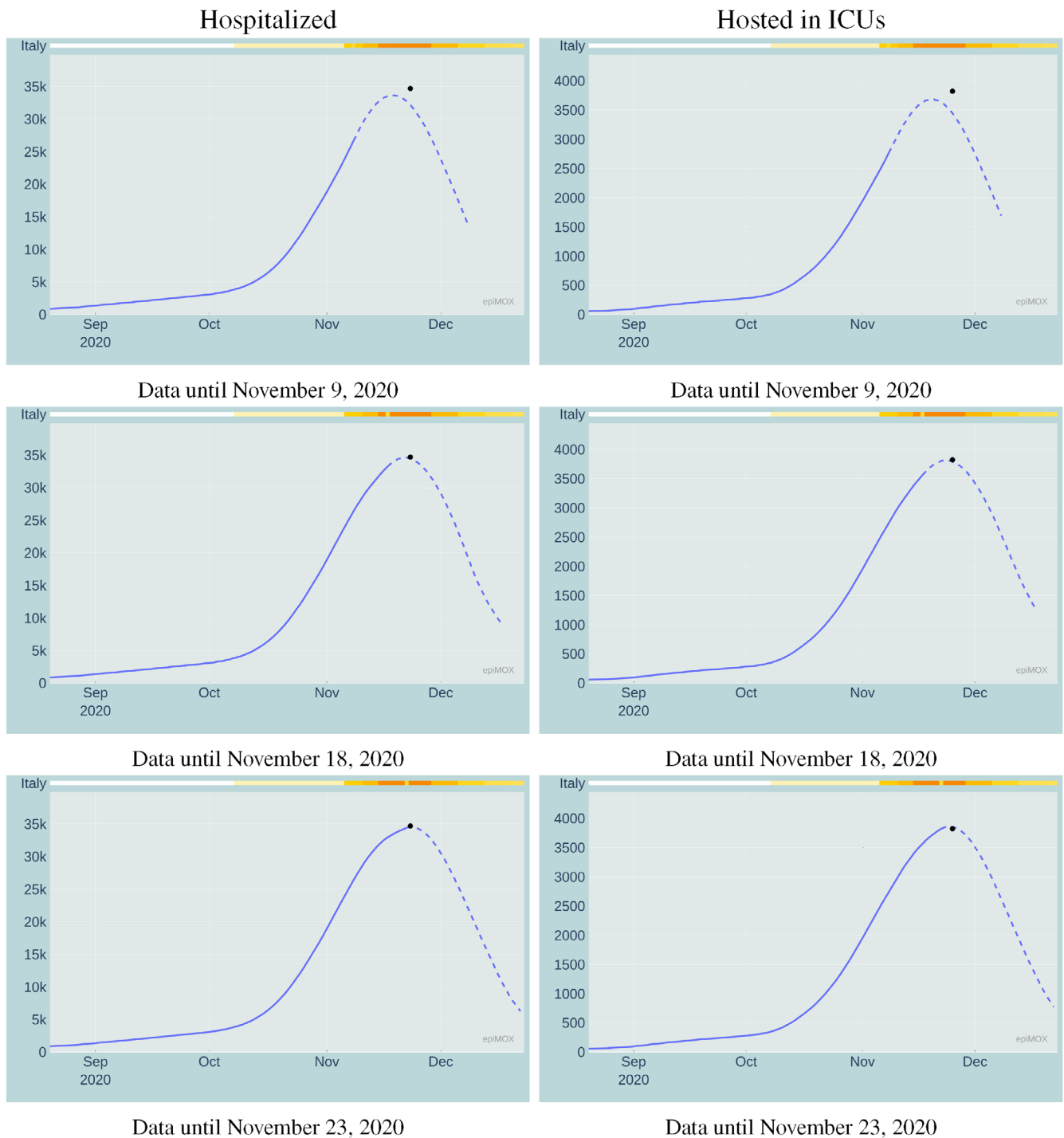
Hospitalized

Hosted in ICUs

Data until November 9, 2020

Data until November 9, 2020

Data until November 18, 2020

Data until November 18, 2020

Data until November 23, 2020

Data until November 23, 2020

FIGURE 15 Prediction of the peak of the *Hospitalized* (left) and *Hosted in ICUs* (right) compartments during the second epidemic wave estimated by the SUIHTER model at different dates prior the peak. The black dot indicates the peak actually occurred for the corresponding compartment (day and numbers)

(see e.g., Reference 18) indicate that the UK variant reached a prevalence of 17.8% in Italy at the beginning of February 2021 and. Due to its higher transmission rate (recently estimated in Italy as 37% higher than the one of the baseline virus[19]), it is going to become the prevalent virus variant in March 2021.

The updated SUIHTER model accounting for the UK virus variant is governed by the following system of ODEs (see Figure 17)

TABLE 2 Forecast of peak values and days of *Hospitalized* and *Hosted in ICUs* compartments for the second epidemic wave in Italy

Day of forecast	Hospitalized		Hosted in ICUs	
	Peak value	Peak date	Peak value	Peak date
November 09, 2020	33,176	November 19, 2020	3618	November 19, 2020
November 18, 2020	34,639	November 22, 2020	3989	November 25, 2020
November 23, 2020	34,527	November 23, 2020	3859	November 24, 2020
Actual peaks	34,697	November 23, 2020	3823	November 25, 2020

(A) Isolated

(B) Hospitalized

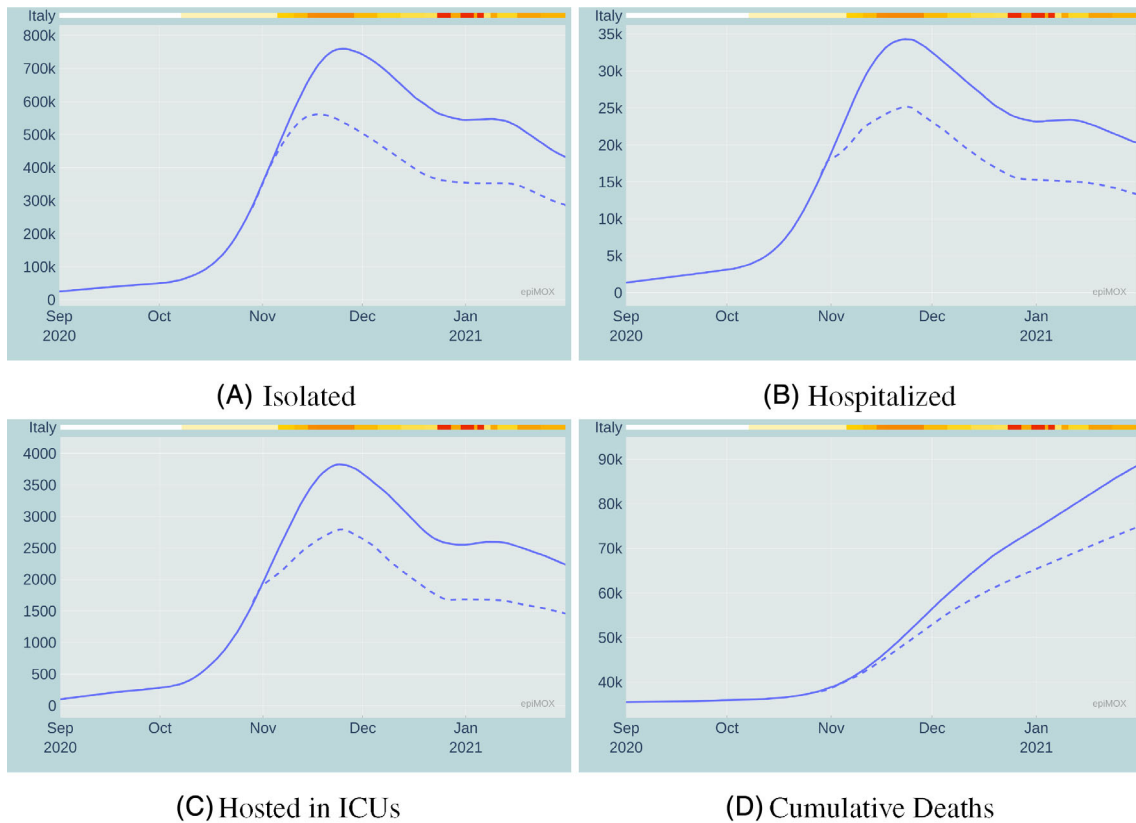(C) Hosted in ICUs

(D) Cumulative Deaths

FIGURE 16 Comparison between the observed time evolution (solid line) of four compartments and a simulated scenario (dashed line) in which the November restrictive measures had been anticipated by 10 days
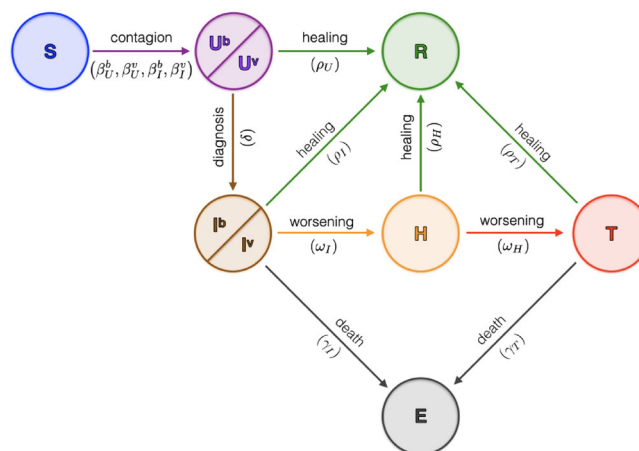

FIGURE 17 Sketch of the compartmental epidemiological model SUIHTER used to account for the UK virus variant

$$\dot{S}(t) = -S(t)\frac{\beta_U^b U^b(t) + \beta_U^v U^v(t) + \beta_I^b I^b(t) + \beta_I^v I^v(t)}{N},$$

$$\dot{U}^b(t) = S(t)\frac{\beta_U^b U^b(t) + \beta_I^b I^b(t)}{N} - (\delta + \rho_U)U^b(t),$$

$$\dot{U}^v(t) = S(t)\frac{\beta_U^v U^v(t) + \beta_I^v I^v(t)}{N} - (\delta + \rho_U)U^v(t),$$

$$\dot{I}^b(t) = \delta U^b(t) - (\rho_I + \omega_I + \gamma_I)I^b(t),$$

$$\dot{I}^v(t) = \delta U^v(t) - (\rho_I + \omega_I + \gamma_I)I^v(t),$$

$$\dot{H}(t) = \omega_I\left(I^b(t) + I^v(t)\right) - (\rho_H + \omega_H)H(t),$$

$$\dot{T}(t) = \omega_H H(t) - (\rho_T + \gamma_T)T(t),$$

$$\dot{E}(t) = \gamma_I\left(I^b(t) + I^v(t)\right) + \gamma_T T(t),$$

$$\dot{R}(t) = \rho_U\left(U^b(t) + U^v(t)\right) + \rho_I\left(I^b(t) + I^v(t)\right) + \rho_H H(t) + \rho_T T(t),$$

(2)

endowed with suitable initial conditions, where the indices b and v respectively denote the base and UK variant splitting of the *Undetected* and *Isolated* compartments and a 37% higher transmission rate for the UK variant is considered, namely

$$\beta_U^v = 1.37\beta_U^b, \qquad \beta_I^v = 1.37\beta_I^b.$$

A comparison between the 30-day forecast obtained using the original SUIHTER model and the updated model accounting for the effect of the UK virus variant starting on February 20, 2021 is presented in Figures 18, 19 and 20 for the *Isolated*, *Hospitalized* and *Hosted in ICUs* compartments, respectively.

The most remarkable differences are noticed on the *Isolated*, the *Hospitalized*, and, at a lesser extent, the *Hosted in ICUs* compartments. This analysis does not account for the vaccination campaign. Moreover, by the time we write this paper, the national spreading of the UK variant is still highly non-uniform, with a pronounced prevalence in a few local environments. The numerical simulation at the national scale provides a homogenization effect which is perhaps rather inaccurate. We highlight that the model accounting for the variant was able to predict the outbreak of the third epidemic wave observed in Italy in late February 2021.

4.4 | 30-day forecast with different scenarios

Modeling the epidemic evolution by a mathematical epidemiological model such as SUIHTER allows not only to analyze backward *what-if* scenario investigating the effectiveness of past political choices (as discussed in section 4.2), but also to forecast the evolution of the epidemic in the near future.
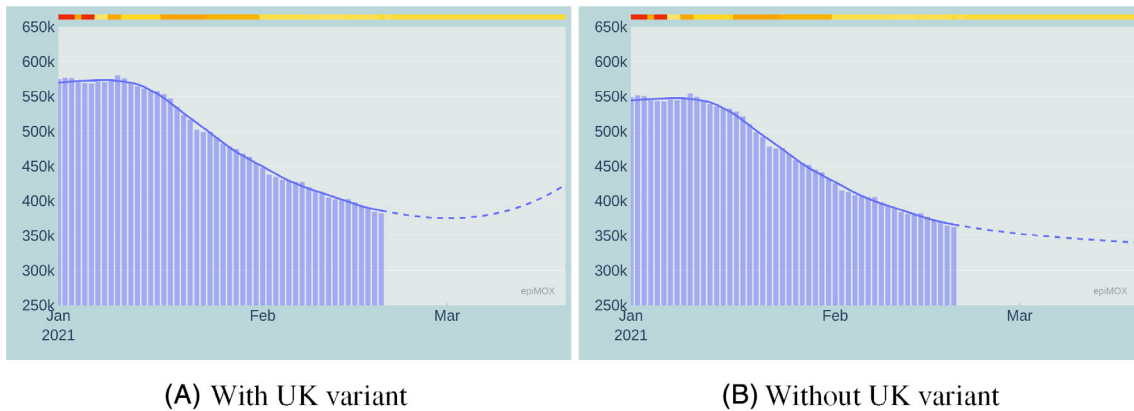

(A) With UK variant   (B) Without UK variant

FIGURE 18 30-day forecast of the *Isolated* compartment with (left) and without (right) the effect of the UK virus variant
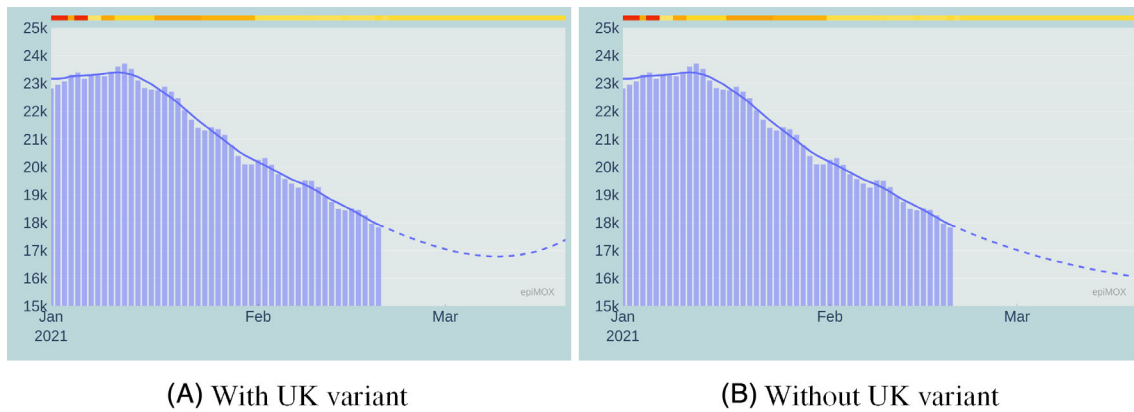
(A) With UK variant

(B) Without UK variant

FIGURE 19   30-day forecast of the *Hospitalized* compartment with (left) and without (right) the effect of the UK virus variant

(A) With UK variant
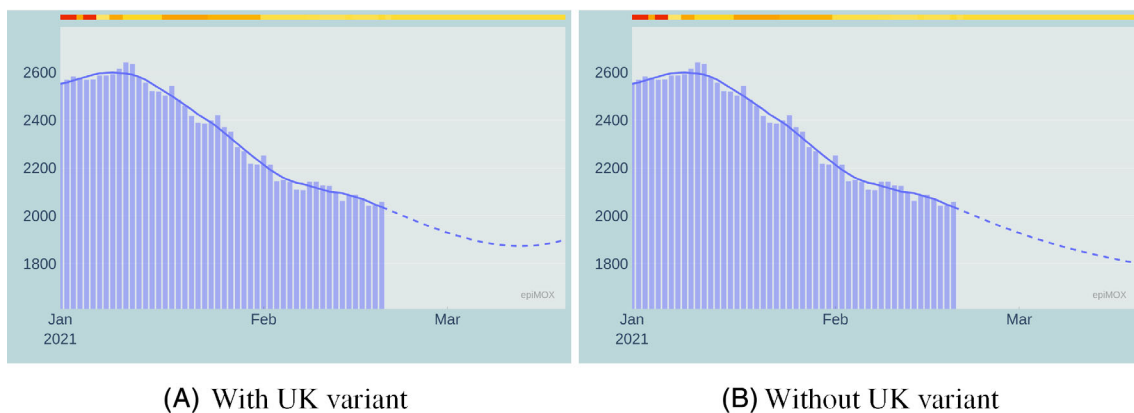
(B) Without UK variant

FIGURE 20   30-day forecast of the *Hosted in ICUs* compartment with (left) and without (right) the effect of the UK virus variant

As previously mentioned, the outbreak of the third epidemic wave has been observed in Italy since the second half of February 2021. At the time of writing, this outbreak is in its exponential growth phase and its evolution will strongly depend on the NPIs that will be imposed in the coming weeks.

We have considered three different scenarios with different level of severity in the NPIs:

- the first scenario considers the restriction currently imposed (on March 7, 2021) in which some limitations are adopted at national level (use of mask, limitation of mobility among regions, limitation on recreational activities, ...), while additional stricter limitations (confinement within municipality limits, distance learning for school and universities, ...) are adopted with increasing level of severity in *yellow*, *orange* and *red* regions)‡;
- the second scenario considers that all the region become *yellow* starting on March 8, 2021;
- the second scenario considers that all the region become *red* starting on March 8, 2021.

A comparison among the time evolution of the *Isolated*, *Hospitalized* and *Hosted in ICUs* compartments for the three scenarios are displayed in Figures 21, 22 and 23, respectively. As expected, when more severe restrictions are imposed the growths of all the infecting compartments are slowed down. The results clearly highlight that the ongoing outbreak that has been activated by the increased transmission rate of the new virus variants could be controlled only should stricter restriction measures be imposed as soon as possible. Without a prompt intervention the risk of a third epidemic wave that will end up to be worse than the previous two is extremely high.

The model results displayed on the dashboard are obtained real-time running the model any time the dashboard is interrogated, thus allowing to switch between different forecast or what-if scenarios.
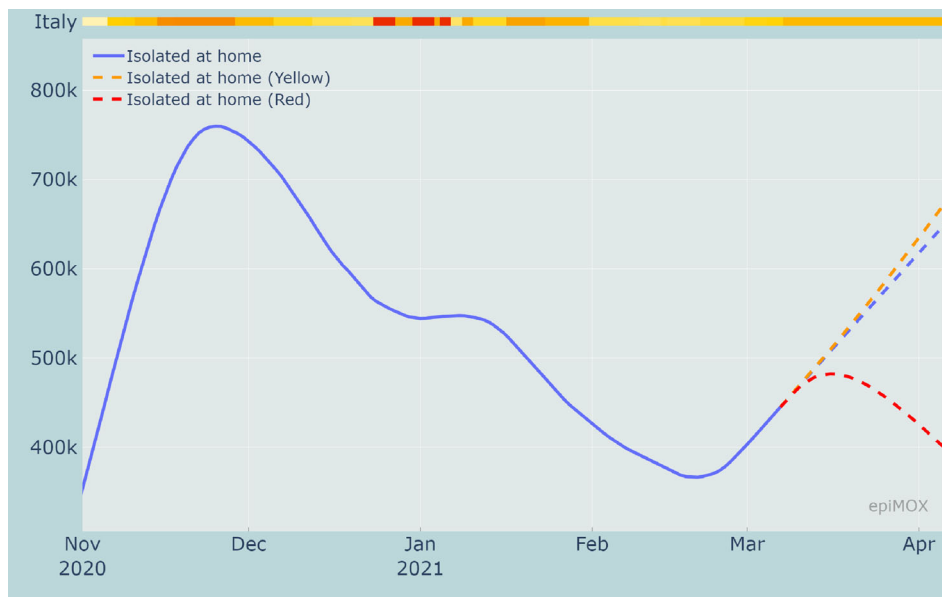
FIGURE 21 30-day forecast of the *Isolated* compartment with current restrictions (at March 7, 2021) and with 3 alternative restriction scenarios
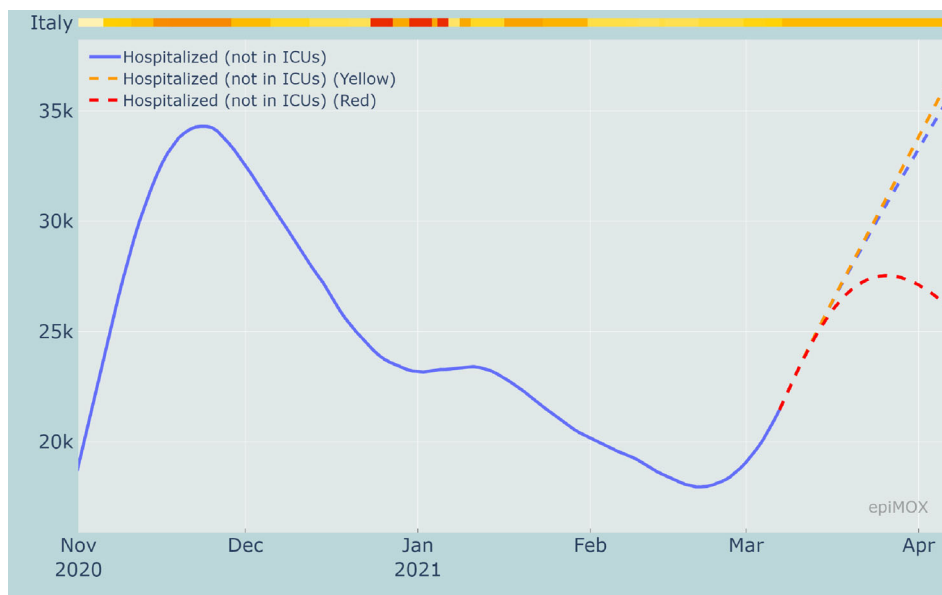

FIGURE 22 30-day forecast of the *Hospitalized* compartment with current restrictions (at March 7, 2021) and with 3 alternative restriction scenarios

As mentioned in Section 4 and discussed in details in Reference 2, the SUIHTER model can be calibrated using a Monte-Carlo Markov Chain (MCMC) approach for a robust estimate of the parameters. For each parameter considered in the calibration, the MCMC calibration starts from a prior, in our case a uniform distribution around an initial guess computed by minimizing with a least-square minimization procedure the difference between data and model results, and provides its posterior probability density function.

Based on the posterior estimate of the model calibration it is possible to supply confidence interval for the time evolution of the different compartments.

The solid curves displayed in Figure 24 represent the time evolution of the mean value of the different compartments in the three considered scenarios, while the shaded region surrounding each curve represents the 95% confidence interval associated with the probabilistic characterization of the model parameters.

FIGURE 23 30-day forecast of the *Hosted in ICUs* compartment with current restrictions (at March 7, 2021) and with 3 alternative restriction scenarios
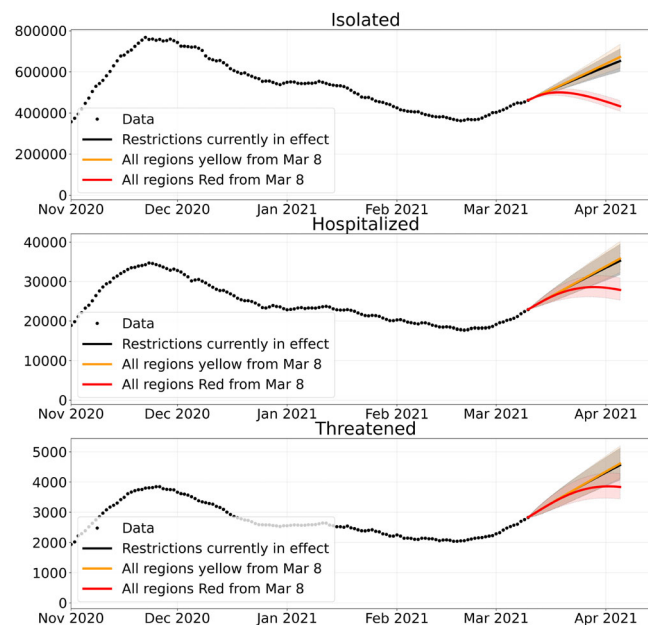

FIGURE 24 30-day forecast of four compartments with current restrictions (at March 7, 2021) and with 2 alternative restriction scenarios and 95% confidence interval

5 | CONCLUSIONS

In this paper, we presented the *epiMOX* dashboard for the analysis of the COVID-19 epidemic that hit northern Italy in early Spring 2020 and that is still severely affecting the entire Country in March 2021. Data were reported for several compartments (*total and daily positive* cases, *isolated at home*, *hospitalized*, *hosted in ICUs*, *cumulative deaths*, *recovered*) that allow to provide a synthetic yet informative description of the epidemic evolution. Several analyses were carried out by means of the dashboard. A careful comparison between the first two epidemic waves (the one of Spring 2020,

and the one that took off in Fall 2020) is made. Finally, we have reported results, mainly concerning predictions, obtained by using the novel epidemiological differential mathematical model SUIHTER. This allowed us to provide short-term forecasts on the epidemic trends, as well to examine some what-if scenarios related to the possible implementations of different NPIs than those that had been actually enforced.

ORCID

Giovanni Ardenghi 🄳 https://orcid.org/0000-0001-5154-5973
Luca Dede' 🄳 https://orcid.org/0000-0002-6558-8277

ENDNOTES

* https://www.gazzettaufficiale.it/eli/gu/2020/11/04/275/so/41/sg/pdf

† DPCM November 04, 2020, https://www.gazzettaufficiale.it/eli/gu/2020/11/04/275/so/41/sg/pdf

‡ A detailed summary of the evolution of the NPIs introduced at the regional level has been collected on this webpage https://it.wikipedia.org/wiki/Gestione_della_pandemia_di_COVID-19_in_Italia

REFERENCES

1. Quarteroni A, Dede' L, Parolini N. *Data analysis and predictive mathematical modeling for COVID-19 epidemic studies*. Cham: Springer International Publishing; 2020:1-7.
2. Parolini N, Dede' L, Antonietti PF, et al. SUIHTER: A new mathematical model for COVID-19. Application to the analysis of the second epidemic outbreak in Italy. *arXiv*. 2021. https://arxiv.org/abs/2101.03369.
3. Savitzky A, Golay MJE. Smoothing and differentiation of data by simplified least squares procedures. *Anal Chem*. 1964;36(8):1627-1639.
4. R. Battiston webpage. https://www.robertobattiston.it/ [Accessed March 12, 2021].
5. Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. *Proc R Soc Lond*. 1927;115(772):700-721.
6. Bertuzzo E, Mari L, Pasetto D, et al. The geography of COVID-19 spread in Italy and implications for the relaxation of confinement measures. *Nat Commun*. 2020;11(1):4264.
7. Brauer F, Castillo-Chavez C, Feng Z. *Mathematical models in epidemiology*. New York, NY: Springer; 2019.
8. Gatto M, Bertuzzo E, Mari L, et al. Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures. *Proc Natl Acad Sci U S A*. 2020;117(19):10484-10491.
9. Giordano G, Blanchini F, Bruno R, et al. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nat Med*. 2020;26:855-860.
10. Hethcote HW. The mathematics of infectious diseases. *SIAM Rev*. 2000;42(4):599-653.
11. Martcheva M. *An introduction to mathematical epidemiology*. Vol 61. New York, NY: Springer; 2015.
12. Piazzola C, Tamellini L, Tempone R. A note on tools for prediction under uncertainty and identifiability of SIR-like dynamical systems for epidemiology. *Math Biosci*. 2021;332:108514.
13. Haario H, Laine M, Mira A, Saksman E. DRAM: efficient adaptive MCMC. *Stat Comput*. 2006;16:339-354.
14. Miles PR. pymcmcstat: a Python package for Bayesian inference using Delayed Rejection Adaptive Metropolis. *J Open Source Softw*. 2019;4(38):1417.
15. Giordano G, Colaneri M, Filippo AD, et al. Vaccination and SARS-CoV-2 variants: how much containment is still needed? A quantitative assessment. *arXiv*. 2021. https://arxiv.org/abs/2102.08704.
16. Chen J, Gao K, Wang R, Wei G. Prediction and mitigation of mutation threats to COVID-19 vaccines and antibody therapies. *Chem Sci*. 2021;12:6929-6948.
17. Public Health England (PHE), NERVTAG note on B.1.1.7 severity. https://www.gov.uk/government/publications/nervtag-update-note-on-b117-severity-11-february-2021
18. Istituto Superiore di Sanità, Prevalenza della variante VOC 202012/01, lineage B.1.1.7 in Italia. Studio di Prevalenza 4–5 Febbraio 2021; 2021.
19. Istituto Superiore di Sanità, Qual è la trasmissibilità della "variante inglese" in Italia?; 2021.

How to cite this article: Parolini N, Ardenghi G, Dede' L, Quarteroni A. A mathematical dashboard for the analysis of Italian COVID-19 epidemic data. *Int J Numer Meth Biomed Engng*. 2021;37(9):e3513. https://doi.org/10.1002/cnm.3513