Modelling the COVID-19 epidemic and the vaccination campaign in Italy by the SUIHTER model

Nicola Parolini [a, *], Luca Dede' [a], Giovanni Ardenghi [a], Alfio Quarteroni [a, b]

[a] MOX, Department of Mathematics, Politecnico di Milano, Italy
[b] Institute of Mathematics, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland

ABSTRACT

Several epidemiological models have been proposed to study the evolution of COVID-19 pandemic. In this paper, we propose an extension of the SUIHTER model, to analyse the COVID-19 spreading in Italy, which accounts for the vaccination campaign and the presence of new variants when they become dominant. In particular, the specific features of the variants (e.g. their increased transmission rate) and vaccines (e.g. their efficacy to prevent transmission, hospitalization and death) are modeled, based on clinical evidence. The new model is validated comparing its near-future forecast capabilities with other epidemiological models and exploring different scenario analyses.

© 2022 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Since the beginning of the COVID-19 pandemic in early 2020, the mathematical community has devoted an unprecedented effort in developing novel mathematical, numerical and statistical tools to investigate the evolution of the most challenging public health threat in decades. For an overview on recent advances in mathematical epidemiology, computational modelling, physics-based simulation, data science, and machine learning applied to the COVID-19 pandemic, we refer the reader to (Kuhl, 2021). Even limiting the focus to the Italian context, several contributions have been proposed to accurately describe the spatio-temporal spreading of the epidemic in Italy (Bertuzzo et al., 2020; Della Rossa et al., 2020; Gatto et al., 2020; Giordano et al., 2020; Loli Piccolomini & Zama, 2020), to forecast its future evolution (Bartolucci et al., 2021; Farcomeni et al., 2021; Parolini et al., 2021a) and to quantify (and possibly optimize) the effects of containing measures, including both pharmaceutical and non-pharmaceutical interventions (NPIs) (Bonifazi et al., 2021; Giordano et al., 2021; Marziano et al., 2021).

As for every other virus, SARS-CoV-2, the virus responsible of COVID-19, changes in time through mutations that generate variants. Variants may differ from the wild-type virus by transmission rates, capability of leading to severe disease, response to vaccination. Given the enormous impact that the COVID-19 pandemic had (and is still having) on global health, economy, environment and society, the appearance and evolution of SARS-CoV-2 have been carefully tracked since the beginning of the

---

* Corresponding author.
  E-mail address: nicola.parolini@polimi.it (N. Parolini).
  Peer review under responsibility of KeAi Communications Co., Ltd.

pandemic by collecting sequences on shared public repositories (PANGO Lineages, 2021; Nextstrain, 2021; GISAID, 2021) and monitoring, in particular, the so-called *variants of interest* (VOI) and *variants of concern* (VOC). In particular, the latter are defined as those variants which may represent an additional risk due to an increased transmissibility, an increased virulence or a decrease in the effectiveness of public health and social measures or available diagnostics, vaccines, therapeutics (Abu-Raddad et al., 2021; Koyama et al., 2020). Since the beginning of the pandemic, four variants have been classified as VOC by the World Health Organization (W. H. Organization, 2021), namely the *Alpha* variant (lineage B.1.1.7), designated on December 18, 2020, first documented in United Kingdom in September 2020, the *Beta* variant (lineage B.1.351), designated on December 18, 2020, first documented in South Africa in May 2020, the *Gamma* variant (lineage P.1), designated on January 11, 2021, first documented in Brazil in November 2020, and the *Delta* variant (lineage B.1.617.2), designated on May 11, 2021, first documented in India in October 2020. Finally, the recent variant *Omicron* was first documented in November 24, 2021 in South Africa and designated on November 26, 2021. It is currently subject of an intense effort by the scientific community to characterize its behaviour in terms of transmissibility, disease severity, immune escape.

When the prevalence of one (or more) of these variants become relevant, the epidemiological models used to track the pandemic evolution should take them into account. Several modelling contributions to describe the effect of variants have been proposed in the literature in the past few months (Dizioli & Radzikowski, 2021; Giordano et al., 2021; Krueger et al., 2021; Ramos et al., 2021).

Another relevant contribution that deserves the attention of the modelling community is the impact of the vaccination campaign that has now reached, in several countries, a large coverage of the population. Different vaccines for COVID-19 have been authorized by the public health organizations and introduced in the market since late 2020 (Hodgson et al., 2021; Jeyanathan et al., 2020). In particular, in the western countries, the most of the vaccination campaign have been carried out using four products: the *Comirnaty* vaccine (from Pfizer/BioNTech), the *Spikevax* vaccine (from Moderna), the *Vaxzevria* vaccine (from AstraZeneca) vaccine, and the *Janssen* vaccine (from Johnson & Johnson). The first two (Comirnaty and Spikevax) are based on the messenger RNA (mRNA) technology and similarly behave in terms of efficacy (Self et al., 2021). AstraZeneca and Janssen vaccines are based on the more traditional viral vector technology. Among the four vaccines, Janssen is the only proposed (at least in the first phase of the vaccination campaign) with a single dose administration, while the others are supplied with two doses separated by a time delay ranging between 3 and 6 weeks for mNRA vaccines and between 4 and 12 weeks for Vaxzevria.

SUIHTER is an epidemiological model, first introduced in (Parolini et al., 2021a), which is designed to conform to the epidemiological data that have been made available by the Italian Authorities since the beginning of the epidemics. The choice of the compartments defining the model has been driven by the assumption that the model should match as close as possible such available data.

In this paper, we propose an extension of the SUIHTER model. This new model is designed to account for the effect of the vaccination campaign, from one side, and the possible coexistence of different virus variants, from the other side. In particular, we consider the administration of two vaccine doses and the presence of the variants *Alpha* and *Delta* that sequentially became dominant in Italy during 2021. A further improvement of the model is its ability to exploit additional set of data that were lately made available by the Italian Authorities, specifically some additional fluxes between compartments (such as new admissions in hospital and ICUs). These information were not available when the original SUIHTER model was introduced. Our analysis does not consider the *Omicron* variant because of the scarce knowledge about its prevalence and transmissibility when this paper was submitted (mid-December 2021).

Several attempts have been made to propose rigorous evaluation strategies to compare the forecast capabilities of different models; see, e.g., (Bracher et al., 2021a; Cramer et al., 2021). Ensemble approaches combining multiple available forecasts from different models have proved to outperform single models in forecasting the early phases of COVID-19 pandemic (Bracher et al., 2021b; Funk et al., 2020; Taylor & Taylor, 2021), as well as influenza (Reich et al., 2019a, 2019b), dengue fever (Johansson et al., 2019), and Ebola (Viboud et al., 2018) outbreaks. The superior performance of ensemble forecast have been shown in different disciplines and they are typically associated to the ability of integrating the information coming from different models producing accurate predictions with well-calibrated uncertainty (Bates & Granger, 1969; Leutbecher & Palmer, 2008), avoiding, at the same time, the risk of delayed (or missing) forecasts or premature interruption of the forecast supply, which may occur when relying on a single specific model.

To evaluate its performance in terms of forecasting capability, the SUIHTER model has joined the European COVID-19 Forecast Hub (European Covid-19 Forecast Hub, 2021) since April 2021. The aim of this collaborative project is to provide decision makers and the general public with reliable estimates of the near-term future evolution of the COVID-19 pandemic in European countries by collecting forecasts from different models into an ensemble. A quantitative comparison based on suitable ranking indicators has been carried out, showing that the SUIHTER model can produce reliable short term forecast on different epidemic quantities of interest.

Moreover, different analyses have been carried out to evaluate at which extent the extended SUIHTER model accounting for variants and vaccination is able to better identify specific epidemic trends (such as new outbreak associated to the growth of a more transmissible virus variant) and how it can be used to quantify the effect of the vaccination campaign (including the effect of the vaccination rate).

The paper is organized as follows: in Section 2 the SUIHTER model including the vaccine compartment is introduced, and it is later extended in Section 2.1 to also account for an emerging variant. The strategy used to model multiple scenarios characterized by different NPIs is presented in Section 3. The numerical results are presented and discussed in Section 4,

including a quantitative validation of the forecasting capabilities of the model and different analyses highlighting the role of the different new features of the model (including new variants and vaccination effects). Finally, in Section 5, we draw our conclusions and we discuss model's limitations and some possible future developments.

2. Mathematical model

The design of the original SUIHTER model, introduced in (Parolini et al., 2021a), relied on the criterion that the epidemiological compartments should match as close as possible the data that were daily made available by the Italian authorities. In the present work, we propose an extended version of the SUIHTER model which is able to exploit a set of additional data that started to be released at a later stage (including, e.g., the number of new admissions in hospitals and ICUs, as well as data classified by age and clinical status). Moreover, the new model accounts for the role of the vaccination campaign that is ongoing in most countries since the end of 2020, by adding three additional compartments collecting vaccinated individuals.

The extended SUIHTER model is defined by the following system of ordinary differential equations:

$$\dot{S}(t) = -S(t)\,\frac{\beta_U\,U(t)}{N} - v_1,$$

$$\dot{U}(t) = (S(t) + \sigma_1\,V_1(t) + \sigma_2\,V_2(t))\,\frac{\beta_U\,U(t)}{N} - (\delta + \rho_U)\,U(t),$$

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

$$\dot{H}(t) = \omega_I\,I(t) - (\rho_H + \omega_H + \gamma_H)\,H(t) + \theta_T\,T(t),$$

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

$$\dot{E}(t) = \gamma_I\,I(t) + \gamma_H\,H(t) + \gamma_T\,T(t),$$ (1)

$$\dot{R}(t) = \rho_U\,U(t) + \rho_I\,I(t) + \rho_H\,H(t) - v_R,$$

$$\dot{V}_1(t) = v_1 - v_2 - \sigma_1\,V_1(t)\,\frac{\beta_U\,U(t)}{N},$$

$$\dot{V}_2(t) = v_2 - \sigma_2\,V_2(t)\,\frac{\beta_U\,U(t)}{N},$$

$$\dot{V}_R(t) = v_R,$$

endowed with suitable initial conditions, where the model compartments are defined as follows (see Fig. 1):

- S: number of *susceptible* (uninfected) individuals;
- U: number of *undetected* (both asymptomatic and symptomatic) infected individuals;
- I: number of infected individuals *isolated* at home;
- H: number of infected *hospitalized* individuals;
- T: number of infected *threatened* individuals hosted in ICUs;
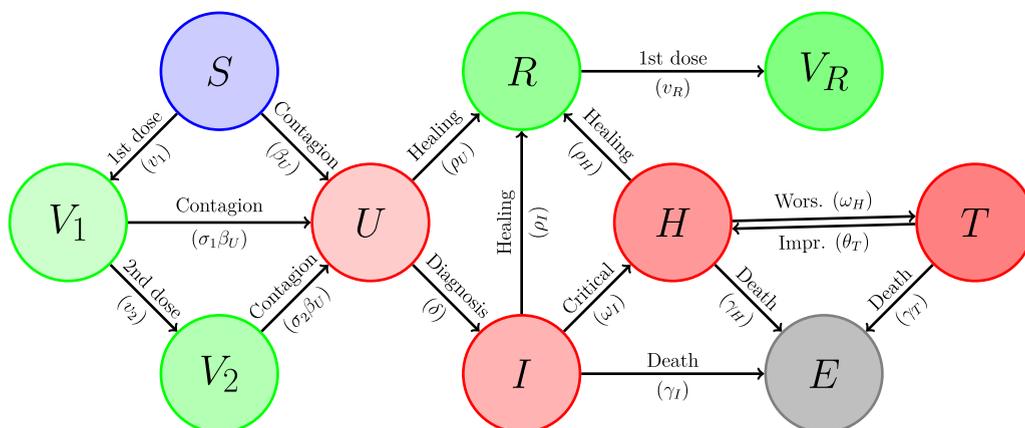- E: number of *extinct* individuals;

Fig. 1. Interactions among compartments in the extended SUIHTER model.

- R: number of *recovered* individuals;
- V_1: number of individuals partially *vaccinated* with only one dose;
- V_2: number of individuals fully *vaccinated* with two doses;
- V_R: number of *recovered* individuals who have been *vaccinated*,

and $N = S + U + I + H + T + E + R + V_1 + V_2 + V_R$ denotes the total population (assumed constant). Here, t measures time in days, compartments report figures in units. The system is numerically solved with the explicit fourth-order Runge-Kutta scheme with time step $\Delta t = 1$ day.

The quantity v_1 denotes the number of susceptible individuals per day who acquire immunity from the first dose. Similarly, v_2 denotes the number of individuals per day which are already vaccinated with one dose and acquire a stronger immunity by the second dose. Moreover, v_R denotes the number of recovered individuals per day who receive a vaccine dose. The number of first doses $w_1(t)$ and second doses $w_2(t)$ administrated in Italy are daily recorded on the Italian Government repository (Commissario straordinario per l'emergenza Covid-19, 2021). The quantity w_1 does not include the number of vaccine doses $w_R(t)$ administrated to (detected) recovered individuals. Unfortunately, these vaccination rates cannot be directly used in system (1) since during the whole epidemic period a great number of cases were not identified so that many recovered individuals have been vaccinated as if they were still susceptible. The number of individuals who recovered before (or without) being detected can be defined as $R_U(t) = R(t) - R_D(t)$, where

$$R_D(t) = \int_{t_0}^{t} (\rho_I I(\tau) + \rho_H H(\tau)) \, d\tau,$$

is the number of individuals who recovered after having been detected and can be obtained by postprocessing the computed compartments I and H. The indicator R_U can be used to estimate the rate of vaccine doses that have been actually administrated to susceptible individuals, by assuming that recovered individuals who were not formerly detected have the same probability of being vaccinated as susceptible individuals. Namely, the rate of first doses administrated to susceptible individuals are estimated as

$$v_1(t) = w_1(t - t_{im}) \frac{S(t)}{S(t) + R_U(t)},$$

where we have assumed that the time required to develop the immunity is $t_{im} = 14$ days. Similarly, the second doses administrated to partially vaccinated individuals (V_1) are only a fraction of the total number of second doses, therefore we get

$$v_2(t) = w_2(t - t_{im}) \frac{S(t)}{S(t) + R_U(t)}.$$

Finally, the total number of doses per day that recovered individuals (either detected or undetected) receive is given by:

$$v_R(t) = w_R(t - t_{im}) + w_1(t - t_{im}) \frac{R_U(t)}{S(t) + R_U(t)}$$

The model can be initialized at any time t_0 prior to the start of the vaccination campaign (December 27, 2020 in Italy) by using the data (for those compartments for which data are available, namely I, H, T, E), null initial values for V_1 and V_2, while, as proposed in (Parolini et al., 2021b), the *Undetected* and *Recovered* compartments are initialized as

$$R(t_0) = \left(\frac{1}{\text{IFR}(t_0)} - 1\right) E(t_0), \qquad U(t_0) = \left(\frac{\text{CFR}(t_0 + d)}{\text{IFR}(t_0)} - 1\right) (I(t_0) + H(t_0) + T(t_0)),$$

where $\text{IFR}(t)$ is the *Infection Fatality Ratio*, $\text{CFR}(t)$ is the time-dependent *Case Fatality Ratio* and $d = 13$ days denotes the confirmation-to-death delay. In (Parolini et al., 2021b), we considered a constant IFR computed as the weighted average of the age-specific IFR estimates weighted by the population age structure, under the assumption of equal attack rates across age-groups, as proposed in (Brazeau et al., 2020). A better estimate of the reference $\text{IFR}(t)$ at a specific time can be computed by considering the variable percentage of each age-group among the total infected over time (data available on (Dati della Sorveglianza integrata COVID-19 in Italia, 2021) since December 8, 2020), namely:

$$\text{IFR}(t) = \sum_{i=1}^{m} q_i(t) \, \text{IFR}_i,$$

where IFR_i denotes the infection fatality ratio for age-group i and $q_i(t)$ is the percentage of infected at time t belonging to age-group i. This correction may be particularly relevant when using the model in a time frame in which the vaccination campaign is not homogeneous across age-groups, so that the percentage of different age-groups among the total infected may vary

significantly. A further improved estimate for IFR(t) can be obtained considering the vaccine effectiveness in reducing the mortality for different age-groups, namely:

$$\text{IFR}(t) = \sum_{i=1}^{M} q_i(t)\left[(1 - v_{i,1}(t) - v_{i,2}(t)) + v_{i,1}(t)\, m_{i,1} + v_{i,2}(t)\, m_{i,2}\right] \text{IFR}_i, \tag{2}$$

where $v_{i,1}(t)$ and $v_{i,2}(t)$ denote the fraction of individuals in age-group i who received one or two doses, respectively, at time t, while $m_{i,1}$ and $m_{i,2}$ denote the reduced probability to die for an isolated infected individual of age-group i who is vaccinated (with one or two doses, respectively) w.r.t. an unvaccinated isolated infected individual in the same age-group.

The time-history of the CFR(t) since the beginning of the epidemic is shown in Fig. 2 (left) where the high values during the first epidemic wave clearly indicates the strong underreporting. In Fig. 2 (right), CFR(t) and reference IFR(t) estimated with (2) (which can be computed based on data available from December 8, 2020) are compared. The effect of the vaccination campaign during year 2021 is clearly visible in the reduction of the reference IFR(t).

Model (1) depends on the following parameters (rates):

- β_U denotes the transmission rate due to contacts between susceptible and undetected infected individuals;
- 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;
- θ_T denotes the rate at which T-individuals improve their health conditions and return to the less critical H compartment;
- δ denotes the rate of detection, relative to undetected infected individuals;
- ρ_U, ρ_I and ρ_H denote the rate of recovery for three classes (U, I and H, respectively) of infected subjects;
- γ_I, γ_H and γ_T denote the mortality rates for the individuals isolated at home, hospitalized and hosted in ICUs, respectively.

Some of the parameters in system (1) are evaluated starting from the available data. In particular, the worsening rates ω_I and ω_H are obtained directly from the new hospitalization and the new admission data as follows

$$\omega_I(t) = \frac{\hat{H}^+(t)}{\hat{I}(t)}, \qquad \omega_H(t) = \frac{\hat{T}^+(t)}{\hat{H}(t)}, \tag{3}$$

where $\hat{H}$ are the daily data provided by the Dipartimento della Protezione Civile (DPC) for hospitalized individuals, while $\hat{H}^+$ and $\hat{T}^+$ are the daily data provided by Istituto Superiore di Sanità (ISS) (Dati della Sorveglianza integrata COVID-19 in Italia, 2021) and DPC (Presidenza del Consiglio dei Ministri, 2021) respectively for new hospitalizations and admissions in ICUs. To reduce the effects of the weekly oscillation associated to irregular data reporting, the values of ω_I and ω_H used in (1) are the 1-week moving averages of the time series estimated in (3).

The detection rate δ is also estimated by assuming that at each time the probability p_D of being detected can be related to the ratio between the infection fatality rate and the time-dependent case fatality rate, namely

$$p_D(t) = \frac{IFR(t)}{CFR(t + d)}.$$

Assuming a mean detection time $t_D = 8$ days, the detection rate is then computed at each time t as

(a) CFR   (b) CFR vs IFR
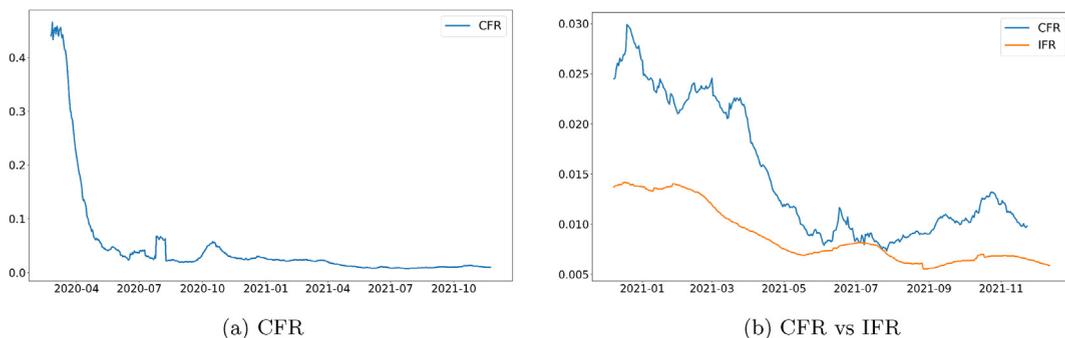
Fig. 2. Time-evolution of the reference CFR(t) since the beginning of the epidemic (left) and a comparison between CFR(t) and IFR(t) accounting for age-dependent attack rates and vaccine efficacy in preventing death since December 2020 (right).

$$\delta(t) = \frac{p_D(t)}{t_D}.$$

To avoid spurious oscillations, the time-history of $p_D(t)$ is also smoothed taking its weekly moving average. The time-history of the probability p_D of being detected at time t, based on the estimate of IFR(t) given in (2), is displayed in Fig. 3 where we can notice that the at mid July 2021, when the lowest incidence was achieved, the probability of being detected was almost 100%.

Given an estimate of δ, if we assume that all *Undetected* individuals can leave the compartment only by either being detected or recovering, we can also estimate the *Undetected* recovery rate ρ_U as

$$\rho_U(t) = \frac{1 - t_D\,\delta(t)}{t_R},$$

where t_R is the mean recovery time for *Undetected*, which is estimated through the calibration process.

At this stage, the effect of the vaccine in the model is given by the susceptibility of vaccinated individuals with respect to non-vaccinated individuals, given by the parameters $\sigma_1 = 0.287$ and $\sigma_2 = 0.115$ for partially and completely vaccinated individuals, respectively. These values are computed based on the estimates reported in (Istituto Superiore di Sanità, 2021a) about vaccine efficacy against infection.

The remaining parameters of the model need to be calibrated to reproduce the epidemic history. Some parameters can be considered as constant in time throughout the pandemic, while we prescribe the other ones are piecewise constant over n time phases $[t_k, t_{k+1}]$, $k = 1, ..., n$ of prescribed duration (typically equal to 15 days) in which the simulation time windows $[t_0, t_F]$ is subdivided. The parameters chosen to be constant in time are: t_R, γ_I and γ_T, while $\beta_{U,k}$, $\rho_{I,k}$, $\rho_{H,k}$, $\theta_{T,k}$ and $\gamma_{H,k}$, $k = 1, ..., n$ are constant on the k-th phase $[t_k, t_{k+1}]$. All these parameters are calibrated through a two-step calibration: a first estimate is obtained with a least square procedure, this estimate is then used as initial condition for a Monte Carlo Markov Chain (MCMC) procedure. The prior is a uniform distribution around the least square estimate with $\pm 30\%$ of the parameter value interval. For a more detailed description of the calibration process, see (Parolini et al., 2021a).

2.1. Accounting for virus variants

With the appearance of virus variants featuring a higher transmission rate than the wild type virus (Campbell et al., 2021), predicting the evolution of the epidemic curve has become more challenging. As a matter of fact, when a more aggressive variant starts spreading, the epidemiological characteristic could change, even drastically. When using the model for forecasting the future evolution of the epidemic in the presence of an emerging variant, it is necessary to include in model (1) the effects of variants. In particular, the model is modified by splitting the *Undetected* compartment, in which the contagion occurs, into two new compartments:

- U^b: number of *Undetected* (both asymptomatic and symptomatic) infected with the base variant, that is the variant being dominant at the initial time of the simulation,
- U^v: number of *Undetected* (both asymptomatic and symptomatic) infected with the variant of concern that is spreading faster among the population.
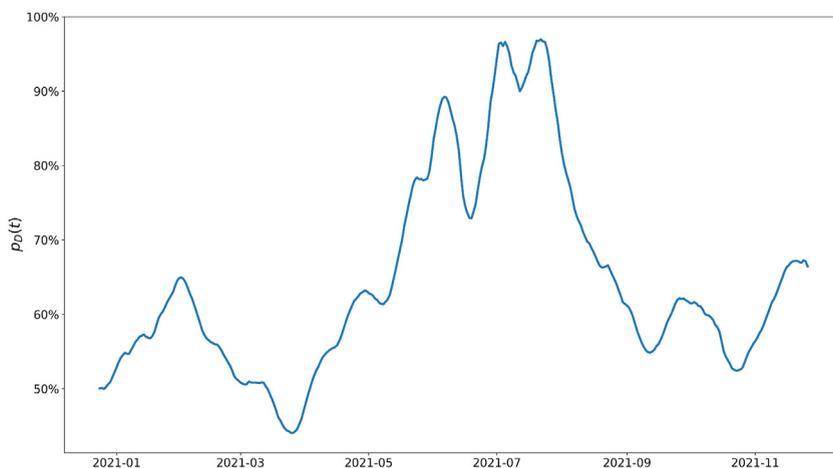
Fig. 3. Time-evolution of the probability p_D of being detected.

Susceptible individuals have different probabilities of being infected by people belonging to each of these two compartments, typically the new variant has a higher transmission rate than the base one. For example the *Delta* variant has been estimated to be around 50% more transmissible than the *Alpha* one (Scientific Advisory Group for Emergencies, 2021), and, in turn, it was previously estimated to be 38% more transmissible than the wild type virus (Tanaka et al., 2021).

The extended SUIHTER model with variants reads:

$$\dot{S}(t) = -S(t)\,\frac{\beta_U^b\,U^b(t) + \beta_U^v\,U^v(t)}{N} - \nu_1,$$

$$\dot{U}^b(t) = \left(S(t) + \sigma_1^b\,V_1(t) + \sigma_2^b\,V_2(t)\right)\frac{\beta_U^b\,U^b(t)}{N} - (\delta + \rho_U)\,U^b(t),$$

$$\dot{U}^v(t) = \left(S(t) + \sigma_1^v\,V_1(t) + \sigma_2^v\,V_2(t)\right)\frac{\beta_U^v\,U^v(t)}{N} - (\delta + \rho_U)\,U^v(t),$$

$$\dot{I}(t) = \delta\,(U^b(t) + U^v(t)) - (\rho_I + \tilde{\omega}_I + \tilde{\gamma}_I)\,I(t),$$

$$\dot{H}(t) = \tilde{\omega}_I\,I(t) - (\rho_H + \tilde{\omega}_H + \gamma_H)\,H(t) + \theta_T\,T(t),$$

$$\dot{T}(t) = \tilde{\omega}_H\,H(t) - (\theta_T + \gamma_T)\,T(t),$$ (4)

$$\dot{E}(t) = \tilde{\gamma}_I\,I(t) + \gamma_H\,H(t) + \gamma_T\,T(t),$$

$$\dot{R}(t) = \rho_U\,(U^b(t) + U^v(t)) + \rho_I\,I(t) + \rho_H\,H(t) - \nu_R,$$

$$\dot{V}_1(t) = \nu_1 - \nu_2 - \sigma_1^b\,V_1\,\frac{\beta_U^b\,U^b(t)}{N} - \sigma_1^v\,V_1\,\frac{\beta_U^v\,U^v(t)}{N},$$

$$\dot{V}_2(t) = \nu_2 - \sigma_2^b\,V_2\,\frac{\beta_U^b\,U^b(t)}{N} - \sigma_2^v\,V_2\,\frac{\beta_U^v\,U^v(t)}{N},$$

$$\dot{V}_R(t) = \nu_R,$$

endowed with suitable initial conditions. After calibrating model (1) on a time range $[t_0, t_0^f]$, model (4) can be initialized at time t_0^f using the parameters of the last phase of the calibration to forecast the epidemic evolution in the presence of an emergent variant. Notice that, when a single virus type is considered, we have $U^v(t) = 0$ and $U^b(t) = U(t)$, thus model (4) reduces to model (1).

Model (4) is initialized at time t_0^f with data from DPC if available. In order to allow the model to account for possible errors in reported data, we introduced a Gaussian error $\varepsilon \sim \mathcal{N}(0, s_k^2)$ on these values, where s_k is the standard deviation of the weekly average data for the compartment k during the whole calibration period. The other compartments, i.e. *Susceptible*, *Undetected base*, *Undetected variant*, *Recovered*, *Vaccinated with one dose*, *Vaccinated with two doses*, are initialized with the values at t_0^f obtained from thande simulation with model (1).

The initial values for U^b and U^v are obtained by splitting the value of U by accounting for the prevalence of the rising variant p_v given by surveillance reports of ISS (Istituto Superiore di Sanità, 2021b). The transmission rate for the base virus β_U^b is the last calibrated parameter β_U at time t_0^f, suitably rescaled to take into account the fact that the fraction p_v of infected individuals already contracted the virus variant. Given the increase in transmissibility f_v of the spreading variant, we obtain the following estimates for the transmission rates at time t_0^f

$$\beta_U^b(t_0^f) = \frac{\beta_U(t_0^f)}{1 + (f_v - 1)p_v},$$ (5)

$$\beta_U^v(t_0^f) = f_v\,\beta_U^b(t_0^f).$$ (6)

The coefficients σ_1^b and σ_2^b accounts for the susceptibility of vaccinated individuals w.r.t. the base version of the virus and their value are the same as σ_1 and σ_2 introduced in Section 1. The coefficients $\sigma_1^v = 0.5$ and $\sigma_2^v = 0.22$ (England, 1000England) instead are modified to take into account the reduced vaccines effectiveness against virus variants. The parameters $\tilde{\omega}_I$, $\tilde{\omega}_H$ and $\tilde{\gamma}_I$ are modified parameters to take into account the vaccines effect in reducing the severity of the disease and the mortality. Based on the vaccines surveillance report of Italian (Istituto Superiore di Sanità, 2021a) and UK governments, we defined the coefficients h_1, h_2, t_1, t_2, m_1, and m_2.

- $h_1 = 0.655$ and $h_2 = 0.417$ denote the probabilities of a isolated infected individual who is vaccinated (with one or two doses, respectively) to be hospitalized w.r.t. an unvaccinated isolated infected individual;
- $t_1 = 0.63$ and $t_2 = 0.6$ denote the probabilities of a hospitalized individual who is vaccinated (with one or two doses, respectively) to be admitted in ICU w.r.t. an unvaccinated hospitalized individual;

- $m_1 = 0.724$ and $m_2 = 0.35$ denote the probabilities of an isolated infected individual who is vaccinated (with one or two doses, respectively) to die w.r.t. an unvaccinated isolated infected individual.

Since in model (4), the infected compartments (U, I, H and T) collect both vaccinated and unvaccinated individuals, in order to quantify the effect of the vaccination on the worsening and mortality parameters (ω_I, ω_H and γ_I), we first estimate the partition of new positives coming from either S, V_1, or V_2 and entering either U_b or U_v compartments as follows

$$u_S^b(t) = \frac{S(t)}{N_b(t)}, \qquad u_1^b(t) = \frac{\sigma_1^b V_1(t)}{N_b(t)}, \qquad u_2^b(t) = \frac{\sigma_2^b V_2(t)}{N_b(t)}, \tag{7}$$

$$u_S^v(t) = \frac{S(t)}{N_v(t)}, \qquad u_1^v(t) = \frac{\sigma_1^v V_1(t)}{N_v(t)}, \qquad u_2^v(t) = \frac{\sigma_2^v V_2(t)}{N_v(t)}, \tag{8}$$

where $N_b(t) = S(t) + \sigma_1 V_1(t) + \sigma_2 V_2(t)$ and $N_b(t) = S(t) + \sigma_1^v V_1(t) + \sigma_2^v V_2(t)$.

The modified worsening and mortality parameters $\tilde{\omega}_I$, $\tilde{\omega}_H$ and $\tilde{\gamma}_I$ including the vaccine effects are then obtained as

$$\tilde{\omega}_I(t) = \omega_I(t_0^f) \frac{u_s(t) + h_1 u_1(t) + h_2 u_2(t)}{u_s(t_0^f) + h_1 u_1(t_0^f) + h_2 u_2(t_0^f)}, \qquad t > t_0^f, \tag{9}$$

$$\tilde{\omega}_H(t) = \omega_H(t_0^f) \frac{u_s(t) + t_1 u_1(t) + t_2 u_2(t)}{u_s(t_0^f) + t_1 u_1(t_0^f) + t_2 u_2(t_0^f)}, \qquad t > t_0^f, \tag{10}$$

$$\tilde{\gamma}_I(t) = \gamma_I(t_0^f) \frac{u_s(t) + m_1 u_1(t) + m_2 u_2(t)}{u_s(t_0^f) + m_1 u_1(t_0^f) + m_2 u_2(t_0^f)}, \qquad t > t_0^f. \tag{11}$$

where the total partition (including both base and variant contributions) of new positives coming from S, V_1, or V_2 is given by

$$u_S(t) = \frac{u_S^b(t) N_b(t) U^b(t) + u_S^v(t) f_v N_v(t) U^v(t)}{N_b(t) U^b(t) + N_v(t) f_v U^v(t)}$$

$$u_1(t) = \frac{u_1^b(t) N_b(t) U^b(t) + u_1^v(t) f_v N_v(t) U^v(t)}{N_b(t) U^b(t) + N_v(t) f_v U^v(t)} \tag{12}$$

$$u_2(t) = \frac{u_2^b(t) N_b(t) U^b(t) + u_2^v(t) f_v N_v(t) U^v(t)}{N_b(t) U^b(t) + N_v(t) f_v U^v(t)}.$$

We notice that, when new variants are not spreading, then the latter equations simply reduce to Eq. (7). Moreover, we point out that the rescaling in (12) using the partition at the initial time t_0^f of the forecast simulation has been used to account for the fact that the values $\omega_I(t_0^f)$, $\omega_H(t_0^f)$ and $\gamma_I(t_0^f)$ have been obtained calibrating the model on data which were already affected by the vaccination campaign (until time t_0^f).

Although more complex models could be devised to deal with the vaccination, for instance by splitting each infected compartment (U, I, H, T) in vaccinated and unvaccinated, we preferred to keep the model as simple as possible, but still able to account for both the reduced susceptibility of vaccinated individuals, as well as the lower risk of hospitalization and death through relations (9)–(11). Moreover, in the present model we are neglecting the waning immunity over time, which has been recently highlighted in several studies (Cohn et al., 2021; Goldberg et al., 2021; Tartof et al., 2021), since the time-frame considered in the simulations (see Section 4) are limited to the Summer 2021.

3. Forecast scenarios

The COVID-19 pandemic put a heavy burden on the Italian health system and caused a huge loss of human lives. During 2020 and early 2021, the Italian Government imposed restrictive measures to the population, in attempts to restrain and to contain the epidemic waves. Different NPIs have been applied since the beginning of the COVID-19 epidemic, ranging from basic containment measures (such as the mandatory use of masks) to drastic ones (full lockdown). All these interventions modified the transmissibility of the virus by reducing the number of contacts between people or the probability to contagion associated to a contact. Within epidemiological models, the effect of these NPIs can therefore be accounted for by acting on the transmission rates $\beta_U^{\{b,v\}}$. We included in the model the possibility of considering different future scenarios, corresponding to different possible NPIs or social events that may also lead to a change in the transmission rate (such as, for instance, school re-openings or holidays). We introduced a transmission coefficient τ that measures the change of the transmission rate at a given time due to imposition of restrictions (NPIs) or other social events and occurrences (schools opening, holidays, etc.). The

coefficient τ is computed for the scenario to simulate, as well for the actual conditions in the last phase of the calibration; see A. The transmission rates $\beta_U^{\{b,v\}}$ are indeed modulated by considering the change of the restriction coefficient as follows

$$\beta_U^{\{b,v\}}(t) = \frac{\tau(t)}{\tau(t_0^f)} \beta_U^{\{b,v\}}(t_0^f), \qquad t > t_0^f. \tag{13}$$

We remark that the transmission coefficient τ is used in a relative fashion for predictions, i.e. changes in NPIs and social activities are encapsulated in the term $\frac{\tau(t)}{\tau(t_0^f)}$ that modulates the predicted transmission rate. Several scenarios can be considered, each one with its own restriction coefficient. Novel restrictions soon to be implemented can be referred to population belonging to a specific age group (e.g. distance learning for schools) or to specific places (e.g. restaurants and bars closed). In order to take into account the heterogeneity of the measures we consider the interactions between people of different age groups in different contexts and modify them according to the scenario considered. Starting from the matrices of average number of contacts by age and context obtained from (Mossong et al., 2008), we associate a level of risk to each age group and context of exposure. SUIHTER is a homogeneous model and does not explicitly accounts for age or exposition context structure, thus the restriction coefficient is obtained from the restriction matrix by a weighted mean on the age groups with weights given by the amount of population in each group. Examples of scenario analyses using this approach to quantify the effect of different NPIs have already been presented and discussed in (Parolini et al., 2021b). The procedure for computing the restriction coefficient τ for a given NPI scenario is described Appendix A. Additional details will be available in a forthcoming paper.

4. Results and discussion

In this section we present the results obtained with the proposed model. In particular, in Section 4.1, we first provide a validation of the forecast capabilities of the model by comparing with other single model forecasts as well as ensemble forecasts. The role of the new features introduced in the model, namely variants and vaccines, are investigated in the numerical results presented in Sections 4.2 and 4.3, respectively. Finally, in Section 4.4, the results of a scenario analysis related to the introduction of Green Pass restrictions are presented and discussed.

4.1. Forecast assessment

The extended SUIHTER model (4) with variants has been included in the European COVID-19 Forecast Hub (European Covid-19 Forecast Hub, 2021) which collects short-term forecasts of COVID-19 cases (i.e. new contagions) and deaths across Europe, obtained by a multitude of infectious disease modelling teams, under the coordination of the European Centre for Disease Prevention and Control.

Every week, the different teams upload their forecasts of the cumulative value of incidence cases and incidence deaths per week for one or several European countries, with a forecast horizon between 1 and 4 weeks. The forecasts are supplied in an interval format, including for each quantity point evaluation and confidence interval in 23 quantiles ranging from 0.01 to 0.99.

For each forecast the absolute error (AE) and the weighted interval score (WIS), for both cases and deaths, are computed. The absolute error is simply the absolute value of the difference between the forecast and the measured data. The weighted interval score is a proper scoring rule proposed in (Bracher et al., 2021a) to evaluate and score forecasts in an interval format. It is a generalization of the absolute error and has three components: dispersion, overprediction and underprediction. Dispersion is a weighted average of the width of prediction intervals for different levels of uncertainty, the higher the level the higher the weight. The overprediction and underprediction are penalties added whenever an observation falls outside a central prediction interval. The penalties depend on how far the observation fall from the interval and the nominal level of the interval. For more details on these metrics, see (Cramer et al., 2021).

The number of forecasts supplied by the different teams may be different since some of them joined the project later than others, or forecasts in certain weeks for some models may be missing. In order to allow a fair comparison, in (Cramer et al., 2021) a relative measure of forecast performance, called *relative WIS* (relWIS), was introduced, taking into account that different teams may not cover the same forecast targets. The relative WIS is obtained performing a *pairwise comparison tournament* in which a score ratio between each pair of models is computed on the set of shared targets. The relative WIS is then obtained as the geometric mean of these ratios and normalized with the corresponding quantity computed for a reference baseline model. The forecast of the baseline model is defined with median given by the data measured in the most recent week, with uncertainty around the median based on changes in the past of the time series (see (Cramer et al., 2021)).

Models with smaller values of relWIS are thus performing better and a value below one means that the model performs better than the baseline model. The relWIS indicator can then be used to score different models.

The same procedure can be used to define a second scoring rule, the *relative absolute error* (relAE), in which the pairwise comparison tournament is performed based on the absolute error instead of the weighted interval score (see (Cramer et al., 2021)).

Table 1

Relative absolute error and relative weighted interval score for the SUIHTER, baseline and ensemble forecast (with 1-week horizon) over the entire forecast history (32 weeks from May 2, 2021 to December 11, 2021).

Model	Cases		Deaths	
	relAE	relWIS	relAE	relWIS
EuroCOVIDhub-baseline	1	1	1	1
EuroCOVIDhub-ensemble	0.59	0.55	0.48	0.42
SUIHTER	0.31	0.47	0.29	0.36

Table 2

Relative absolute error and relative weighted interval score for the SUIHTER and ensemble forecast (with 1 to 4-week horizons) over the last 10 weeks (from August 28, 2021 to October 30, 2021).

Model	Horizon	Cases		Deaths	
		relAE	relWIS	relAE	relWIS
EuroCOVIDhub-ensemble	1	0.72	0.69	0.78	0.40
	2	1.02	0.93	0.59	0.32
	3	1.16	1.00	0.52	0.30
	4	1.35	1.15	0.64	0.38
SUIHTER	1	0.62	0.60	0.73	0.36
	2	0.93	0.89	0.49	0.27
	3	1.13	1.21	0.48	0.28
	4	1.21	1.39	0.49	0.31

The 1-week horizon forecast obtained with the SUIHTER model has been uploaded weekly on the European COVID-19 Forecast Hub (European Covid-19 Forecast Hub, 2021) since April 2021. The performance metrics over the entire forecast history (32 weeks from May 2, 2021 to December 11, 2021) are reported in Table 1.

The table reports a comparison between the 1-week forecast obtained with the SUIHTER model, the baseline forecast and the `EuroCOVIDhub-ensemble` forecast obtained using all the forecast upload on the hub for each target. By definition, the baseline forecast has a relWIS and relAE equal to one. As discussed in the introduction, ensemble forecasts are usually expected to be more accurate and robust than those produced by single models. The results presented in Table 1 clearly show that the SUIHTER model is able to produce short-term forecasts which outperforms the ensemble forecasts for both cases and deaths. Moreover, the results reported for Italy on (European Covid-19 Forecast Hub, 2021) show that, according to the relWIS scoring indicator, the SUIHTER model ranked second on cases forecast (behind the MUNI-ARIMA model and second on deaths forecast (behind the LANL-Growth Rate model) among 24 models.

For the 10 forecasts in the time range from August 28, 2021 to November 2, 2021, multiple horizon forecasts (ranging from 1 to 4 weeks) have also been evaluated and compared with the `EuroCOVIDhub-ensemble`. The results are presented in Table 2. As expected both the relative weighted interval score and the relative absolute error increase as longer time horizon are considered. It is worth noticing that the good performances of the SUIHTER model are confirmed for short term predictions of cases forecast (up 2-weeks horizons), where the SUIHTER results are found to be better than both the baseline and ensemble forecasts in terms of absolute error and weighted interval score. Moreover, superior performances of the SUIHTER model for deaths forecast are found even for longer time horizons (up to 4 weeks), where very low values of both relAE and relWIS are obtained.

In Figs. 4 and 5 we report the forecasts with the predicted median value and 95% prediction interval along with the actual data of cases and deaths, respectively. The dates on the x-axis denote the target end date, i.e. the last day of the week for which the forecast has been made. We can note that for the cases forecast the data fall within the prediction intervals (with only 2 exceptions for longer-time forecasts). Concerning deaths forecasts, the model fails to supply a reliable forecast in one case (see Fig. 5 (f)). However, we should point out that the reported data for the week ending on October 23, 2021 were biased by the large number of deaths (28) not previously reported that have been added that week to correct the data.

Remark 1. *The forecasts produced with the extended SUIHTER model (4) and submitted to the European COVID-19 Forecast Hub were obtained considering (i) only the Alpha variant from May 2, 2021 to June 12, 2021, (ii) both the Alpha and Delta variants from June 13, 2021 to August 6, 2021, (iii) only the Delta variant from August 7, 2021 to December 11, 2021.*

4.2. Effect of new variants

During the spreading of a new variant it is fundamental to consider that it is way more transmissible than the prevalent one. For this reason we tried to predict the evolution of the epidemic when the *Delta* variant was spreading in Italy in the time period between July 5, 2021 and August 5, 2021, eventually becoming the prevalent one. We considered a 50% increase in transmissibility ($f_v = 1.5$) for the Delta variant w.r.t. the Alpha variant (Scientific Advisory Group for Emergencies, 2021). We
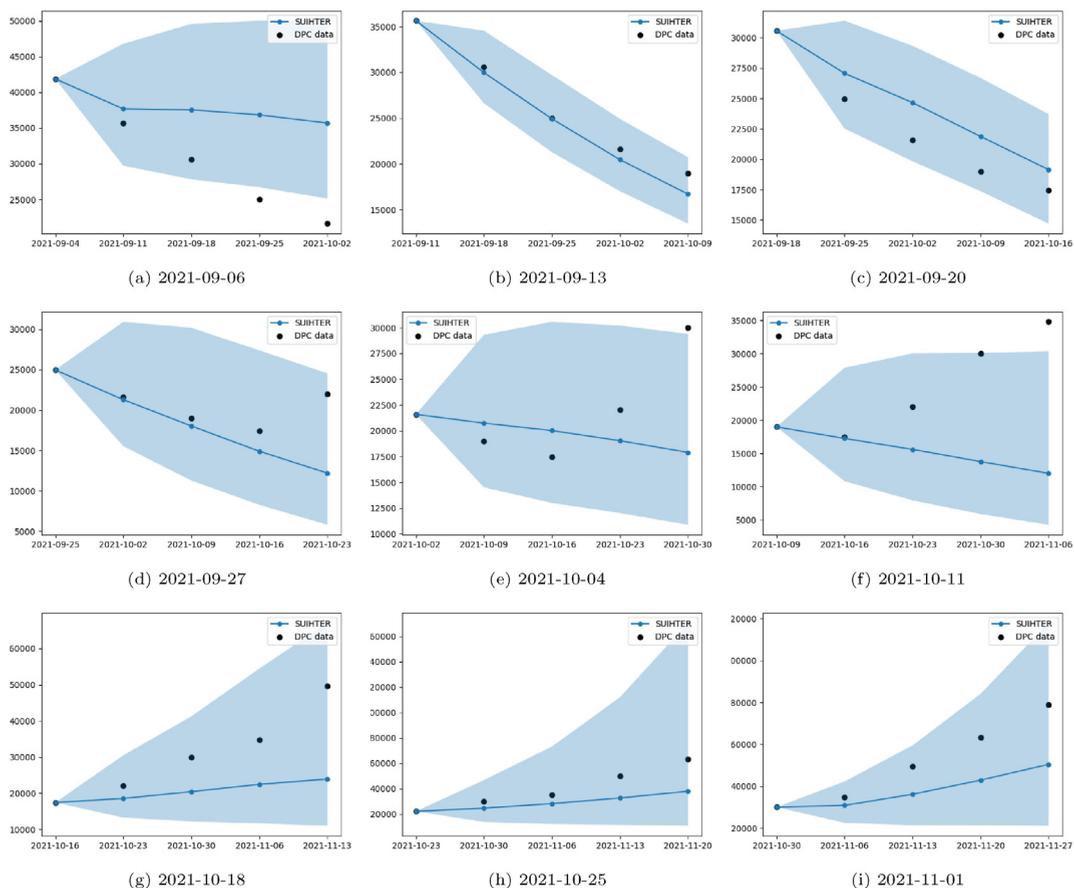
Fig. 4. Weekly new cases forecasts computed with the SUIHTER model on a 4-weeks horizon with 95% confidence interval along with actual data provided by DPC.

performed two simulations, the first one with model (4) described in section 2.1, initialized with the Delta prevalence data provided by the ISS surveillance reports (Istituto Superiore di Sanità, 2021b), namely $p_v = 0.31$. The second simulation was made with $p_v = 0$, thus assuming that no new variant was spreading. In this case the calibrated transmission rate β_U is not rescaled but used as is. The model parameters have been calibrated taking into account data available from February 20, 2021 to July 5, 2021. The two simulations are shown in Fig. 6 where it is possible to appreciate that considering the *Delta* variant the model was able to predict the rise of the epidemic curve occurred in July, which would not be predicted if the role of the variant would have been neglected.

Note that this case features the effect of summer holidays with people gathering in tourism regions and contracting the virus. This was accounted for in the simulations by introducing a summer scenario as detailed in Section 3, resulting in a 42.7% increase on β_U, i.e. such that $\frac{\tau(t)}{\tau(t_0^t)} = 1.427$ in Eq. (13). Should we have neglected the spreading of the *Delta* variant (see Fig. 6), the summer re-opening scenario alone would not be enough to justify the number of cases observed. Similarly, we simulated for the same period, with the same set of parameters, but without accounting for the summer scenario, thus considering only the effect of the *Delta* variant. The result is shown in Fig. 7 alongside with the simulation including both the *Delta* variant and the summer scenario. Also in this case, the *Delta* variant alone could not make the epidemic curve rise as it has been observed, thus implying that the phenomenon witnessed in July was very likely due to a combination of the spreading of the *Delta* variant as well as the growth in the number social contacts due to summer vacations.

4.3. Effects of the vaccination

The evolution of the COVID-19 epidemic in 2021 has been strongly influenced by the introduction of effective vaccines. In the present section we present a set of numerical results to illustrate how the extended SUIHTER model introduced in Section 2 can improve the forecast quality. To show the effect of the presence of vaccinated compartments, we simulated the trend of the epidemic starting from May 22, 2021 and lasting for two months until July 22, 2021. The parameters are obtained by calibrating the SUIHTER model on the DPC data from February 20, 2021 to May 22, 2021 including the vaccinated compartments too. Then, the model has been recalibrated over the same period of time, on the same data, but neglecting the
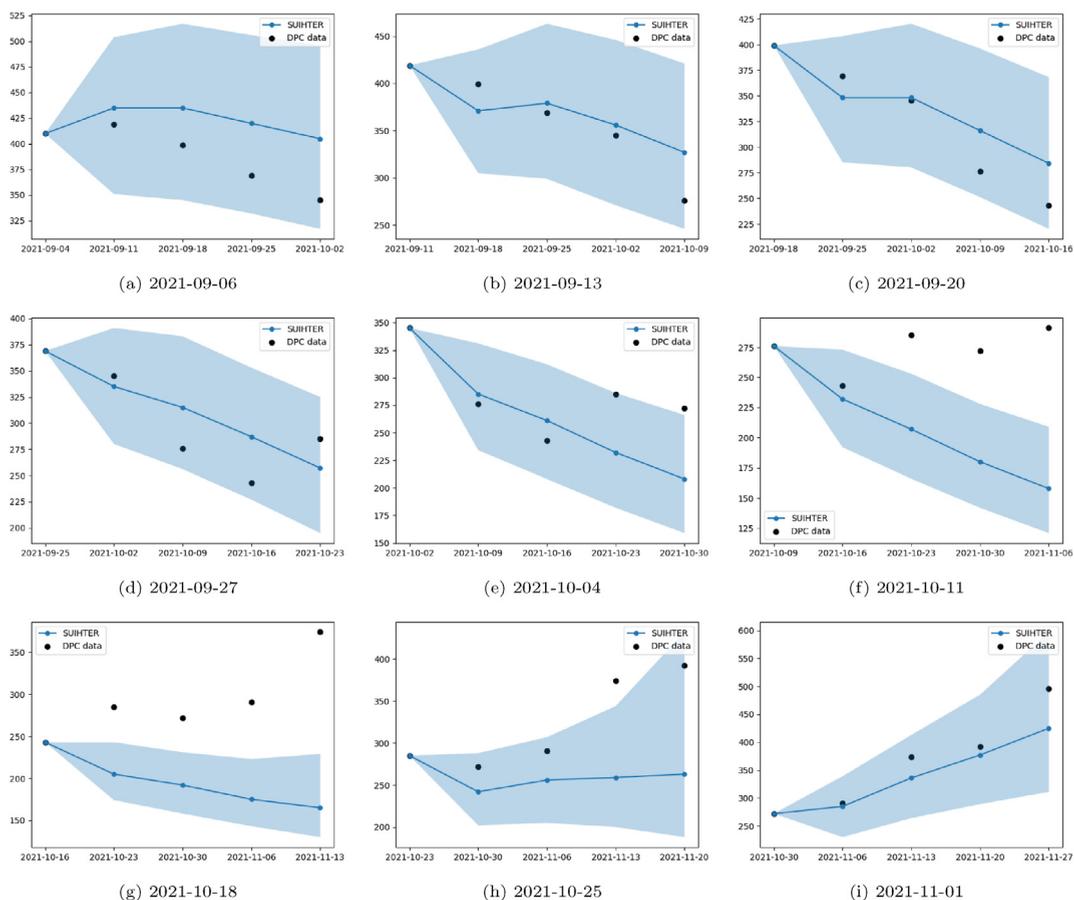
Fig. 5. Deaths forecasts computed with the SUIHTER model on a 4-weeks horizon with 95% confidence interval along with actual data provided by DPC.
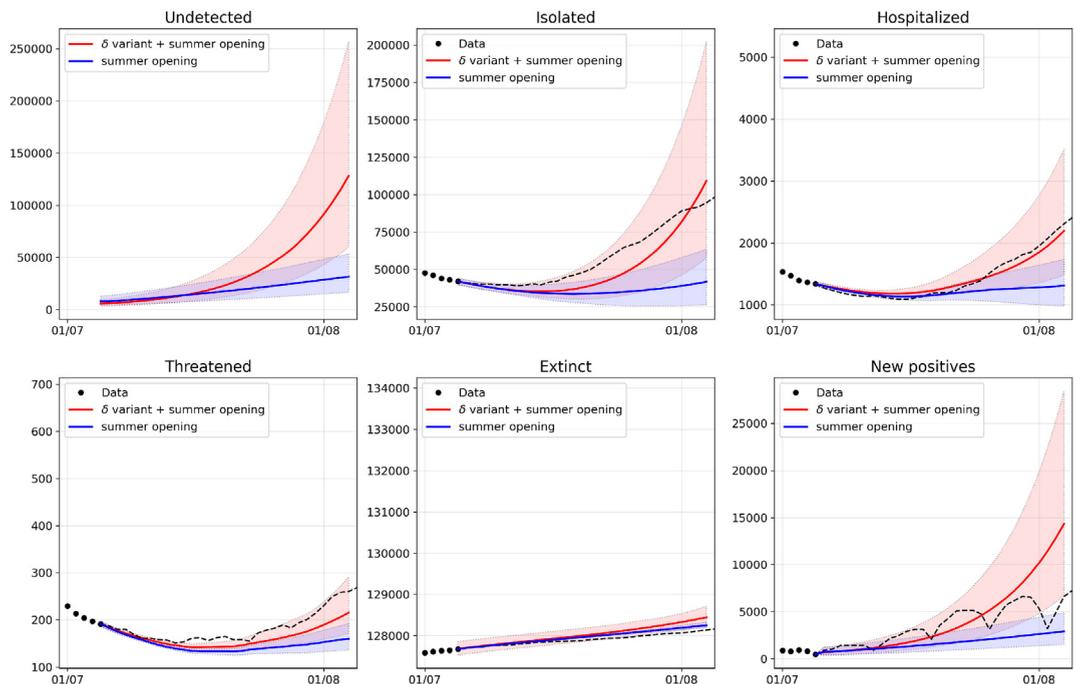

Fig. 6. Simulations performed from 5 July to 5 August with (red) and without (blue) taking into account the spreading of the highly infectious *Delta* variant.
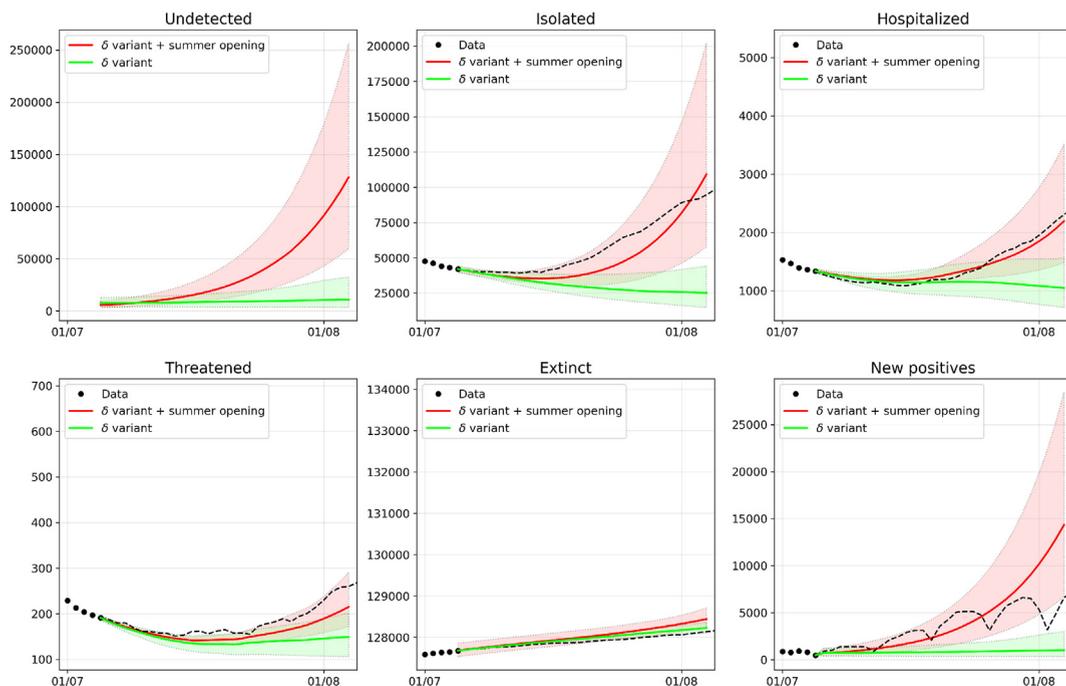
Fig. 7. Simulations performed from 5 July to 5 August with (red) and without (green) taking into account the summer openings resulting in an increased number of social contacts.
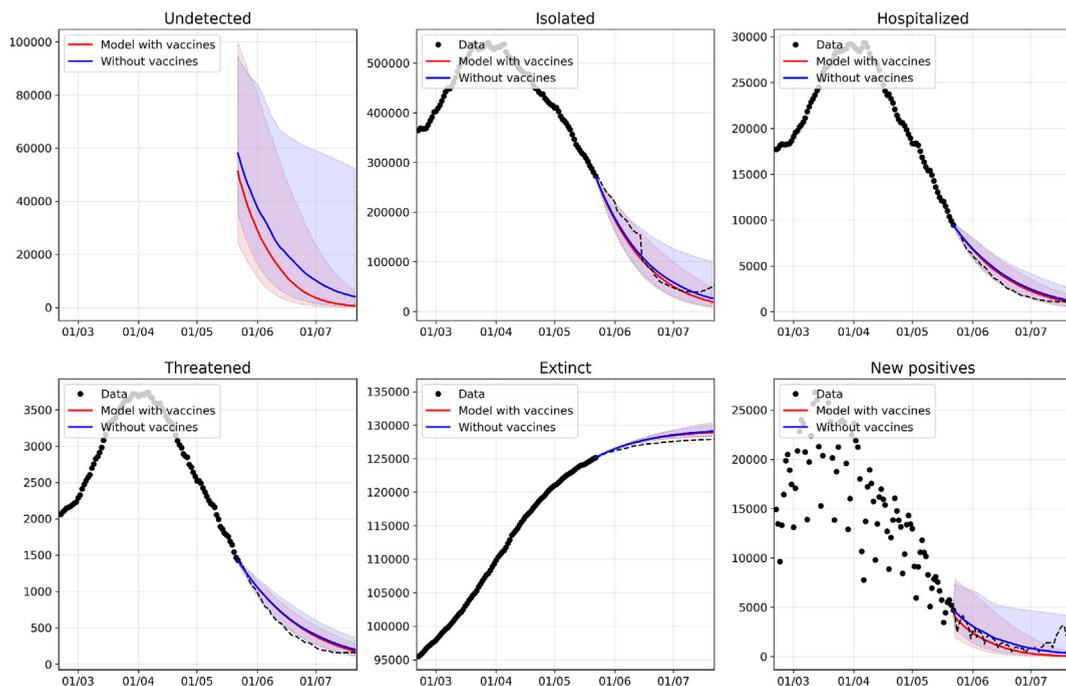

Fig. 8. Comparison between the models with vaccines (red) and without vaccines (blue in a forecast analysis covering the time interval between May 22, 2021 and July 22, 2021.

presence of vaccinated compartments (setting to zero both v_1 and v_2 as well as the initial values of V_1 and V_2). The results are reported in Fig. 8. Even though the two simulations yield almost the same values for most of the compartments, the one without vaccines produces a higher number of deaths. This is due to the fact that it includes the observed effect of vaccines directly in the calibrated transmission rates so obtaining a similar result on positive cases but it does not consider the
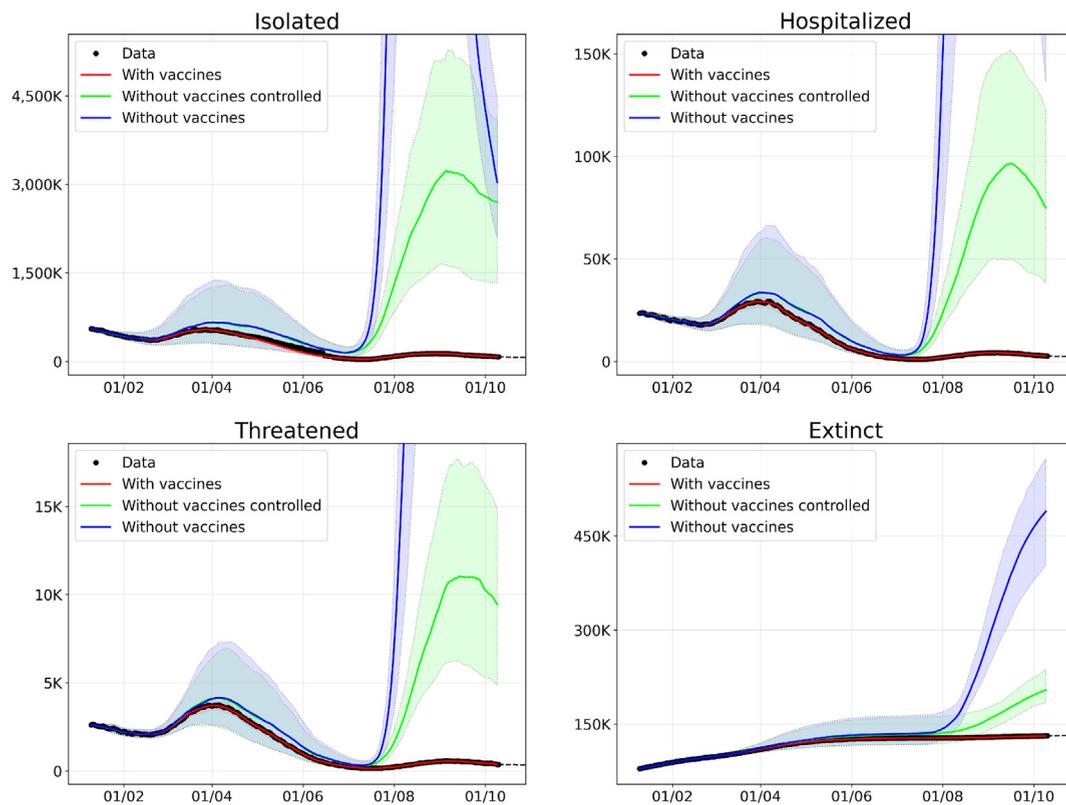
Fig. 9. Simulations of the epidemic during vaccination campaign with what-if scenarios neglecting the presence of vaccinated compartments.

mortality reduction given by vaccines, which is instead considered in model (4). Furthermore, model (4) can give us an estimation of how positives split into unvaccinated, partially vaccinated with one dose or fully vaccinated with second dose.

To show the effect of the vaccines on the model and estimate how the epidemic would have evolved without the vaccines, we compared three simulations obtained with the same set of parameters calibrated from January 10, 2021 to October 10, 2021. The first simulation is obtained with model (1) described in Section 2, used also for calibration, and thus taking into account vaccination compartments and reproduces faithfully the epidemic history. The simulation is then repeated with the same set of parameters removing the vaccinated compartments. In Fig. 9, we see that the epidemic curves have an uncontrolled exponential increase from the beginning of July, corresponding to summer openings and spreading of the *Delta* variant. Obviously this scenario in unrealistic as some containment measures would have been applied before reaching a critical level of new cases and hospital beds occupancy. The third analyzed scenario (in green in Fig. 9) considers the imposition of automatic restrictions when critical incidence levels are reached. Three thresholds are set, based on the guidelines prescribed by the Italian government.[1] The first restrictions are the ones imposed in yellow zones and are applied once the weekly incidence is above 50 cases per 100 000 population. When the incidence reaches 150 cases per 100 000 population, orange zone restrictions are applied and finally red zone restrictions are imposed once the limit of 250 cases per 100 000 population has been exceeded. Even though in this scenario the number of cases has been greatly reduced, the containment measures are not enough to prevent a big epidemic wave in the summer. Analyzing these scenarios we can speculate on how things would have gone without the vaccination campaign and on how many human lives have been saved.

4.4. Scenario analysis: adoption of Green Pass

Since August 6, 2021, people are required to hold the so-called Green Pass for certain activities, a document that certifies to have been vaccinated, having been tested negative in the previous 48 h, or recovered from COVID-19. The Green Pass started to be required for accessing indoor restaurants, long-distance travel, and cultural and leisure activities. Starting September 1, 2021, the Green Pass has been also required to all the personnel working at school (from grade 1 to universities) as well as to university students. More recently, on October 15, 2021, this rule has been extended to all working categories. The objective of

[1] Decreto Legge 23 luglio 2021, n. 105, https://www.gazzettaufficiale.it/eli/id/2021/09/30/21A05687/sg.
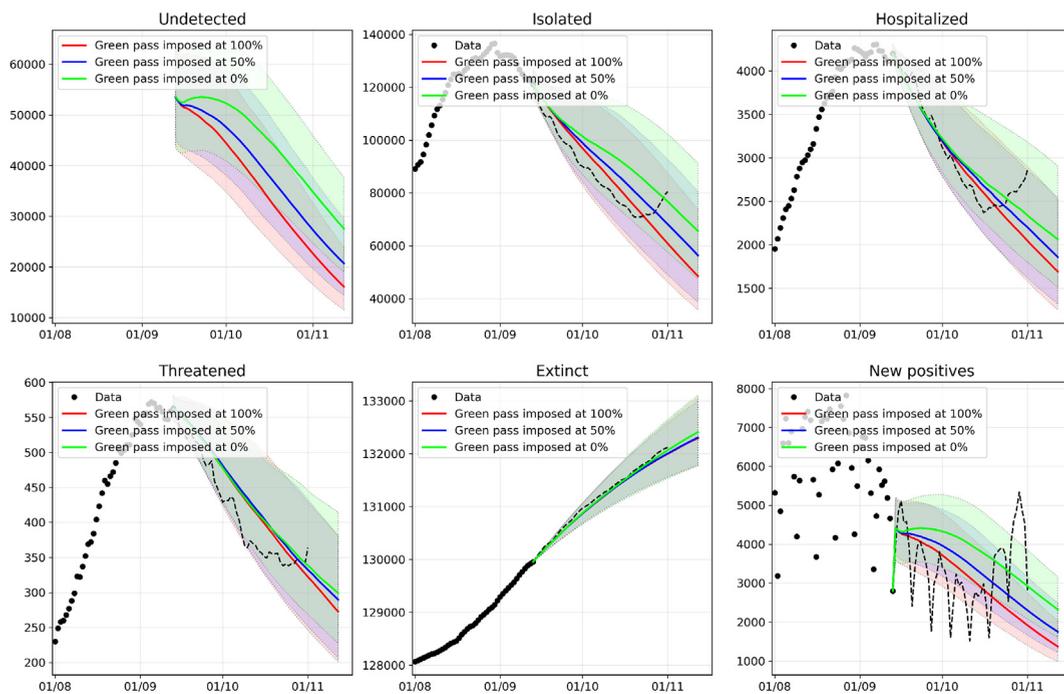
Fig. 10. Scenario analysis comparing the effect of different levels of imposition of Green Pass in schools.

the measure is clearly to control the spread of the epidemic and to encourage a further extension of the vaccination campaign to those people which are not yet vaccinated.

In this context, the SUIHTER model has been applied to estimate the effect of the introduction of the Green Pass at school, with the two-fold objectives:

1. Estimating whether the school re-opening in mid September 2021 would have produced a new epidemic wave (as happened in September 2020);
2. Quantifying at which extent a strict control on the Green Pass certification would have been critical to guarantee a control of the epidemic.

A scenario analysis has been carried out calibrating the SUIHTER model with the data available until September 10, 2021 and performing a two-months forecast until November 10. In order to account for the effect of the adoption of the Green Pass in schools, we computed two different contact matrices for each scenario: one for the vaccinated people (assuming all the people having the Green Pass from vaccination only) and the other for susceptible individuals. From these matrices, we obtained two different restriction coefficients, leading to two β_U rates, one for the compartment S, accounting for unvaccinated people, and thus without Green Pass and one for V_1 and V_2. We considered three different scenarios with increased β_U due to school re-openings, assuming the Green Pass is made compulsory for high schools and university students and staff:

1. the Green Pass is strictly enforced so that unvaccinated individuals are not entering schools and universities;
2. the Green Pass is mildly enforced resulting in 50% of unvaccinated individuals still entering schools and universities;
3. there are no controls on Green Pass so that everyone can enter schools and universities, even though they are not vaccinated.

The results are shown in Fig. 10. In none of the three cases analyzed the increase of β_U is enough to produce a new epidemic wave. However, there is a significant difference on the number of positive cases. This difference may have been important in view of the new epidemic wave that occurred in Italy (as in the rest of Europe during the Fall 2021, since it allowed to start from a lower incidence level the new exponential growth phase.

5. Conclusions

In this paper, we have presented a new epidemiological model derived from the original SUIHTER model introduced in (Parolini et al., 2021a) for the analysis of the COVID-19 epidemic in Italy. The range of applicability of the model has been

extended to cover new aspects of the epidemic that became indispensable to be considered for an accurate description of the phenomenon, namely the emergence of new virus variants and the role of the vaccination campaign.

In particular, the new model accounts for the presence of different virus variants to describe the transition phases in which the wild-type virus (or the variant which is predominant at a given time) is replaced by a new variant characterized by a higher transmission rate. This feature has to be accounted for in order to capture variant outbreaks, such as the one observed in Italy in Summer 2021 when the Delta variant increased its prevalence from 22% to 95% in one month.

The variant model can also be combined with the vaccine model by considering different transmission rate reduction factors (which define to the efficacy of the vaccination in reducing the transmission) for wild type virus or specific variants.

The reduced transmission is not the only way vaccines contributes in the model. The worsening rates also account for the increasing number of vaccinated individuals, by including the consequent reduced probability of developing a serious disease once vaccinated.

The role of the vaccination in controlling the evolution of the epidemic in Italy during the Spring and Summer 2021 has been quantified through a *what-if* scenario analysis. The results indicate that, without the vaccination campaign, the Summer outbreak (mostly due to the Delta variant) would have stroke Italy with an unprecedented force and, even with strict NPIs restrictions, a new epidemic wave, even stronger than the previous ones, would have likely occurred.

The reliability of the results produced by the SUIHTER model, in particular concerning forecast analysis, have also been investigated. The competitive performances of the SUIHTER model in short-term forecasting have been quantitatively verified by comparison with other single model forecasts that joined the European COVID-19 Forecast Hub project (European Covid-19 Forecast Hub, 2021). These results highlight the capability of our model to supply reliable near-future forecasts of cases up to two weeks horizon), while longer-term horizon scenarios (up to four weeks) can be successfully covered when considering forecasts of deaths, thanks to the rich compartmental structure of the SUIHTER model which includes the number of individuals who are hospitalized and hosted in ICUs.

The limitation of the original SUIHTER model, which are shared by several models involving a large number of compartments, were discussed in details in (Parolini et al., 2021a). Most of them, in particular the homogeneous nature of the model which does not include spatial heterogeneity neither age stratification, also apply to the extension presented in the present paper. The range of applicability of the new model is broader since it allows to simulate and forecast the epidemic evolution in the presence of virus variants and vaccines. The proposed approach to deal with emerging virus variants is rather general and may applied to any new variant of concern for which an estimate on the increased transmission rate (relative to previous variants) is available. The vaccination model can be further extended to include the reduction of the vaccine efficacy over time and the additional *booster* dose that should be able to guarantee a longer vaccine immunity (Bar-On et al., 2021).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

A. Determining the transmission coefficient for NPI scenarios

The transmission coefficient τ introduced in Section 3 (Eq. (13)) encapsulates in a single scalar value the effect of different NPIs (distancing measures, restriction on leisure activities, distance learning) or social occurrences and events (school opening and closing, school breaks and holidays) that may be accounted in forecasting the evolution of the epidemics. Since the NPIs are often involving specific age-groups – like schools activities – as well as specific contexts of exposition, we evaluate the transmission coefficient τ based on a multi-age multi-group characterization of the contact matrix. We consider the POLYMOD contact matrices for Italy proposed in (Mossong et al., 2008), counting the average number of contacts that individuals from different age-groups (0–4, 5–9, 10–14, 15–19, 20–24, 25–29, 30–34, 35–39, 40–44, 45–49, 50–54, 55–59, 60–64, 65–69, 70+ for a total of $N_a = 15$ groups) have in the different contexts: home, school, work, transport, leisure and other. We remark that, with respect to the standard table, we use data on the composition of households obtained from ISTAT (for 2019) (ISTAT, 2021), to further split the contacts taking place at home into a *family* context (in which contacts between family members are counted) and a *house* context (in which contacts occurring at home among individuals who are not family members are counted); see (Dede' et al., 2021). This has the advantage to account for NIPs limiting visits of non-family members to other households, a restriction that was often enforced in Italy during 2020 and early 2021. As example, we limit here to report in Table 3 the average number of contacts by age group in the $N_c = 7$ contexts of exposition.

Table 3
Average number of contacts in Italy by age group and social context of exposition.

Ages	Total	Family	Home	School	Work	Transport	Leisure	Other
0−4	16.54	2.30	2.19	5.27	0.00	0.98	3.06	2.75
5−9	20.49	2.27	2.34	8.87	0.00	1.12	4.53	1.37
10−14	27.38	2.21	2.22	11.98	0.05	1.35	5.62	3.80
15−19	29.28	2.05	2.54	13.22	0.05	1.74	6.83	2.87
20−24	22.15	1.49	2.02	1.17	4.49	0.96	7.23	4.80
25−29	21.00	1.04	2.43	2.23	5.21	1.13	6.30	2.66
30−34	18.03	1.26	2.29	0.85	3.92	0.76	5.24	3.72
35−39	21.25	1.75	2.63	0.68	7.78	1.05	3.92	3.45
40−44	22.35	1.63	2.25	2.53	7.00	0.67	4.48	3.79
45−49	19.27	1.50	1.49	2.61	8.24	0.88	1.93	2.64
50−54	22.30	1.38	1.37	5.54	8.05	0.52	2.02	3.41
55−59	18.27	1.11	1.77	1.41	4.60	0.68	3.62	5.06
60−64	18.43	0.91	2.37	1.07	6.05	0.87	3.53	3.63
65−69	12.74	0.71	2.39	0.55	0.48	0.95	3.33	4.33
70+	10.55	0.71	2.53	0.06	1.04	0.22	4.22	1.77

Let us indicate by c_{ij}^k the (average) number of contacts of individuals in age group i with individuals in age group j in the context of exposition k, where $i, j = 1, \ldots, N_a = 15$ and $k = 1, \ldots, N_c = 7$. Then, Table 3 represents $\tilde{c}_i^k = \sum_{j=1}^{N_a} c_{ij}^k$, i.e. the average number of contacts of individuals in age group i with other individuals in the context of exposition k. The transmission coefficient from individuals in the age group i to other individuals in context of exposition k reads

$$\tilde{\tau}_i^k = \tilde{c}_i^k p_i^k,$$

for $i = 1, \ldots, N_a = 15$ and $k = 1, \ldots, N_c = 7$, where p_i^k is the probability that a contact of a infectious individual in age group i is able to expose any other individual in context k. Specifically, we determine p_i^k as

$$p_i^k = r_{a,i} f_i^k \left(r_{c,i}^k, s_i^k \right),$$

where: $r_{a,i}$ is the transmission risk associated to age group i ($r_{a,i} = 0.34$ for $i = 1, 2, 3$, $r_{a,i} = 1$ for $i = 4, \ldots, 13$, and $r_{a,i} = 1.47$ for $i = 1, 2, 3$ (Comitato Tecnico Scientifico, 2021); $r_{c,i}^k$ is the risk of transmission associated to individuals in the age group i within the exposition context k (for example contacts occurring in context $k = 1$ "family" are riskier than those occurring in $k = 4$ "work"); $s_i^k \in [0, 1]$ accounts for NPIs and social occurrences for individuals in age group i and in context k (for example, during school holidays $s_i^3 = 0$ for all $i = 1, \ldots, N_a = 15$, instead if NPIs prescribe students in high schools attending only at 50% in presence, then $s_i^3 = 0.5$); the functions f_i^k combine the coefficients $r_{c,i}^k$ and s_i^k based on age group i and context k. These functions are designed to account for different subcontexts of exposition within a given k, like it occurs for $k = 6$ ("leisure") in restaurants, bars, sports, etc … or for $k = 4$ ("work") for different working environments as healthcare, manufacturing, etc … The function f_i^k is also used to account for the mismatch between age groups and grades in school or other contexts; for example, NIPs regarding schools are prescribed in Italy based on grades, not age. For more details on the definition of the transmission coefficient $\tilde{\tau}_i^k$ we refer the interested reader to (Dede' et al., 2021).

Finally, the transmission coefficient τ used in Eq. (13) is obtained as

$$\tau = \sum_{k=1}^{N_c} \sum_{i=1}^{N_a} \varsigma_i \, \tilde{\tau}_i^k,$$

where ς_i is the percentage of Italian population within the age group i (ISTAT, 2021).

References

Abu-Raddad, L. J., Chemaitelly, H., & Butt, A. A. (2021). Effectiveness of the bnt162b2 covid-19 vaccine against the b.1.1.7 and b.1.351 variants. *New England Journal of Medicine, 385*(2), 187−189. https://doi.org/10.1056/NEJMc2104974. pMID: 33951357. arXiv:.

Bar-On, Y. M., Goldberg, Y., Mandel, M., Bodenheimer, O., Freedman, L., Kalkstein, N., Mizrahi, B., Alroy-Preis, S., Ash, N., Milo, R., & Huppert, A. (2021). Protection of BNT162b2 vaccine booster against Covid-19 in Israel. *New England Journal of Medicine, 385*(15), 1393−1400. https://doi.org/10.1056/NEJMoa2114255. pMID: 34525275. arXiv:.

Bartolucci, F., Pennoni, F., & Mira, A. (2021). A multivariate statistical approach to predict COVID-19 count data with epidemiological interpretation and uncertainty quantification. *Statistics in Medicine, 40*(24), 5351−5372. https://doi.org/10.1002/sim.9129. https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.9129

Bates, J. M., & Granger, C. W. J. (1969). The combination of forecasts. *Journal of the Operational Research Society, 20*(4), 451−468. https://doi.org/10.1057/jors.1969.103

Bertuzzo, E., Mari, L., Pasetto, D., Miccoli, S., Casagrandi, R., Gatto, M., & Rinaldo, A. (2020). The geography of COVID-19 spread in Italy and implications for the relaxation of confinement measures. *Nature Communications, 11*, 4264.

Bonifazi, G., Lista, L., Menasce, D., Mezzetto, M., Pedrini, D., Spighi, R., & Zoccoli, A. (2021). *Study on the effects of the restrictive measures for containment of the COVID-19 pandemic on the reproduction number R_t in Italian regions.* arXiv:2106.02603.

Bracher, J., Ray, E. L., Gneiting, T., & Reich, N. G. (2021a). Evaluating epidemic forecasts in an interval format. *PLoS Computational Biology, 17*, Article e1008618. https://dx.plos.org/10.1371/journal.pcbi.1008618.

Bracher, J., Wolffram, D., Deuschel, J., Görgen, K., Ketterer, A., Ullrich, A., … Schienle, M. (2021b). *Short-term forecasting of COVID-19 in Germany and Poland during the second wave - a preregistered study.* medRxiv. https://doi.org/10.1101/2020.12.24.20248826. https://www.medrxiv.org/content/early/2021/01/11/2020.12.24.20248826

Brazeau, N. F., Verity, R., Jenks, S., Fu, H., Whittaker, C., Winskill, P., Dorigatti, I., Walker, P., Riley, S., Schnekenberg, R. P., Hoeltgebaum, H., Mellan, T. A., Mishra, S., Unwin, H. J. T., Watson, O. J., Cucunubá, Z. M., Baguelin, M., Whittles, L., Bhatt, S., … Okell, L. C. (2020). COVID-19 infection fatality ratio: Estimates from seroprevalence. Imperial College London *Tech. Rep., 34.* https://doi.org/10.25561/83545.

Campbell, F., Archer, B., Laurenson-Schafer, H., Jinnai, Y., Konings, F., Batra, N., & de Waroux, O. L. P. (2021). Increased transmissibility and global spread of SARS-CoV-2 variants of concern as at June 2021. *Euro Surveillance, 26*(24), Article 2100509.

Cohn, B. A., Cirillo, P. M., Murphy, C. C., Krigbaum, N. Y., & Wallace, A. W. (2021). *Breakthrough SARS-CoV-2 infections in 620,000 U.S. Veterans, February 1, 2021 to August 13, 2021.* arXiv:https://www.medrxiv.org/content/early/2021/10/14/2021.10.13.21264966.full.pdf. medRxiv. https://doi.org/10.1101/2021.10.13.21264966 https://www.medrxiv.org/content/early/2021/10/14/2021.10.13.21264966.

Cramer, E. Y., Ray, E. L., Lopez, V. K., Bracher, J., Brennen, A., Rivadeneira, A. J. C., … Reich, N. G. (2021). *Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the US.* medRxiv. https://doi.org/10.1101/2021.02.03.21250974

Comitato Tecnico Scientifico, Presidenza del Consiglio dei Ministri, Dipartimento della Protezione Civile, Italia, Verbale 57 della riunione del 22 Aprile 2020, https://emergenze.protezionecivile.gov.it/it/sanitarie/coron<https://raw.githubusercontent.com/pcm-dpc/COVID-19-Verbali-CTS/master/2020-04/covid-19-cts-verbale-057-20200422.pdf (last accessed December 16, 2021).

Commissario straordinario per l'emergenza Covid-19, 2021. Presidenza del Consiglio dei Ministri, https://github.com/italia/covid19-opendata-vaccini (last accessed October 8, 2021).

Dede', L., Parolini, N., Quarteroni, A., & Villani, G. (2021). *A novel multi-age and multi-group compartmental model for the COVID-19 epidemic in Italy.* Tech. rep., in preparation.

Della Rossa, F., Salzano, D., Di Meglio, A., De Lellis, F., Coraggio, M., Calabrese, C., Guarino, A., Cardona-Rivera, R., De Lellis, P., Liuzza, D., Lo Iudice, F., Russo, G., & di Bernardo, M. (2020). A network model of Italy shows that intermittent regional strategies can alleviate the COVID-19 epidemic. *Nature Communications, 11*(5106).

Dizioli, A., & Radzikowski, A. (2021). A pandemic forecasting framework: An application of risk analysis. *IMF Working Papers, 2021*(226), A001. https://doi.org/10.5089/9781513594842.001.A001. https://www.elibrary.imf.org/view/journals/001/2021/226/article-A001-en.xml

Dati della Sorveglianza integrata COVID-19 in Italia, 2021 in Italia, https://www.epicentro.iss.it/coronavirus/sars-cov-2-dashboard (last accessed December 13, 2021).

P. H. England, COVID-19 vaccine surveillance report Week 27, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachmentTNQDotTNQ/data/file/1000512/VaccineTNQDotTNQ/surveillanceTNQDotTNQ/reportTNQDotTNQ/-TNQDotTNQ/weekTNQDotTNQ/27.pdf (last accessed October 19, 2022).

Farcomeni, A., Maruotti, A., Divino, F., Jona-Lasinio, G., & Lovison, G. (2021). An ensemble approach to short-term forecast of COVID-19 intensive care occupancy in Italian regions. *Biometrical Journal, 63*(3), 503–513. https://doi.org/10.1002/bimj.202000189. https://onlinelibrary.wiley.com/doi/abs/10.1002/bimj.202000189

Funk, S., Abbott, S., Atkins, B. D., Baguelin, M., Baillie, J. K., Birrell, P., … ISARIC4C Investigators. (2020). *Short-term forecasts to inform the response to the Covid-19 epidemic in the UK.* medRxiv. https://doi.org/10.1101/2020.11.11.20220962. https://www.medrxiv.org/content/early/2020/12/04/2020.11.11.20220962

Gatto, M., Bertuzzo, E., Mari, L., Miccoli, S., Carraro, L., Casagrandi, R., & Rinaldo, A. (2020). Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures. *Proceedings of the National Academy of Sciences, 117*(19), 10484–10491. https://doi.org/10.1073/pnas.2004978117

Giordano, G., Blanchini, F., Bruno, R., Colaneri, P., Di Filippo, A., Di Matteo, A., & Colaneri, M. (2020). Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nature Medicine*, 1–6.

Giordano, G., Colaneri, M., Filippo, A. D., Blanchini, F., Bolzern, P., Nicolao, G. D., Sacchi, P., Colaneri, P., & Bruno, R. (2021). Modeling vaccination rollouts, SARS-CoV-2 variants and the requirement for non-pharmaceutical interventions in Italy. *Nature Medicine, 27*, 993–998. https://doi.org/10.1038/s41591-021-01334-5

Goldberg, Y., Mandel, M., Bar-On, Y. M., Bodenheimer, O., Freedman, L., Haas, E. J., Milo, R., Alroy-Preis, S., Ash, N., & Huppert, A. (2021). Waning immunity after the BNT162b2 vaccine in Israel. *New England Journal of Medicine.* https://doi.org/10.1056/NEJMoa2114228

Hodgson, S. H., Mansatta, K., Mallett, G., Harris, V., Emary, K. R. W., & Pollard, A. J. (2021). What defines an efficacious COVID-19 vaccine? A review of the challenges assessing the clinical efficacy of vaccines against SARS-CoV-2. *The Lancet Infectious Diseases, 21*(2), e26–e35. https://doi.org/10.1016/S1473-3099(20)30773-8. https://www.sciencedirect.com/science/article/pii/S1473309920307738

GISAID, 2021. https://www.gisaid.org/(last accessed November 20, 2021).

ISTAT, 2021. Istituto Italiano di Statisica, Banche dati e sistemi informativi, https://www.istat.it/it/dati-analisi-e-prodotti/banche-dati (last accessed December 16, 2021).

European Covid-19 Forecast Hub, 2021. https://covid19forecasthub.eu/(last accessed November 2, 2021)..

Istituto Superiore di Sanità. Bollettino sorveglianza integrata COVID-19 del 14 luglio 2021. https://www.epicentro.iss.it/coronavirus/bollettino/Bollettino-sorveglianza-integrata-COVID-19TNQDotTNQ/14-luglio-2021.pdf

Istituto Superiore di Sanità. (2021b). *Monitoraggio delle varianti del virus SARS-CoV-2 di interesse in sanità pubblica in Italia.* https://www.epicentro.iss.it/coronavirus/sars-cov-2-monitoraggio-varianti-indagini-rapide (last accessed October 16, 2021).

Jeyanathan, M., Afkhami, S., Smaill, F., Miller, M. S., Lichty, B. D., & Xing, Z. (2020). Immunological considerations for COVID-19 vaccine strategies. *Nature Reviews Immunology, 20*(10), 615–632. https://doi.org/10.1038/s41577-020-00434-6

Johansson, M. A., Apfeldorf, K. M., Dobson, S., Devita, J., Buczak, A. L., Baugher, B., … Chretien, J.-P. (2019). An open challenge to advance probabilistic forecasting for dengue epidemics. *Proceedings of the National Academy of Sciences, 116*, 24268–24274. https://doi.org/10.1073/PNAS.1909865116. https://www.pnas.org/content/116/48/24268

Koyama, T., Weeraratne, D., Snowdon, J. L., & Parida, L. (2020). Emergence of drift variants that may affect COVID-19 vaccine development and antibody treatment. *Pathogens, 9*(5). https://doi.org/10.3390/pathogens9050324. https://www.mdpi.com/2076-0817/9/5/324

Krueger, T., Gogolewski, K., Bodych, M., Gambin, A., Giordano, G., Cuschieri, S., Czypionka, T., Perc, M., Petelos, E., Rosińska, M., & Szczurek, E. (2021). *Assessing the risk of COVID-19 epidemic resurgence in relation to the Delta variant and to vaccination passes.* medRxiv. https://doi.org/10.1101/2021.05.07.21256847. https://www.medrxiv.org/content/early/2021/07/18/2021.05.07.21256847

Kuhl, E. (2021). *Computational epidemiology: Data-driven modeling of Covid-19.* Springer Nature.

Leutbecher, M., & Palmer, T. N. (2008). Ensemble forecasting. *Journal of Computational Physics, 227*(7), 3515–3539. https://doi.org/10.1016/j.jcp.2007.02.014. https://www.sciencedirect.com/science/article/pii/S0021999107000812

Loli Piccolomini, E., & Zama, F. (2020). Monitoring Italian COVID-19 spread by a forced SEIRD model. *PLoS One, 15*(8), 1–17. https://doi.org/10.1371/journal.pone.0237417

Marziano, V., Guzzetta, G., Rondinone, B. M., Boccuni, F., Riccardo, F., Bella, A., Poletti, P., Trentini, F., Pezzotti, P., Brusaferro, S., Rezza, G., Iavicoli, S., Ajelli, M., & Merler, S. (2021). Retrospective analysis of the Italian exit strategy from COVID-19 lockdown. arXiv:https://www.pnas.org/content/118/4/e2019617118.full.pdf *Proceedings of the National Academy of Sciences, 118*(4). https://doi.org/10.1073/pnas.2019617118 https://www.pnas.org/content/118/4/e2019617118.

Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., … Edmunds, W. J. (2008). Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Medicine, 5*(3), e74.

Nextstrain, 2021 : Real-time tracking of pathogen evolution, https://nextstrain.org/(last accessed November 20, 2021).

Parolini, N., Dede', L., Antonietti, P. F., Ardenghi, G., Manzoni, A., Miglio, E., Pugliese, A., Verani, M., & Quarteroni, A. (2021a). Suihter: a new mathematical model for COVID-19. application to the analysis of the second epidemic outbreak in Italy. *Proceeding of The Royal Society A, 477*. https://doi.org/10.1098/rspa.2021.0027

Parolini, N., Ardenghi, G., Dede', L., & Quarteroni, A. (2021b). A mathematical dashboard for the analysis of Italian COVID-19 epidemic data. *International Journal for Numerical Methods in Biomedical Engineering, e3513*. https://doi.org/10.1002/cnm.3513

PANGO Lineages, 2021: Latest epidemiological lineages of SARS-CoV-2, https://cov-lineages.org/(last accessed November 20, 2021).

Dipartimento della Protezione Civile, Presidenza del Consiglio dei Ministri, Italia, https://github.com/pcm-dpc/COVID-19 (last accessed January 16, 2021).

Ramos, A. M., Vela-Pérez, M., Ferrández, M. R., Kubik, A. B., & Ivorra, B. (2021). Modeling the impact of SARS-CoV-2 variants and vaccines on the spread of COVID-19. *Communications in Nonlinear Science and Numerical Simulation, 102*, Article 105937. https://doi.org/10.1016/j.cnsns.2021.105937. https://www.sciencedirect.com/science/article/pii/S1007570421002495

Reich, N. G., Brooks, L. C., Fox, S. J., Kandula, S., McGowan, C. J., Moore, E., Osthus, D., Ray, E. L., Tushar, A., Yamana, T. K., Biggerstaff, M., Johansson, M. A., Rosenfeld, R., & Shaman, J. (2019a). A collaborative multiyear, multimodel assessment of seasonal influenza forecasting in the United States. *Proceedings of the National Academy of Sciences, 116*(8), 3146—3154. https://doi.org/10.1073/pnas.1812594116. https://www.pnas.org/content/116/8/3146

Reich, N. G., McGowan, C. J., Yamana, T. K., Tushar, A., Ray, E. L., Osthus, D., Kandula, S., Brooks, L. C., Crawford-Crudell, W., Gibson, G. C., Moore, E., Silva, R., Biggerstaff, M., Johansson, M. A., Rosenfeld, R., & Shaman, J. (2019b). Accuracy of real-time multi-model ensemble forecasts for seasonal influenza in the u.s. *PLoS Computational Biology, 15*, Article e1007486. https://doi.org/10.1371/journal.pcbi.1007486. https://dx.plos.org/10.1371/journal.pcbi.1007486

Scientific Advisory Group for Emergencies. (2021). *Consensus statement on COVID-19*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachmentTNQDotTNQ/data/file/993321/S1267TNQDotTNQ/SPI-M-OTNQDotTNQ/ConsensusTNQDotTNQ/Statement.pdf.

Self, W. H., Tenforde, M. W., Rhoads, J. P., Gaglani, M., Ginde, A. A., Douin, D. J., … IVY Network. (2021). Comparative effectiveness of Moderna, Pfizer-BioNTech, and Janssen (Johnson & Johnson) vaccines in preventing COVID-19 hospitalizations among adults without immunocompromising conditions. *MMWR Morb Mortal Wkly Rep, 70*, 1337—1343. https://doi.org/10.15585/mmwr.mm7038e1external icon. March—August 2021.

Tanaka, H., Hirayama, A., Nagai, H., Shirai, C., Takahashi, Y., Shinomiya, H., … Ogata, T. (2021). Increased transmissibility of the SARS-CoV-2 Alpha variant in a Japanese population. *International Journal of Environmental Research and Public Health, 18*(15), 7752.

Tartof, S. Y., Slezak, J. M., Fischer, H., Hong, V., Ackerson, B. K., Ranasinghe, O. N., Frankland, T. B., Ogun, O. A., Zamparo, J. M., Gray, S., Valluri, S. R., Pan, K., Angulo, F. J., Jodar, L., & McLaughlin, J. M. (2021). Effectiveness of mRNA BNT162b2 COVID-19 vaccine up to 6 months in a large integrated health system in the USA: A retrospective cohort study. *The Lancet, 398*(10309), 1407—1416. https://doi.org/10.1016/S0140-6736(21)02183-8. https://www.sciencedirect.com/science/article/pii/S0140673621021838

Taylor, J. W., & Taylor, K. S. (2021). Combining probabilistic forecasts of COVID-19 mortality in the United States. *European Journal of Operational Research*. https://doi.org/10.1016/j.ejor.2021.06.044. In press https://www.sciencedirect.com/science/article/pii/S0377221721005609.

Viboud, C., Sun, K., Gaffey, R., Ajelli, M., Fumanelli, L., Merler, S., Zhang, Q., Chowell, G., Simonsen, L., & Vespignani, A. (2018). The RAPIDD ebola forecasting challenge: Synthesis and lessons learnt. *Epidemics, 22*, 13—21. https://doi.org/10.1016/j.epidem.2017.08.002. https://www.sciencedirect.com/science/article/pii/S1755436517301275

W. H. Organization, Tracking SARS-CoV-2 variants, https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/(last accessed November 9, 2021).