OXFORD

Subject Section

COVIDHunter: An Accurate, Flexible, and Environment-Aware Open-Source COVID-19 Outbreak Simulation Model

Mohammed Alser*, Jeremie S. Kim, Nour Almadhoun Alserr, Stefan W. Tell, and Onur Mutlu*

ETH Zurich, Zurich 8006, Switzerland

*To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Abstract

Motivation: Early detection and isolation of COVID-19 patients are essential for successful implementation of mitigation strategies and eventually curbing the disease spread. With a limited number of daily COVID-19 tests performed in every country, simulating the COVID-19 spread along with the potential effect of each mitigation strategy currently remains one of the most effective ways in managing the healthcare system and guiding policy-makers. We introduce *COVIDHunter*, a flexible and accurate COVID-19 outbreak simulation model that evaluates the current mitigation measures that are applied to a region and provides suggestions on what strength the upcoming mitigation measure should be. The key idea of COVIDHunter is to quantify the spread of COVID-19 in a geographical region by simulating the average number of new infections caused by an infected person considering the effect of external factors, such as environmental conditions (e.g., climate, temperature, humidity) and mitigation measures.

Results: Using Switzerland as a case study, COVIDHunter estimates that the policy-makers need to keep the current mitigation measures for at least 30 days to prevent demand from quickly exceeding existing hospital capacity. Relaxing the mitigation measures by 50% for 30 days increases both the *daily* capacity need for hospital beds and *daily* number of deaths exponentially by an average of $23.8\times$, who may occupy ICU beds and ventilators for a period of time. Unlike existing models, the COVIDHunter model accurately monitors and predicts the daily number of cases, hospitalizations, and deaths due to COVID-19. Our model is flexible to configure and simple to modify for modeling different scenarios under different environmental conditions and mitigation measures.

Availability: https://github.com/CMU-SAFARI/COVIDHunter

Contact: alserm@ethz.ch, omutlu@ethz.ch

Supplementary information: Supplementary data is available at *Bioinformatics* online.

1 Introduction

Coronavirus disease 2019 (COVID-19) is caused by SARS-CoV-2 virus, which was first detected in Wuhan, the capital city of Hubei Province in China, in early December 2019 (Du Toit, 2020). Since then, it has rapidly spread to nearly every corner of the globe and has been declared a pandemic in March 2020 by the World Health Organization (WHO). As of January 2021, COVID-19 has since resulted in more than 96 million laboratory-confirmed cases around the world, and has killed nearly 2.2% of the infected population. As there are currently no anti-SARS-CoV-2-specific drugs or effective vaccines widely available to everyone, early detection and isolation of COVID-19 patients remain essential for effectively curbing the disease spread. As a result, many countries across the world have implemented unprecedented lockdown and social distancing measures, affecting millions of people. Regardless of the availability and affordability

of COVID-19 testing, it is still extremely challenging to detect and isolate COVID-19 infections at early stages due to three key issues. 1) It is very difficult to accurately identify the initial contraction time of COVID-19 for a patient. This is because COVID-19 patients can develop symptoms between 2 to 14 days (or longer in a few cases) after exposure to the new coronavirus (Lauer *et al.*, 2020; Li *et al.*, 2020). This variable delay is referred to as the virus' *incubation period*. 2) The coronavirus genome can exhibit rapid genetic changes in its nucleotide sequence, which may occur during viral cell replication, within the host body, or during transmission between hosts (Andersen *et al.*, 2020). This genetic diversity affects the virus virulence, infectivity, transmissibility, and evasion of the host immune responses (Phan, 2020; Pachetti *et al.*, 2020; Toyoshima *et al.*, 2020). 3) The situation becomes even worse as the coronavirus can survive and therefore remain *infectious* outside the host, on common surfaces such as metal, glass, and banknotes (both paper and polymer) at room temperature for *up to 28 days* (Kampf *et al.*, 2020; Riddell *et al.*, 2020).

Simulating the spread of COVID-19 has the potential to mitigate the effects of the three key issues, help to better manage the healthcare system, and provide guidance to policy-makers on the effectiveness of various (current, planned or discussed) social distancing and mitigation measures. To this end, many COVID-19 simulation models are proposed (e.g., (Tradigo *et al.*, 2020; Russell *et al.*, 2020; Ashcroft *et al.*, 2020)), some of which are announced to *assist* in decision-making for policy-makers in countries such as the United Kingdom (ICL (Flaxman *et al.*, 2020)), United States (IHME (Reiner *et al.*, 2020)), and Switzerland (IBZ (Huisman *et al.*, 2020)). These models tend to follow one of two key approaches. (1) Evaluating the current actual epidemiological situation by accounting for reporting delays and under-reporting due to inefficiencies such as low number of COVID-19 tests. (2) Evaluating the current and future epidemiological situation by simulating the COVID-19 outbreak *without* relying on the observed (laboratory-confirmed) number of cases in simulation.

The first approach, taken by the IBZ (Huisman *et al.*, 2020), LSHTM (Russell *et al.*, 2020), and (Ashcroft *et al.*, 2020) models, is *not* mainly used for prediction purposes as it reflects the epidemiological situation with about two weeks of time delay (due to its dependence on observed COVID-19 reports). The IBZ model (Huisman *et al.*, 2020) estimates the daily reproduction number, R, of SARS-CoV-2 from observed COVID-19 incidence time series data after accounting for reporting delays and under-reporting using the numbers of confirmed hospitalizations and deaths. The R number describes how a pathogen spreads in a particular population by quantifying the average number of new infections caused by each infected person at a given point in time. The LSHTM model (Russell *et al.*, 2020) adjusts the daily number of observed COVID-19 cases by accounting for under-reporting (uncertainty) using both deaths-to-cases ratio estimates and correcting for delays between case confirmation (i.e., laboratory-confirmed infection) to death.

The second approach, taken by ICL (Flaxman *et al.*, 2020) and IHME (Reiner *et al.*, 2020) models, usually requires a large number of various input parameters and assumptions. IHME (Reiner *et al.*, 2020) model requires input parameters such as testing rates, mobility, social distancing policies, population density, altitude, smoking rates, self-reported contacts, and mask use. This model makes two key assumptions: 1) the infection fatality rate (IFR), which indicates the rate of people that die from the infection is taken using data from the Diamond Princess Cruise ship and New Zealand and 2) the decreasing fatality rate is reflective of increased testing rates (identifying higher rates of asymptomatic cases). ICL (Flaxman *et al.*, 2020) model requires input parameters such as the daily number of confirmed deaths, IFR, mobility rates from Google, age- and country-specific data on demographics, patterns of social contact, and hospital availability. This model makes three key assumptions: 1) age-specific IFRs observed in China and Europe are the same across every country, 2) the number of confirmed deaths is equal to the true number of COVID-19 deaths, and 3) the change in transmission rates is a function of average mobility trends.

To our knowledge, there is currently no model capable of accurately monitoring the current epidemiological situation and predicting future scenarios while considering a reasonably low number of parameters and accounting for the effects of environmental conditions, as we summarize in Table 1. The low number of parameters provides four key advantages: 1) allowing flexible (easy-to-adjust) configuration of the model input parameters for different scenarios and different geographical regions, 2) enabling short simulation execution time and simpler modeling, 3) enabling easy validation/correction of the model prediction outcomes by adjusting fewer variables, and 4) being extremely useful and powerful especially during the early stages of a pandemic as many of the parameters are unknown. Simulation models need to consider the fact that the environmental conditions (e.g., air temperature) affect pathogen infectivity (Fares, 2013; Kampf *et al.*, 2020; Riddell *et al.*, 2020; Xu *et al.*, 2020) and simulating this effect helps to provide accurate estimation of the epidemiological situation.

Our **goal** in this work is to develop such a COVID-19 outbreak simulation model. To this end, we introduce *COVIDHunter*, a simulation model that evaluates the current mitigation measures (i.e., non-pharmaceutical intervention or NPI) that are applied to a region and provides insight into what strength the upcoming mitigation measure

should be and for how long it should be applied, while considering the potential effect of environmental conditions. Our model accurately forecasts the numbers of infected and hospitalized patients, and deaths for a given day, as validated on historical COVID-19 data (after accounting for under-reporting). The **key idea** of COVIDHunter is to quantify the spread of COVID-19 in a geographical region by calculating the daily reproduction number, R, of COVID-19 and scaling the reproduction number based on changes in both mitigation measures and environmental conditions. The R number changes during the course of the pandemic due to the change in the ability of a pathogen to establish an infection during a season and mitigation measures that lead to lower number of of susceptible individuals. COVIDHunter simulates the entire population of a region and assigns *each* individual in the population to a stage of the COVID-19 infection (e.g., from being healthy to being short-term immune to COVID-19) based on the scaled R number. Our model is flexible to configure and simple to modify for modeling different scenarios as it uses *only* three input parameters, two of which are time-varying parameters, to calculate the R number. Whenever applicable, we compare the simulation output of our model to that of four state-of-the-art models currently used to inform policy-makers, IBZ (Huisman *et al.*, 2020), LSHTM (Russell *et al.*, 2020), ICL (Flaxman *et al.*, 2020), and IHME (Reiner *et al.*, 2020).

The **contributions** of this paper are as follows:

- We introduce COVIDHunter, a flexible and validated simulation model that evaluates the current and future epidemiological situation by simulating the COVID-19 outbreak. COVIDHunter accurately forecasts for a given day 1) the reproduction number, 2) the number of infected people, 3) the number of hospitalized people, 4) the number of deaths, and 5) number of individuals at each stage of the COVID-19 infection. COVIDHunter evaluates the effect of different current and future mitigation measures on the COVIDHunter's five numbers.

- As a case study, we statistically analyze the relationship between temperature and number of COVID-19 cases in Switzerland. We find that for each $1°C$ rise in daytime temperature, there is a 3.67% decrease in the daily number of confirmed cases. We demonstrate how *considering* the effect of climate (e.g., daytime temperature) on COVID-19 spread significantly improves the prediction accuracy.

- Compared to IBZ, LSHTM, ICL, and IHME models, COVIDHunter achieves more accurate estimation, provides no prediction delay, and provides ease of use and high flexibility due to the simple modeling approach that uses a small number of parameters.

- Using COVIDHunter, we demonstrate that the spread of COVID-19 in Switzerland is still active (i.e., $R > 1.0$) and curbing this spread requires maintaining the same strength of the currently applied mitigation measures for at least another 30 days.

- We release the well-documented source code of COVIDHunter and show how easy it is to flexibly configure for *any* scenario and extend for different measures and conditions than we account for.

2 Methods

2.1 Overview

The primary purpose of our COVIDHunter model is to *monitor and predict* the spread of COVID-19 in a flexibly-configurable and easy-to-use way, while accounting for changes in mitigation measures and environmental conditions over time. We employ a three-stage approach to develop and deploy this model. (1) The COVIDHunter model predicts the daily R value based on *only* three input parameters to maintain both quick simulation and high flexibility in configuring these parameters. Each input parameter is configured based on either existing research findings or user-defined values. Our model allows for *directly* leveraging existing models that study the effect of *only* mitigation measures (or *only* environmental conditions) on the spread of COVID-19, as we show in Section 2.2. (2) The COVIDHunter model predicts the number of COVID-19 cases based on the predicted R number. COVIDHunter simulates the entire population of a region and labels each individual according to different stages of the COVID-19 infection timeline. Each stage has a different degree of infectiousness and contagiousness. The model simulates these stages for each individual to maintain accurate predictions. (3) The COVIDHunter model predicts the number of hospitalizations and deaths based on both

Table 1. Comparison to other models used to inform government policymakers, as of January 2021.

Model	Open Source	Well- Documented[#]	Accounting for Weather Changes	Low Number of Parameters	Reported COVID-19 Statistics
COVIDHunter (this work)	✓	✓	✓	✓	✓ (R, cases, hospitalizations, and deaths)
IBZ (Huisman *et al.*, 2020)	✓	✗	✗	✓	✗ (only R)
LSHTM (Russell *et al.*, 2020)	✓	✗	✗	✓	✗ (only cases)
ICL (Flaxman *et al.*, 2020)	✓	✓	✗	✗	✓ (R, cases, hospitalizations, and deaths)
IHME (Reiner *et al.*, 2020)	✓[*]	✗	✗	✗	✗ (cases, hospitalizations, and deaths)

[#] Based on each model's GitHub page (all models are available on GitHub).   [*] The available packages are configured *only* for the IHME infrastructure.

the predicted number of cases and the R number. Next, we explain the COVIDHunter model in detail.

2.2 How does the COVIDHunter Model Work?

The COVIDHunter model predicts the dynamic value of R for a population at a given day while considering three key factors: 1) the transmissibility of an infection into a susceptible host population, 2) mitigation measures (e.g., lockdown, social distancing, and isolating infected people), and 3) environmental conditions (e.g., air temperature). Our model calculates the time-varying R number using Equation 1 as follows:

$$R(t) = R0 * (1 - M(t)) * C_e(t) \qquad (1)$$

The R number for a given day, t, is calculated by multiplying three terms: 1) the base reproduction number ($R0$) for the subject virus, 2) one minus the mitigation coefficient (M), for the given day t and 3) the environmental coefficient (C_e) for the given day t.

The $R0$ number quantifies the transmissibility of an infection into a susceptible host population by calculating the expected average number of new infections caused by an infected person in a population with no prior immunity to a specific virus (as a pandemic virus is by definition novel to all populations). Hence, the $R0$ number represents the transmissibility of an infection at only the beginning of the outbreak assuming the population is not protected via vaccination. Unlike the R number, $R0$ number is a fixed value and it does not depend on time. The R number is a time-dependent variable that accounts for the population's reduced susceptibility. The $R0$ number for the COVID-19 virus can be obtained from several existing studies (such as in (Hilton and Keeling, 2020; Chang *et al.*, 2020; Shi *et al.*, 2020; de Souza *et al.*, 2020; Rahman *et al.*, 2020)) that estimate it by modeling contact patterns during the first wave of the pandemic.

The mitigation coefficient (M) applied to the population is a time-dependent variable and it has a value between 0 and 1, where 1 represents the strongest mitigation measure and 0 represents no mitigation measure applied. In different countries, mitigation measures take different forms, such as social distancing, self-isolation, school closure, banning public events, and complete lockdown. These measures exhibit significant heterogeneity and differ in timing and intensity across countries (Hale *et al.*, 2020; Davies *et al.*, 2020). Quantifying the mitigation measures on a scale from 0 to 1 across different countries is challenging. *The Oxford Stringency Index* (Hale *et al.*, 2020) maintains a twice-weekly-updated index that takes values from 0 to 100, representing the severity of nine mitigation measures that are applied by more than 160 countries. Another study (Brauner *et al.*, 2020) estimates the effect of *only* seven mitigation measures on the R number in 41 countries. We can *directly* leverage such studies for calculating the mitigation coefficient on a given day after changing the scale from 0:100 to 0:1 by dividing each value of, for example, the Oxford Stringency Index by 100.

The environmental coefficient (C_e) is a time-dependent variable representing the effect of external environmental factors on the spread of COVID-19 and it has a value between 0 and 2. Several related viral infections, such as the Influenza virus, human coronavirus, and human respiratory, already show notable seasonality (showing peak incidences during *only* the winter (or summer) months) (Moriyama *et al.*, 2020; Fisman, 2012). The seasonal changes in temperature, humidity, and ultraviolet light affect the pathogen infectiousness outside the host (Fares, 2013; Kampf *et al.*, 2020; Riddell *et al.*, 2020; Xu *et al.*, 2020). However, the indoor environmental conditions are usually well-controlled throughout the year, where human behavior and number of households can be the major contributor to the spread of the COVID-19 (Moriyama

et al., 2020). There are currently several studies that demonstrate the strong dependence of the transmission of SARS-CoV-2 virus on one or more environmental conditions, *even after* controlling (isolating) the impact of mitigation measures and behavioral changes that reduce contacts. Several studies have demonstrated increased infectiousness by a country-dependent fixed-rate with each $1°C$ fall in daytime temperature (Xie and Zhu, 2020; Prata *et al.*, 2020). Another study supports the same temperature-infectiousness relationship, but it also finds that before applying any mitigation measures, a one degree drop in relative humidity shows increased infectiousness by a rate lower ($2.94\times$ less) than that of temperature (Wang *et al.*, 2020).

One of the most comprehensive studies that spans more than 3700 locations around the world is *HARVARD CRW* (Xu *et al.*, 2020). It finds the statistical correlation between the relative changes in the R number and both weather (temperature, ultraviolet index, humidity, air pressure, and precipitation) and air pollution (SO2 and Ozone) after controlling the impact of mitigation measures. The study provides a *CRW Index* that has a value from 0.5 to 1.5. The percentage difference between any two consecutive values provided by the CRW Index represents the effect that both weather and air pollutants have on the R number. For example, a drop in the CRW Index by 10% in a given location points to a 10% reduction in the R number due to weather changes and air pollutants. Our model enables applying *any of these studies* by adjusting our environmental coefficient on a given day, as we experimentally demonstrate in Section 3. For example, if the COVIDHunter user chooses to consider the HARVARD CRW study, and the CRW Index shows, for example, a 10% drop compared to its immediately preceding data point, then the environmental coefficient of COVIDHunter should be 0.9 so that the R value decreases by also 10%. Next, we explain how our model forecasts the number of COVID-19 cases based on Equation 1.

2.3 Predicting the Number of COVID-19 Cases

COVIDHunter tracks the number of infected and uninfected persons over time by clustering the population into four main categories: HEALTHY, INFECTED, CONTAGIOUS, and IMMUNE. The model initially considers the entire population as uninfected (i.e., HEALTHY). For each simulated day, the model calculates the R value using Equation 1 and decides how many persons can be infected during that day. The day when the first case of infection in a population introduced is defined by the user. For each newly infected person (INFECTED), the model maintains a counter that counts the number of days from being infected to being contagious (CONTAGIOUS). Several COVID-19 case studies show that *presymptomatic* transmission can occur 1–3 days before symptom onset (Wei *et al.*, 2020; Slifka and Gao, 2020). COVID-19 patients can develop symptoms mostly after an incubation period of 1 to 14 days (the median incubation period is estimated to be 4.5 to 5.8 days) (Lauer *et al.*, 2020; Li *et al.*, 2020). We calculate the number of days of being contagious after being infected as a random number with a Gaussian distribution that has user-defined lowest and highest values. Each contagious person may infect N other persons depending on mobility, population density, number of households, and several other factors (Ferguson *et al.*, 2020). We calculate the value of N to be a random number with a Gaussian distribution that has the lowest value of 0 and the highest value determined by the user. If N is greater than the R number (i.e., the target number of infections for that day has been reached), further infections are curtailed preventing overestimation of N by infecting *only* R persons. Once the contagious person infects the desired number of susceptible persons, the status of the contagious person becomes immune (IMMUNE). The immune

status indicates that the person has immunity to reinfection due to either vaccination or being recently infected (Lumley *et al.*, 2020).

Our model also simulates the effect of infected travelers (e.g., daily cross-border commuters within the European Union) on the value of R. These travelers can initiate the infection(s) at the beginning of the pandemic. If such infected travelers are absent (due to, for example, emergency lockdown) from the target population, the virus would die out once the value of R decreases below 1 for a sufficient period of time. Both the number and percentage of infected travelers entering a region are configurable in our model. The percentage of incoming infected travelers is *not* affected by the changes in the local mitigation measures, as these travelers were infected abroad.

Our model predicts the *daily* number of COVID-19 cases for a given day t, as follows:

$$Daily\_Cases(t) = \sum_{n=0}^{T_{INF}(t)} N(n) + \sum_{m=0}^{U_{CON}(t)} N(m) \quad (2)$$

where T_{INF} is the daily number of infected travelers that is a user-defined variable, $N()$ is a function that calculates the number of persons to be infected by a given person as a random number with a Gaussian distribution, and U_{CON} is the daily number of contagious persons calculated by our model.

2.4 Predicting the Number of COVID-19 Hospitalizations and Deaths

There are currently two key approaches for calculating the estimated number of both hospitalizations and deaths due to COVID-19: 1) using historical statistical probabilities, each of which is unique to each age group in a population (Bhatia and Klausner, 2020; Bi *et al.*, 2020) and 2) using historical COVID-19 hospitalizations-to-cases and deaths-to-cases ratios (Kobayashi *et al.*, 2020). We choose to follow a modified version of the second approach as it does *not* require 1) clustering the population into age-groups and 2) calculating the risk of each individual using the given probability, which both affect the complexity of the model and the simulation time.

The number of COVID-19 hospitalizations for a given day, t, can be calculated as follows:

$$Daily\_Hospitalizations(t) = Daily\_Cases(t) * X * C_X \quad (3)$$

where $Daily\_Cases(t)$ is calculated using Equation 2 and X is the hospitalizations-to-cases ratio that is calculated as the average of daily ratios of the number of COVID-19 hospitalizations to the laboratory-confirmed number of COVID-19 cases. As the *true* number of cases is unknown due to lack of population-scale testing, it is extremely difficult to make accurate estimates of the *true* number of COVID-19 hospitalizations. As such, we assume a fixed multiplicative relationship between the number of laboratory-confirmed cases and the *true* number of cases. We use the user-defined correction coefficient, C_X, of the hospitalizations-to-cases ratio to account for such a multiplicative relationship.

The number of COVID-19 deaths for a given day t can be calculated as follows:

$$Daily\_Deaths(t) = Daily\_Cases(t) * Y * C_Y \quad (4)$$

where $Daily\_Cases(t)$ is calculated using Equation 2 and Y is the deaths-to-cases ratio, which is calculated as the average of daily ratios of the number of COVID-19 deaths to the number of COVID-19 laboratory-confirmed cases. The observed number of COVID-19 deaths can still be less than the *true* number of COVID-19 deaths due to, for example, under-reporting. We use the user-defined correction coefficient, C_Y, to account for the under-reporting. One way to find the *true* number of COVID-19 deaths is to calculate the number of excess deaths. The number of excess deaths is the difference between the observed number of deaths during time period and expected (based on historical data) number of deaths during the same time period. For this reason, C_Y may not necessarily be equal to C_X.

2.5 Model Validation

We can validate our model using two key approaches. 1) Comparing the daily R number predicted by our model (using Equation 1) with the daily reported official R number for the same region. 2) Comparing the daily number of COVID-19 cases predicted by our model (using Equation 2) with the daily number of laboratory-confirmed COVID-19 cases. As of January 2021, we have already witnessed one year of the pandemic, which provides us several observations and lessons. The most obvious source of uncertainty, affecting *all* models, is that the *true* number of persons that are previously infected or currently infected is *unknown* (Wilke and Bergstrom, 2020). This affects the accuracy of the reported R number since it is calculated as, for example, the ratio of the number of cases for a week (7-day rolling average) to the number of cases for the preceding week. Adjusting the parameters of our model to fit the curve of the number of confirmed cases is likely to be highly uncertain. The publicly-available number of COVID-19 hospitalizations and deaths can provide more reliable data.

For these reasons, we decide to use a combination of reported numbers of cases, hospitalizations, and deaths for validating our model using three key steps. 1) We leverage the more reliable data of reported number of hospitalizations (or deaths) to estimate the *true* number of COVID-19 cases using the ratio of number of laboratory-confirmed hospitalizations (or deaths) to the number of laboratory-confirmed cases during the second wave of the COVID-19 pandemic. We assume that the COVID-19 statistics during the second wave is more accurate than that during the first wave because generally more testing is performed in the second wave. 2) We consider a multiplicative relationship between the *true* number of COVID-19 cases and that estimated in step 1. In our experimental evaluation (Section 3), we use the *true* number of COVID-19 cases calculated using different multiplicative factor values (we refer to them as *certainty rate levels*) as a ground-truth for validating our model. A certainty rate of, for example, 50% means that the *true* number of COVID-19 cases is actually *double* that calculated in step 1. 3) We use our model to calculate both the daily R number (using Equation 1) and the number of COVID-19 cases (using Equation 2). We fix the two terms of Equation 1, $R0$ and C_e, using publicly-available data for a given region and change the third term, M, until we fit the curve of the number of cases predicted by our model to the ground-truth plot calculated in step 2. We use the same methodology to validate our predicted numbers of hospitalizations and deaths with different certainty rate levels as we show in Section 3 and the Supplementary Excel File[1].

2.6 Flexibility and Extensibility of the COVIDHunter Model

We especially build COVIDHunter model to be flexible to configure and easy to extend for representing any existing or future scenario using different values of the three terms of Equation 1, 1) $R0$, 2) $M(t)$, 3) $C_e(t)$, in addition to several other parameters such as the population, number of travelers, percentage of expected infected travelers to the total number of travelers, and hospitalizations- or deaths-to-cases ratios. Our modeling approach acts across the overall population without assuming any specific age structure for transmission dynamics. It is still *possible* to consider each age group separately using individual runs of COVIDHunter model simulation, each of which has its own parameter values adjusted for the target age group. The COVIDHunter model considers each location independently of other locations, but it also accounts for potential movement between locations by adjusting the corresponding parameters for travelers. By allowing most of the parameters to vary in time, t, the COVIDHunter model is capable of accounting for any change in transmission intensity due to changes in environmental conditions and mitigation measures over time. As we explain in Section 2.2, the flexibility of configuring the environmental coefficient and mitigation coefficient allows our proposed model to control for location-specific differences in population density, cultural practices, age distribution, and time-variant mitigation responses in each location. Our modeling approach considers a single strain of the COVID-19 virus by using a single base reproduction

---

[1] https://github.com/CMU-SAFARI/COVIDHunter/blob/main/Evaluation_Results/SimulationResultsForSwitzerland.xlsx

number, $R0$. It is possible to consider multiple virus strains by running the model simulation multiple times, each of which considers one of the strains individually. The model can be extended to consider multiple virus strains by replacing the $R0$ number by multiple $R0$ numbers that represent the different strains (Reichmuth *et al.*, 2021).

3 Results

We evaluate the daily 1) R number, 2) mitigation measures, and 3) numbers of COVID-19 cases, hospitalizations, deaths. We also evaluate the daily numbers of `HEALTHY`, `INFECTED`, `CONTAGIOUS`, and `IMMUNE` in the Supplementary Excel File[2]. We compare the predicted values to their corresponding observed values whenever possible. We provide a comprehensive treatment of *all* datasets, models, and evaluation results with different model configurations in the Supplementary Materials and the Supplementary Excel Files[3].

3.1 Determining the Value of Each Variable in the Equations

We use Switzerland as a use-case for all the experiments. However, our model is not limited to any specific region as the parameters it uses are completely configurable. To predict the R number, we use Equation 1 that requires three key variables. We set the base reproduction number, $R0$, for the SARS-CoV-2 in Switzerland as 2.7, as shown in (Hilton and Keeling, 2020). We choose two main approaches for setting the value of the time-varying environmental coefficient variable (C_e). 1) Performing statistical analysis for the relationship between the daily number of COVID-19 cases and average daytime temperature in Switzerland. As we provide in the Supplementary Materials, Section 1, our statistical analysis shows that each $1°C$ rise in daytime temperature is associated with a 3.67% (t-value = -3.244 and p-value = 0.0013) decrease in the daily number of confirmed COVID-19 cases. We refer to this approach as Cases-Temperature Coefficient (CTC). 2) Applying the *HARVARD CRW* (Xu *et al.*, 2020) (CRW in short), which provides the statistical relationship between the relative changes in the R number and both weather factors and air pollutants after controlling for the impact of mitigation measures. We change the daily mitigation coefficient, $M(t)$, value based on the ratio of number of confirmed hospitalizations to the number of confirmed cases with two certainty rate levels of 100% and 50%, as we explain in detail in Section 2.5. This helps us to take into account uncertainty in the observed number of COVID-19 cases, hospitalizations, and deaths. We set the minimum and maximum incubation time for SARS-CoV-2 as 1 and 5 days, respectively, as 5-day period represents the median incubation period worldwide (Lauer *et al.*, 2020; Li *et al.*, 2020). We set the population to 8654622. We empirically choose the values of N, the number of travelers, and the ratio of the number of infected travelers to the total number of travelers to be 25, 100, and 15%, respectively.

3.2 Evaluating the Expected Number of COVID-19 Cases for Model Validation

As the exact true number of COVID-19 cases remains unknown (due to, for example, lack of population-scale COVID-19 testing), we expect the true number of COVID-19 cases in Switzerland to be higher than the observed (laboratory-confirmed) number of cases. We calculate the expected true number of cases based on both numbers of deaths and hospitalizations, as we explain in Section 2.5. To account for possible missing number of COVID-19 deaths, we consider the excess deaths instead of observed deaths. We calculate the excess deaths as the difference between the observed weekly number of deaths in 2020 and 5-year average of weekly deaths. We find that X (hospitalizations-to-cases ratio) and Y (deaths-to-cases ratio, using excess death data) to be 3.526% and 2.441%, respectively, during the second wave of the pandemic in Switzerland. We choose the second wave to calculate the values of X and Y as

Switzerland has increased the daily number of COVID-19 testing by $5.31\times$ (21641/4074) on average compared to the first wave. We calculate the expected number of cases on a given day t with certainty rate levels of 100% and 50% based on hospitalizations by dividing the number of hospitalizations at t by X and $X/2$, respectively, as we show in Figure 1. We apply the same approach to calculate the expected number of cases on a given day t with certainty rate levels of 100% and 50% based on deaths using Y and $Y/2$, respectively.

Based on Figure 1, we make two key observations. 1) The plot for the expected number of cases calculated based on the number of deaths is shifted forward by 10-20 days (15 days on average) from that for the expected number of cases calculated based on the number of hospitalizations. This is due to the fact that each hospitalized patient usually spends some number of days in hospital before dying of COVID-19. We do not observe a significant time shift between the plot of the expected number of cases calculated based on the number hospitalizations and the plot of observed (laboratory-confirmed) cases. 2) The expected number of cases calculated based on the number of hospitalizations is on average $1.99\times$ higher than the expected number of cases calculated based on the number of deaths (after accounting for the 15-day shift) for the same certainty rate. This is expected as not all hospitalized patients die.

We conclude that both numbers of hospitalizations and deaths can be used for estimating the true number of COVID-19 cases after accounting for the time-shift effect.
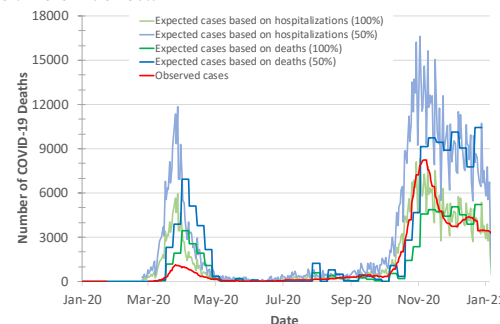

Fig. 1. Observed (officially reported) and expected number of COVID-19 cases in Switzerland during the year of 2020. We calculate the expected number of cases based on both the hospitalizations-to-cases and deaths-to-cases ratios for the second wave. We assume two certainty rate levels of 50% and 100%.

3.3 Observed and Predicted R number of SARS-CoV-2

We calculate the predicted R number using our model (Equation 1) and compare it to the observed official R number and the R number of two state-of-the-art models, ICL and IBZ, for the two years of 2020 and 2021. We configure COVIDHunter using the following configurations: 1) CTC as environmental condition approach, 2) certainty rate levels of 50% and 100%, and 3) mitigation coefficient value of 0.7. All our scripts are provided in our GitHub page. We consider the mean R number provided by the ICL model. We consider the median R number calculated by the IBZ model based on observed number of hospitalized patients. IBZ provides the predicted (after mid of December 2020) R number as the mean of the estimates from the last 7 days.

Based on Figure 2, we make three key observations. 1) COVIDHunter predicts the changes in R number much (4-13 days) earlier than that predicted by ICL model, which leads to a more accurate prediction. The R number predicted by COVIDHunter (with a certainty rate level of 50%) is on average $1.56\times$ less than that predicted by ICL model, IBZ model, and the observed official R number. Using a certainty rate level of 100%, COVIDHunter predicts the R number to be close in value to the observed R number. 2) Our model predicts that the current R number is still higher than 1 (1.137 and 1.023 using certainty rate levels of 50% and 100%, respectively) during January 2021. This indicates that the spread of the SARS-CoV-2 virus is still active and it causes exponential increase in number of new cases. 3) Our model predicts that if we keep the same mitigation measure strength as that of January 2021 for at least 30 days (M(t)= 0.7), then the R number would drop by 18.2% (R= 0.929 and 0.836 for certainty rate levels of 50% and 100%, respectively). However, if the mitigation measures that are applied nationwide in Switzerland are relaxed

by 50% (M(t)= 0.35) for only 30 days (22 January to 22 February 2021), then the R number increases by at least $2.17\times$.

We conclude that COVIDHunter's estimation of the R number is more accurate than that calculated by the ICL and IBZ models, as validated by the currently observed R number.
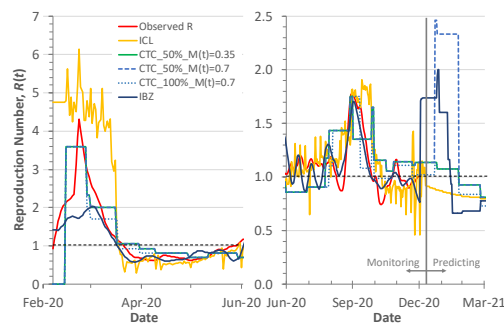

Fig. 2. Observed and predicted reproduction number, $R(t)$, for the two years of 2020 and 2021. We use CTC environmental condition approach, certainty rate levels of 50% and 100%, and mitigation coefficient values of 0.35 and 0.7 for COVIDHunter. We compare COVIDHunter's predicted R number to the observed R number and two state-of-the-art models, ICL and IBZ. The horizontal dashed line represents $R(t) = 1.0$.

3.4 Evaluating the Mitigation Measures

We evaluate the mitigation coefficient, $M(t)$, which represents the mitigation measures applied (or to be applied) in Switzerland from January 2020 to May 2021. We use two different environmental condition approaches, CRW and CTC. We assume two certainty rate levels of 50% and 100% to account for uncertainty in the observed number of cases. We use five mitigation coefficients, $M(t)$, values of 0.35, 0.4, 0.5, 0.6, and 0.7 for each configuration of COVIDHunter during 22 January to 22 February 2021. We compare the evaluated mitigation measures to that evaluated by the Oxford Stringency Index (Hale *et al.*, 2020), as we provide in Figure 3. We also evaluate the mitigation coefficient when we ignore the effect of environmental changes (i.e., by setting C_e=1 in Equation 1), while maintaining the same number of COVID-19 cases of that provided with a certainty rate level of 50%.

Based on Figure 3, we make four key observations. 1) Excluding the effect of environmental changes from the COVIDHunter model, by setting C_e=1 in Equation 1, leads to an inaccurate evaluation of the mitigation measures. For example, during the summer of 2020 (between the two major waves of 2020), COVIDHunter (`WithoutCTC_50%`) evaluates the mitigation coefficient to be as high as 0.6. This means that the mitigation measures (*only* mandatory of wearing mask on public transport) applied during the summer of 2020 are *only* 14% more relaxed compared to the mitigation measures (e.g., closure of schools, restaurants, and borders, ban on small and large events) applied during the first wave, which is implausible. This highlights the importance of considering the effect of external environmental changes on simulating the spread of COVID-19. Unfortunately, environmental change effects are *not* considered by *any* of the IBZ, LSHTM, ICL, and IHME models, which we believe is a serious shortcoming of these prior models. 2) A drop by 3% (as we observe during the mid of November 2020) to 30% (as we observe during the end of August 2020) in the strength of the mitigation measures for a certain period of time (10 to 20 days) is enough to double the predicted number of COVID-19 cases. 3) We evaluate the strength of the mitigation measures applied in Switzerland to be usually (65% of the time) up to 80% to 131% higher than that provided by the Oxford Stringency Index. 4) The strength of the mitigation measures has changed 11 times during the year of 2020, each of which is maintained for at least 9 days and at most 66 days (32 days on average).

We conclude that considering the effect of environmental changes (e.g., daytime temperature) on the spread of COVID-19 improves simulation outcomes and provides accurate evaluation of the strength of the past and current mitigation measures.
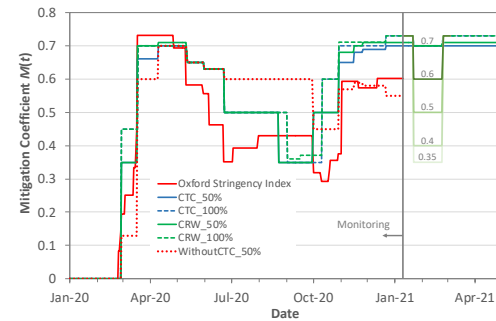

Fig. 3. Predicted strength of the mitigation measures (mitigation coefficient, $M(t)$) applied in Switzerland from January 2020 to May 2021 provided by Oxford Stringency Index and COVIDHunter. We use two different environmental condition approaches, CRW and CTC. We assume two certainty rate levels of 50% and 100%. We use five mitigation $M(t)$ values of 0.35, 0.4, 0.5, 0.6, and 0.7 for each configuration of our model during 22 January to 22 February 2021. The plot called `WithoutCTC_50%` represents the evaluation of the current mitigation measures while ignoring the effect of environmental changes.

3.5 Evaluating the Predicted Number of COVID-19 Cases

We evaluate COVIDHunter's *predicted* daily number of COVID-19 cases in Switzerland. We compare the predicted numbers by our model to the observed numbers and those provided by three state-of-the-art models (ICL, IHME, and LSHTM), as shown in Figure 4. We calculate the observed number of cases as the expected number of cases with a certainty rate level of 100% (as we discuss in Section 3.2). We use three default configurations for the prediction of the ICL model: 1) strengthening mitigation measures by 50%, 2) maintaining the same mitigation measures, and 3) relaxing mitigation measures by 50% which we refer to as `ICL+50%`, `ICL`, and `ICL-50%`, respectively, in Figures 4, 5, and 6. We use the mean numbers reported by the IHME model that represents the most relaxed mitigation measures, called as "no vaccine" by the IHME model. We use the median numbers reported by the LSHTM model.

Based on Figure 4, we make four key observations. 1) Our model predicts that the number of COVID-19 cases reduces significantly (less than 600 daily cases) within March 2021 if the same strength of the currently applied mitigation measure is maintained for at least 30 days. If the authority decides to relax the mitigation measures to the lowest strength that has been applied during the year of 2020 (i.e., $M(t) = 0.35$), then the daily expected number of cases increases by an average of $29.6\times$ and $23.8\times$ (up to 288,827 daily cases) using the CRW and CTC environmental approaches, respectively. We provide a comprehensive evaluation for the effect of different mitigation coefficient values on the number of cases in the Supplementary Materials, Section 2. 2) COVIDHunter predicts the number of COVID-19 cases to be equivalent to that predicted by the IHME model during the second wave with a certainty rate level of 50%. However, during the first wave, the predictions of the IHME model matches the expected number of cases using a certainty rate level of 100%. This means that, unlike our model, the IHME model considers the laboratory-confirmed cases to be as if the tests are done at a population-scale during the first wave, which is very likely incorrect. This is in line with a recent study (Ioannidis *et al.*, 2020) that demonstrates the high inaccuracy of the IHME model. 3) Overall, our model predicts on average $1.7\times$ and $1.9\times$ smaller number of COVID-19 cases than that predicted by ICL model using CTC and CRW approaches, respectively, and a certainty rate of 50%. This suggests that the multiplicative relationship between the confirmed number of cases and the true number of cases can be represented by a certainty rate of 22% to 33%, which our model can easily account for. The ICL model also shows that there is a sharp drop in the daily number of cases after 13 November 2020, which corresponds to a $1.6\times$, $1.4\times$, and $1.3\times$ increase in the Oxford Stringency Index, CRW coefficient, and CTC coefficient, respectively, applied on 30 October 2020 as we show in Figure 3. 4) The number of COVID-19 cases estimated by the LSHTM model during the first wave is 1) on average 24% less than that estimated by COVIDHunter and 2) 10 days late from that predicted by COVIDHunter, IHME, and ICL. The prediction of the LSHTM model during the second wave is not available by the model's pre-computed projections.

We conclude that COVIDHunter provides more accurate estimation of the number of COVID-19 cases, compared to IHME (which provides

inaccurate estimation during the first wave) and ICL (which provides over-estimation), with a complete control over the certainty rate level, mitigation measures, and environmental conditions. Unlike LSHTM, COVIDHunter also ensures no prediction delay.
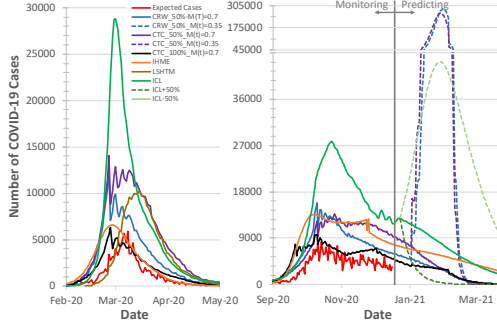

Fig. 4. Observed and predicted number of COVID-19 cases by our model and other three state-of-the-art models. We use two different environmental condition approaches, CRW and CTC with two certainty rate levels of 50% and 100%. We use two mitigation coefficient, $M(t)$, values of 0.35 and 0.7 for each configuration of our model during 22 January to 22 February 2021.

3.6 Evaluating the Predicted Number of COVID-19 Hospitalizations

We evaluate COVIDHunter's *predicted* daily number of COVID-19 hospitalizations in Figure 5. We use the observed official number of hospitalizations as is. Using the number of cases calculated with Equation 2, we find X (hospitalizations-to-cases ratio) to be 4.288% and 2.780%, using CRW and CTC, respectively, during the second wave.

We make five key observations based on Figure 5. 1) The number of hospitalizations calculated by COVIDHunter with a certainty rate level of 50% matches that calculated by the IHME model. However, IHME model provides a 10-12-day late prediction compared to that provided by COVIDHunter and the ICL model. 2) The ICL model predicts the number of hospitalizations to be $5\times$ and $7\times$ higher than that predicted by COVIDHunter during the first wave ($9.3\times$ and $8.1\times$ during the second wave), using the CTC and CRW approaches, respectively, for evaluating the environmental conditions and a certainty rate of 50%. This suggests that the ICL model provides $10\times$ and $18.6\times$ higher number of hospitalizations compared to the observed number of hospitalizations, during first and second waves, respectively, which is highly unlikely and overestimated. 3) COVIDHunter with a certainty rate level of 100% predicts the number of cases to perfectly fit the curve of the observed number of hospitalizations, reaching up to 257 hospitalized patients a day. 4) Our model predicts that the number of COVID-19 hospitalizations reduces with stricter mitigation measures maintained for at least 30 days. Relaxing the mitigation measures by 50% (M is changed from 0.7 to 0.35) *exponentially* increases the number of hospitalizations by an average of $29.6\times$ and $23.8\times$, reaching up to 12385 new daily hospitalized patients, as predicted by COVIDHunter using CRW and CTC environmental condition approaches, respectively. This is in line with what the ICL model (ICL-50%) predicts, when ICL model is configured to 50% relaxation in the mitigation measures. 5) The use of the CTC approach for determining the environmental coefficient value yields a slightly different number of hospitalizations compared to that provided by the use of the CRW approach. This is expected as the CTC approach considers only the monthly average change in temperature, whereas the CRW approach considers the daily change in *several* environmental conditions.

We conclude that 1) unlike the IBZ and LSHTM models, COVIDHunter is able to predict the number of hospitalizations and 2) COVIDHunter provides more accurate estimation of the number of hospitalizations compared to that calculated by ICL (which provides overestimation) and IHME (which provides late estimation). COVIDHunter predicts the number of COVID-19 hospitalizations in a simple, convenient and flexible way that requires calculating only the daily number of cases and the hospitalization-to-cases ratio, C_X.
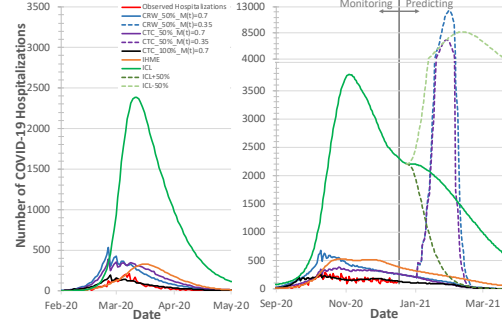

Fig. 5. Observed and predicted number of COVID-19 hospitalizations. We use two different environmental condition approaches, CRW and CTC with two certainty rate levels of 50% and 100%. We use two mitigation coefficient values, $M(t)$, of 0.35 and 0.7 for each configuration of our model during 22 January to 22 February 2021.

3.7 Evaluating the Predicted Number of COVID-19 Deaths

We evaluate COVIDHunter's *predicted* daily number of COVID-19 deaths in Figure 6 after accounting for the 15-day shift (as we discuss in Section 3.2). We calculate the observed number of deaths as the number of excess deaths (Section 2.4) to account for uncertainty in reporting COVID-19 deaths. Using the number of cases calculated using Equation 2, we find Y (deaths-to-cases ratio, using excess death data) to be 2.730% and 1.739%, using CRW and CTC, respectively, during the second wave.

We make three key observations based on Figure 6. 1) COVIDHunter with a certainty rate of 100% predicts the number of deaths to perfectly fit the three curves of the observed number of *excess* deaths, ICL deaths, and IHME deaths, reaching up to 160 hospitalized patients a day. During the second wave, the ICL curve is shifted (late prediction) by 5-10 days from that of other models. 2) Similar to what we observe for the number of hospitalizations, our model predicts that the number of COVID-19 deaths significantly reduces with stricter mitigation measures maintained for at least the upcoming 30 days. Relaxing the mitigation measures by 50% ($M(t)$ is changed from 0.7 to 0.35) *exponentially* increases the death toll by an average of $29.6\times$ and $23.8\times$, reaching up to 7885 new daily deaths, as predicted by COVIDHunter using CRW and CTC environmental condition approaches, respectively. 3) During the first wave, the use of a certainty rate of 50% provides $2.55\times$ and $2.1\times$ ($2.36\times$ and $1.52\times$ during the second wave) higher number of deaths compared to that provided by ICL and IHME models, when COVIDHunter uses CRW and CTC environmental condition approaches, respectively.

We conclude that 1) unlike the IBZ and LSHTM models, COVIDHunter is able to predict the number of deaths, 2) COVIDHunter predicts the number of deaths to be similar to that predicted by the ICL and IHME models. Yet, COVIDHunter provides more accurate estimation of other COVID-19 statistics (R, number of cases and hospitalizations) compared to ICL and IHME, as we comprehensively evaluate in the previous sections, and 3) COVIDHunter requires calculating only the daily number of cases and the deaths-to-cases ratio, C_Y, to predict the daily number of deaths.
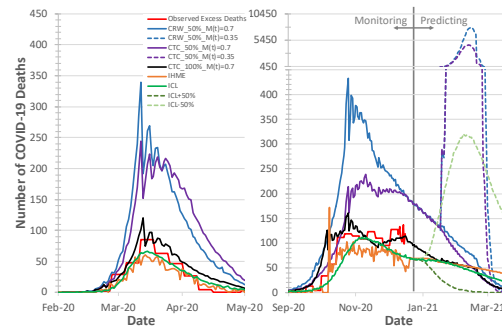

Fig. 6. Observed and predicted number of COVID-19 deaths. We use two different environmental condition approaches, CRW and CTC with two certainty rate levels of 50% and 100%. We use two mitigation coefficient values, $M(t)$, of 0.35 and 0.7 for each configuration of our model during 22 January to 22 February 2021.

4 Summary and Future Work

We demonstrate that we can monitor and predict the spread of COVID-19 in an easy-to-use, flexible, and validated way using our new simulation model, COVIDHunter. We show how to flexibly configure our model for any scenario and easily extend it for different mitigation measures and environmental conditions. The use of a small number of variables in our model enables a simple and flexible yet powerful way of adapting our model to different conditions for a given region. We demonstrate the importance of considering the effect of environmental changes on the spread of COVID-19 and how doing so can greatly improve simulation accuracy. COVIDHunter flexibly offers the ability to directly make the best use of existing models that study the effect of one or both of environmental conditions and mitigation measures on the spread of COVID-19.

We benchmark our model against major alternative models of the COVID-19 pandemic that are used to assist governments. Compared to these models, COVIDHunter achieves more accurate estimation, provides no prediction delay, and provides ease of use and high flexibility due to the simple modeling approach that uses a small number of parameters. Using COVIDHunter, we demonstrate that the spread of COVID-19 in Switzerland (as a case study) is still active (i.e., $R > 1.0$) and curbing this spread requires maintaining the same strength of the currently applied mitigation measures for at least another 30 days. Using COVIDHunter ($CTC\_100\%\_M(t)=0.7$) on 7 January 2021, we predicted that on 27 January 2021 the number of cases, hospitalizations, and deaths will drop by 19%, 20%, and 30%, respectively. The predicted drop is in line with the observed official number of cases, hospitalizations, and deaths (as shown by the Federal Office of Public Health in Switzerland www.covid19.admin.ch) but with different ratios (41%, 59%, and 49%, respectively). We believe the difference between the observed and the COVIDHunter's predicted numbers of cases, hospitalizations, and deaths is due to one or more of the following reasons: 1) The lack of population-scale COVID-19 testing, 2) the use of a more stricter mitigation measure than $M(t) = 0.7$, and 3) the lack of information about ground truth on number of COVID-19 cases, hospitalizations, and deaths. We provide insights on the effect of each change in the strength of the applied mitigation measure on the number of daily cases, hospitalizations, and deaths. We make all the data, statistical analyses, and a well-documented model implementation publicly and freely available to enable full reproducibility and help society and decision-makers to accurately and openly review the current situation and estimate future impact of decisions.

We suggest and plan at least five main directions/additions to further improve the predictive power and benefits of our COVIDHunter model. 1) Clustering the population based on age-groups. This has potential different effects on, for example, population, environmental conditions, mitigation measures (Bhatia and Klausner, 2020; Bi *et al.*, 2020). 2) Considering vaccinated persons as another new category of persons in a population. 3) Considering reinfection after immunity (Lumley *et al.*, 2020). 4) Considering the average number of households (or population density), as well as other potential population-level effects, while calculating the number of new infected persons caused by an infected person. 5) Considering different strains of the COVID-19 virus by allowing for multiple base reproduction numbers. Our goal is to update COVIDHunter with such improvements and capabilities while keeping its simplicity, ease of use, and flexibility of its modeling strategy.

References

Andersen, K. G., Rambaut, A., Lipkin, W. I., *et al.* (2020). The proximal origin of SARS-CoV-2. *Nature medicine*, **26**(4), 450–452.

Ashcroft, P., Huisman, J. S., Lehtinen, S., *et al.* (2020). COVID-19 infectivity profile correction. *Swiss Medical Weekly*, **150**(3132).

Bhatia, R. and Klausner, J. (2020). Estimating individual risks of COVID-19-associated hospitalization and death using publicly available data. *PloS one*, **15**(12), e0243026.

Bi, Q., Wu, Y., Mei, S., *et al.* (2020). Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *The Lancet Infectious Diseases*, **20**(8), 911–919.

Brauner, J. M., Mindermann, S., Sharma, M., *et al.* (2020). Inferring the effectiveness of government interventions against COVID-19. *Science*.

Chang, S. L., Harding, N., Zachreson, C., *et al.* (2020). Modelling transmission and control of the COVID-19 pandemic in Australia. *Nature Communications*.

Davies, N. G., Kucharski, A. J., Eggo, R. M., *et al.* (2020). Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital

services in the UK: a modelling study. *The Lancet Public Health*.

de Souza, W. M., Buss, L. F., da Silva Candido, D., *et al.* (2020). Epidemiological and clinical characteristics of the early phase of the COVID-19 epidemic in Brazil. *Nature Human Behaviour*, **4**, 856–865.

Du Toit, A. (2020). Outbreak of a novel coronavirus. *Nature Reviews Microbiology*, **18**(3), 123–123.

Fares, A. (2013). Factors influencing the seasonal patterns of infectious diseases. *International journal of preventive medicine*, **4**(2), 128.

Ferguson, N., Laydon, D., Nedjati-Gilani, G., *et al.* (2020). Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. *Imperial College London*, **10**, 77482.

Fisman, D. (2012). Seasonality of viral infections: mechanisms and unknowns. *Clinical Microbiology and Infection*, **18**(10), 946–954.

Flaxman, S., Mishra, S., Gandy, A., *et al.* (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, **584**(7820), 257–261.

Hale, T., Petherick, A., Phillips, T., and Webster, S. (2020). Variation in government responses to COVID-19. *Blavatnik school of government working paper*, **31**.

Hilton, J. and Keeling, M. J. (2020). Estimation of country-level basic reproductive ratios for novel Coronavirus (SARS-CoV-2/COVID-19) using synthetic contact matrices. *PLOS Computational Biology*, **16**(7), 1–10.

Huisman, J. S., Scire, J., Angst, D. C., *et al.* (2020). Estimation and worldwide monitoring of the effective reproductive number of SARS-CoV-2. *medrxiv*.

Ioannidis, J. P., Cripps, S., and Tanner, M. A. (2020). Forecasting for COVID-19 has failed. *International journal of forecasting*.

Kampf, G., Todt, D., Pfaender, S., and Steinmann, E. (2020). Persistence of coronaviruses on inanimate surfaces and their inactivation with biocidal agents. *Journal of Hospital Infection*, **104**(3), 246–251.

Kobayashi, T., Jung, S.-m., Linton, N. M., *et al.* (2020). Communicating the Risk of Death from Novel Coronavirus Disease (COVID-19). *Journal of Clinical Medicine*, **9**(2).

Lauer, S. A., Grantz, K. H., Bi, Q., *et al.* (2020). The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Annals of internal medicine*, **172**(9), 577–582.

Li, Q., Guan, X., Wu, P., *et al.* (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *New England Journal of Medicine*.

Lumley, S. F., O'Donnell, D., Stoesser, N. E., *et al.* (2020). Antibody status and incidence of SARS-CoV-2 infection in health care workers. *New England Journal of Medicine*.

Moriyama, M., Hugentobler, W. J., and Iwasaki, A. (2020). Seasonality of respiratory viral infections. *Annual review of virology*, **7**.

Pachetti, M., Marini, B., Benedetti, F., *et al.* (2020). Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *Journal of Translational Medicine*, **18**, 1–9.

Phan, T. (2020). Genetic diversity and evolution of SARS-CoV-2. *Infection, genetics and evolution*, **81**, 104260.

Prata, D. N., Rodrigues, W., and Bermejo, P. H. (2020). Temperature significantly changes COVID-19 transmission in (sub) tropical cities of Brazil. *Science of the Total Environment*, page 138862.

Rahman, B., Sadraddin, E., and Porreca, A. (2020). The basic reproduction number of SARS-CoV-2 in Wuhan is about to die out, how about the rest of the World? *Reviews in Medical Virology*, page e2111.

Reichmuth, M., Hodcroft, E., Riou, J., *et al.* (2021). Transmission of SARS-CoV-2 variants in Switzerland. [https://ispmbern.github.io/covid-19/va riants/index.pdf](https://ispmbern.github.io/covid-19/variants/index.pdf). Accessed: 2021-1-30.

Reiner, R. C., Barber, R. M., Collins, J. K., *et al.* (2020). Modeling COVID-19 scenarios for the United States. *Nature Medicine*.

Riddell, S., Goldie, S., Hill, A., *et al.* (2020). The effect of temperature on persistence of SARS-CoV-2 on common surfaces. *Virology journal*, **17**(1), 1–7.

Russell, T. W., Golding, N., Hellewell, J., *et al.* (2020). Reconstructing the early global dynamics of under-ascertained COVID-19 cases and infections. *BMC medicine*, **18**(1), 1–9.

Shi, Q., Hu, Y., Peng, B., *et al.* (2020). Effective control of SARS-CoV-2 transmission in Wanzhou, China. *Nature medicine*, pages 1–8.

Slifka, M. K. and Gao, L. (2020). Is presymptomatic spread a major contributor to COVID-19 transmission? *Nature Medicine*, **26**(10), 1531–1533.

Toyoshima, Y., Nemoto, K., Matsumoto, S., *et al.* (2020). SARS-CoV-2 genomic variations associated with mortality rate of COVID-19. *Journal of human genetics*, **65**(12), 1075–1082.

Tradigo, G., Guzzi, P. H., Kahveci, T., and Veltri, P. (2020). A method to assess COVID-19 infected numbers in Italy during peak pandemic period. In *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 3017–3020. IEEE.

Wang, J., Tang, K., Feng, K., and Lv, W. (2020). High temperature and high humidity reduce the transmission of COVID-19. *Available at SSRN 3551767*.

Wei, W. E., Li, Z., Chiew, C. J., *et al.* (2020). Presymptomatic Transmission of SARS-CoV-2—Singapore, January 23–March 16, 2020. *Morbidity and Mortality Weekly Report*, **69**(14), 411.

Wilke, C. O. and Bergstrom, C. T. (2020). Predicting an epidemic trajectory is difficult. *Proceedings of the National Academy of Sciences*, **117**(46), 28549–28551.

Xie, J. and Zhu, Y. (2020). Association between ambient temperature and COVID-19 infection in 122 cities from China. *Science of the Total Environment*, **724**, 138201.

Xu, R., Rahmandad, H., Gupta, M., *et al.* (2020). The Modest Impact of Weather and Air Pollution on COVID-19 Transmission. *medRxiv*.

Supplementary Material for

COVIDHunter: An Accurate, Flexible, and Environment-Aware Open-Source COVID-19 Outbreak Simulation Model

Mohammed Alser, Jeremie S. Kim, Nour Almadhoun Alserr,
Stefan W. Tell, and Onur Mutlu

ETH Zurich, Zurich 8006, Switzerland

1 Statistical Relationship Between Temperature and Number of COVID-19 Cases

The purpose of this study is to explore the relationship between the daily new confirmed COVID-19 case counts or death counts and temperature in Switzerland. We obtain the daily number of confirmed COVID-19 cases and deaths in Switzerland from official reports of the Federal Office of Public Health (FOPH) in Switzerland [1] starting from March 2020 until January 2020. We obtain the air temperature data from the Federal Office of Meteorology and Climatology (MeteoSwiss) in Switzerland [2]. We calculate the daily average air temperature during the same time period (March 2020 to December 2020) for all the 26 cantons in Switzerland.

To evaluate the correlation between the temperature data and the number of daily confirmed COVID-19 cases or the daily counts of death, we use a generalized additive model (GAM). GAM is usually used to calculate the linear and non-linear regression models between meteorological factors (e.g., temperature, humidity) with COVID-19 infection and transmission [3, 4, 5]. Our analyses are performed with R software version 4.0.3., where $p-value < 0.05$ is considered statistically significant. Our model attempts to represent the linear behavior of the growth curve of the counts of the new confirmed cases or deaths in Switzerland. Therefore, we can test the hypothesis of whether there is a significant negative correlation between the COVID-19 confirmed daily case or death counts and temperature.

The results demonstrate a significant negative correlation between temperature and COVID-19 daily case and death counts. Specifically, the relationship is linear for the average temperature in the range from 1-26°C. Based on Figure S1, we make two key observations. 1) For each 1°C rise in temperature, there is a 3.67% (t-value = 3.244 and p-value = 0.0013) decrease in the daily number of COVID-19 confirmed cases (Figure S1(a)). 2) For each 1°C rise in temperature, there is a 23.8% decrease in the daily number of COVID-19 deaths (t-value = 9.312 and p-value = 0.0), as shown in Figure S1(b).
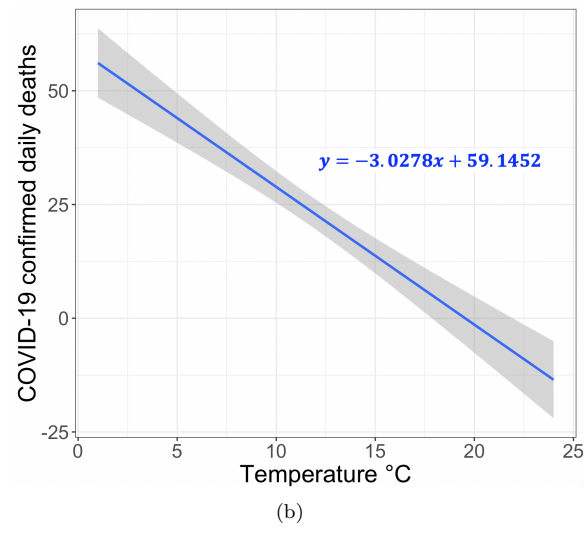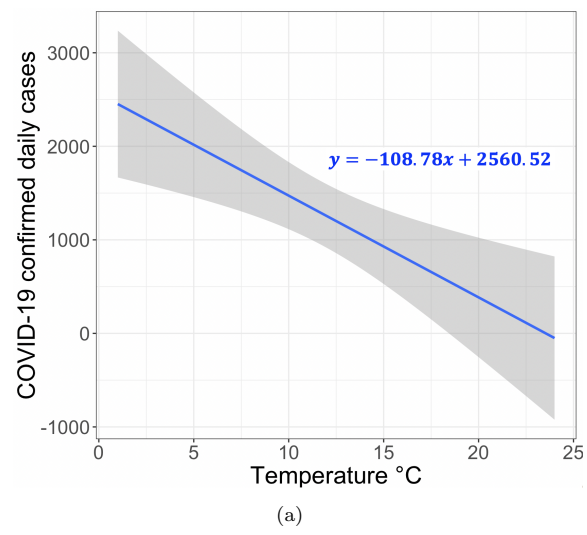
Figure S1: Correlation between temperature and COVID-19 confirmed (a) case count and (b) death count in 26 cantons of Switzerland.

2 Evaluating the Effect of Different Mitigation Coefficient Values on COVIDHunter's Predicted Number of Cases, Hospitalizations, and Deaths

Using COVIDHunter, we predict the number of COVID-19 cases, hospitalizations, and deaths during 22 January to 31 March 2021. We show the maximum and the average of daily number of COVID-19 cases, hospitalizations, and deaths over 22 January to 31 March 2021 in Figures S2 and S3, respectively. We use two environmental condition approaches, CRW and CTC, with a certainty rate level of 50%. We assume five mitigation coefficient, $M(t)$, values of 0.35, 0.4, 0.5, 0.6, and 0.7 for each configuration of COVIDHunter during 22 January to 22 February 2021.

This range of mitigation coefficient values covers the lowest (i.e., $M(t)$=0.35) and the highest (i.e., $M(t)$=0.7) strengths of mitigation measures that have been applied during the year of 2020.

Based on Figures S2 and S3, we make three key observations. 1) COVIDHunter predicts that the *maximum* of daily number of COVID-19 cases, hospitalizations, and deaths over 22 January to 31 March 2021 would be 4972, 213, and 136, respectively, using CRW and $M(t)$=0.7, as we show in Figure S2(a-c). Using our environmental condition approach, CTC, and $M(t)$=0.7, the *maximum* of daily number of COVID-19 cases over 22 January to 31 March 2021 would be 7580 and the *maximum* of daily number of COVID-19 hospitalizations and deaths would be almost same as that calculated by COVIDHunter with CRW, as we show in Figure S2(d-f). 2) Relaxing the mitigation measures by 50% (M is changed from 0.7 to 0.35) *exponentially* increases the *maximum* of daily number of cases, hospitalizations, and deaths by 58×, reaching up to 288827, 12385, and 7885, respectively, as predicted by COVIDHunter with the CRW approach (Figure S2(a-c)). Using the CTC appraoch and $M(t)$=0.35, COVIDHunter predicts an *exponential* increase in the *maximum* of daily number of cases, hospitalizations, and deaths by *only* 34.5×, as we show in Figure S2(a-c). This is expected as the CTC approach considers only the drop in temperature rather than the average effect of many environmental conditions as the CRW approach does. 3) Relaxing the mitigation measures by 50% (M is changed from 0.7 to 0.35) causes the *daily* number of cases, hospitalizations, and deaths to *exponentially* increase by an average of 29.6× and 23.8× over 22 January to 31 March 2021 using CRW and CTC environmental approaches, respectively, as we show in Figure S3.

We conclude that COVIDHunter provides flexible evaluation of the effect of different strength of the past and current mitigation measures on the number of COVID-19 cases, hospitalizations, and deaths. COVIDHunter evaluates the applied mitigation measures with high flexibility of configuring the environmental coefficient and mitigation coefficient, which helps society and decision-makers to accurately review the current situation and estimate future impact of decisions.
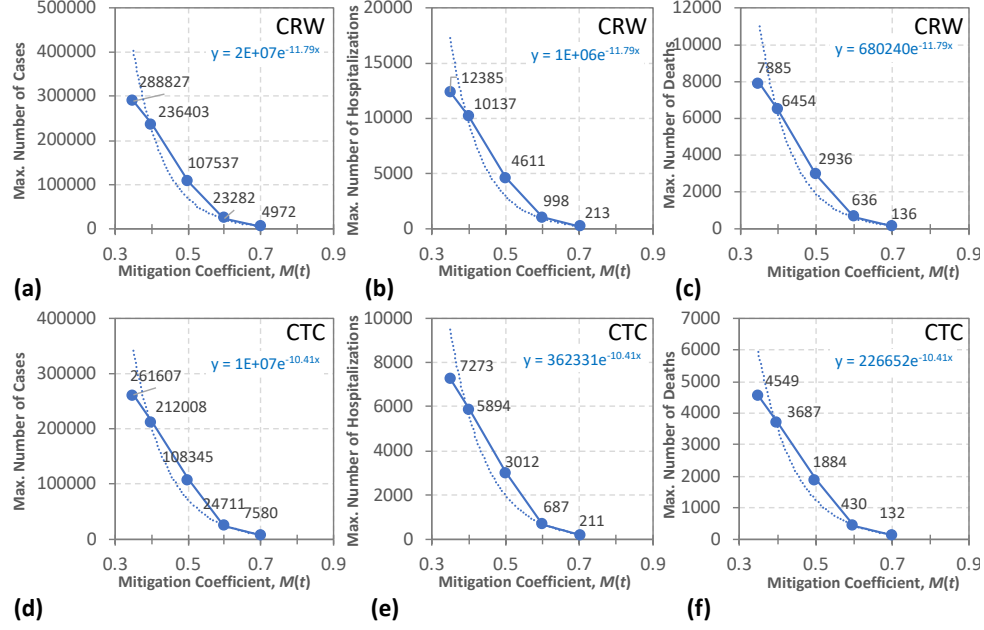
Figure S2: The maximum of daily number of COVID-19 cases, hospitalizations, and deaths as predicted by COVIDHunter over 22 January to 31 March 2021. We use five mitigation coefficient, $M(t)$, values of 0.35, 0.4, 0.5, 0.6, and 0.7 for each configuration of our model during 22 January to 22 February 2021. We use two different environmental condition approaches, CRW **(a)-(c)** and CTC **(d)-(f)** with a certainty rate level of 50%. Dashed line represents exponential model fit to data.
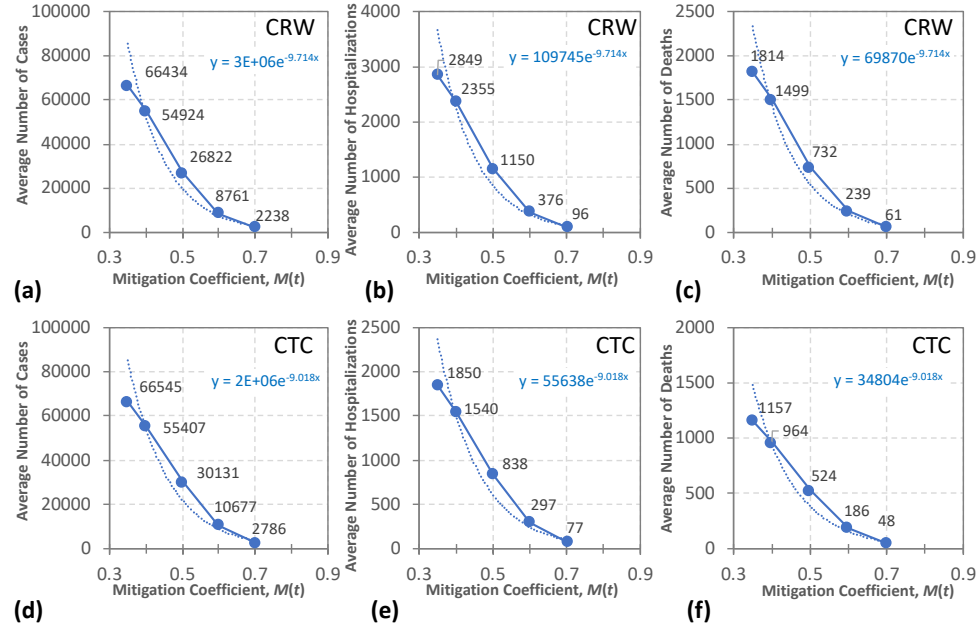

Figure S3: The average of daily number of COVID-19 cases, hospitalizations, and deaths as predicted by COVIDHunter over 22 January to 31 March 2021. We use five mitigation coefficient, $M(t)$, values of 0.35, 0.4, 0.5, 0.6, and 0.7 for each configuration of our model during 22 January to 22 February 2021. We use two different environmental condition approaches, CRW **(a)-(c)** and CTC **(d)-(f)** with a certainty rate level of 50%. Dashed line represents exponential model fit to data.

3   Evaluated Datasets

Our experimental evaluation uses a large number of different real datasets, including 1) daily R number values, 2) observed daily number of COVID-19 cases, 3) observed daily number of COVID-19 hospitalizations, 4) observed daily number of COVID-19 deaths, 5) number of excess deaths, 6) the estimated strength of mitigation measures as calculated by the Oxford Stringency Index, 7) estimation of COVID-19 statistics as calculated by existing state-of-the-art simulation models, ICL, IHME, LSHTM, and IBZ, from seven different sources as we list below. The raw datasets are provided in the Supplementary Excel File[1] and it can be also obtained from the original sources as we list below:

- Observed COVID-19 statistics (R number values and number of cases, hospitalizations, and deaths)

    - Official reports (January 7, 2021): https://www.covid19.admin.ch/en/overview
    - Smoothed data (January 7, 2021): https://ourworldindata.org/coronavirus/country/switzerland?country=~CHE

- Excess deaths:

    - Information: https://www.bfs.admin.ch/bfs/en/home/statistics/health/state-health/mortality-causes-death.html
    - Direct link (January 7, 2021): https://www.bfs.admin.ch/bfs/en/home/statistics/health/state-health/mortality-causes-death.assetdetail.12607335.html

- Oxford Stringency Index

    - https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker#data

- Imperial College London (ICL) Model:

    - Information: https://mrc-ide.github.io/global-lmic-reports/
    - Direct link (January 15, 2021): https://github.com/mrc-ide/global-lmic-reports/raw/master/data/2021-01-30_v7.csv.zip

- Institute for Health Metrics and Evaluation (IHME) Model:

    - Information: https://mrc-ide.github.io/global-lmic-reports/
    - Direct link (January 15, 2021): http://www.healthdata.org/covid/data-downloads

- The London School of Hygiene Tropical Medicine (LSHTM) Model:

    - Information: https://cmmid.github.io/topics/covid19/global_cfr_estimates.html
    - Direct link (January 15, 2021): https://raw.githubusercontent.com/cmmid/cmmid.github.io/master/topics/covid19/reports/under_reporting_estimates/under_ascertainment_estimates.csv

- The Theoretical Biology Group at ETH Zurich (IBZ) Model:

    - Information: https://ibz-shiny.ethz.ch/covid-19-re-international/
    - Direct link (January 15, 2021): https://github.com/covid-19-Re/dailyRe-Data

References

[1] Coronavirus - The Federal Office of Public Health in Switzerland. https://www.bag.admin.ch/bag/en/home/krankheiten/ausbrueche-epidemien-pandemien/aktuelle-ausbrueche-epidemien/novel-cov.html. Accessed: 2020-12-31.

[2] Switzerland forecast - The Federal Office of Meteorology and Climatology MeteoSwiss. https://www.meteoswiss.admin.ch/home.html?tab=overview. Accessed: 2020-12-31.

---

[1]https://github.com/CMU-SAFARI/COVIDHunter/blob/main/Evaluation_Results/ComparisonToOtherModels.xlsx

[3] Jiangtao Liu, Ji Zhou, Jinxi Yao, Xiuxia Zhang, Lanyu Li, Xiaocheng Xu, Xiaotao He, Bo Wang, Shihua Fu, Tingting Niu, et al. Impact of meteorological factors on the COVID-19 transmission: A multi-city study in China. *Science of the Total Environment*, page 138513, 2020.

[4] David N Prata, Waldecy Rodrigues, and Paulo H Bermejo. Temperature significantly changes COVID-19 transmission in (sub) tropical cities of Brazil. *Science of the Total Environment*, page 138862, 2020.

[5] Jingui Xie and Yongjian Zhu. Association between ambient temperature and COVID-19 infection in 122 cities from China. *Science of the Total Environment*, 724:138201, 2020.