

# Model uncertainty and decision making: Predicting the Impact of COVID-19 Using the CovidSim Epidemiological Code

Wouter Edeling<sup>1</sup>, Hamid Arabnejad<sup>2</sup>, Robert C Sinclair<sup>3</sup>, Diana Suleimenova<sup>2</sup>, Krishnakumar Gopalakrishnan<sup>3</sup>, Bartosz Bosak<sup>4</sup>, Derek Groen<sup>2</sup>, Imran Mahmood Qureshi Hashmi<sup>2</sup>, Daan Crommelin<sup>1, 5</sup>, and Peter V Coveney <sup>\*3, 6</sup>

<sup>1</sup>Scientific Computing Group, Centrum Wiskunde & Informatica, Amsterdam, The Netherlands

<sup>2</sup>Department of Computer Science, Brunel University London, London, UK

<sup>3</sup>Centre for Computational Science, University College London, London, UK

<sup>4</sup>Poznań Supercomputing and Networking Center, Poznań, Poland

<sup>5</sup>Korteweg-de Vries Institute for Mathematics, University of Amsterdam, Amsterdam, The Netherlands

<sup>6</sup>Informatics Institute, University of Amsterdam, Amsterdam, The Netherlands

3rd November 2020

\*Corresponding author; email: [p.v.coveney@ucl.ac.uk](mailto:p.v.coveney@ucl.ac.uk)

## Abstract

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV2) virus has rapidly spread worldwide since December 2019, and early modelling work of this pandemic has assisted in identifying effective government interventions. The UK government relied in part on the CovidSim model developed by the MRC Centre for Global Infectious Disease Analysis at Imperial College London, to model various non-pharmaceutical intervention strategies, and guide its government policy in seeking to deal with the rapid spread of the COVID-19 pandemic during March and April 2020. CovidSim is subject to different sources of uncertainty, namely parametric uncertainty in the inputs, model structure uncertainty (*i.e.*, missing epidemiological processes) and scenario uncertainty, which relates to uncertainty in the set of conditions under which the model is applied. We have undertaken an extensive parametric sensitivity analysis and uncertainty quantification of the current CovidSim code. From the over 900 parameters that are provided as input to CovidSim, we identified a key subset of 19 parameters to which the code output is most sensitive. We find that the uncertainty in the code is substantial, in the sense that imperfect knowledge in these inputs will be magnified to the outputs, up to the extent of ca. 300%. Most of this uncertainty can be traced back to the sensitivity of three parameters. Compounding this, the model can display significant bias with respect to observed data, such that the output variance does not capture this validation data with high probability. We conclude that quantifying the parametric input uncertainty is not sufficient, and that the effect of model structure and scenario uncertainty cannot be ignored when validating the model in a probabilistic sense.

# 1 Introduction

CovidSim is an individual-based simulation code developed by the MRC Centre for Global Infectious Disease Analysis at Imperial College London. It is a modified version of an earlier model designed to support pandemic influenza planning [12], and has now been used to explore various non-pharmaceutical interventions (NPI) with the aim of reducing the transmission of the coronavirus, as documented in the key paper [11], denoted as ‘Report 9’. CovidSim played an important role in the United Kingdom in reorienting UK Government policy from herd immunity to a strategy focused on suppression of the viral infection [23]. It should be noted, however, that many competitor models exist. Notable examples include the work performed at the London School of Hygiene & Tropical Medicine; see for instance [24, 3], and especially [4], where the effect of different NPIs in the UK is modelled.

CovidSim creates a network of individuals, located in areas defined by high-resolution population density data. In the model, contacts with other individuals can be made within four different place types, namely within the household, at schools, university and work places. It is possible to model a combination of different NPIs, namely general social distancing (SD), social distancing for those over 70 years of age (SDOL70), home isolation of suspected cases (CI), voluntary home quarantine (HQ) and place closure of universities and schools (PC); see Table 2 of Report 9 [11]. CovidSim contains over 900 input parameters, which are mainly located in two input files. In addition, a small number of parameters that define certain characteristics of the intervention scenario one wishes to study are supplied via the command line.

We have investigated the reproducibility of the code, as also done previously [1, 8]. That said, we especially focus on CovidSim’s robustness under uncertainty in the input parameters. By ‘robustness’ in this context we mean the extent to which the code amplifies uncertainties from the input to the output. Thus, our main aim is to take the model as given and examine the uncertainty in its predictions when its parameters are treated as random variables instead of deterministic inputs. A dimension-adaptive sampling method [15], is employed to be able to handle the high-dimensional input space. These type of anisotropic sampling methods adaptively exploit a possible low effective dimension, where only a subset of all inputs have a significant impact on the model output. A wide range of domains have seen application of such dimension-adaptive samplers, *e.g.* computational electromagnetism [25], finance [21, 16] or natural convection problems [13], to name just a few. We only perform a limited validation study, by examining the ability of the predicted output distribution to envelop the observed COVID-19 death count, conditional on a predefined intervention scenario.

Due to the large number of inputs, one cannot hope to obtain an accurate, data-informed value of all parameters in play. Moreover, regarding CovidSim’s influential status, and its likely use in future COVID-19 predictions, it is therefore important to assess the impact of parametric uncertainty on the model output.

## 2 Sources of uncertainty

The predictions of most computational models are affected by uncertainty from a variety of different sources. We identify the following three sources of uncertainty in CovidSim, namely parametric uncertainty, model structure uncertainty and scenario uncertainty. This breakdown is not uncommon; see *e.g.* [5, 28, 7].

Parametric uncertainty arises due to imperfect knowledge of the model input parameters  $\xi \in \mathbb{R}^d$ , described in Section 3. Model structure uncertainty is more fundamental, as it relates to uncertainty about the appropriate mathematical structure of the model, denoted by  $\mathcal{M}$ . One can think of missing epidemiological processes that are not implemented in CovidSim; see the discussion in Section 4.2.1. Finally, a scenario  $\mathcal{S}$  is the set of conditions under which a model  $\mathcal{M}(\xi)$  is applied. In the case of CovidSim,  $\mathcal{S}$  includes the choice of NPI scenarios, the initialisation of the model, and the well-known reproduction number  $R_0$ . Note that the actual implementation of  $\mathcal{S}$  will be parameterised as well, and that we could technically lump these parameters in with  $\xi$ . However, the scenario parameters are of a different nature than the ‘internal’ inputs  $\xi$ , and treating  $\mathcal{S}$  as a separate category mirrors the way in which the results were presented in Report 9, which showed results for different NPIs and  $R_0$  values.

Thus, if we denote  $q$  as the predicted output quantity of interest, we have  $q = q(\xi, \mathcal{M}, \mathcal{S})$ , where all three arguments are uncertain. As noted, our main goal is to quantify the impact of parametric uncertainty. By treating the inputs as random variables with probability density function  $p(\xi)$ , our mean prediction is given by

$$\mathbb{E}[q \mid \mathcal{M}, \mathcal{S}] := \int_{\Omega_\xi} q(\xi, \mathcal{M}, \mathcal{S}) p(\xi) d\xi, \quad (1)$$

where  $\Omega_\xi$  is the support of  $p(\xi)$ . The uncertainty in the prediction (1) can be represented by either the corresponding variance or confidence intervals. It is important to note that our results are conditional on  $\mathcal{M}$  and  $\mathcal{S}$ . We are not in a position to change the former, and we illustrate the importance of scenario uncertainty by repeating the parametric uncertainty analysis for two different scenarios.

## 3 CovidSim parameters

In this section we describe how we arrived at our selection of input parameters that we vary as part of our uncertainty quantification study. We have divided the parameters present in the input files into three groups:

- (i) Group 1: intervention parameters, these are parameters meant to slow down the viral infection, which can still be varied for a fixed  $\mathcal{S}$ , *e.g.* “Length of time households are quarantined.”, when HQ is part of the selected NPI set.
- (ii) Group 2: disease parameters, related to the characteristics of COVID-19, *e.g.* “Latent period.”
- (iii) Group 3: spatial / geographic parameters, parameters which apply to the properties of the network, *e.g.* “Relative transmission rate for place types.”

The purpose of this classification was to direct initial, exploratory uncertainty quantification (UQ) and sensitivity analysis (SA) campaigns<sup>1</sup> on a coherent subset of parameters. The final UQ campaign (see Section 4), contains parameters from all three groups.

Before starting the UQ analysis, we first performed a parameter study, using in part expert domain knowledge from the CovidSim team at Imperial College, to reduce the number of inputs. We focus on a scenario  $\mathcal{S}$  based on the suppression release in the Report 9 folder on GitHub [29], using the intervention setting PC\_CI\_HQ\_SD. In this case we have a total of 940 parameters. Note that some input parameters are vectors, in which case we counted each entry as a separate input parameter. On top of our own initial selection, over the course of our analysis we have received feedback from the developers of CovidSim as to the inclusion of a given parameter in the UQ study. This led to the following breakdown, where we classify each input into one of three categories, based on received suggestions:

1. Do not vary: 76% of total
2. Vary: 6% of total
3. Optional: 18% of total

Many of the parameters are (currently) not used in the case of COVID-19 simulation, such as numerous vaccination parameters. For more details, see the ‘parameter list’ folder in [39] for the full list with all input parameters, their default values and reasons for inclusion or exclusion from the Imperial College CovidSim team. Optional parameters were suggested as potential candidates, but have certain drawbacks when used as random variables, *e.g.* one can argue that parameters which are fitted to data are better left unperturbed.

Although we made our own considerations and decisions as to which parameters to include in the UQ study, the large number of parameters in play requires expert knowledge to make a suitable initial selection. A total of 60 of these parameters were included in a UQ campaign at some point. These are displayed separately in [39]. We choose uninformative uniform distributions to reflect our lack of knowledge in the most-likely values of these inputs, with bounds based on either data or expert knowledge.

<sup>1</sup>By a ‘campaign’ we mean a single forward propagation step of uncertainty from the input to the output.

## 4 Uncertainty quantification results

We use EasyVVUQ [33, 41] from the Verified Exascale Computing for Multiscale Applications (VECMA) toolkit [18], for propagating the input uncertainties through CovidSim. To interface CovidSim with EasyVVUQ, templates from the CovidSim input files are generated. In the process a single file is generated, which contains all inputs, with their types and default values specified. Simply counting the number of entries in this file allows us to exactly determine the number of parameters present in the code, which is how we arrived at the aforementioned number of 940 inputs.

Each of the  $d$  input parameters  $\xi_i$  is given an independent probability density function (pdf), *i.e.*,  $\xi_i \sim p(\xi_i)$ . Next, a  $d$ -dimensional sampling plan from the joint pdf  $\prod_i p(\xi_i)$  is created, after which CovidSim is evaluated at each input point. To handle the high-dimensional input space, we refine the sampling plan in a dimension-adaptive manner, the details of which can be found in the Supplementary Materials A.

We have performed a limited analysis on the original closed-source version of the code. The majority of our sensitivity analysis and uncertainty quantification efforts lie with the current updated open-source release of CovidSim.

### 4.1 Original version

With respect to the original version, we have been able to achieve exact reproducibility of the results in the influential Report 9 [11], when running this within an Azure cloud environment. Other authors have also found CovidSim to be reproducible [8, 1]. Attempts to run the code on other computing resources failed; we could not reproduce the same results on Linux machines and since this version is no longer supported it was not investigated further.

### 4.2 Current version

We apply dimension-adaptive UQ and SA on the aforementioned suppression release within the Report 9 folder on GitHub [29]. As mentioned, we performed separate exploratory UQ campaigns on the parameters belonging to the three groups of Section 3. Any input that was selected at least once for refinement during the dimension-adaptive sampling was included in the final, large-scale UQ campaign. This led to a total of 19 final parameters; see Supplementary Materials B.

Important scenario parameters are the reproduction number  $R_0$ , and two ‘trigger’ parameters, which are specified via the command line. In the case of modelling a suppression strategy, the SD and PC interventions are triggered when the weekly number of new intensive care unit (ICU)

cases exceeds the value supplied by the first trigger. Likewise, they are suspended when this metric drops below the second specified trigger [11]. The results below are conditional on the selected NPI measures, as well as fixed values for  $R_0$  and the ICU triggers.

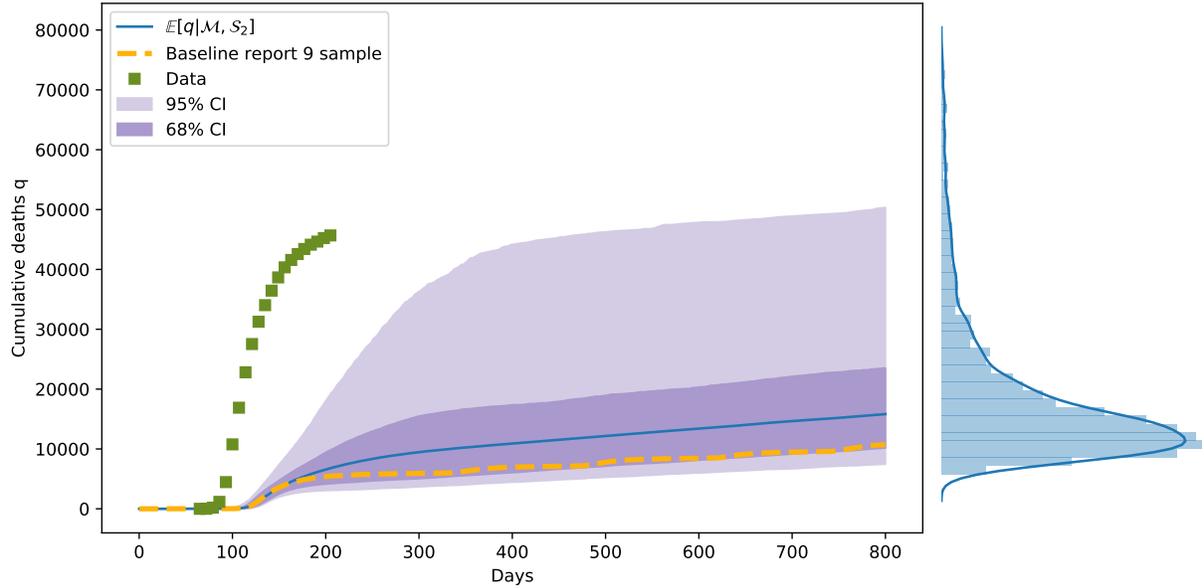
#### 4.2.1 Confidence intervals

We consider two different PC\_CI\_HQ\_SD suppression scenarios, as this class is the closest to actual NPIs that were implemented in the UK. Figure 1(a) shows the 68% and 95% confidence intervals of the cumulative death prediction, for scenario  $\mathcal{S}_1$  with  $R_0 = 2.4$ , and on/off ICU triggers of 60/15. This is one of the scenarios considered in Report 9. In addition, the pdf of the total death count after 800 days is plotted. The latter shows clear non-Gaussian behaviour, with a heavy tail towards a higher death count. The corresponding Report 9 total death count is 8700 [11]. The current version, which now supports averaging over stochastic realizations, predicts 9500 [31]. Our mean prediction of Eqn. (1) is almost double this amount. The Report 9 predictions are still captured by the distribution (at approximately the lower boundary of the 68% confidence interval), but the distribution also supports low-probability events which are about 5–6 times higher than those given in Report 9, at the upper bound of the 95% confidence interval.

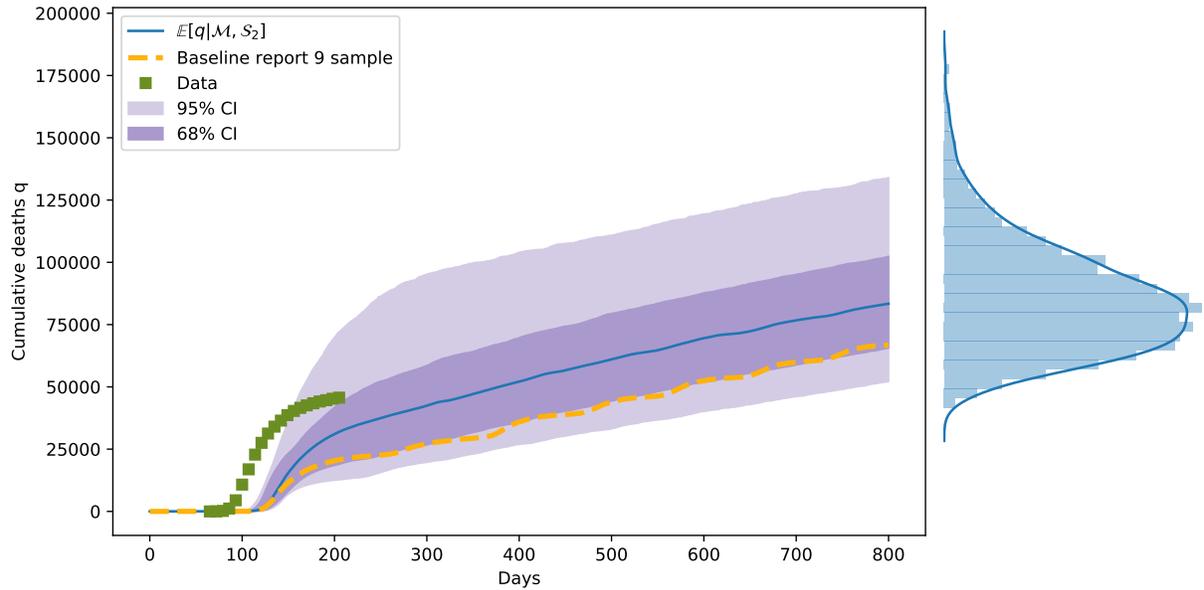
Note that the output distribution conditioned on  $\mathcal{S}_1$  clearly under predicts the observed death count in the UK, which is also plotted in Figure 1(a). We therefore selected  $\mathcal{S}_2$  using parameters from Report 9 that gave the highest predicted mortality, *i.e.*,  $R_0 = 2.6$  and ICU on/off triggers of 400/300 cases. The results are displayed in Figure 1(b). The deterministic Report 9 prediction is still located at the 68% confidence interval lower border. However, the total death count pdf is notably less skewed, although still not exactly symmetric.

In both cases, the bulk of the probability mass remains within the same order of magnitude as the predictions of Report 9, despite the fact that we vary 19 influential parameters. That said, Figure 1 clearly indicates that the results are very sensitive to  $\mathcal{S}$ . As noted, we also plot the observed death count validation data from [19] in both sub figures, which are clearly not captured well by the output distributions, although scenario  $\mathcal{S}_2$  does perform better than  $\mathcal{S}_1$ . It is also clear from Figure 1 that, in both cases, the rate of infection starts too slowly, and that it must be assumed that the epidemic started earlier than suggested in Report 9, which is in line with the findings of [1]. Hence, if one aims to validate CovidSim in a probabilistic sense (*i.e.*, obtaining a distribution which captures validation data with high probability), it is crucial to either tune the scenario parameters, or to quantify the scenario uncertainty.

Finally, let us emphasize that the authors of Report 9 did not claim that their parametrization at the time would be able to match the death count data of the coming months. The main message was that it would “be necessary to layer multiple interventions, regardless of whether suppression or mitigation is the overarching policy goal” [11], and it also showed that doing nothing at all would have disastrous consequences.



(a) Scenario  $\mathcal{S}_1$ :  $R_0 = 2.4$ , ICU on/off triggers 60/15.



(b) Scenario  $\mathcal{S}_2$ :  $R_0 = 2.6$ , ICU on/off triggers 400/300.

**Figure 1** The mean cumulative death prediction for scenario  $\mathcal{S}_1$  (top) and  $\mathcal{S}_2$  (bottom), plus confidence intervals, and at the right of each figure, the pdf of the total death count after 800 days. These results were obtained using a computational budget of 3000 CovidSim evaluations per scenario. Day 0 corresponds to January 1st, 2020. In addition, we plot the observed cumulative death count data for the UK (green squares) in both figures, obtained from [19]. The first data point is at March 6th 2020, which corresponds to day 66. The striped line is a single sample from CovidSim (current release), run with the baseline parameter values of Report 9.

### 4.2.2 Model structure uncertainty

We also reiterate that there is uncertainty in the model structure  $\mathcal{M}$ , as a different model might have given a better fit to the data, while still conditioned on the same scenario of the preceding section. For instance, during the pandemic it has become apparent that the COVID-19 spread in hospitals and care homes constituted a significant fraction of the overall spread, particularly in the UK [31]. The spread in these locations, which are not explicitly modelled in CovidSim, may also be a reason why the number of cases initially forecast with CovidSim was lower than the number that occurred in reality. Although precautions have been taken to reduce this spread, and the availability of personal protective equipment has improved, incorporating these factors will still be important for those models that need to be validated against data from the start of the pandemic.

Other missing epidemiological processes which might become important for future predictions are face masks and contact tracing. In March 2020, the beneficial effects of wearing face masks was still heavily contested [10]. However, research is now available that suggests that wearing a face mask reduces viral spread when coughing [40], and that it correlates on the population level with a reduced case incidence [26].

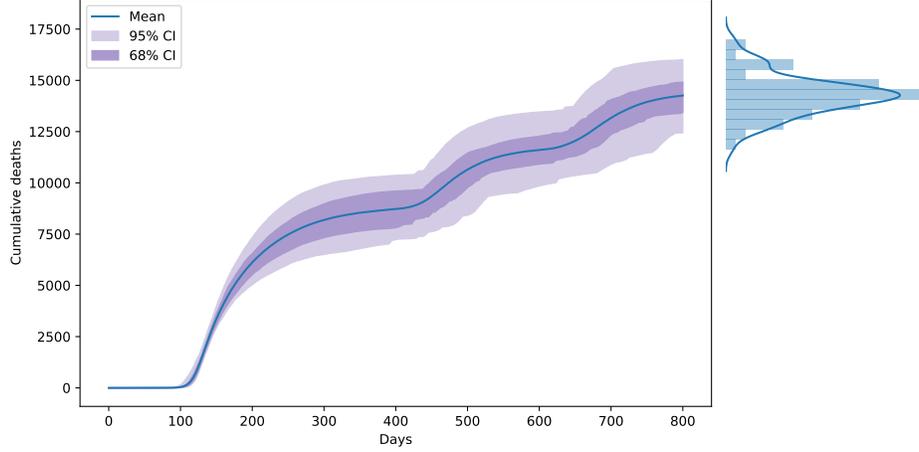
In many countries with low case prevalence, contact tracing is used to reduce the spread of COVID-19. Contact tracing capability was very limited in the UK during March 2020, but it has now improved and could be incorporated in future models. Here, the quality and extent of contact tracing are important, as imperfect contact tracing has a strongly reduced benefit [22].

### 4.2.3 Random seeds

CovidSim is stochastic, with 4 random seeds, specified via the command line. Two random seeds are used in the creation of the network of individuals mentioned in Section 1. The remaining seeds affect the interactions between individuals, controlling how they become infected and propagate infection. The role of the random seeds in the code is of some significance, but they do not play as large a role as the dominant parameters shown in the preceding section. Specifically, we varied the 4 random seeds, keeping all other parameters fixed, and compared the amount of output variance we obtain compared to varying the parameters with fixed seeds. The uncertainty due to the seeds is significantly smaller, see Figure 2. In light of these results, we do not vary the random seeds in our parametric uncertainty analysis.

### 4.2.4 Sensitivity analysis

With sensitivity analysis, the aim is to apportion the uncertainty of the model output to specific (combinations of) input parameter uncertainties. To this end, Sobol indices measure the fraction of the output variance that each combination of input parameters is responsible for, when



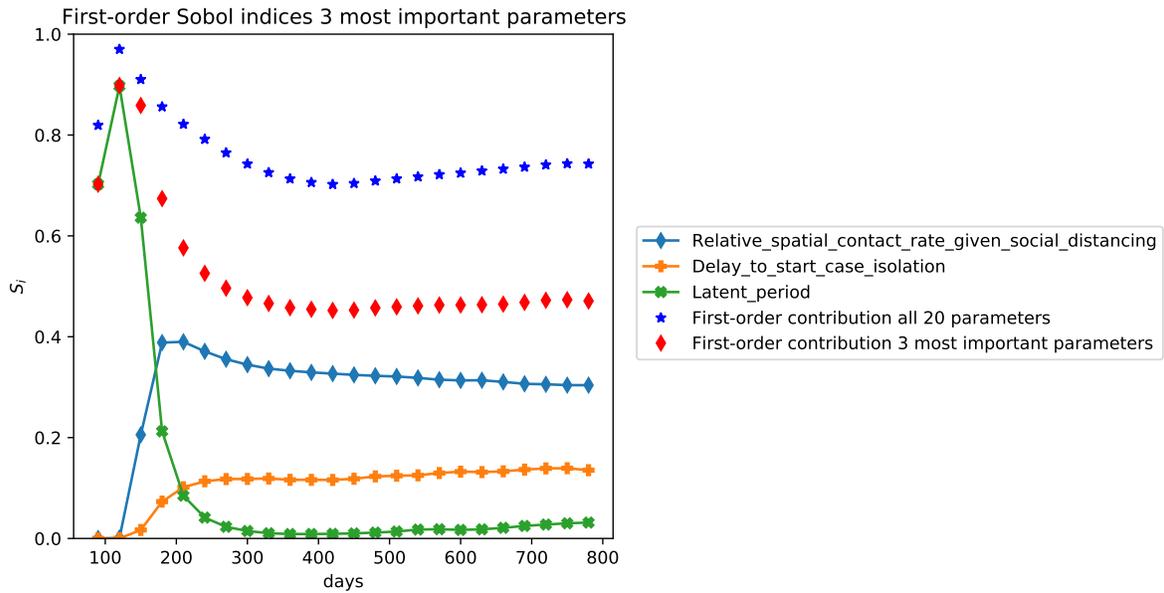
**Figure 2** The confidence intervals for the predicted cumulative deaths under scenario  $\mathcal{S}_1$ , varying the random seeds only. The seeds were sampled on a standard tensor grid sampling plan of 81 points. The variance is significantly smaller than in Figure 1(a), in which 20 input parameters were varied.

given a distribution on the inputs [34]. They can be computed in a post-processing step, once the input uncertainties are propagated through the computational model [20]; see the Supplementary Materials A.2.

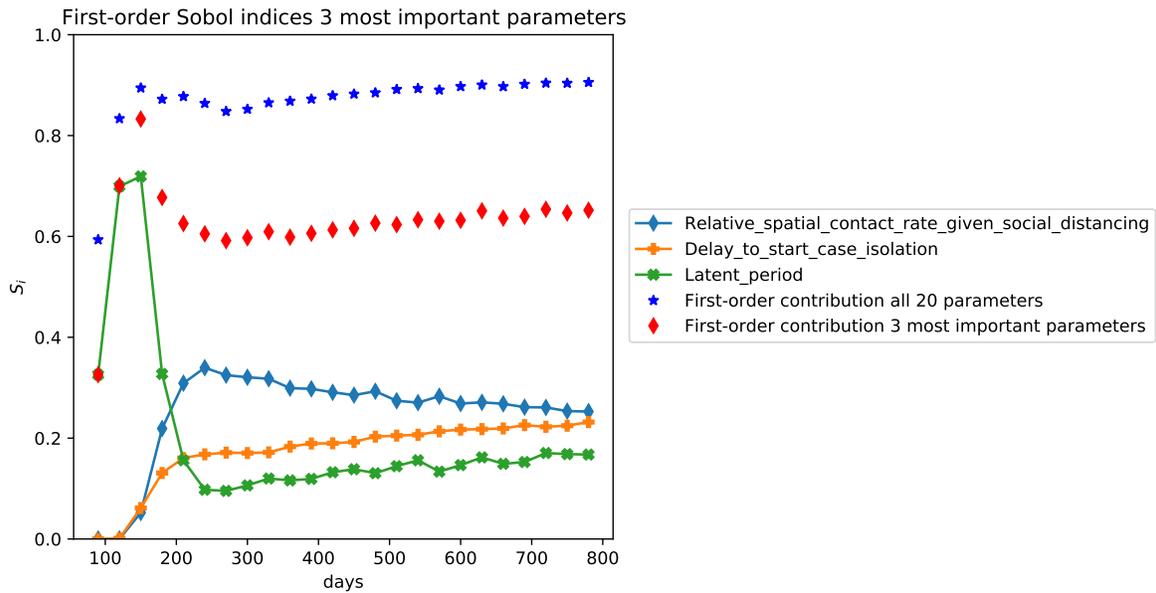
The first-order Sobol indices  $S_i$  are defined as  $S_i := \mathbb{V}[q_i]/\mathbb{V}[q] \in [0, 1]$ , for  $i = 1, \dots, d$ . Here  $\mathbb{V}[q]$  is the total output variance, and  $\mathbb{V}[q_i]$  is the partial variance attributed to one particular input parameter [34]. Figure 3 displays the three  $S_i$  with the highest values for scenarios  $\mathcal{S}_1$  and  $\mathcal{S}_2$ ; see the Supplementary Materials D for more results. The Sobol indices are plotted against time, which shows that in the beginning the latent period is the most influential, although only for a short amount of time. The latent period is the period in which a patient has no symptoms and is not infectious. A longer latent period therefore means that the rate of disease spread is slower in this early exponential growth stage, when there are still relatively few cases present.

The second important parameter is the ‘Relative spatial contact rate given social distancing’ parameter, which indicates the assumed effectiveness of social distancing. Finally, the third parameter (in both scenarios), to dominate the variance is the ‘Delay to start case isolation’.

In Figure 3 we also plot the sum of all 19 first-order Sobol indices. This shows that first-order effects, *i.e.*, the fraction of the variance obtained by varying individual parameters, account for a little under 80% in the case of  $\mathcal{S}_1$ , and roughly 90% of the variance for  $\mathcal{S}_2$ . Conversely, interaction effects between parameters therefore account for no more than 10–20% in our chosen scenarios. We also show the sum of the first-order indices of just the three most important parameters, *i.e.*, those actually plotted in Figure 3, which already accounts for roughly 50% and 67% of the observed variance in cumulative deaths for  $\mathcal{S}_1$  and  $\mathcal{S}_2$  respectively.



(a) Scenario  $S_1$ :  $R_0 = 2.4$ , ICU on/off triggers 60/15.



(b) Scenario  $S_2$ :  $R_0 = 2.6$ , ICU on/off triggers 400/300.

**Figure 3** The first-order Sobol indices for two scenarios, of the same 3 most dominant parameters, plotted against time at one month intervals. It shows the fraction of the variance that each parameter is responsible for, over time. In addition, we show the sum of all 19 first-order indices (blue stars). The sum of the 3 most dominant parameters is also shown (red diamonds).

Finally, we note that although  $\mathcal{S}_1$  and  $\mathcal{S}_2$  share the same three most influential parameters, the ranking differs from the fourth parameter onward. More details can be found in Section D.2 of the Supplementary Materials.

#### 4.2.5 Uncertainty amplification

Although we based our input distributions (see Table 2 of Supplementary Materials B), on a combination of available data and expert knowledge, (in general) a certain level of ambiguity remains with respect to the choice of input distribution. In the Supplementary Materials C, we therefore devise a measure that examines the *amplification* of uncertainty in the outputs, with respect to a given set of input distributions. This relative measure of output-to-input variability is based on the coefficients of variation ratio (*CVR*), and is given by

$$CVR := CV(\bar{q})/CV(\bar{\xi}) = \left( \frac{1}{N} \sum_{n=1}^N \frac{\sigma_{q_n}}{\mu_{q_n}} \right) \bigg/ \left( \frac{1}{d} \sum_{i=1}^d \frac{\sigma_{\xi_i}}{\mu_{\xi_i}} \right). \quad (2)$$

A coefficient of variation is a dimensionless quantity that measures the variability of a random variable with respect to its mean, and is defined as the standard deviation over the mean ( $\sigma/\mu$ ). In (2),  $CV(\bar{q})$  and  $CV(\bar{\xi})$  are the mean coefficients of variation of the output  $q \in \mathbb{R}^N$  and input  $\xi \in \mathbb{R}^d$ , respectively. The results for CovidSim using (2) are displayed in Table 1, which shows that the uncertainty in the input is amplified by almost a factor of 3 for scenario 1. In contrast, CovidSim is more robust under  $\mathcal{S}_2$ , in which case the same input uncertainty is still amplified to the output, although now by a factor of 2.

scenario	$CV(\bar{\xi})$	$CV(\bar{q})$	<i>CVR</i>
$\mathcal{S}_1$	0.1950	0.6097	3.13
$\mathcal{S}_2$	0.1950	0.3872	1.99

**Table 1** The mean coefficients of variation for the input and output, and the ratio of the two, which serves as our robustness score.

## 5 Summary of results

Conditional on a given scenario  $\mathcal{S}$ , we found that the Report 9 predictions are captured by the parametric uncertainty at the lower bound of the 68% confidence interval. Most of the probability mass contains predictions with an order of magnitude similar to those predictions. However, the pdf

of the total death count is skewed, and can support low probability events with a predicted death count that is about 5 to 6 times higher at the upper bound of the 95% confidence interval.

The amount of output uncertainty will clearly depend upon the variance in the chosen input distributions. For this reason we use a measure that looks at the relative variability between the output and the input. We find that CovidSim amplifies the input uncertainty, to the extent of 300%, depending on the chosen NPI scenario. Despite this amplification of uncertainty, the distribution of the output does not envelope available validation data well for the two scenarios we considered. We do note, however, that the predictions will be very sensitive to the chosen scenario  $S$ , which therefore must be tuned if one wishes to validate CovidSim against available data; see *e.g.* [1].

The Sobol sensitivity analysis revealed that only three parameters are responsible for (up to) 67% of the observed output variance. Hence, if one wishes to reduce the amount of output uncertainty, refining the estimates of these three input distributions based on observational data would be a clear priority.

## 6 Conclusion

The code is reproducible in original and new forms; it has a colossal 940 parameters, from which we selected 60 that are relevant for our assessment of uncertainty. Out of these 60, we extracted 19 input parameters that contribute most to the variance. Our analysis shows that a mere three cover almost two-thirds of the variance induced by these 19 parameters. CovidSim can amplify uncertainties present in the inputs to those manifested in outputs to the tune of approximately 300% for certain NPI scenarios. The limited data now available to compare against the model's predictions from March 2020 exhibit discrepancies on a scale that is well beyond the bounds of reasonable probabilistic model outcomes, obtained by propagating parametric input uncertainty.

To retrofit the model's outputs with the observed data requires additional *post-hoc* tuning of certain parameters that control the scenario in which the model is applied. In addition, CovidSim omits a mathematical description of various processes now known to play an important role in the spread of Covid-19.

All these issues need to be addressed in seeking to provide a more quantitative albeit still strongly probabilistic version of the code which might be suitable for its future application in healthcare and governmental decision-making. Our findings exemplify how sensitivity analysis and uncertainty quantification can help improve model development efforts, and in this case support the creation of epidemiological forecasting with quantified uncertainty.

As an alternative to retrofitting the scenario parameters, one could attempt to quantify the uncertainty related to the scenario the model is applied in. One such potential route for future research could involve creating cheap surrogate models for CovidSim, *e.g.* in the stochastic space of

the identified most influential parameters, which opens up the possibility of Bayesian inference [27]. Repeating the procedure for a discrete set of scenario parameters then allows for the combined estimation of the parametric and the scenario uncertainty using Bayesian ensemble methods; see *e.g.* [28, 7].

## Acknowledgements

This project was undertaken on behalf of the Royal Society’s RAMP initiative. The work was funded as part of the European Union Horizon 2020 research and innovation programme under grant agreements #800925 (VECMA project, [www.vecma.eu](http://www.vecma.eu)) and #823712 (CompBioMed2 Centre of Excellence, [www.compbiomed.eu](http://www.compbiomed.eu)), as well as the UK EPSRC for the UK High-End Computing Consortium (EP/R029598/1). We are grateful to Mike Cates, Graeme Ackland and Ken Rice for helpful discussions while this work was in progress. Finally, we would like to thank Prof. Neil Ferguson for his support of our investigation and his response to our questions.

## Data availability

Figures 1(a) and 1(b) display publicly available cumulative death count data for the UK, obtained from [19]. Furthermore, the ‘parameter list’ folder in [39] contains a spreadsheet with all input parameters, their default values and reasons for inclusion or exclusion from the Imperial College CovidSim team.

## Code availability

The version of EasyVVUQ that was used to generate our results has been pushed to a separate, publicly available GitHub branch, see [38]. Likewise, the FabSim3 interface between EasyVVUQ and CovidSim, that was used to execute the ensembles on the PNSC Eagle supercomputer, can be found here [39]. This repository contains the ‘covid\_automated.py’ file, which is the main Python script, located in the ‘final\_adaptive\_covid\_easyvвуq’ folder.

## References

- [1] G.J. Ackland, K. Rice, B.M. Wynne, and V. Martin. The long term predictions from imperial college covidsim report 9. *medRxiv*, doi: <https://doi.org/10.1101/2020.06.18.20135004>, 2020.

- [2] G.T. Buzzard. Global sensitivity analysis using sparse grid interpolation and polynomial chaos. *Reliability Engineering & System Safety*, 107:82–89, 2012.
- [3] A. Clark, M. Jit, C. Warren-Gash, B. Guthrie, H.H.X. Wang, S.W. Mercer, C. Sanderson, M. McKee, C. Troeger, K.L. Ong, et al. Global, regional, and national estimates of the population at increased risk of severe COVID-19 due to underlying health conditions in 2020: a modelling study. *The Lancet Global Health*, 2020.
- [4] N.G. Davies, A.J. Kucharski, R.M. Eggo, A. Gimma, W.J. Edmunds, T. Jombart, K. O’Reilly, A. Endo, J. Hellewell, E.S. Nightingale, et al. 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*, 2020.
- [5] D. Draper. Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):45–70, 1995.
- [6] R.P. Dwight, S.G.L. Desmedt, and P.S. Omrani. Sobol indices for dimension adaptivity in sparse grids. In *Simulation-Driven Modeling and Optimization*, pages 371–395. Springer, 2016.
- [7] W.N. Edeling, P. Cinnella, and R.P. Dwight. Predictive RANS simulations via Bayesian Model-Scenario Averaging. *Journal of Computational Physics*, 275:65–91, 2014.
- [8] S. Eglén. Codecheck certificate 2020-010. [https://zenodo.org/record/3865491#.XuPW\\_y-ZPGI](https://zenodo.org/record/3865491#.XuPW_y-ZPGI).
- [9] J. Feinberg and H.P. Langtangen. Chaospy: An open source tool for designing methods of uncertainty quantification. *Journal of Computational Science*, 11:46–57, 2015.
- [10] S. Feng, C. Shen, N. Xia, W. Song, M. Fan, and B.J. Cowling. Rational use of face masks in the covid-19 pandemic. *The Lancet Respiratory Medicine*, 8(5):434–436, 2020.
- [11] N. Ferguson, D. Laydon, G. Nedjati-Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunubá Perez, G. Cuomo-Dannenburg, et al. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand.
- [12] N.M. Ferguson, D.A.T. Cummings, C. Fraser, J.C. Cajka, P.C. Cooley, and D.S. Burke. Strategies for mitigating an influenza pandemic. *Nature*, 442(7101):448–452, 2006.
- [13] B. Ganapathysubramanian and N. Zabaras. Sparse grid collocation schemes for stochastic natural convection problems. *Journal of Computational Physics*, 225(1):652–685, 2007.
- [14] T. Gerstner and M. Griebel. Numerical integration using sparse grids. *Numerical Algorithms*, 3(18):209–232, 1998.

- [15] T. Gerstner and M. Griebel. Dimension-adaptive tensor-product quadrature. *Computing*, 71(1):65–87, 2003.
- [16] M. Griebel and M. Holtz. Dimension-wise integration of high-dimensional functions with applications to finance. *Journal of Complexity*, 26(5):455–489, 2010.
- [17] D. Groen, A.P. Bhati, J. Suter, J. Hetherington, S.J. Zasada, and P.V. Coveney. Fabsim: facilitating computational research through automation on large-scale and distributed e-infrastructures. *Computer Physics Communications*, 207:375–385, 2016.
- [18] D. Groen, R.A. Richardson, D.W. Wright, V. Jancauskas, R. Sinclair, P. Karlshofer, M. Vassaux, H. Arabnejad, T. Piontek, P. Kopta, et al. Introducing VECMAtk-verification, validation and uncertainty quantification for multiscale and HPC simulations. In *International Conference on Computational Science*, pages 479–492. Springer, 2019.
- [19] GOV.UK Coronavirus (COVID19) in the UK. Deaths in United Kingdom. <https://coronavirus-staging.data.gov.uk/deaths>.
- [20] J.D. Jakeman, M.S. Eldred, G. Geraci, and A. Gorodetsky. Adaptive multi-index collocation for uncertainty quantification and sensitivity analysis. *International Journal for Numerical Methods in Engineering*, 121(6):1314–1343, 2020.
- [21] K.L. Judd, L. Maliar, S. Maliar, and R. Valero. Smolyak method for solving dynamic economic models: Lagrange interpolation, anisotropic grid and adaptive domain. *Journal of Economic Dynamics and Control*, 44:92–123, 2014.
- [22] M.J. Keeling, T.D. Hollingsworth, and J.M. Read. Efficacy of contact tracing for the containment of the 2019 novel coronavirus (covid-19). *J Epidemiol Community Health*, 2020.
- [23] K. Kelland. Sobering coronavirus study prompted Britain to toughen its approach. <https://www.reuters.com/article/us-health-coronavirus-britain-research/sobering-coronavirus-study-prompted-britain-to-toughen-its-approach-idUSKBN2141EP>.
- [24] A.J. Kucharski, T.W. Russell, C. Diamond, Y. Liu, J. Edmunds, S. Funk, R.M. Eggo, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The Lancet. Infectious Diseases*, 20(5):553, 2020.
- [25] D. Loukrezis, U. Römer, and H. De Gerssem. Assessing the performance of Leja and Clenshaw-Curtis collocation for computational electromagnetics with random input data. *International Journal for Uncertainty Quantification*, 9(1), 2019.
- [26] W. Lyu and G.L. Wehby. Community Use Of Face Masks And COVID-19: Evidence From A Natural Experiment Of State Mandates In The US. *Health Affairs*, page 10.1377/hlthaff, 2020.

- [27] Y. Marzouk and D. Xiu. A stochastic collocation approach to Bayesian inference in inverse problems. *Communications in Computational Physics*, 6(4):826–847, 2009.
- [28] P.D. Meyer, M. Ye, M.L. Rockhold, S.P. Neuman, and K.J. Cantrell. Combined estimation of hydrogeologic conceptual model, parameter, and scenario uncertainty with application to uranium transport at the Hanford Site 300 Area. Technical report, Pacific Northwest National Lab.(PNNL), Richland, WA (United States), 2007.
- [29] MRC Centre for Global Infectious Disease Analysis. COVID-19 CovidSim Model - Report 9 folder. <https://github.com/mrc-ide/covid-sim/tree/master/report9>.
- [30] A. Narayan and J.D. Jakeman. Adaptive Leja sparse grid constructions for stochastic collocation and high-dimensional approximation. *SIAM Journal on Scientific Computing*, 36(6):A2952–A2983, 2014.
- [31] Office for National Statistics. Deaths involving COVID-19 in the care sector, England and Wales Deaths involving COVID-19 in the care sector, England and Wales: deaths occurring up to 12 June 2020 and registered up to 20 June 2020 (provisional). <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/articles/deathsinvolvingcovid19inthecaresectorenglandandwales/latest>.
- [32] Poznan Supercomputing and Networking Center HPC WIKI . The Eagle supercomputer. <https://wiki.man.poznan.pl/hpc/index.php?title=Eagle>.
- [33] R.A. Richardson, D.W. Wright, W. Edeling, V. Jancauskas, J. Lakhilili, and P.V. Coveney. EasyVVUQ: A Library for Verification, Validation and Uncertainty Quantification in High Performance Computing. *Journal of Open Research Software*, 8(1), 2020.
- [34] I.M. Sobol. On sensitivity estimation for nonlinear mathematical models. *Matematicheskoe Modelirovanie*, 2(1):112–118, 1990.
- [35] I.M. Sobol. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55(1-3):271–280, 2001.
- [36] B. Sudret. Polynomial chaos expansions and stochastic finite element methods. *Risk and Reliability in Geotechnical Engineering*, pages 265–300, 2015.
- [37] Bruno Sudret. Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering and System Safety*, 93(7):964–979, 2008.
- [38] VECMA Consortium. EasyVVUQ software - CovidSim branch. <https://github.com/UCL-CCS/EasyVVUQ/tree/CovidSim>.
- [39] VECMA Consortium. FabSim3 - EasyVVUQ interface for CovidSim. <https://github.com/arabnejad/FabCovidsim/tree/dev>.

- [40] S. Verma, M. Dhanak, and J. Frankenfield. Visualizing the effectiveness of face masks in obstructing respiratory jets. *Physics of Fluids*, 32(6):061708, 2020.
- [41] D.W. Wright, R.A. Richardson, W.N. Edeling, J. Lakhili, R.C. Sinclair, V. Jancauskas, D. Suleimenova, B. Bosak, M. Kulczewski, T. Piontek, P. Kopta, I. Chirca, H. Arabnejad, O.O. Luk, O. Hoenen, J. Węglarz, D. Crommelin, D. Groen, and P.V. Coveney. Building Confidence in Simulation: Applications of EasyVVUQ. *Advanced Theory and Simulations*, 3(8), 2020.
- [42] D. Xiu and G.E. Karniadakis. The Wiener–Askey polynomial chaos for stochastic differential equations. *SIAM Journal on Scientific Computing*, 24(2):619–644, 2002.