The Impact of Uncertainty on the CovidSim Pandemic Code

INI Event: The Role of Uncertainty in Mathematical Modelling of Pandemics

February 8th, 2022

Wouter Edeling, Hamid Arabnejad, Robbie Sinclair, Diana Suleimenova, Krishnakumar Gopalakrishnan, Bartosz Bosak, Derek Groen, Imran Mahmood, Daan Crommelin, and Peter Coveney.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 800925.

CovidSim: model

- CovidSim: an individual-based epidemiological code (Imperial College).
- Modified from an earlier influenza version.
- March 2020: predict effect of Non-Pharmaceutical Interventions (NPIs) to reduce spread of COVID19 in UK.

- Creates a network of individuals based on population density data.
- Individual interact within places:
  - Household
  - Schools
  - Universities
  - Workplaces

- Influential model: key paper (Report 9 [1]) was (in part) responsible for reorienting UK policy from herd immunity to suppression.

- We performed uncertainty analysis on behalf of UK RAMP team.

[1] Ferguson, Neil, et al. "Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand." (2020).

CovidSim: uncertainty

- CovidSim is uncertain:

  1) Parametric uncertainty: input parameter values $\boldsymbol{\xi} \in \mathbb{R}^d$ are not known exactly.

  2) Model-form uncertainty: what is the appropriate mathematical structure of the model? For instance: missing epidemiological processes or missing places types.

  3) Scenario uncertainty: uncertainty in the scenario $\mathcal{S}$ under which the model $\mathcal{M}$ is applied. For instance: initial conditions or selected NPI measures.

- The output Quantity of Interest (QoI): $q = q\left(\boldsymbol{\xi}, \mathcal{M}, \mathcal{S}\right)$

- We quantified the uncertainty in q due to parametric uncertainty, and indirectly due to scenario uncertainty.

CovidSim: parametric robustness

- Specifically, we looked at robustness:
  - How much is parametric uncertainty amplified from the input to the output?
- Involves computing mean and variance of q:

$$\mathbb{E}\left[q \mid \mathcal{M}, \mathcal{S}\right] := \int_{\Omega_{\boldsymbol{\xi}}} q\left(\boldsymbol{\xi}, \mathcal{M}, \mathcal{S}\right) p\left(\boldsymbol{\xi}\right) \mathrm{d}\boldsymbol{\xi},$$

$$\mathbb{V}\left[q \mid \mathcal{M}, \mathcal{S}\right] = \int_{\Omega_{\boldsymbol{\xi}}} \left(q - \mathbb{E}[q \mid \mathcal{M}, \mathcal{S}]\right)^2 p(\boldsymbol{\xi}) d\boldsymbol{\xi}$$

- Important: results are conditional on the model and application scenario.
- Problem 1: $\boldsymbol{\xi} \in \mathbb{R}^d$ with $d = 940$.
- Problem 2: $p\left(\boldsymbol{\xi}\right)$ is not given, but must be chosen.

CovidSim: input dimension

- Problem 1: $\boldsymbol{\xi} \in \mathbb{R}^d$ with $d = 940$.

  - Many are not relevant, manual selection: 60 inputs left
  - Divided these into 3 groups:

  1) Intervention parameters, e.g. "Household compliance w/ quarantine"
  2) Biological disease parameters, e.g. "Latent period".
  3) Spatial / Geographic parameters, e.g. "Relative place contact rates".

  - Perform preliminary UQ on each group & identify important inputs.
  - Final UQ campaign: include all identified important inputs: 19 inputs.

- Problem 2: $p\left(\boldsymbol{\xi}\right)$ is not given, but must be chosen.

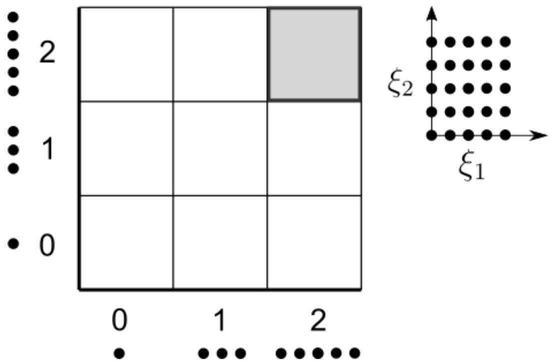
  - We used expert opinion: CovidSim developers.
  - We used uniform inputs due to lack of knowledge.

CovidSim: input dimension

- Final campaign: 19 inputs.
  - This is still quite high!
  - EasyVVUQ: (Quasi) Monte Carlo, Polynomial Chaos, Stochastic Collocation.
  - None are ideal: we implemented a dimension-adaptive SC sampler [2].

- Stochastic Collocation:
  - Polynomial expansion
  - Basic building blocks: 1D quadrature rules of m points.

  - Standard SC: extend to d dimensions
  - via a single tensor product.
  - Computational cost: m^d

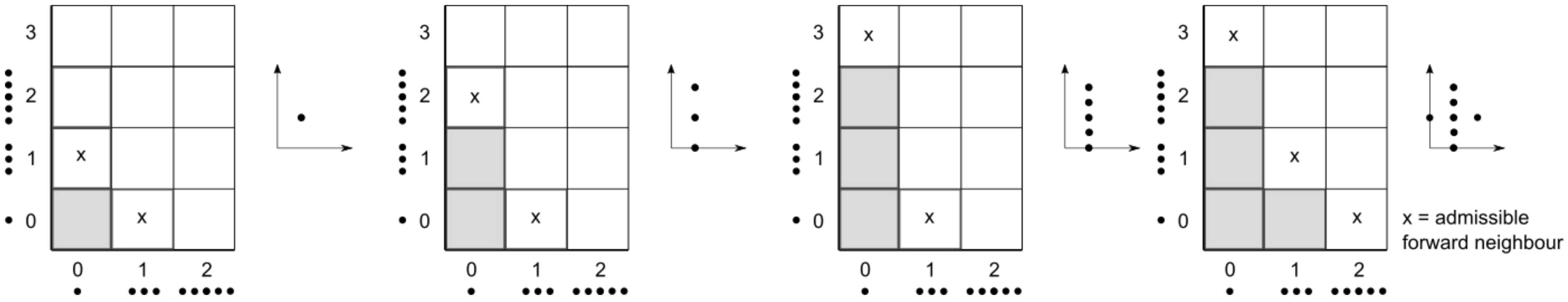


$$q(\boldsymbol{\xi}) \approx \tilde{q}(\boldsymbol{\xi}) = \sum_{j_1=1}^{m_1} \cdots \sum_{j_d=1}^{m_d} q(\xi_{j_1}, \cdots, \xi_{j_d})\, a_{j_1}(\xi_1) \otimes \cdots \otimes a_{j_d}(\xi_d)$$

[2] Gerstner, Thomas, and Michael Griebel. "Dimension–adaptive tensor–product quadrature." Computing 71.1 (2003): 65-87.

Curse of dimensionality

- For $d = 19, \; m^d$ is far too expensive.
- Dimension-adaptive sampling: postpone the curse of dimensionality
- Basic idea:
  - **Initialize**: start with a single sample.
  - **Look ahead**: evaluate code in 'candidate directions'
  - **Rank order**: compute error metric for all directions
  - **Adapt**: only add direction with highest error to sampling plan

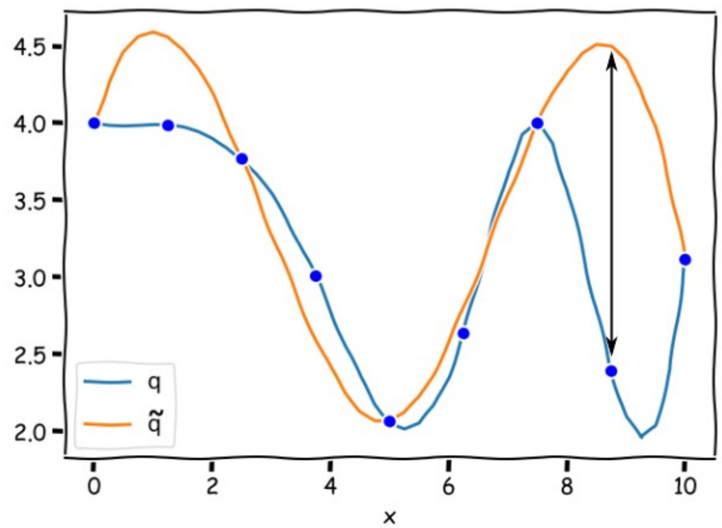


Error measure

- **Rank order**: compute error metric for all directions

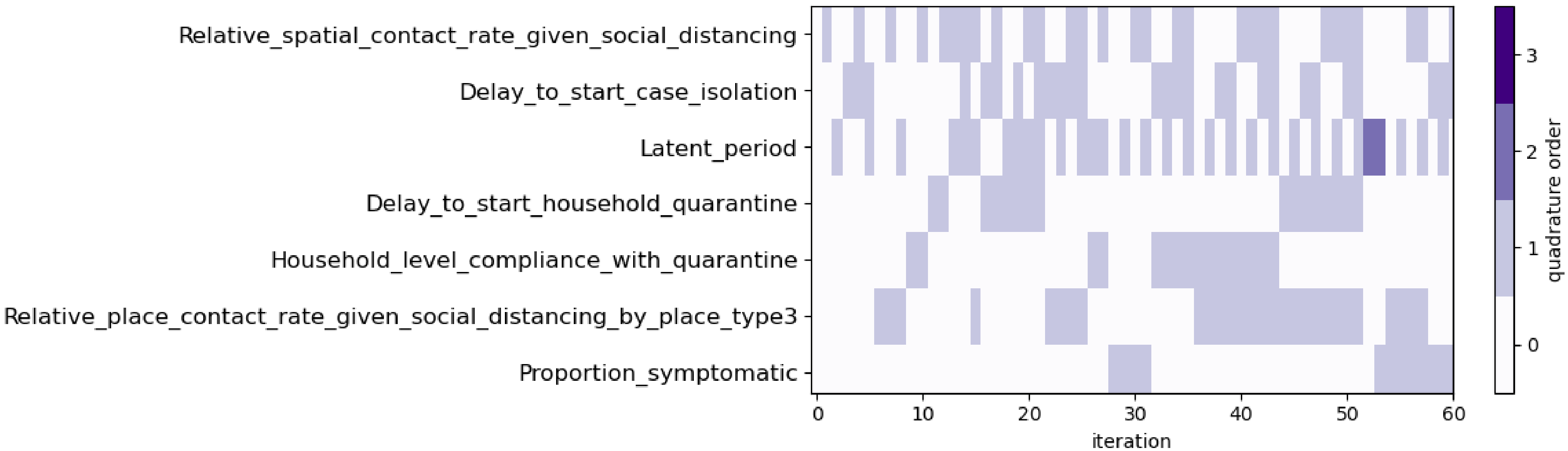
- Common error choice: hierarchical surplus
  - Difference between code q and SC interpolation $\widetilde{q}$ at new candidate inputs

$$\Delta\left(\boldsymbol{\xi}_j^{(l)}\right) := q\left(\boldsymbol{\xi}_j^{(l)}\right) - \tilde{q}_\Lambda\left(\boldsymbol{\xi}_j^{(l)}\right), \quad \boldsymbol{\xi}_j^{(l)} \in X_l \backslash X_\Lambda.$$



Adaptation

- Visualization of adaptation for CovidSim:

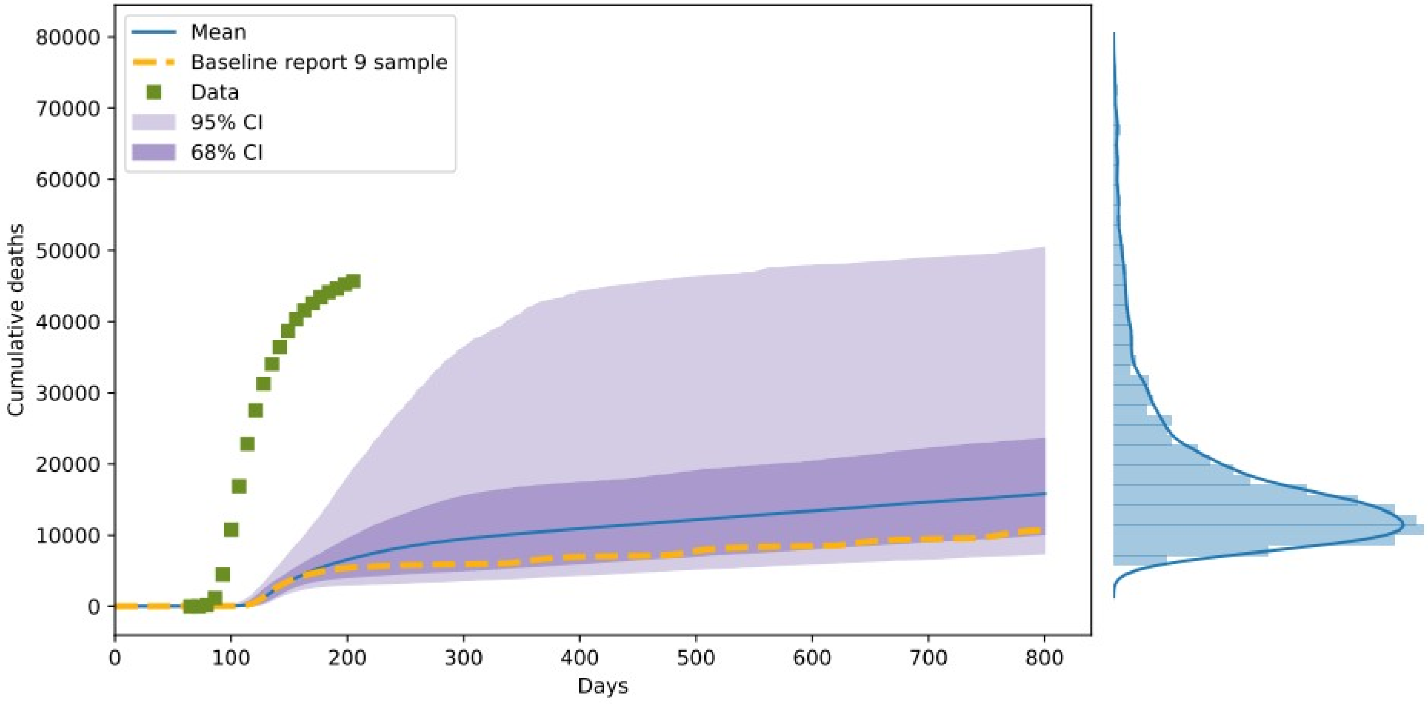


- After 60 iterations, 7 / 19 inputs are refined to quadrature order > 0.

- Each iteration: new EasyVVUQ ensemble of CovidSim samples.
  - Used FabSim3 to submit EasyVVUQ jobs to the PSNC Eagle machine.
  - Each CovidSim sample used 28 cores.

Results: confidence intervals

- Confidence intervals cumulative deaths, for a given scenario $\mathcal{S}_1$ from Report 9:
  - $R_0 = 2.4$, turn NPIs on: weekly new ICU patients > 60, turn NPIs off < 15.

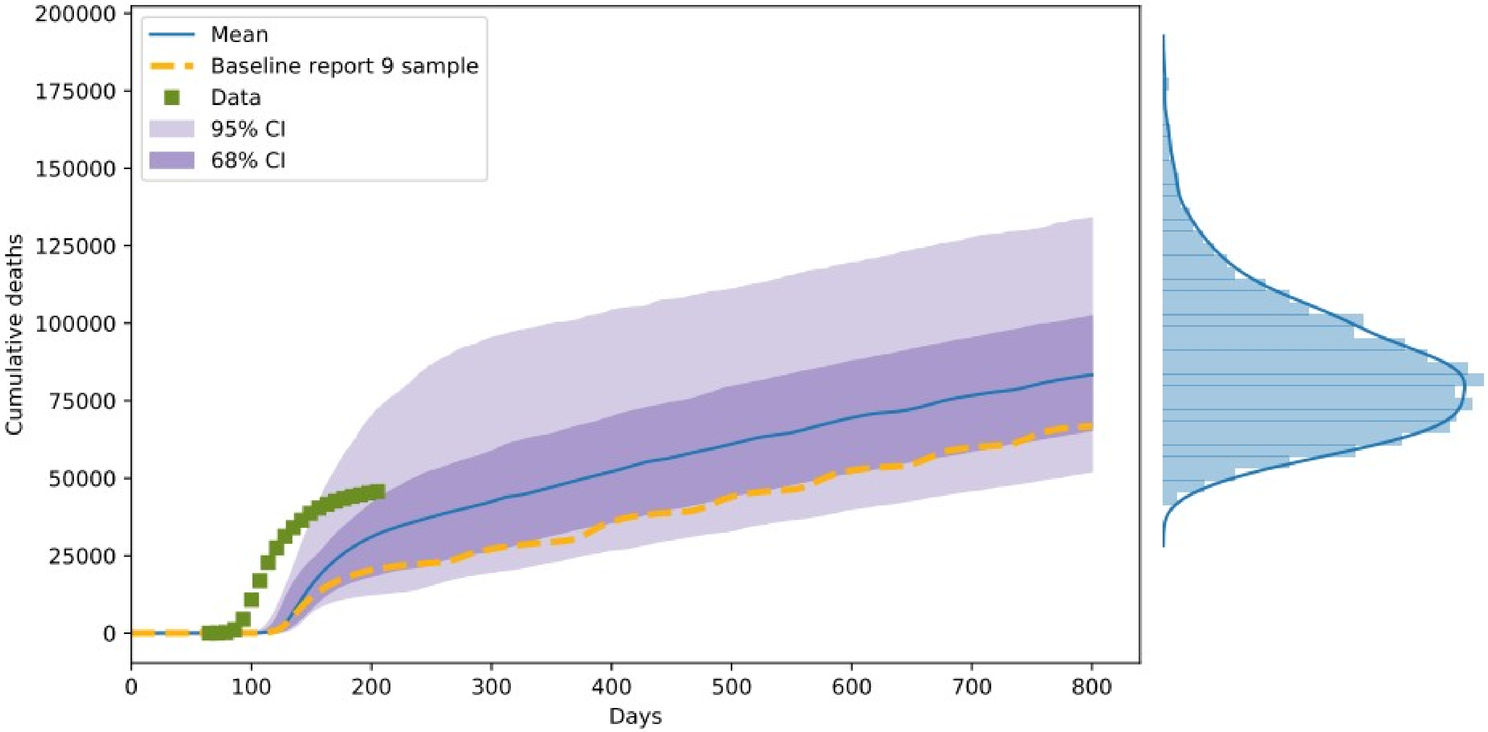


- Despite large variance, does not capture data: <u>uncertainty in initial condition is missing.</u> With post-hoc tuning bias can be fixed [4].

[4] Ackland, Graeme J., et al. "The long term predictions from imperial college CovidSim Report 9." medRxiv (2020).

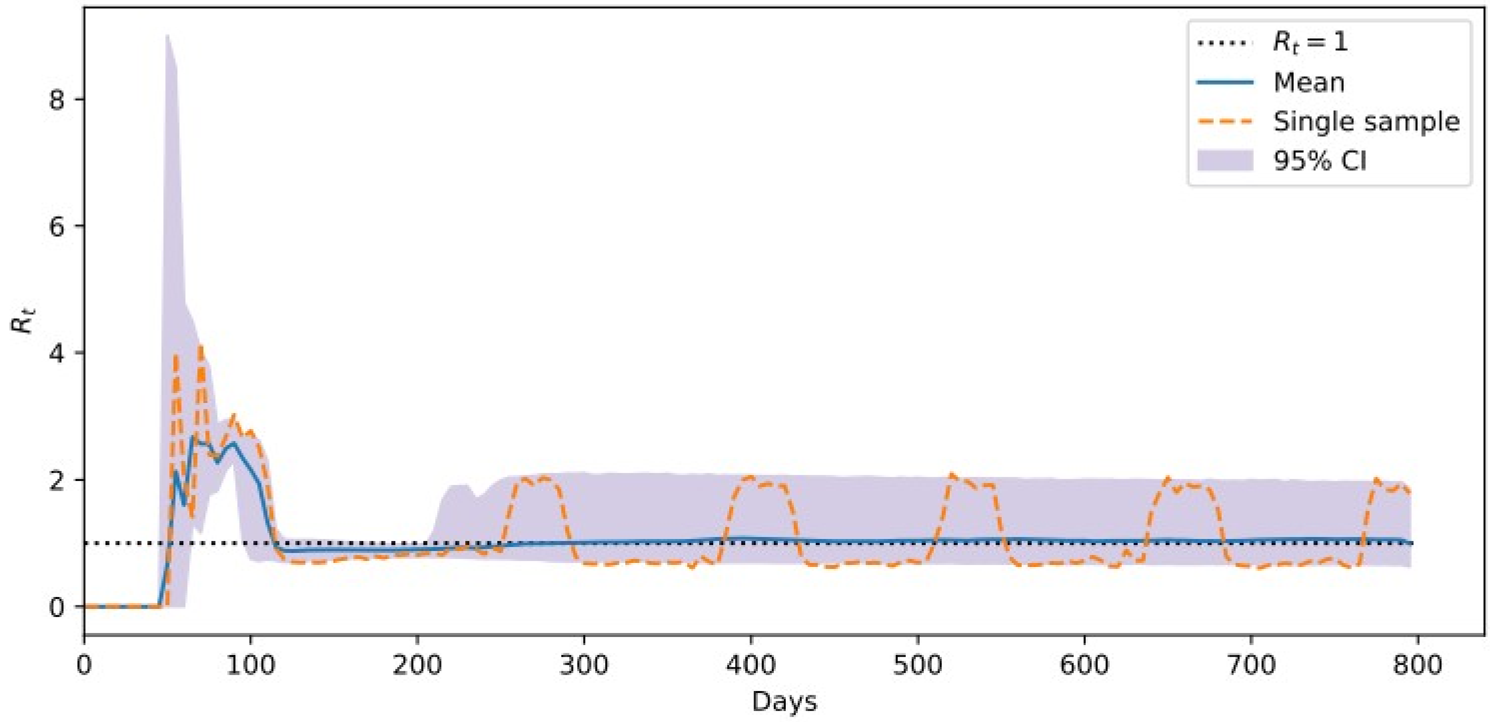
Results: confidence intervals

- Confidence intervals, for a different scenario $\mathcal{S}_2$:
  - $R_0 = 2.6$, turn NPIs on: weekly new ICU patients > 400, turn NPIs off < 300.



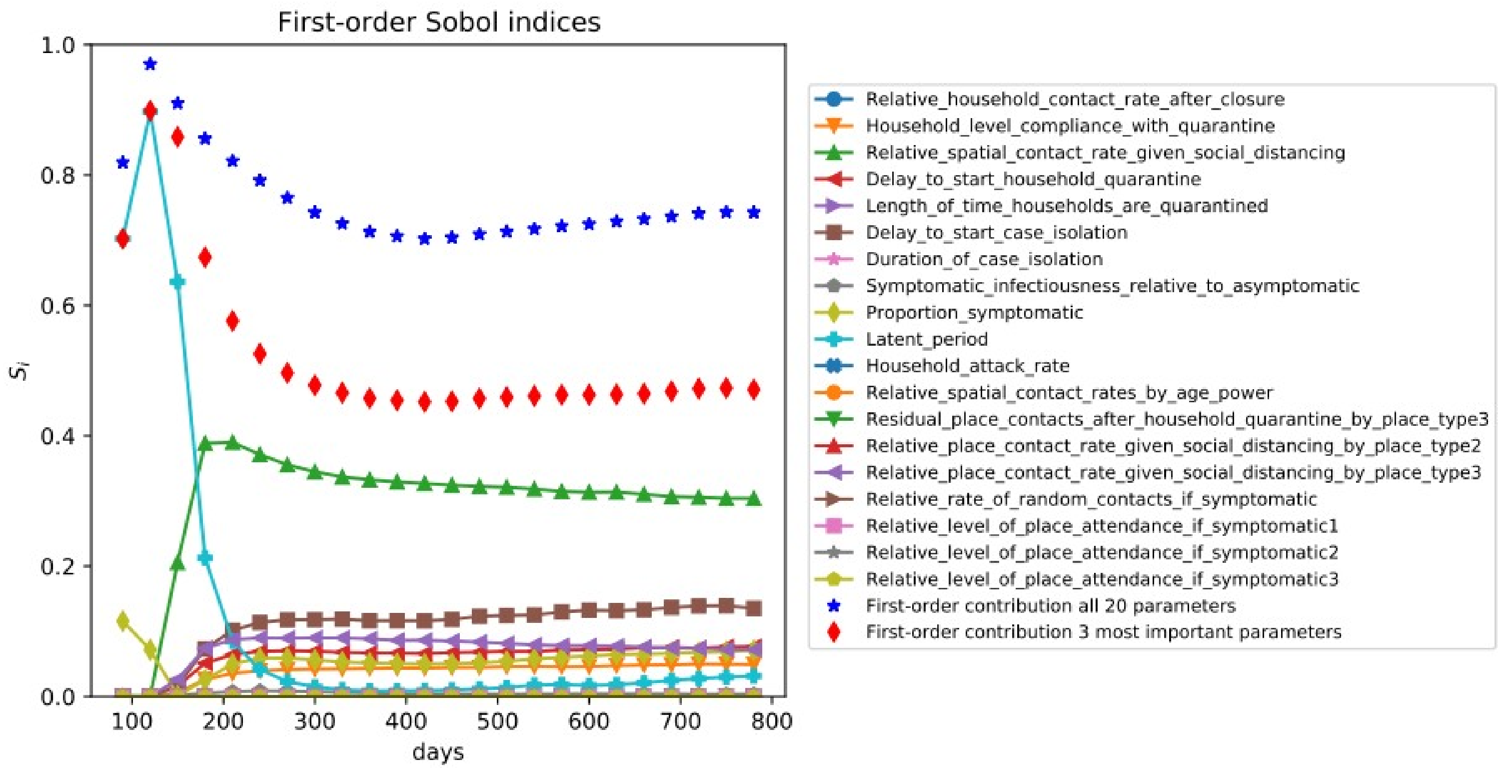
Results: confidence intervals

- Confidence intervals, for a different QoI: R_t



Results: Sobol indices

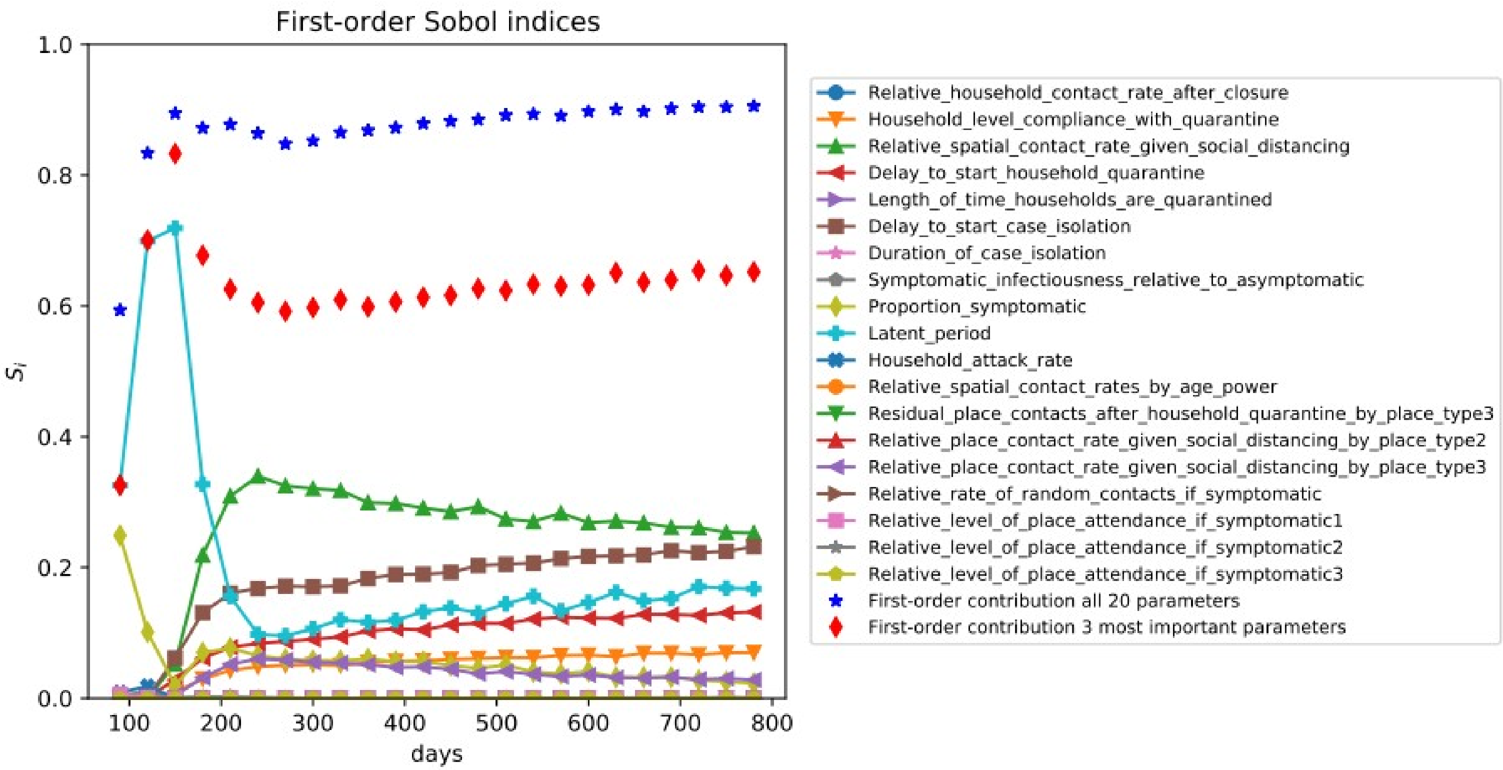
- First-order Sobol indices for scenario $\mathcal{S}_1$:
  - First-order index: measures fraction of output variance attributed to each input parameter. Computed using method of [3].



[3] Jakeman, John D., et al. "Adaptive multi-index collocation for uncertainty quantification and sensitivity analysis." International Journal for Numerical Methods in Engineering 121.6 (2020): 1314-1343.

Results: Sobol indices

- First-order Sobol indices for scenario $\mathcal{S}_2$: qualitatively similar.
  - Overall, NPI-related parameters are most important
  - Note that just 3 inputs already cover > 60% of the variance.



Results: robustness

ECMA

- Robustness: examine the amplification (or damping) of uncertainty.
- General question: *is the observed output variance large compared to the "amount" of uncertainty injected at the input?*
- Sobol indices do not measure this, we look at the Ratio of the average input – output Coefficients of Variation (CVR)
- CV = standard deviation / mean, a dimensionless measure of variability.

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

- CVR > 1: amplification of uncertainty, CVR < 1: damping.

Results: robustness

- CovidSim results, for scenario $\mathcal{S}_1$ and $\mathcal{S}_2$ and cumulative death QoI:

scenario	$CV\left(\bar{\boldsymbol{\xi}}\right)$	$CV\left(\bar{q}\right)$	CVR
$\mathcal{S}_1$	0.1950	0.6097	3.13
$\mathcal{S}_2$	0.1950	0.3872	1.99

- Both scenarios amplify uncertainty, by a factor of 3 and 2.

Conclusion

- We performed adaptive parametric UQ on CovidSim.
- It has 940 parameters, which we iteratively brought down to 19.
- Of the 19, just 3 already cover > 60% of the output variance.
- It is also important to:
  - Quantify the scenario uncertainty as well.
  - Possibly, use an ensemble of models to estimate the model-choice uncertainty.
  - Investigate calibration / data assimilation.

Article

ECMA

In case you wish to read the paper:

Edeling, Wouter and Arabnejad, Hamid and Sinclair, Robbie and Suleimenova, Diana and Gopalakrishnan, Krishnakumar and Bosak, Bartosz and Groen, Derek and Mahmood, Imran and Crommelin, Daan and Coveney, Peter V. (2021). *The impact of uncertainty on predictions of the CovidSim epidemiological code*. Nature Computational Science, 1(2), 128-135.

www.nature.com/articles/s43588-021-00028-9

nature computational science

Research also featured in:

- Nature news & views: Kathy Leung, *Quantifying the uncertainty of CovidSim*
- Nature news: David Adam, *Simulating the pandemic: What COVID forecasters can learn from climate models*
- UK science museum: Roger Highfield, *Corona virus: virtual pandemics*

Thank you