Uncertainty Quantification Patterns

UKAEA NEPTUNE kickoff meeting
18-01-2021

Wouter Edeling
CWI Amsterdam

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

VECMA project

- **VECMA** = Verified Exascale Computing for Multiscale Applications (www.vecma.eu)

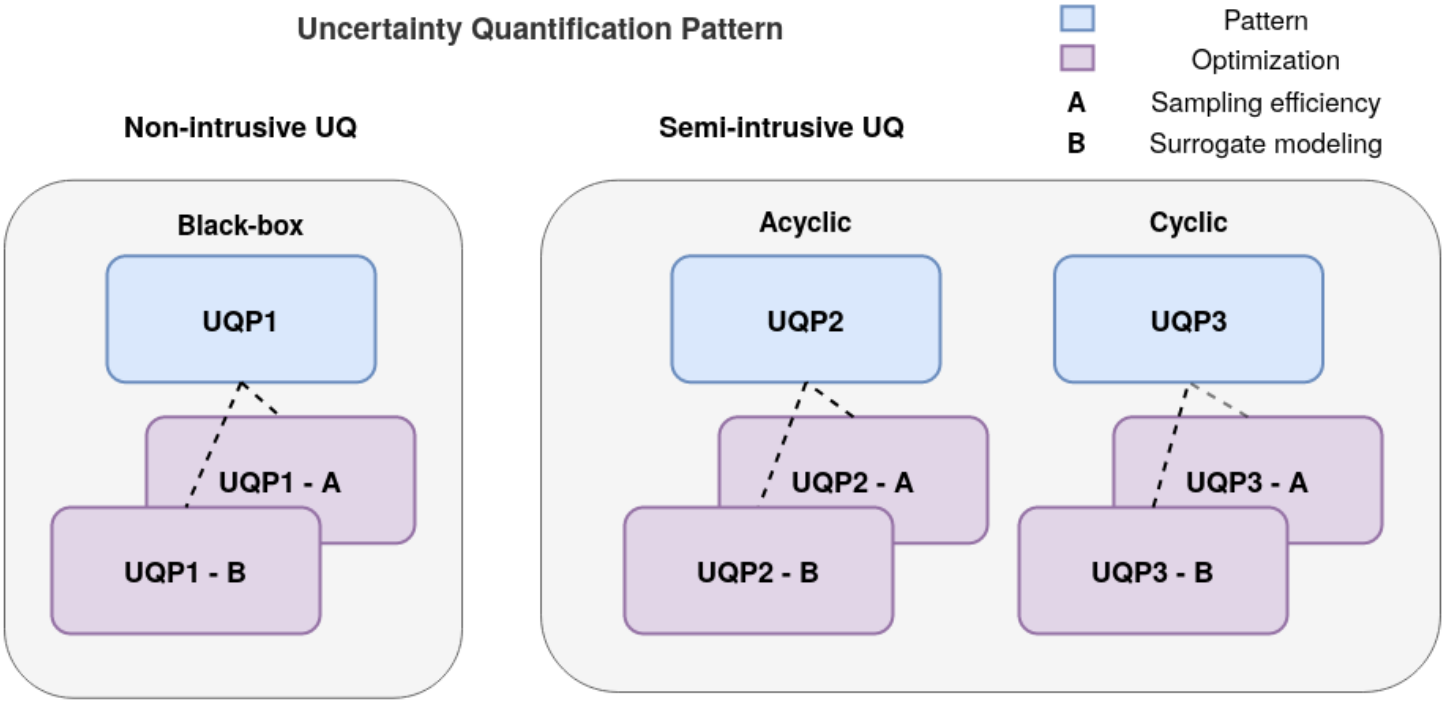
- **Main goals**:
  - Automate VVUQ for diverse multiscale applications.
  - Exploit peta / (emerging) exascale HPC resources.

- **Application partners**
  - Fusion (Max Planck Institute of Plasma Physics)
  - Material Science (UCL)
  - Biomedicine (UCL, UvA)
  - Migration (Brunel University)
  - Climate Modelling (CWI)
  - HPC partners: Bull Atos, PSNC, LRZ

- **Main deliverable: VECMA toolkit**
  - Surrogate models for small-scale processes.
  - Forward uncertainty propagation.
  - Tools for submitting ensembles on HPC machines.

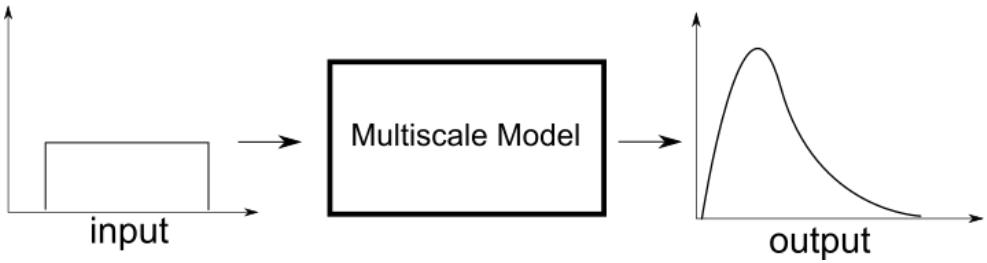
UQ Patterns

- UQ Pattern (**UQP**): generic UQ templates for **forward uncertainty propagation**
  - 2 levels of intrusion: non-intrusive & semi-intrusive
  - 2 types of multiscale model: acyclic & cyclic
  - 2 types of optimization: efficient sampling (A) & surrogate models (B)

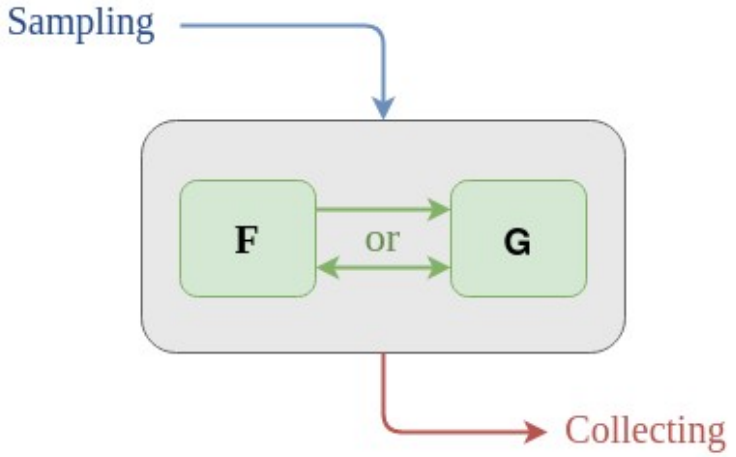


UQP1

- Forward UQ: propagate uncertainties from the inputs to the output:

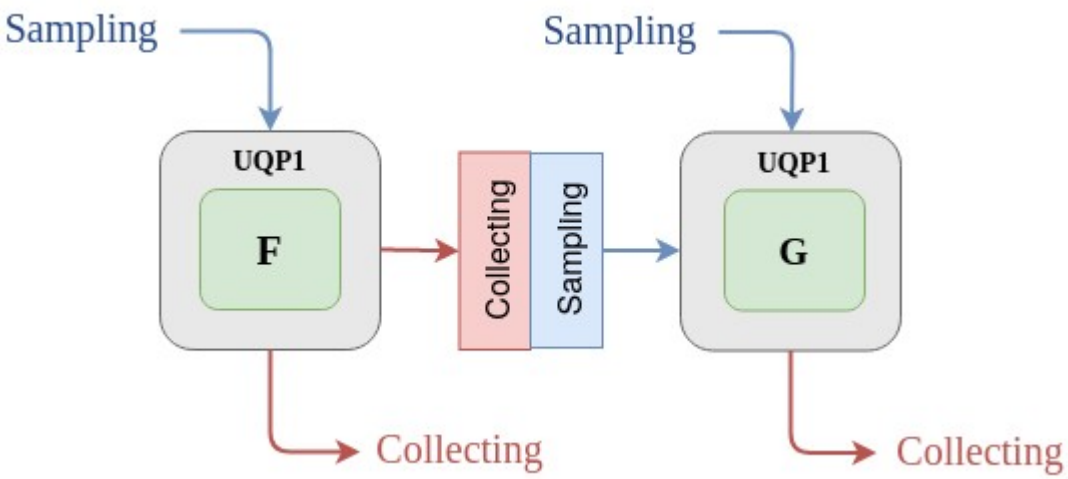


- A multiscale model is a **collection** of different submodels or codes.
- UQP1: ignore this, treat as black box:
  - Fully non-intrusive
  - Methods: Monte Carlo, stochastic collocation
  - Later: SC with efficient sampling (UQP1-A)



UQP2

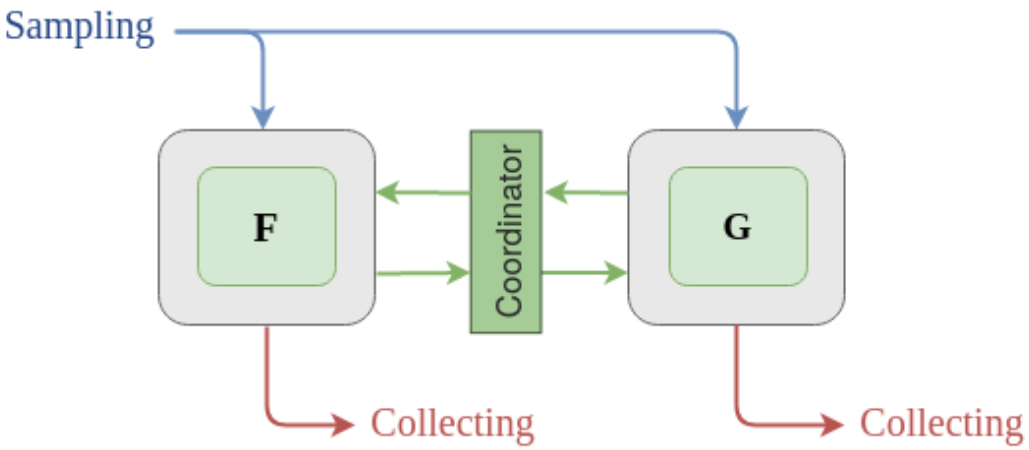
- UQP2: **acyclic** multiscale models
  - Uncertainty flows in one direction, no feedback
  - Semi-intrusive: don't modify submodels, do modify connection between submodels



  - If we replace submodel G with a surrogate: UQP2-B.

UQP3

- UQP3: **cyclic** multiscale models
  - Information is exchanged between submodels.
  - Also semi intrusive.



  - Cyclic multi-scale models with efficient sampling/surrogates (UQP3-A/B):

    A. Nikishova, *Semi-intrusive uncertainty propagation for multiscale models, J. Comp. Sci., 35, 80–90, 2019.*

UQP speedup

- You can derive potential speedups of different UQPs
- Speedup of UQP2-A (wrt UQP1):

$$\frac{C_{\mathrm{UQP1}_\mu}}{C_{\mathrm{UQP2\text{-}A}_\mu}} = \frac{n^{d_M+d_\mu} C_\mu}{n^{d_\mu+d_y} C_\mu} = n^{d_M-d_y}.$$

- For more information on the UQPs and their speedups:

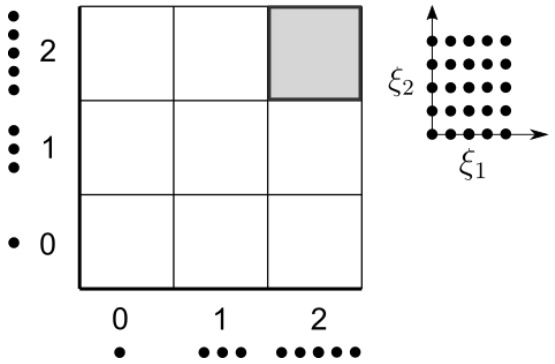
D. Ye, L. Veen. A. Nikishova, *Uncertainty Quantification Patterns for Multiscale Models*, Roy Soc Phil Trans A (accepted), 2021.

UQP1

- Let $q(\boldsymbol{\xi})$ be the output of some computational model.
- This talk: assess the impact of uncertainties in $\boldsymbol{\xi}$ on q
- Common techniques: (Quasi) MC, Polynomial Chaos, Stochastic Collocation:

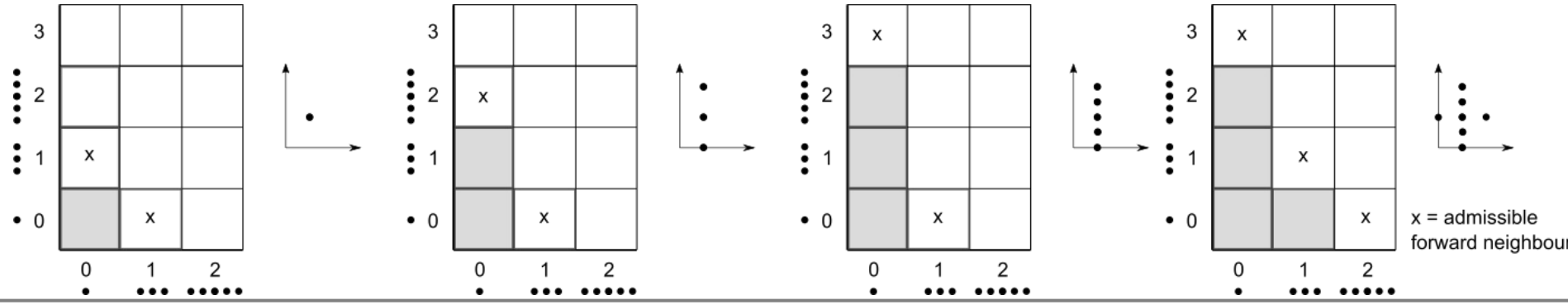
$$q(\boldsymbol{\xi}) \approx \tilde{q}(\boldsymbol{\xi}) = \sum_{j_1=1}^{m_1} \cdots \sum_{j_d=1}^{m_d} q(\xi_{j_1}, \cdots, \xi_{j_d})\, a_{j_1}(\xi_1) \otimes \cdots \otimes a_{j_d}(\xi_d)$$

- Basic building blocks: **1D quadrature rules**
- Extended to d dimensions via tensor product
- Computational cost: m^d



Curse of dimensionality

- For many computational models: $\boldsymbol{\xi} \in \mathbb{R}^d$ with $d \gg 10$
- In which case m^d is far too expensive
- UQP1-A: **dimension-adaptive sampling** (Gerstner & Griebel (2003))
- 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



Ensemble execution

- ‘Look-ahead’ step: evaluate code in new directions: ensemble computation

- One possibility: FabSim3
  - Python3 automation toolkit.
  - Handles data transfer to and from HPC machine
  - One-line commands (command line or API)

```
# run the UQ ensemble
fab.run_uq_ensemble(config, campaign.campaign_dir, script='Gray_Scott',
                    machine="eagle_vecma", PilotJob = False)
```

- Other options: QCG-PilotJob or FabSim3 + QCG-PilotJob

- QCG-PJ: a single job allocation containing many smaller jobs (HTC)
  - Circumvents limitations on number of jobs

EasyVVUQ

- Everything but job execution is handled by **EasyVVUQ**.
- EasyVVUQ: computes
  - distributions,
  - moments
  - Sobol sensitivity indices
- Jalal will talk about this.

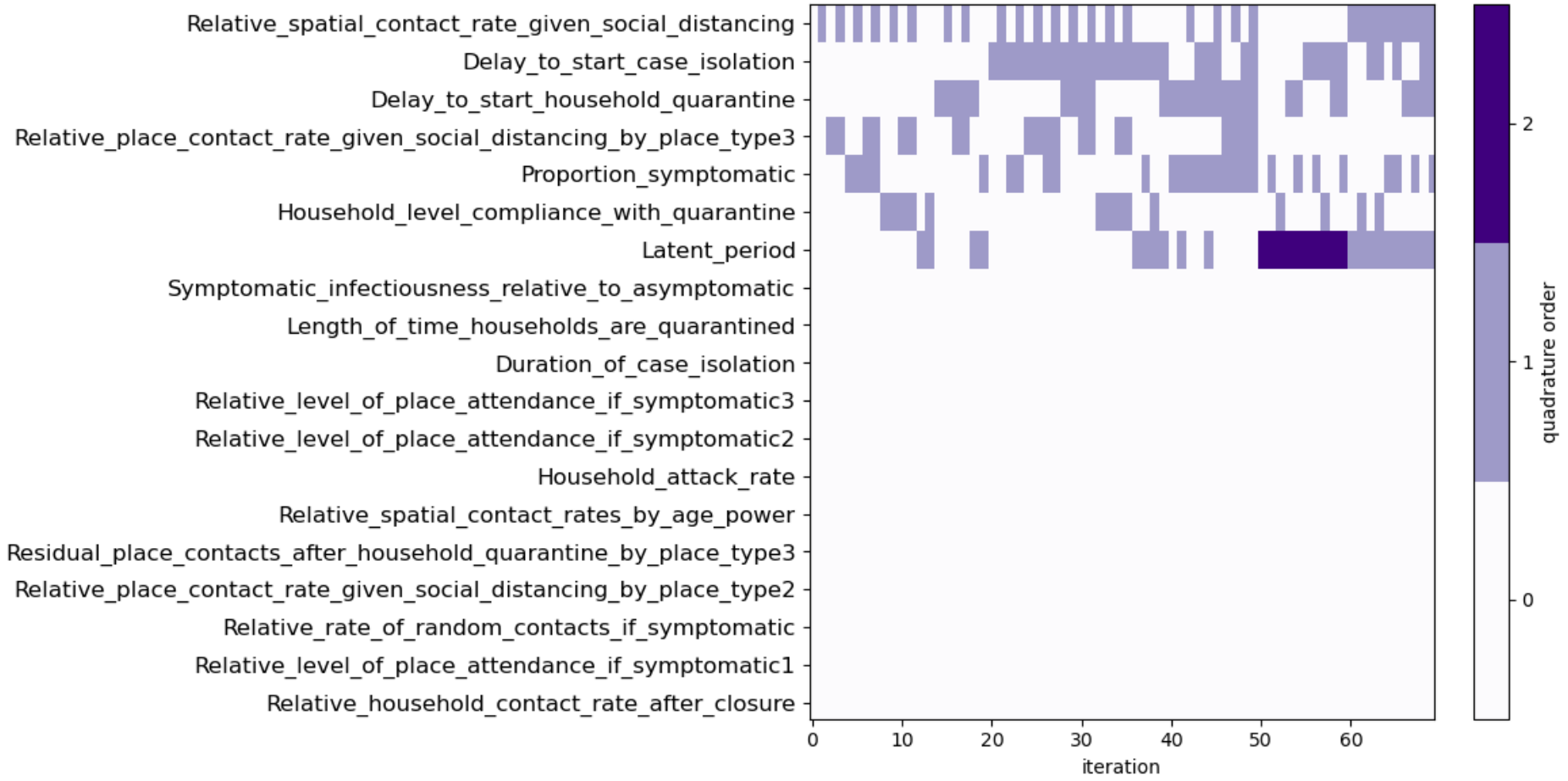
```
#perform look-ahead step
sampler.look_ahead(analysis.l_norm)

#####################################
# Ensemble execution using FabSim3 #
#####################################

# run the UQ ensemble
fab.run_uq_ensemble(config, campaign.campaign_dir, script='ohagan',
                    machine="eagle_vecma", skip=skip, PilotJob = False)

#wait for jobs to complete and check if all output files are retrieved
#from the remote machine
fab.verify(config, campaign.campaign_dir,
           campaign._active_app_decoder.target_filename,
           machine="eagle_vecma", PilotJob=False)

#run the UQ ensemble
fab.get_uq_samples(config, campaign.campaign_dir,
                   number_of_samples=sampler._number_of_samples,
                   skip=skip,
                   machine='eagle_vecma')

#########################################
# End ensemble execution using FabSim3 #
#########################################

#adaptation step
analysis.adapt_dimension('f', data_frame, method='surplus')
```

UQP1-A example: CovidSim

- CovidSim model from Imperial College (Ferguson, 2020)
  - Models effect of Non Pharmaceutical Interventions.
- On request of Royal Society's RAMP team we estimated parametric uncertainty.
- Computational cost single sample: max 10 minutes
  - Using 1 node with 28 cores at PSNC Eagle supercomputer.
- 940 input parameters, of which we identified 60 as interesting to vary
- Grouped these into 1) intervention, 2) disease and 3) geographic inputs.
- 3 separate UQ campaigns: every input that was refined at least once goes in the final campaign.
- Final UQ campaign: 19 input parameters.

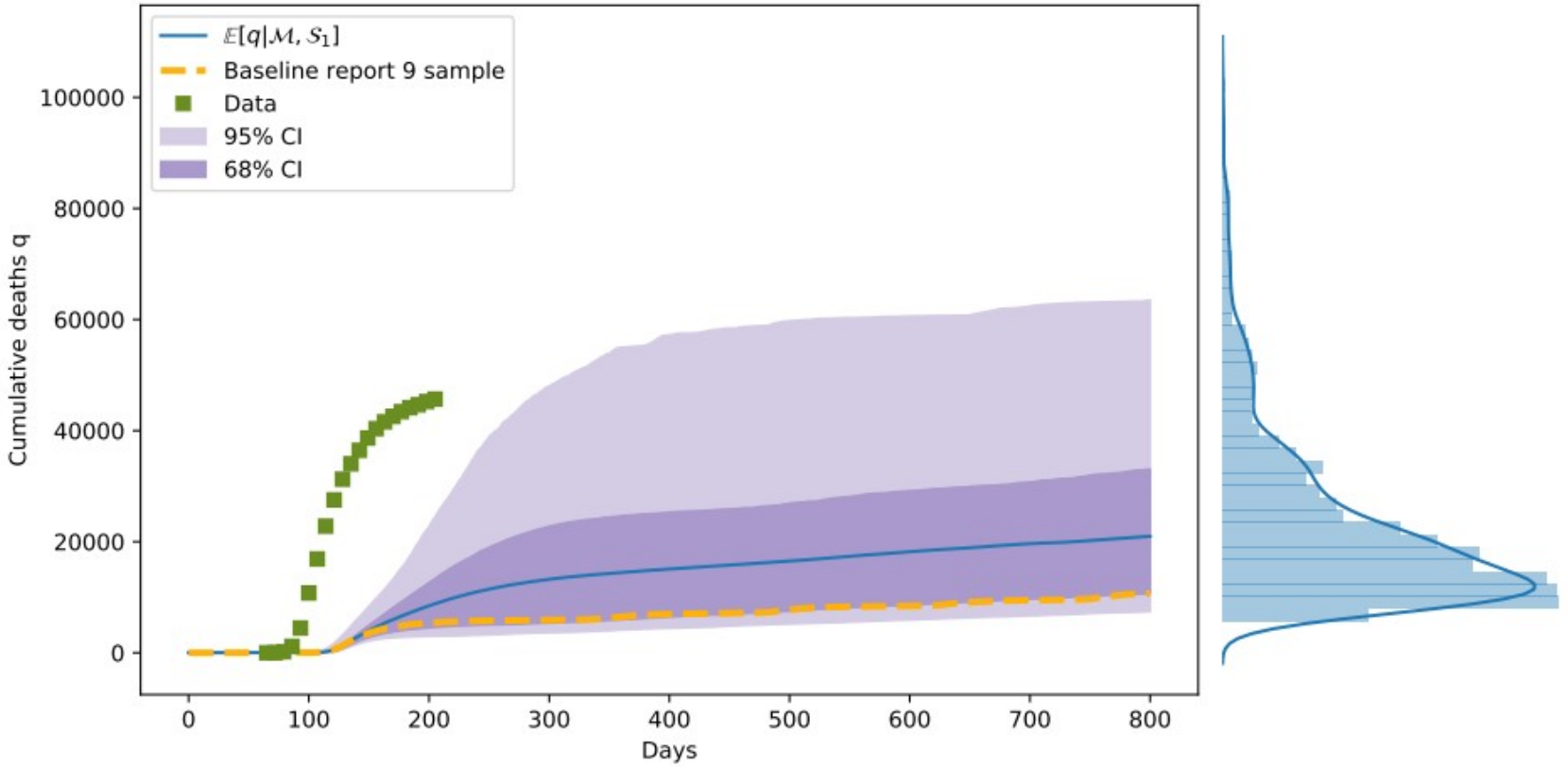
UQP1-A: adaptive grid refinement

- 19 parameters: chose uniform input distribution.
- Refinement: effective dimension of 7 during the 1st 70 iterations:



UQP1-A: CovidSim

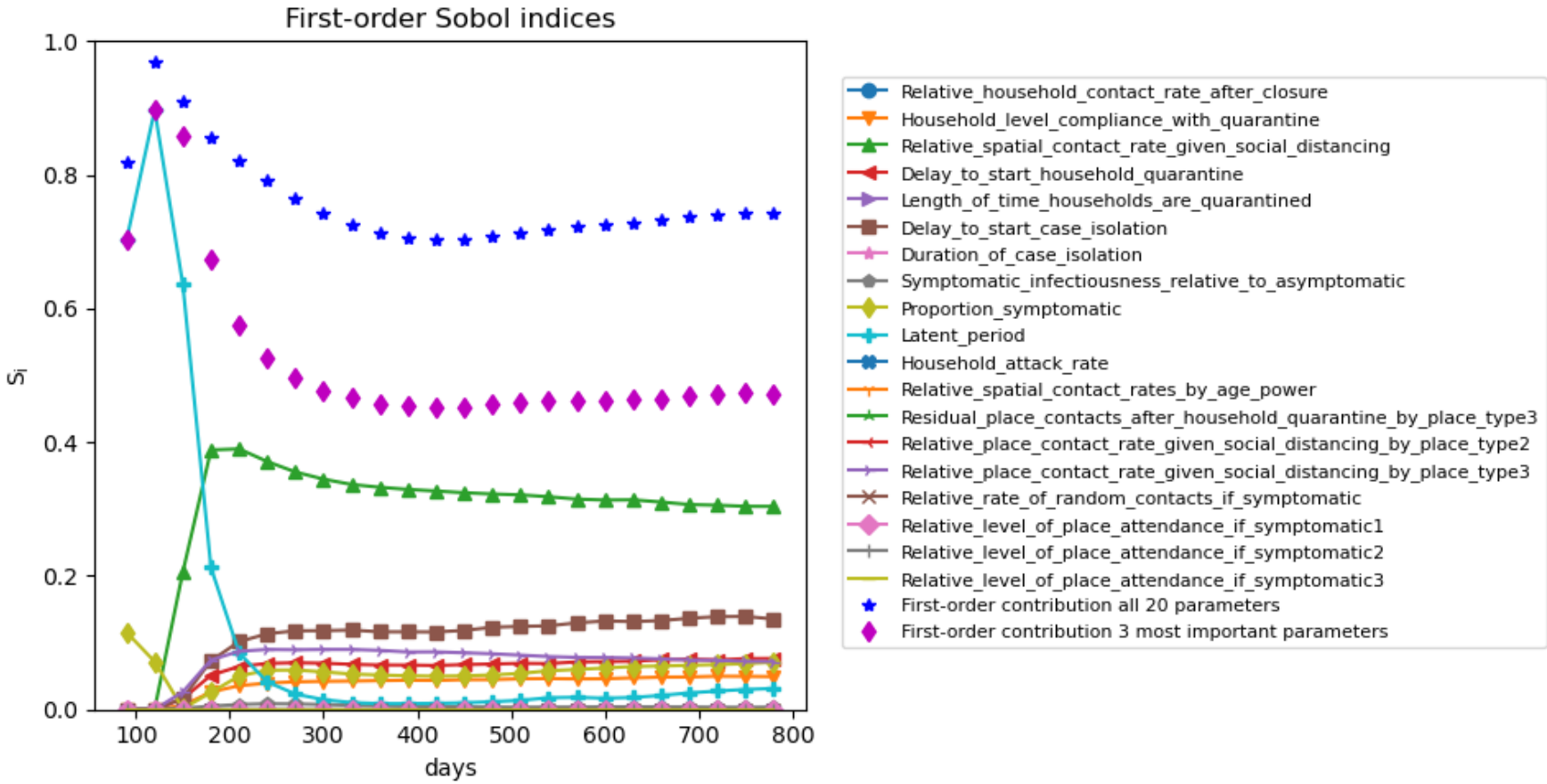
- Output distribution of cumulative death count



W. Edeling, H. Arabnejad, R. Sinclair et al., *The Impact of Uncertainty on Predictions of the CovidSim Epidemiological Code*, Nat. Comp. Sci. (accepted) 2021.

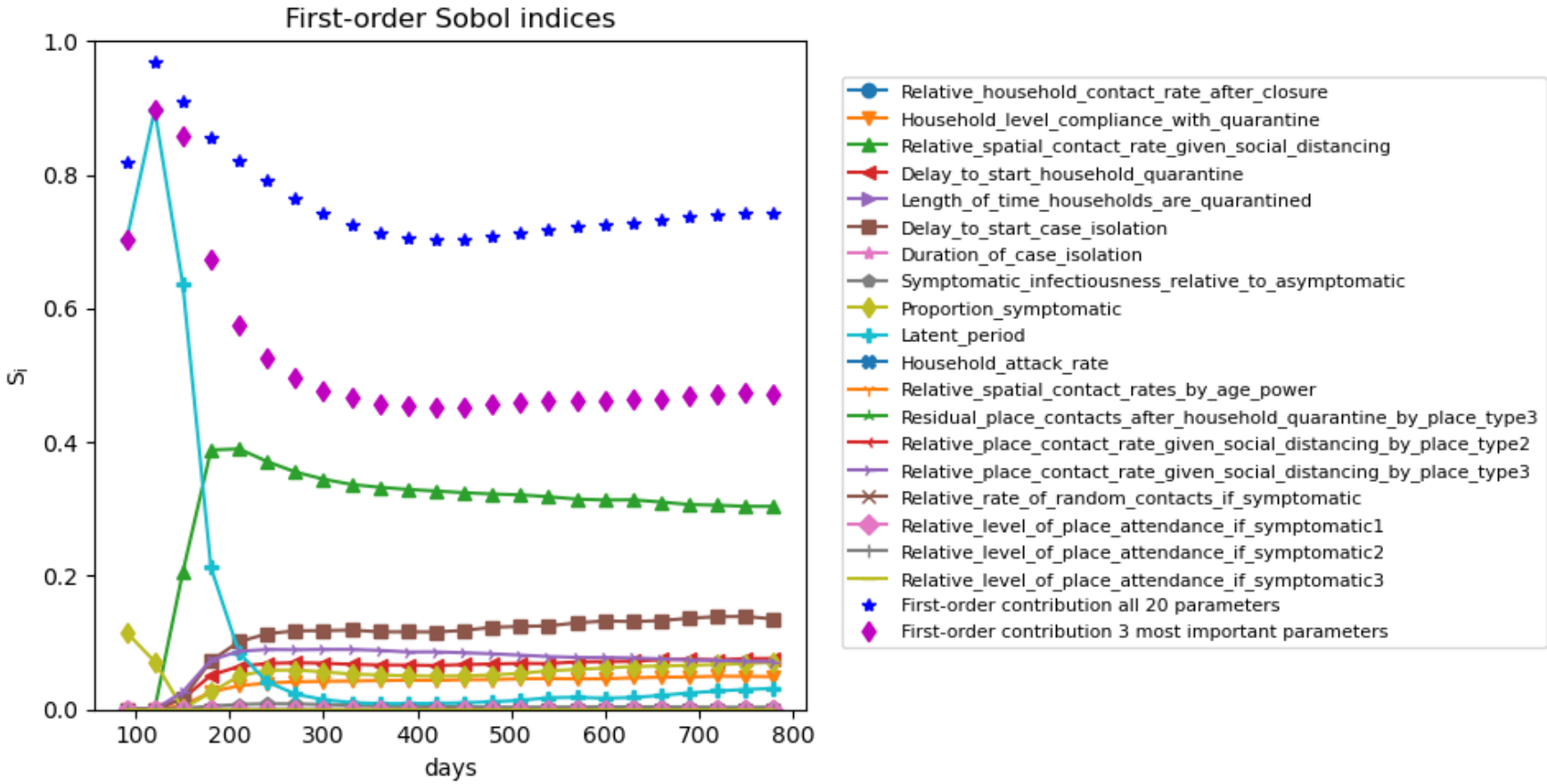
UQP1-A: Sensitivity analysis

- Sobol sensitivity indices can be obtained in a post processing step.
- First-order index $S_i \in [0, 1]$: fraction of the output variance the i-th input is responsible for:



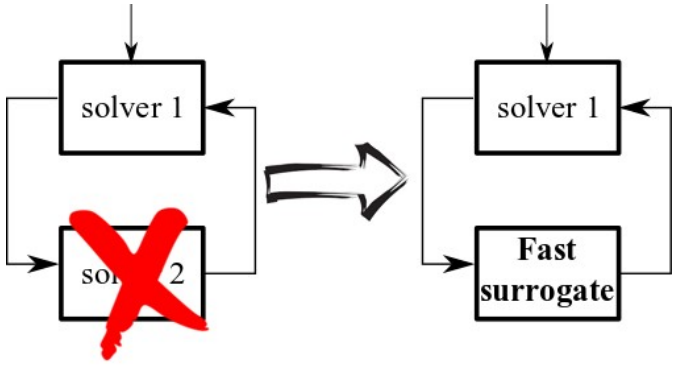
UQP1-A: Sensitivity analysis

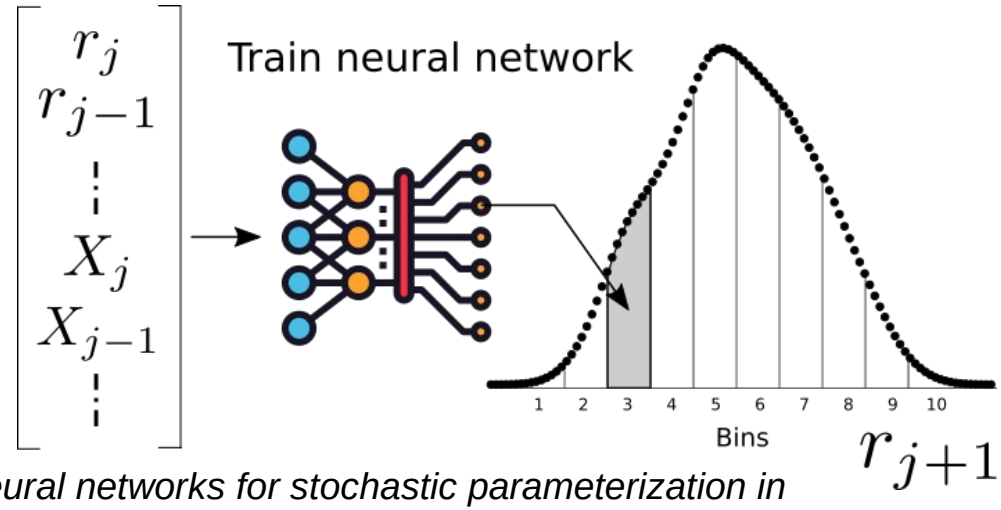
- Sobol sensitivity indices can be obtained in a post processing step.
- First-order index $S_i \in [0, 1]$: fraction of the output variance the i-th input is responsible for:



UQP3-B

- UQP3-B: cyclic multiscale model with a surrogate for the micro model.
- One type of surrogate: conditional resampling of training data via neural net.
- WE are in the process of building a toolbox for different (UQP3) surrogates: EasySurrogate
- Application to Lorenz 96:





D. Crommelin, W. Edeling, *Resampling with neural networks for stochastic parameterization in multiscale systems, Phys D (submitted), 2021.*

Questions?