D4.4: Report on application use cases

Due Date	14 December 2021
Delivery	03 December 2021
Submission of updated version	N/A
Lead Partner	MPG
Dissemination Level	Public
Status	Final
Approved	Executive Board
Version	V0.7

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

DOCUMENT INFO

Date and version number	Author	Comments
30.06.2021 v0.1	Onnie Luk	Outline
06.10.2021 v0.2	David Coster	Expanded the outline to section headers and some introductory text; added text for the fusion application
02.11.2021 v0.3	VECMA application partners	Text from the WP4 partners
02.11.2021 v0.4	David Coster	Moved from google docs to word
02.11.2021 v0.5	David Coster	Further refinement; version sent to WP4 members and for internal review
17.11.2021 v0.6	David Coster	Modification to address comments from the internal reviewers.
30.11.2021 v0.7	David Coster	Modifications to address comments from the project PI.

CONTRIBUTORS

- Onnie Luk – Author
- David Coster – Author and editor
- Diana Suleimenova, Derek Groen, Hamid Arabnejad – Author
- Maxime Vassaux, James Suter, Shunzhou Wan, Werner Muller – Author
- Wouter Edeleing – Author
- Michal Kulczewski – Author
- Alfonso Santiago – Author
- Dongwei Ye, Valeria Krzhizhanovskaya – Author
- Maxime Vassaux, James Suter, Xuanye Gu – Internal Reviewers
- Peter Coveney – Project Principal Investigator

TABLE OF CONTENTS

1 Executive Summary

This deliverable, D4.4 ("Report on Application Use Cases") covers work on the Application Use Cases (Task 4.5) in section 2.1 and the work on applying the VVUQ toolkit to external applications (Task 4.6) in section 2.2. It builds on previous WP4 deliverables and relies on references to those deliverables.

The different VECMA applications have explored a range of Uncertainty Quantification Patterns (UQPs) and Verification and Validation Patterns (VVPs) and demonstrated that these patterns apply across a large variety of fields. Monte-Carlo, Polynomial Chaos Expansion (PCE) and Stochastic Collocation (SC) have all been used to analyse the uncertainty in the Quantities of Interest (QoIs) driven by the uncertainties in the inputs or the inherent uncertainty of the process. Of particular interest is the use of adaptive methods which allows for UQ analysis to be applied for a much larger set of varying parameters.

Various components of the VECMA VVUQ toolkit have been successfully used by external partners on their applications. Different partners have used FabSim3, EasyVVUQ and QCG-PJ to apply UQ to their applications. The Hackathons have proved to be particularly useful in this context.

2 VECMA tools usage by internal and external application

Section 2.1 covers work on the Application Use Cases (Task 4.5, led by UCL), while section 2.2 covers work on applying the VVUQ toolkit to external applications (Task 4.6, led by UVA).

2.1 VECMA exemplar applications

The first WP4 deliverable, **D4.1: Report on Application Software Readiness** identified four applications (fusion, materials, Binding-Affinity-Calculator and In-Stent-Restenosis) as being on the fast-track, and four (Continuum Blood Flow, Cell-based Blood Flow, Climate and Multiscale Migration Prediction) as being on the deep track.

The second, **D4.2: Report on the implementation of nonintrusive VVUQ techniques** discussed the use of non-intrusive VVUQ techniques for fusion applications using Polynomial Chaos; In-stent restenosis (2D) application using Quasi-Monte Carlo; climate application and multiscale migration prediction application using Stochastic Collocation; materials application and binding affinity calculator application using parameter sweeps and ensemble bootstrapping.

The third, **D4.3: Report on the implementation of nonintrusive and intrusive VVUQ techniques** provided details about CovidSim; FACS: Flu and Coronavirus Simulator; Dutch-Covid Model; Multiscale Migration Prediction; Climate; Fusion; In-stent restenosis 3D (ISR3D); Materials; and Binding Affinity Calculator (BAC) with more details in the appendix for many of the applications.

This section provides more details for the applications, building on the information already provided in these earlier deliverables and provides the outcome performed in **Task 4.5: Application Use Cases** whose description is:

Demonstrate the applicability of the VECMA tools implemented in Tasks 4.2 and 4.3 to real use cases drawn from our diverse applications portfolio, through evaluating the effectiveness of the designed UQ and V&V primitives in that context. All application domains will provide such a real-case evaluation, and at least two applications will be sufficiently advanced by M30 to allow for in-depth analysis of performance and the extent to which the VVUQ toolkit has aided these applications in becoming "actionable".

2.1.1 In-Stent Restenosis 2D and 3D (ISR2D & ISR3D)

In-Stent Restenosis 2D/3D are multiscale computational models to simulate post-stenting neointima growth in the blood vessel. For the fast-track application ISR2D, both non-intrusive (NI) and semi-intrusive (SI) UQ methods have been applied (reported in **D4.2**). Multiple UQPs have been implemented including: UQP1, UQP1-B, UQP3-B.

For the non-intrusive UQ method, the entire model was regarded as a black-box and the uncertainty propagation was investigated with quasi-Monte Carlo (qMC) method with Sobol sequence. In the UQP1-B, a surrogate model based on Gaussian process regression was developed and subsequently replaced the original ISR2D in the quasi-Monte Carlo simulations. In the semi-intrusive uncertainty quantification (UQP3-B), the most expensive submodel, blood flow solver, is replaced with a surrogate model. We developed multiple surrogate models for the blood flow simulation, including a physics-simplified method (Phys), nearest neighbour method (DD I & II) and convolutional neural network (CNN). See [Nikishova 2019 and Ye 2021] for details on surrogates. The surrogate models then replaced the blood flow submodel in the quasi-Monte Carlo simulations.

A summary of uncertainty estimation of the quantities of interest by different methods is presented in Table 1**.** The NI estimations have the smallest error in the estimation of the mean and the restenosis ratio. The SI with CNN results has a smaller error than some other methods for each estimator. All the SI and NI results show a statistically significant underestimation of the mean value (two-value t-test, p value < 0.01)

Table 1 Comparison of the estimates of means and standard deviation of neointimal growth and restenosis ratio with qMC, SI and NI methods. The indicated error is the absolute difference from the reference qMC value. The four surrogate models for SIUQ are data-driven model I (DD I), data-driven model II (DD II), physics surrogate model (Phys) and convolutional neural network model (CNN).

UQ method	Micro model	Mean estimation $\times$ 10^{-1} (mm^2)		Standard deviation $\times$ 10^{-2} (mm^2)		Restenosis ratio(%)	
		Value	Error	Value	Error	Value	Error
qMC	LBM[a]	3.04	0	7.41	0	12.2	0
SI	DD I[b]	2.88	0.16	7.43	0.02	8.3	3.9
SI	DD II[b]	2.79	0.25	7.51	0.10	6.3	5.9
SI	Phys[b]	2.26	0.78	7.98	0.57	1.1	11.1
SI	CNN	2.82	0.22	7.54	0.13	7.7	4.5
NI	/	2.97	0.07	7.58	0.17	10.3	1.9

The execution times and resulting speedups of the SI and NI methods relative to the qMC method are evaluated and shown in Table 2. Because of the light surrogate model, the SI approach with CNN was approximately seven times faster than black-box qMC, an improvement of more than a factor three over the nearest-neighbour interpolation based surrogate model. The simplified physics model was

even faster, but was also the least accurate one, while the SI with CNN based surrogate model provided the best uncertainty quantification and sensitivity estimates among the four surrogates.

Table 2 Comparison of the computational time and corresponding speedup of different approaches. The time value indicates the mean computational time obtained over 1024 samples. τ_micro is the execution time of the micro model (LBM/surrogate models) in one ISR2D

UQ method	Micro model	$\mathcal{T}_{ISR}$ (min)	$\mathcal{T}_{ISR^*}$ (min)	$\mathcal{T}_{micro}$ (min)	$\mathcal{T}_{train}$ (min)	$\mathcal{T}_{sample}$ (min)	N	Speedup of UQ
qMC	LBM [a]	89.4	/	74.9	/	/	1024	1
SI	DD I [b]	/	50.9	37.8	/	894	1024	1.72
SI	DD II [b]	/	14.6	2.05	/	447	1024	5.94
SI	Phys [b]	/	11.9	0.08	/	/	1024	7.51
SI	CNN	/	12.8	0.26	9.9	447	1024	6.75
NI	/	/	0.17	/	4.3	1.1×10^4	1024	6.82

Based on the result from [Nikishova 2019 and Ye 2021], one can see that uncertainty and sensitivity analysis of SI were as good as the NI by comparing the SI and NI results at similar computational efficiency. SI has the additional advantage of granularity, since only part of the model is replaced by the surrogate. This means that the parameters of the submodels not replaced by the surrogate can be varied and studied without changing the surrogate, as long as the replaced micro model is not affected. For example, in the case of the ISR2D model, different parameters and rule sets for cell behaviour can be used with the existing surrogate model for flow. On the other hand, using a NI model for a different biological ruleset would require essentially building a new NI surrogate, which would incur a significant computational cost. In general, both SI and NI approaches performed well. The SI approach is more suitable for cyclic multiscale simulations as it retains the framework of the simulation and can obtain the training data for the surrogate model at a relatively low cost.

For the deep track application ISR3D, non-intrusive UQ was performed (reported in **D4.3**). Due to the expensive computational cost, a surrogate model based on Gaussian process regression was developed and deployed in the quasi-Monte Carlo simulations of UQ. The results of UQ are shown in Figure 1 and the speedup of UQP1-B (with surrogate model) compared to the UQP1 is shown in Table 3. Different from ISR2D, the blood flow submodel and the SMC submodel share almost the same amount of computational cost in a single run of ISR3D. Therefore, the semi-intrusive UQ method is not ideal as the maximum speedup possible is only 2 with a surrogate model for the submodels.

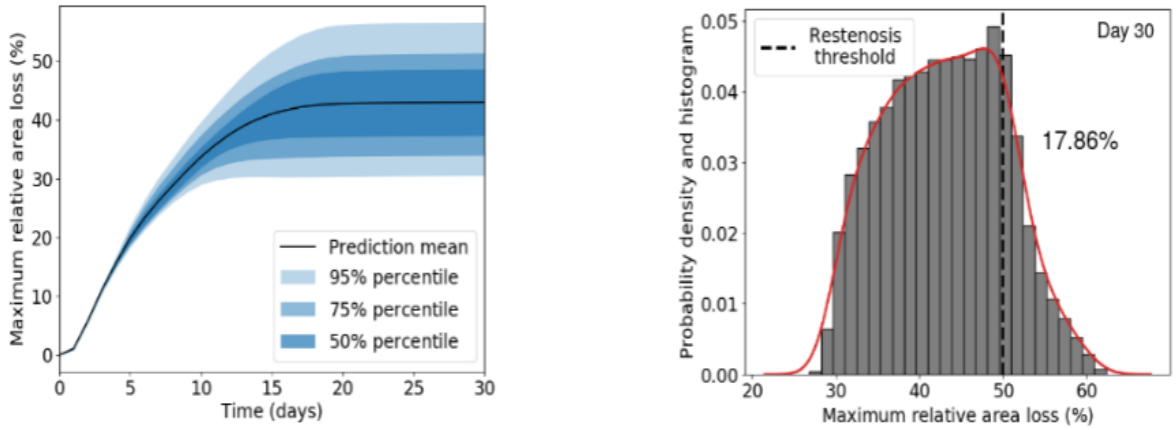


Figure 1 Mean, 50%, 75% and 95% percentile of the maximum relative area loss over time with quasi-Monte Carlo method and corresponding histogram and probability density function at day 30.

Table 3 Computation cost of ISR3D model and surrogate model and its corresponding speedup.

UQ method	$\mathcal{T}_{\text{ISR}}$ (core hour)	$\mathcal{T}_{\text{train}}$ (core hour)	$\mathcal{T}_{\text{sample}}$ (core hour)	N_{UQ}	Speedup of UQ
qMC	585.1	/	/	/	1
NI	6.1×10^{-8}	3.1×10^{-2}	585.1×512	5×10^5	976.6

2.1.2 Climate

We have applied various fast- and deep track UQ tools to problems related to weather and climate modelling. Most of these have already been described in D4.3, which we only briefly recap here:

1. We applied a fast-track non-intrusive method to the computationally expensive Dutch Atmospheric Large Eddy Simulation. Here we looked at the uncertainty in both physical and numerical parameters [Jansson_2021]. The number of parameters was relatively low, such that we could apply stochastic collocation, and get a much faster rate of convergence compared to Monte Carlo sampling.
2. In the deep track we looked at scalability in an algorithmic context. We developed the reduced surrogate method [Edeling_2020], which compresses the data requirement for surrogate models of multiscale systems by several orders of magnitude. We applied this to a 2D ocean model.
3. A related deep track activity is replacing the computational model for the small scales by a surrogate. We developed new stochastic machine-learning based surrogate models for this purpose and applied them to a simplified atmospheric model [Crommelin_2021].

Coupling a small-scale surrogate to a large-scale physical model is not guaranteed to be stable. In D4.3 we already alluded to using new `online' learning strategies to deal with this issue. Since that report, we have combined online learning in combination with the aforementioned reduced surrogates. The methodology is described in more detail in the forthcoming deliverable D2.3. Here we will just briefly show some new results.

Figure 2 shows the time evolution of two QoIs from the 2D ocean model. First, we have the global energy (E(t)) and the global entropy (Z(t)). These are common QoI which are computed from spatially varying large-scale vorticity fields. The figure shows the results from both a high-resolution reference model (dots), and a low-resolution model with small-scale reduced surrogate (solid lines). Note that these lines are indistinguishable from each other. This is because we apply continual online learning here. Without this, the QoI paths will diverge (due to chaos), and we have also observed that the system with a surrogate can become unstable if only offline learning is applied. We see the results of Figure 2 therefore as a promising first step towards coupled surrogate - physical multiscale systems, with a dedicated learning strategy that takes the coupling between the surrogate model and the physical system into account. All methods mentioned here are implemented in EasySurrogate.

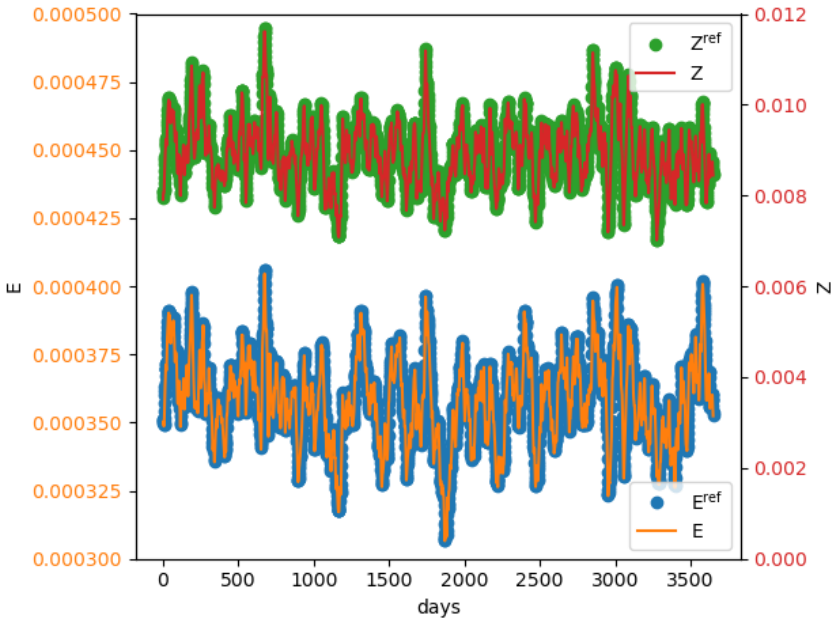


Figure 2 The time evolution of the enstrophy (Z) and energy (E), extracted from the large-scale vorticity fields of the 2D Navier-Stokes equations. The dots represent the values of the large-scale PDE with a reference small-scale model, and the solid lines indicate the solution extracted from the large-scale PDE with a reduced small-scale surrogate, subject to online learning.

2.1.3 Binding Affinity Calculator (BAC) / NAMD

The Binding Affinity Calculator (BAC) application as described in D4.3 has been equipped solely with non-intrusive UQ by means of adaptive Stochastic Collocation (see Figure 3). Intrusive UQ methods have not been considered due to the unrealistic implementation costs associated.

For a realistic BAC simulation, the number of parameters is very large. There are ~16 000 energy terms in the system we are studying here, excluding the terms for all the water molecules. These energy terms contain ~40 000 parameters. Only limited studies have been performed to quantify uncertainties from force field parameters, using relatively simple models such as TIP4P water molecules and/or focusing on a small subset of parameters such as those for the Lennard–Jones potential or the atomic radius and charge parameters. While quantification of the uncertainties from all the force field parameters is beyond the scope of this work, we note that the above studies show that the prediction uncertainty arising from parameters may be larger than statistical simulation uncertainty. To quantify the uncertainty associated with force field parameters, more advanced sampling techniques (*e.g.* Active Subspaces) or in-depth implementation of intrusive methods will have to be considered.

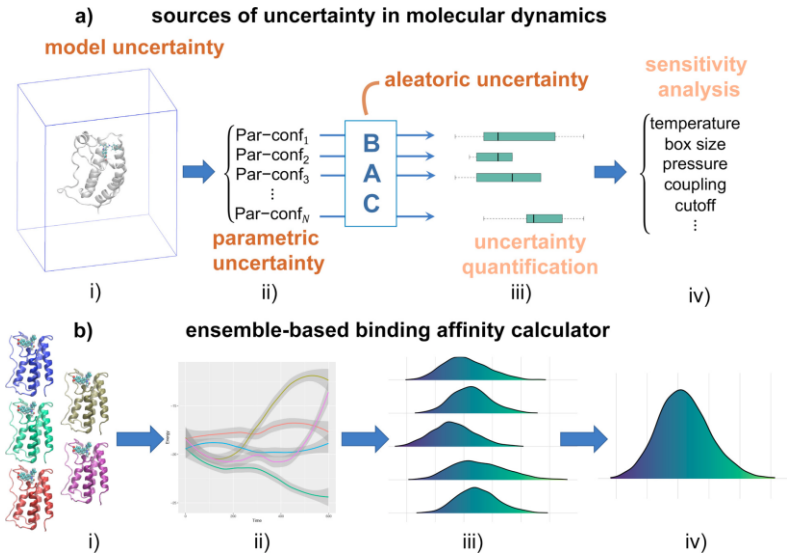


Figure 3 Sources of uncertainty and quality of predictions in molecular simulations for ensemble-based binding affinity calculations. (a) The types of uncertainties in the simulation (i) and the settings of parametric configurations (ii) are responsible for the uncertainty in predicted binding affinities (iii). Sensitivity analysis determines input parameters that most substantially impact predicted binding energy variability (iv). (b) The random errors are dealt with by ensemble approaches, in which multiple replicas (i) are simulated from initially close conformations. Neighbouring trajectories in the "underlying" phase space diverge exponentially fast (ii), generating different distributions for a quantity of interest (iii). The number of replicas used to perform ensemble averaging (iv) varies, depending on the required accuracy and the power of the available computational resources.

We have *a priori* restricted the number of uncertain inputs: a 14-dimensional space is still too large to sample with standard SC or polynomial chaos expansions, while simple Monte Carlo is known to have a slow convergence rate. For this reason, we employ a dimension-adaptive variant of the SC sampler. The exponential increase with the number of inputs d limits practical applications of the standard SC method to less than about 10 uncertainty parameters. Non-adaptive SC for 14-parameters would have required m^{14} ensembles of MD simulations, with m the quadrature order.

Our UQ campaign resulted in simulating 123 ensembles of MD simulations, with quadrature required up to order 6 for the most influential parameter (*e.g.* temperature, see Figure 4). We considered the adaptive SC method partially converged, in the sense of the relative error variance, but not the relative error mean (see Figure 5). With quadrature order being refined for only a very limited number of parameters (temperature, box size and compressibility) and having exceeded our computational budget (2,000,000 CPUhs on SuperMUC-NG), we in turn limited the campaign to 10 iterations.

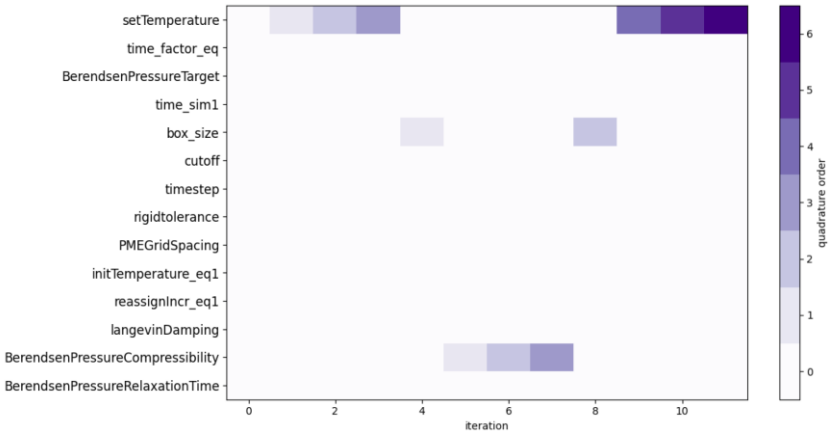


Figure 4 Refinement of quadrature order for each parameter of the BAC application with the iterations of the adaptive SC process.

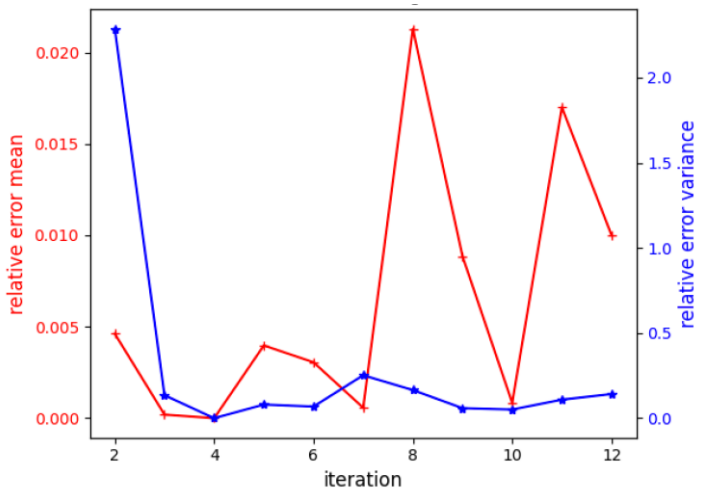


Figure 5 Evolution of the relative error mean (red) and variance (blue) with the iterations of the adaptive SC sampling process.

The use of the toolkit enabled us to demonstrate that the current practice of running one or only a small number of replicas of a molecular dynamics simulation is far from sufficient to control uncertainty as we had hypothesised. Small number of replicas does not enable one to control the error in the quantities of interest, as is achieved in a statistically robust manner by ensembles. Furthermore, the toolkit enabled us to demonstrate that the distributions of properties predicted using classical molecular dynamics cannot be assumed to be Gaussian but need to be assessed in each case, particularly when long-range interactions are involved. Conversely, the findings enabled by the use of the toolkit apply to classical molecular dynamics simulation in general, including to all forms of free energy estimation made using it. In conclusion, if we wish to produce actionable results from molecular dynamics simulations, whatever the predicted quantity of interest, we must invoke ensembles for which the use of modern supercomputers is essential.

2.1.4 Multiscale Migration Prediction (MMP)

Multiscale Migration Prediction (MMP) is a deep-track application aiming to predict or forecast the distribution of forcibly displaced people that escape violent conflicts and arrive to camps in neighbouring countries. It is a complex phenomenon modelled using an agent-based simulation tool, namely Flee, which is based on a set of autonomous decision-making agents with their environmental rules [Suleimenova_2017]. Flee requires a range of input parameters to execute forced displacement instances that affect simulation output [Suleimenova_2021a]. We provided the Complex UQ technique UQP-A ("Sampling Efficiency"), verification and validation approaches to MMP in D4.3. The multiscale nature of the application is derived from a coupling of a Flee micro scale model with a macro model, the integration with a conflict generator, as well as the integration of external data sources with different time scales (*e.g.* weather and food security). Specifically, we investigated precipitation and river discharge levels for the HiDALGO EU Horizon project affecting the movement speed of forcibly displaced people and their decision to remain in their current location or traverse through other routes for the South Sudan conflict between 2016-2017 [Jahani_2021].

We investigated the sensitivity of the input parameters using Stochastic Collocation (SC) with Sobol's method and presented our results in D4.3. Using sensitivity analysis, we identified the influential parameters in Flee affecting simulation output. We also modified Flee by introducing new parameters and changing the logical structure of the algorithm. These modifications led to an improvement in the simulation output, i.e., a decrease in the mean total error across conflict situations [Suleimenova_2021a].

Moreover, we compared the baseline SC results against the output obtained using a Polynomial Chaos Expansion (PCE) sampler with a polynomial order of 2. We determined the Sobol sensitivity indices for input parameters, which corresponded to previously identified parameters across four African countries as illustrated in Figure 6.

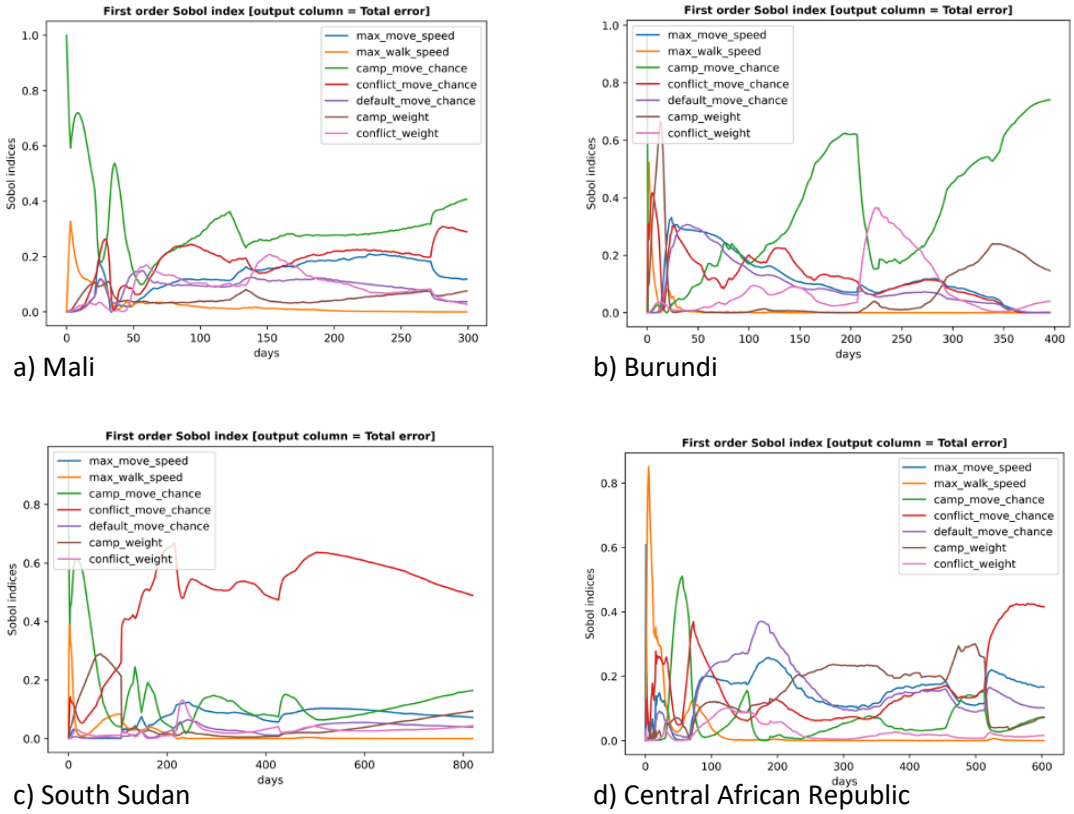


Figure 6 Input parameter exploration results for seven parameters of forced migration across four African countries (a-d) using PCE with a polynomial order of 2.

To establish a good estimation of sensitivity indices, we performed another set of analysis using PCE with a polynomial order of 3. The first-order indices of seven parameters for the Burundi conflict are relatively comparable against the polynomial order of 2 as demonstrated in Figure 7.

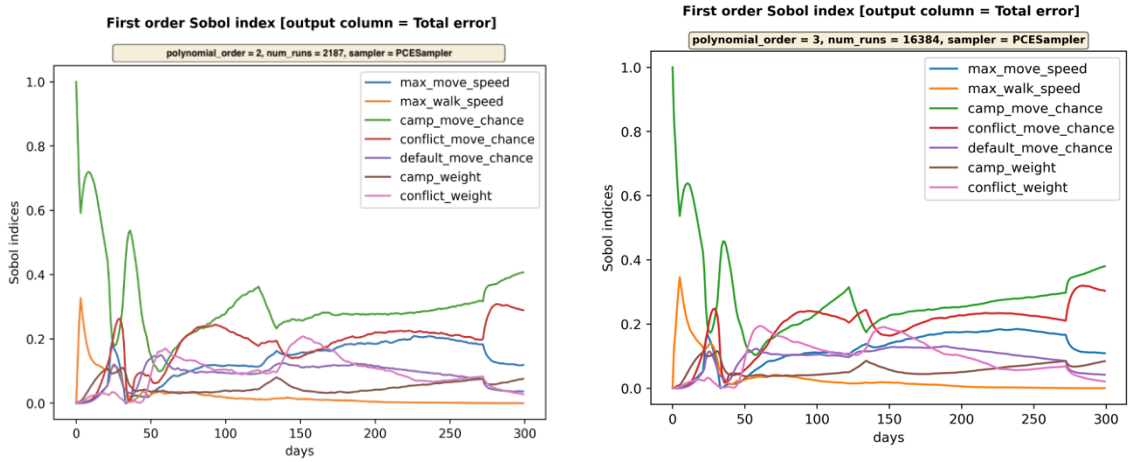


Figure 7 Comparison between the first-order Sobol indices for seven input parameters of the Burundi simulation with polynomial orders of 2 (left plot) and 3 (right plot).

In addition, we analysed polynomial orders for SC and PCE samplers focusing on the influential parameters, namely **max_move_speed**, **camp_move_chance** and **conflict_move_chance**, to verify the rate convergence (see Figure 8). The number of runs increased from 8 to 1000 for each polynomial

order starting from 1 to 9. We provide the detailed description of SA in application to forced displacement including the installation, parameter exploration, execution and analysis of obtained results in our recent publication [Suleimenova_2021b].

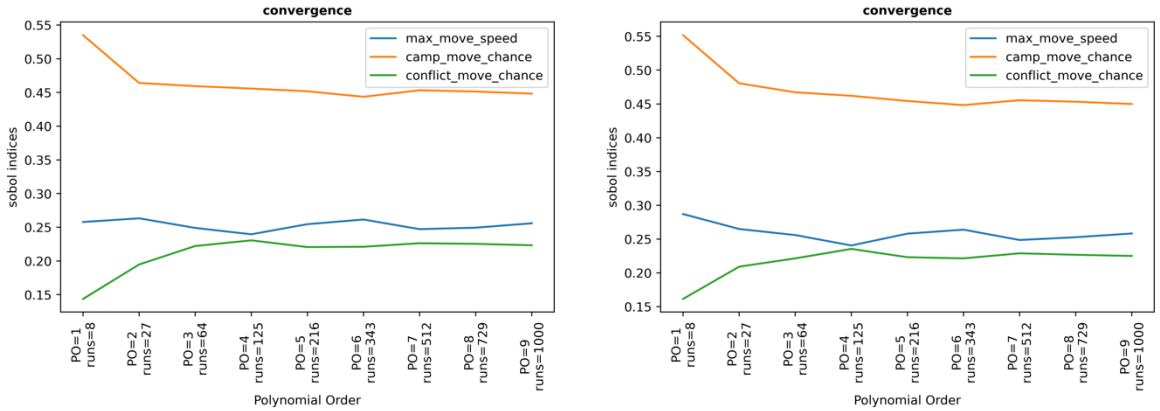


Figure 8 The convergence of the first-order Sobol indices for the influential parameters of Flee using SC (left plot) and PCE (right plot) samplers.

2.1.5 Multiscale Fusion Workflow (MFW)

The fusion multiscale workflow combines a transport code working on large time- and space-scales, an equilibrium code operating on the large space-scale and a turbulence code operating on small time- and space-scales to calculate the performance of the core part of a tokamak, which is a key metric for implementing fusion as a terrestrial power source. A variety of approaches have been used in the fusion application as described in **D4.3: Report on the implementation of nonintrusive and intrusive VVUQ techniques**, section **4.5 Fusion**, including

- UQP1 (treating the entire workflow as a black-box)
- UQP2-A (Semi-Intrusive acyclic)
- metamodeling (UQP2-B, not yet generalized to UQP3-B)
- VVP4 for comparing the experimental and simulation results

In addition to the main workflow incorporating the expensive turbulence code, a variant where a much less expensive proxy is used has allowed for exploratory research on UQ techniques. As a tutorial example, an even more stripped-down fusion app was developed which replaces the entire workflow with a Python program (solving a similar but much simpler problem); which has allowed for an even more extensive exploration of the limits of various UQ techniques.

The fusion-app allowed for a large number of samples to be simulated, and

- showed some deficiencies in the implementation of some parts of the EasyVVUQ Tool Kit
    - calculation of Sobol indices was very expensive (fixed by changing the algorithm to use the coefficients of the PCE polynomials)
    - management of a large number of samples
        - creating many directories at the same level (fixed by creating a directory hierarchy)
        - scaling of the database access (algorithm reworked to O(n))

- that PCE can require very large amounts of memory for high orders (SC does not require this, so SC is to be preferred at the moment for cases with a large number of samples)
- that the analysis phase for PCE and SC can be very expensive in terms of time (not yet addressed since it is better to find methods avoiding a very large number of samples)

The following table shows timing results in applying EasyVVUQ to the fusion-app for 5 varying parameters.

Table 4 Timings for PCE and SC with 5 varying parameters for the fusion mini-app. DASK was used to calculate the samples in SLURM batch queues and the set-up time for this was comparable to the calculation, resulting in the time taken for this phase to be roughly constant.

		PCE			**SC**		
Order	# Samples	Total	Running the samples	Analysing the results	Total	Running the samples	Analysing the results
1	32	89.95	87.54	1.75	43.91	42.78	0.41
2	243	68.59	63.55	3.13	23.24	19.05	2.76
3	1024	38.15	22.88	10.12	42.24	24.58	12.76
4	3125	85.15	22.77	47.77	82.45	21.67	45.95
5	7776	333.14	43.56	256.18	228.59	33.31	162.27
6	16807	1395.55	80.72	1242.13	677.02	64.13	541.31

We see that for higher numbers of samples (corresponding to higher polynomial orders), the time taken for the analysis phase is dominant. This might not be true for other cases since the fusion-app has been designed to bring down the costs of running a sample, but calculating the samples is parallelized whereas the analysis phase is not (yet).

If we look at a fusion-app case with 10 varying parameters (using SC), we see that the analysis costs become completely dominant:

Table 5 Timings for SC with 10 varying parameters for the fusion mini-app.

Order	# Samples	Total	Running the samples	Analysing the results
1	1024	101.63	83.88	14.97
2	59049	15,202.35	4,881.22	10,180.64
3	1048576	3,580,672.00	101,575.42	3,476,360.00

Even though some parallelization of the analysis phase could be implemented (and perhaps should be in the future), other methods of reducing the cost are better.

Applying adaptive methods to the analysis, the costs can be substantially reduced: to achieve comparable accuracy of the surrogate, adaptive SC took about 6 hours versus the 40+ days for the full 3rd order SC, and required 1245 samples rather than the 1 048 576 for the full 3rd order SC.

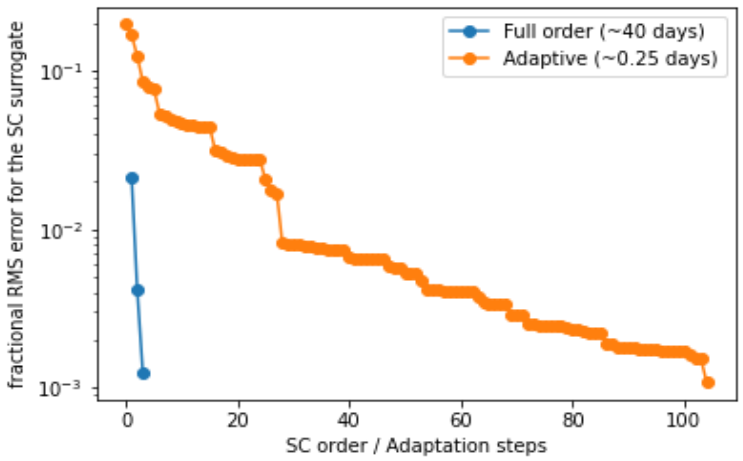


Figure 9 Comparison of the fractional RMS error for predicting the QOI. The abscissa is the PCE order for the "Full order" case, and the number of adaptation steps for the "Adaptive" case.

More details can be found in [Luk_2018], [Luk_2019a], [Luk_2019b], [Groen_2019], [Groen_2021], [Lakhlili_2020], [Richardson_2020], [Wright_2020], [Bosak_2021], [Coster_2021], [Luk_2021a], [Luk_2021b], [Suleimenova_2021b]

2.1.6 Materials

The materials application has been equipped with non-intrusive and semi-intrusive VVUQ techniques. In D4.2 section 2.1.4.1, we presented a fast-track example of using the "sweepsampler" VVUQ element of EasyVVUQ to determine the sensitivity to the calculation of stress in a molecular dynamics simulation of an epoxy network to the size of the simulation box. It was found that Young's modulus (the gradient of stress vs strain) of an ensemble of different sized simulations of epoxy polymers varied considerably as a function of simulation size, with confidence intervals converging over a box size of 4nm. The "sweepsampler" functionality is non-intrusive and allowed for an ensemble of "replica" runs to be generated for each group of input parameters, on which a bootstrap analysis can then be performed, using the EnsembleBoot analysis element (also from EasyVVUQ).

Another example of using fast-track non-intrusive VVUQ techniques for research in materials concerned the aggregation tendency of graphene oxide (GO) flakes in different polymeric environments [Suter_2020]. We used coarse-grained molecular dynamics and newly developed accurate models of GO to determine the structures formed by GO flakes on the microscale. Two hydrophilic polymers, poly-ethylene glycol (PEG) and poly-vinyl alcohol (PVA), are used to illustrate the thermodynamically stable morphologies of GO and relevant dispersion mechanisms. The workflow utilised ensemble simulations to reliably predict the structures formed by GO, with replicas differing in their initial velocities which are generated randomly and independently from a Maxwell–Boltzmann distribution (i.e. a Monte-Carlo UQP). Each simulation was then run for an extended simulated annealing run to explore configuration space. The simulations were started from either a stacked configuration, where all flakes are aggregated, or a dispersed configuration. We defined a classification metric to determine whether the flakes had formed a morphology we could describe as dispersed, intercalated or aggregated/encapsulated (see Figure 10). The classification metric used the distances between each atom within a flake to atoms in other flakes. On an individual atom level we can define whether the atom is in an "aggregated", "dispersed" or "intercalated" environment by using the

distance to atoms in other flakes, and for the flake as a whole we can examine the percentage of the flake atoms in each category. For reliable results, we require the error in our atom percentages to be small. We studied the behaviour of GO flakes with different degrees of oxidation on the GO surface (defined by the C:O ratio, for which we considered C:O = 10.0, 5.0, 2.5 and no oxidation, *i.e.* graphene). For more details see Reference [Suter_2020].

We chose 8 replicas for each scenario of chemical composition or physical starting structure to initially probe the system. We evaluated if this was a large enough sample by measuring our confidence in a quantity of interest using such an ensemble size. Figure 10 shows the 95% confidence interval for whether a flake assembles in PEG. We can see from Figure 10 that apart from flakes with a C:O ratio of 5 and 10, the trend of going from aggregated to dispersed as the oxygen content increases is clearly resolved with the sampling conducted. We can be confident, then, that our trend and results are replicable, and we do not need to increase the sample size further; we had already reached convergence with 8 replicas.

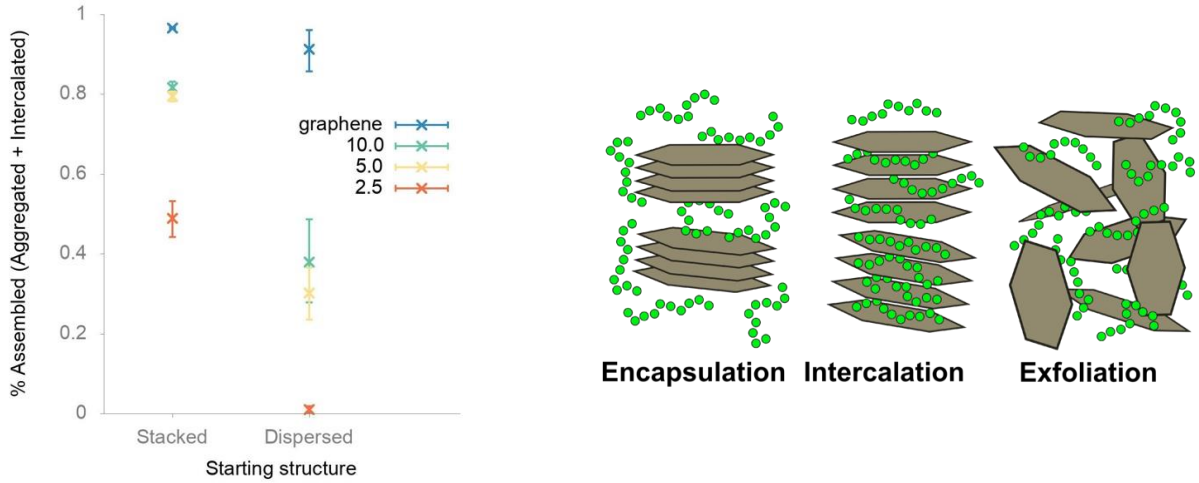


Figure 10 (left) Representation of the three different configurations GO flakes can form in a polymeric environment. (right) The mean measured aggregation of flakes in PEG for different starting structures and amount of oxidation. Error bars are a 95% confidence interval calculated by a bootstrap method. Some systems have very little variance in the measured aggregation and the error bars are smaller than the points shown here (e.g. the graphene stacked system).

For deep-track applications in Materials, we have developed non-intrusive and intrusive VVUQ techniques for SCEMa (Simulation Coupling Environment for Materials). SCEMa is a multiscale simulation tool to predict the properties of materials based on their underlying chemistry and nanoscale structure. SCEMa consists of an implementation of the Heterogeneous Multiscale Method coupling Deal.II (Finite Element Method) and LAMMPS (Molecular Dynamics). It enables simulations coupling semi-concurrently the evolution of an atomistic and a continuum system. The evolution of the continuum system drives the mechanical evolution of the periodic homogeneous atomistic replicas. A description of SCEMa has been given in D4.3, section 2.3.8. In summary, we made use of non-intrusive UQP1 to analyze the uncertainty associated with SCEMa simulations. We have shown that neural network based surrogates converge faster than Gaussian processes based surrogates in terms of distribution of global force output (see Figure 11). We also attempted to build surrogate models of the molecular dynamics models to apply semi-intrusive UQP3B. The non-intrusive UQ relied on the building of a Gaussian Process (GP) surrogate model which showed good agreement with SCEMa's predictions of the force applied on the dogbone sample to strain it. Further, SCEMa was

enhanced by a clustering algorithm enabling to avoid unnecessary redundant MD simulations, based on a redundancy threshold chosen according to the desired accuracy. In turn, we applied the VVP2 (Level of Refinement) to verify the convergence of the clustering algorithm with respect to the similarity threshold parameter.

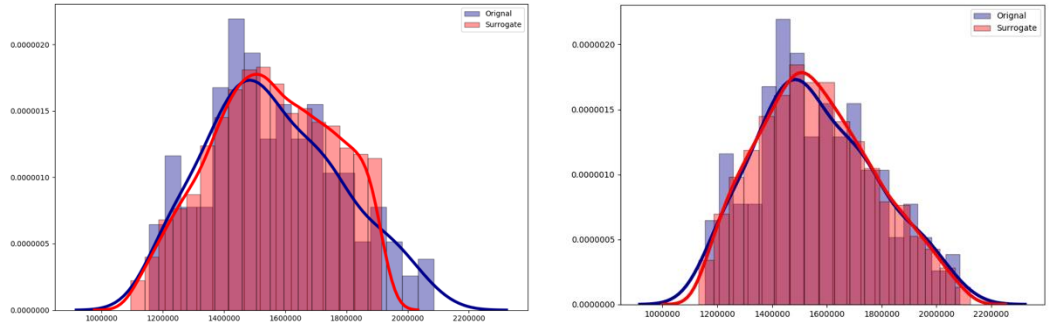


Figure 11 Comparison of surrogate modelling methods efficiency to non-intrusively estimate UQ in SCEMa's output: (left) Gaussian processes based surrogate and (right) neural network based surrogate. The distributions reflect the variability in global force prediction using SCEMa when applying a fixed positive strain on an epoxy bulk.

2.2 External applications

This section covers the work performed in **Task 4.6: Apply the knowledge gained in T4.2 and T4.3 to additional multiscale workflows**:

> *Implement the VVUQ toolkit in additional multiscale application workflows provided by at least two external groups.*

The external applications covered are Neptune (UK fusion community), Alya, UrbanAir and three Coronavirus simulations.

2.2.1 Neptune

During the three hackathons in the presence of VECMAtk developers, the NEPTUNE community represented an important part of the participants in the 3 hackathons. We saw 5 to 10 participants from the NEPTUNE project join the online hackathon meetings out of the 25 to 30 participants. We report the contributions of each application team (BOUT++, Nektar++, EPOCH) during the hackathons in the following section.

Three application teams were able to perform non-intrusive UQ of their single-scale model simulations. All teams were able to apply SC and PCE to compute sensitivity and uncertainty associated with up to four parameters. There is a consensus on using non-intrusive surrogates in the long term which will need to be refined and improved to handle a larger number of input parameters.

Below is reported the UQ attempts entirely and directly performed by the three applications teams.

2.2.1.1 Nektar++

Investigations of Nektar++ consisted in implementing UQ of single-scale simulations of a heat transport model providing a relationship between the quantity of interest (QoI) and two dimensionless numbers.

In this preliminary work PCE, SC, and GP were used to fit various QoIs using EasyVVUQ only (not EasySurrogate).

$$\begin{aligned}\frac{1}{Pr}\left(\frac{\partial \mathbf{u}}{\partial t}+\mathbf{u}\cdot\nabla\mathbf{u}\right) &= -\nabla p + Ra\,T\,\hat{\mathbf{y}}+\nabla^2\mathbf{u}\\ \left(\frac{\partial T}{\partial t}+\mathbf{u}\cdot\nabla T\right) &= \nabla^2 T\\ \nabla\cdot\mathbf{u} &= 0.\end{aligned}$$

The two input parameters, Pr and Ra, were respectively the Prandtl number, that is the ratio of kinematic viscosity to thermal diffusivity (fluid properties), and the Rayleigh number, that is the dimensionless temperature difference. Ra was varied in the range [1.0x10^1, 3.2x10^4] with log-uniform distribution; Pr varied in the range [1,10] (typical values for experiments with air and water) with uniform distribution.

For the steady-state problem, the Nektar++ team led by Ed Threlfall was able to perform SA and constructed PCE and SC surrogates, both using fifth-order polynomials during the hackathons (see Figure 12).

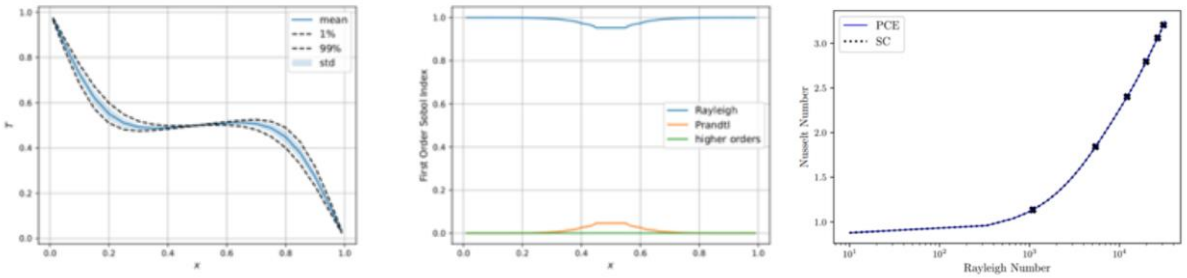


Figure 12 UQ of the horizontal temperature profile halfway up the cavity (left), SA (first-order Sobol indices) of the two parameters (centre), PCE and SC surrogate models of the Nusselt number (right).

Time-dependent simulations were also investigated using larger values of Ra. A GP surrogate model for time series was constructed using EasyVVUQ based on 12 samples varying only the parameter Ra. The surrogate model was in excellent agreement for quiescent cases (solution smooth) but issues were revealed in the case of higher Rayleigh numbers such as a persistent offset (see Figure 13). More work is needed to design optimal surrogates for chaotic time series, in particular surrogates that predict the correct phase diagram of the many scaling regimes in convective turbulence. This may require better tuning of kernel parameters for the GP surrogate; such methods will be discussed in the next section.

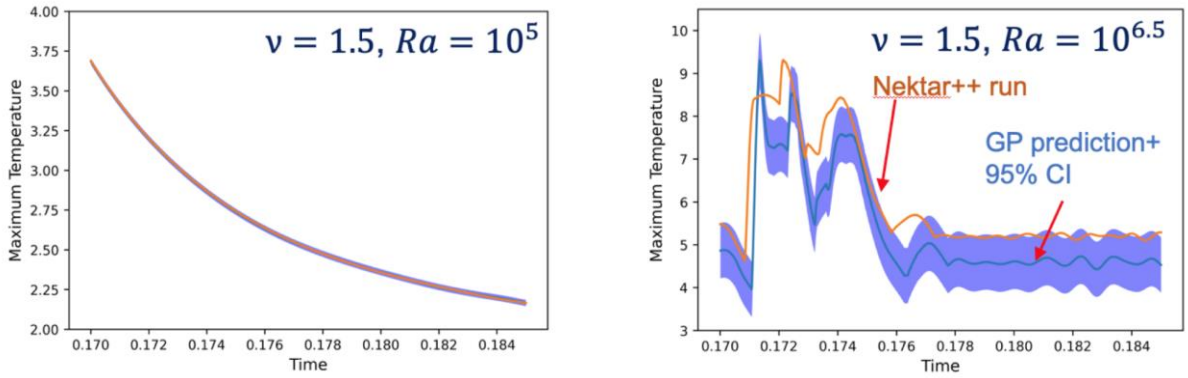


Figure 13 Fitting of the maximum wall temperature evolution for different regimes using a GP surrogate (Matérn ν=1.5 in kernel, and $Ra=10^5$ or $Ra=10^{6.5}$).

2.2.1.2 BOUT++

Investigations of BOUT++ consisted in implementing UQ of single-scale simulations of a heat conduction 1D model and a plasma physics 2D model. In this preliminary work PCE, SC, and GP were used to fit various QoIs using EasyVVUQ but also EasySurrogate in this case. The work was carried by a team consisting of Joseph Parker, Peter Hill, Ben Dudson and collaborators. The 1D model focuses on the evolution of the temperature field expressed as different QoIs: T(x, tend), T(x0, t), log[T(x, tend)], varying the initial temperature and a single diffusivity parameter χ (see Figure 14):

$$\frac{\partial T}{\partial t} = \chi \frac{\partial^2 T}{\partial x^2}$$

Issues of negative values were reported but later were found to be an artefact of polynomial fitting to steep temperature gradients with respect to parameter variations near zero. The issue was fixed by using higher-order polynomial fitting in PCE and SC. Adaptive SC was also used to circumvent the curse of dimensionality, anticipating performing UQ in high-dimension parametric spaces.

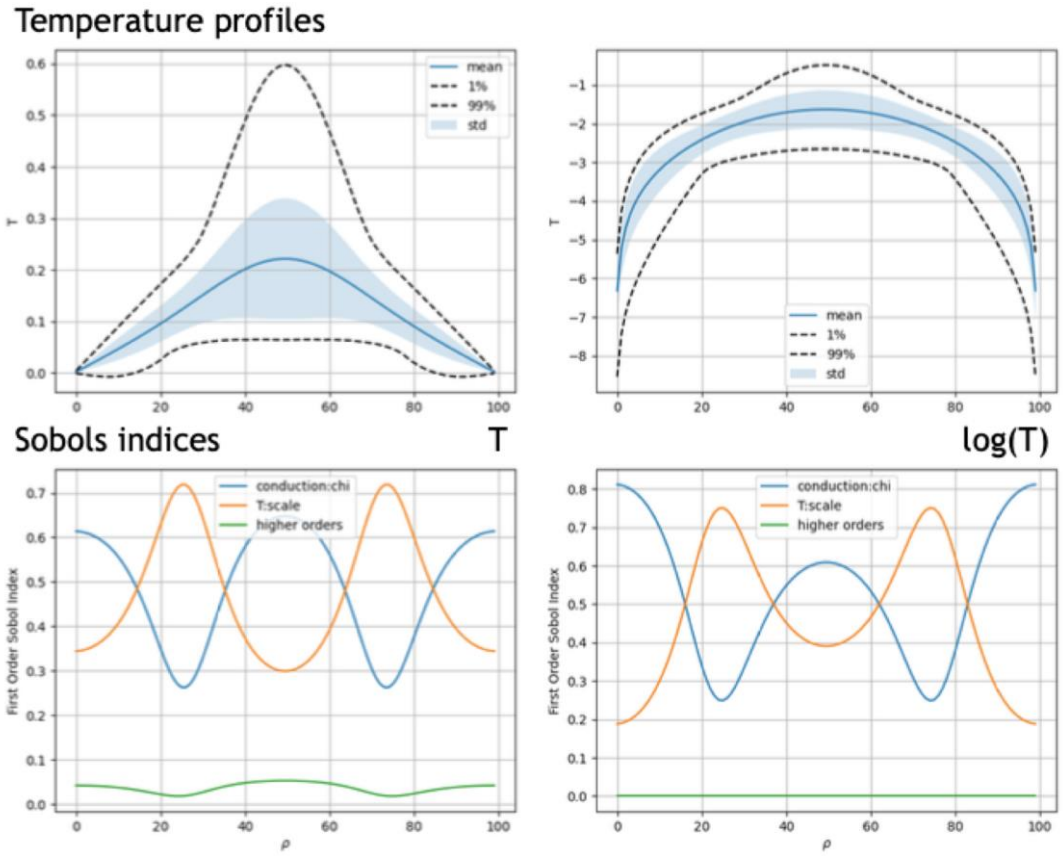


Figure 14 UQ (top) and SA (bottom) of the temperature profile expressed either as T (left) or log(T) (right).

A second campaign was then implemented to quantify uncertainty in the evolution of density perturbations in 2D plasma, varying four parameters: background density, temperature, and 2 dissipation parameters. Third-order PCE would have required 1296 cases, while SC only 256, the latter option was therefore retained.

In later hackathons, a surrogate model was used to determine numerical parameters, and in particular numerical (non-physical) parameters for the design of optimal simulations. The simulated model consisted of the time advance of hyperbolic PDE with elliptic PDE solved every time step:

$$\frac{\partial n}{\partial t} = -\{\phi, n\} + 2\frac{\partial n}{\partial z} + D_n \nabla^2 n$$

$$\frac{\partial \Omega}{\partial t} = -\{\phi, \Omega\} + \frac{2}{n}\frac{\partial n}{\partial z} + \frac{D_\Omega}{n}\nabla^2 \Omega$$

$$\nabla^2 \phi = \Omega$$

These equations were solved using nested solvers, namely using CVODE for time and multigrid for spatial integrations [Hindmarsh 2005]. The surrogate model was built and trained for the aforementioned model which featured a 7-dimensional parameter space and a non-smooth dependence of behaviour on parameters. The QoI in the surrogate model is the error at a given timestep, that is max(Ea, αEr) based on the absolute Ea, the relative Er error tolerances (Ea, Er ∈ [10−15,1]) and α a representative value of ||x||. The first step consisted of using PCE and adaptive SC to generate the surrogate models. Both methods achieved qualitatively similar results, but adaptive SC required many fewer code evaluations (130 vs 441). Both types of surrogates provided a

qualitatively good model of errors at moderate order but over-fitted the noisier data at high order. One main issue was that the accuracy of the surrogate models stopped increasing even though more data was provided.

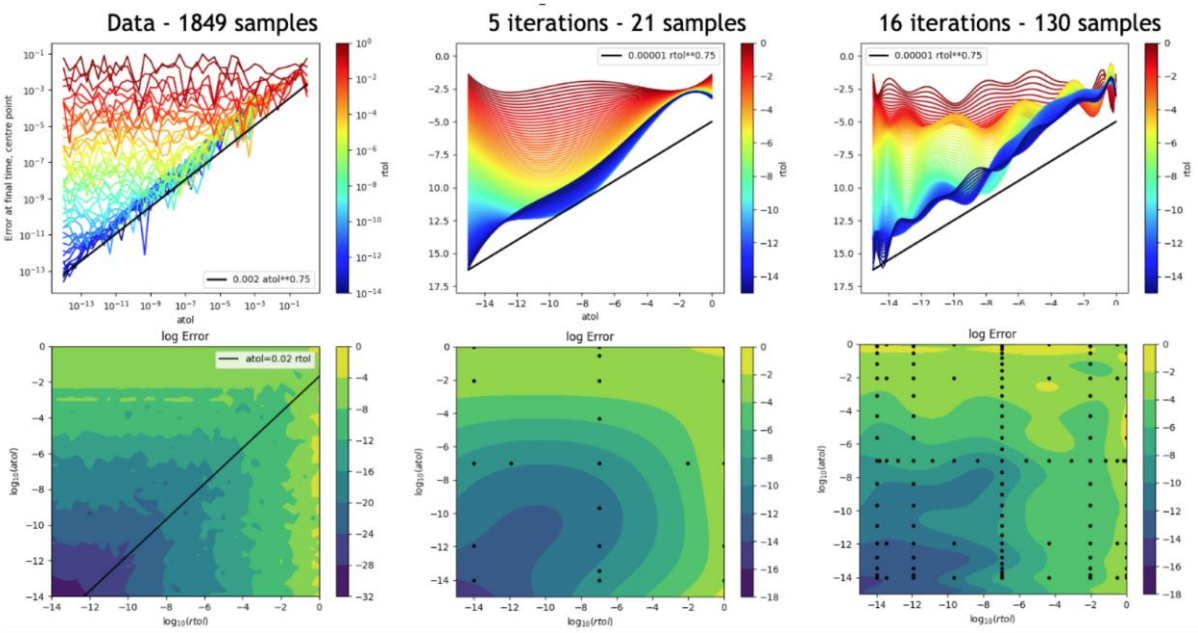


Figure 15 CVODE - 2D scan with adaptive SC. Training data was obtained using a 2D manual parameter sweep (left), error model prediction at partial convergence (middle), at full convergence (right) of adaptive SC algorithm.

Two more advanced surrogate modelling methods were subsequently considered: Artificial Neural Network (ANN) with EasySurrogate, and GPs with the SKLearn library in EasyVVUQ; both capabilities were added to the VECMA Toolkit during the period of this project. ANNs were able to emulate the data correctly, at the cost of training on approximately 50% of the dataset. By contrast, GPs were able to emulate data using only 5% of the dataset, albeit the results were highly sensitive to the sampling points. Adaptive sampling of training data for GP surrogate models might reduce such sensitivity while preserving the need for the smaller parts of the dataset. This is an open question requiring further investigation.

2.2.1.3 EPOCH

Investigations led by Tom Goffrey and collaborators consisted in implementing a surrogate model for stimulated Raman scattering (SRS) in laser-plasma interactions. The surrogate was trained using data produced by the EPOCH code, a mini-app version of which is currently being used in the NEPTUNE project. The EPOCH model simulations are typical of particle-based simulations yielding chaotic systems. Simulations featured 5 to 10 parameters and PCE via EasyVVUQ was applied to build the surrogate models. Custom encoders and decoders were developed for the campaign and execution featuring the QCG-PJ was implemented. A proof of concept sensitivity analysis on the performance of simulated laser-driven implosions was completed. Further conclusions regarding UQ of particle-based methods using EPOCH were not obtained due to time constraints during the course of the hackathons and the project.

Nonetheless, in the meantime, we completed our investigations of standard classical molecular dynamics simulations, another exemplar particle-based simulation method [Vassaux_2021]. Our work

is the first and only fully detailed and thorough analysis of a particle-based code using modern UQ methods. Many-body particle-based simulations are chaotic systems. We showed that the uncertainty arises from a combination of (i) the input parameters and (ii) the intrinsic stochasticity of the method controlled by the random seeds. We performed a sensitivity analysis, which revealed that, out of a total of 175 parameters, just six dominate the variance in the code output. The sensitivity analysis computed first and higher-order Sobol indices, which respectively highlight the individual and combined influence of the parameters. We showed that simulations of free energies dampen the input uncertainty, in the sense that the variation around the mean output free energy is less than the variation around the mean of the assumed input distributions if the output is ensemble-averaged over the random seeds. Without such ensemble averaging, the predicted free energy is five times more uncertain. The distribution of the predicted properties is thus strongly dependent upon the random seed. Owing to this substantial uncertainty, robust statistical measures of uncertainty in molecular dynamics simulation, and more widely we anticipate particle-based simulation methods, require the use of ensembles in all contexts.

2.2.2 Alya: simulations as regulatory evidence

2.2.2.1 Background

Over 5 million people suffer heart failure (HF) in the U.S. alone, with ~1 million new cases annually [Benjamin_2019]. From these patients, about a 10% is in Stage D [Fang_2015] condition, being heart transplant the gold standard treatment. The limited organs availability is making left ventricular assist devices (LVADs) a leading treatment option, with a ~90% 1-year survival rates [Jorde_2014]. LVADs are centrifugal or axial pumps apically implanted that help support the heart to reach the required cardiac output (CO) to sustain life.

There is evidence [Bartoli_2018] that inflammation is associated with LVAD use, combined with the endothelial lesion and the abnormal flow patterns [Rossini_2020] are the three composing parts of the Virchow's triad [Lowe_2003] for thrombus formation. The local flow conditions influence the type of thrombus created. White thrombus are formed in regions with high velocities and high shear stresses that lead to platelet activation [Varga-Szabo_2008] and fibrin aggregation. On the contrary, red thrombus are created by stagnant and slow recirculating flows with low shear stresses that lead to an aggregation of all blood components [Zhao_2008, Tan_2003]. While the latest LVADs generation have a reduced white thrombus formation due to the novel magnetic and hydrodynamic rotors, the patients still suffer thromboembolic events. The reason for this is that the abnormal LV flow patterns combined with the low shear stresses suggest the LV as a relevant site for red thrombus formation.

While there is an extensive number of publications dealing with multiple LVAD factors like ventricular size [Chivukula_2019], cannula implantation position [Prisco_2017], implantation depth [Ong_2013, Liao_2018, Chivukula_2020] or angulation [Neidlin_2021], none of them provide credibility evidence as suggested in the recent ASME V&V40 [ASME_2018], neither guided by the historical V&V20 [ASME_2009] specifically designed for computational fluid dynamics (CFD) more than 10 years ago. The reason for this is, most probably, that such a validation requires a thorough comparison of the simulation results against experiments and hundreds of executions of then numerical model, what

involves a large computational cost. This work follows the V&V40 pipeline for a computational model of a benchtop LV-LVAD system to quantify intraventricular flow patterns.

2.2.2.2 Methods

The bench experiments were performed with the San Diego State University (SDSU) cardiac simulator (CS). This CS is a mock circulation loop of the heart and the circulatory system with an apically implanted LVAD that has been reported previously in [Wong 2014, Garcia 2008]. It involves a silicone LV based on an idealised geometry, immersed in a water-filled tank and connected to an external circulatory loop mimicking the systemic circulation. The tank is fully watertight, so when the piston pump generates negative pressure, the LV expands to the end diastolic volume (EDV). Two beating modes and three pump speeds are used for six validation points. The condition 22[%] @ 68.42[bpm] has EF = 22[%] and HR = 68.42[bpm] with end systolic volume (ESV)=180[cm^3] and EDV=230[cm^3]. The condition 17[%] @ 61.18[bpm] has EF = 17[%] and HR = 61.18[bpm] with ESV=180[cm^3] and EDV=216.86[cm^3]. The pump speeds used for the validation points are 0[rpm], 8k[rpm] and 11k[rpm].

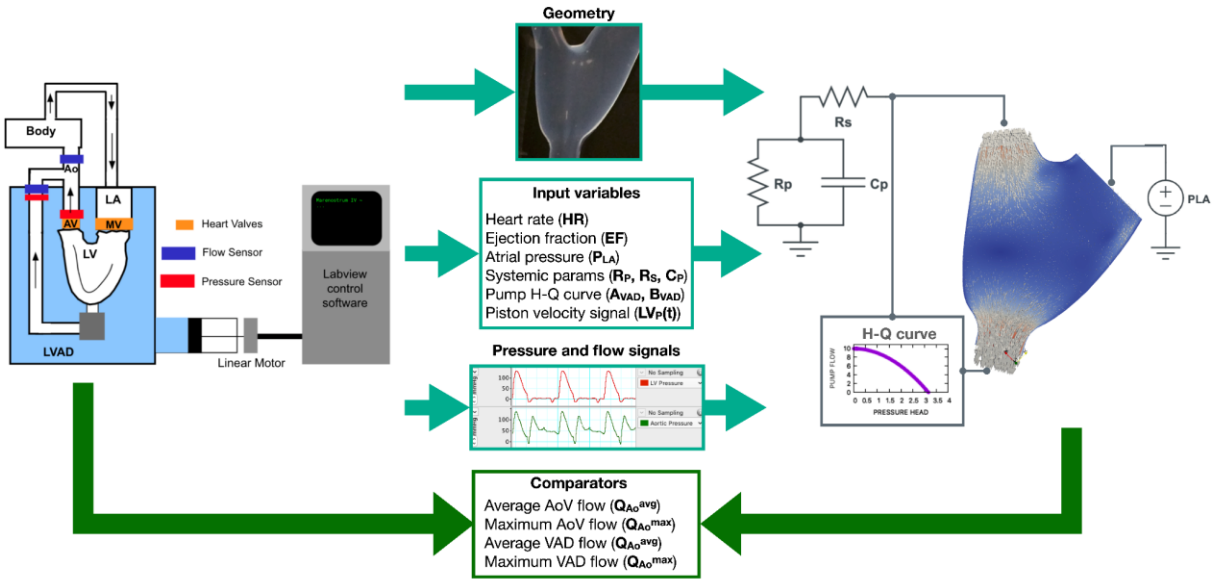


Figure 16 Summary of the experimental-simulation system used for the validation. The simulation domain is created from the silicone ventricle CAD. Multiple input variables are used and afterwards sampled like the heart rate (HR), the ejection fraction (EF), the left atrial pressure (PLA), the systemic characterisation variables (RP,RS,CP), the pump parameters (AVAD, BVAD) and the pressure signal for ventricular deformation LVp(t). Those inputs are feed to the numerical model and the comparators extracted for the Uncertainty quantification.

The computational domain is created from the exact same computer geometry used to manufacture the silicone ventricle. To obtain a computationally inexpensive and accurate way of deforming the ventricle, a unidirectional FSI approach is used to deform the LV (similarly as [Liao 2018]). A pressure is imposed in the external solid domain which afterwards deforms the CFD domain between the ESV and the EDV. Once the simulation pipeline is completed, the input files are modified to work as a template. EasyVVUQ and FabSim3 (in a fork called FabAlya) are used to sample the inputs, execute the instances of the solver and retrieve the results. This is used for the Sensitivity Analysis (SA) and Uncertainty Quantification (UQ).

2.2.2.3 VVUQ plan

The V&V 40 [ASME_2018] standard provides a framework for assessing the relevance and adequacy of the completed VVUQ activities.

- **Question of interest:** For an apically implanted LVAD, does the numerical model that includes as inputs: (a) the pump H-Q performance curve, (b) the heart rate (HR), and (c) the pre-LVAD implantation Ejection Fraction (EF); produce flows and velocity fields that agree with the bench experiment?
- **Context of Use (CoU):** The heart-LVAD computational model may be used to assist in the preclinical design and development of LVAD, by characterising the intraventricular flows for a given pump performance curve. The presented credibility evidence consist of: code and numerical verification by computing the observed rate of convergence in a manufactured solution and a mesh convergence study; UQ with mixed aleatory-epistemic inputs using validation against a bench experiment with six operating conditions. The heart-LVAD computational model will then be used to characterise ventricular flows and derived QoIs, but by no means replacing animal experiments or clinical trials.
- **Model influence:** Although the numerical test will augment the evidence provided by the bench test, animal testing and clinical trials are still required. Therefore the model influence can be categorised as low as it only supports the evidence and it doesn't solely rely on this computational evidence.
- **Decision consequence:** The model is only intended to augment the bench test experiment information related with intra-LV flow fields and not to extract any clinical-related conclusion. Despite this, the model can model the device design in operation points in between the operation points. Therefore the decision consequence is categorised as medium.
- **Model risk assessment:** As the model influence has been categorised as "low" and the decision consequence as "medium", the LV-LVAD model is categorised with a risk of 2 on the 1-5 scale, therefore requiring a mid to low level goals in the VVUQ plan.

These goals are defined in Table 6. The steps to achieve them are:

1. Provide verification evidence: software quality assurance (SQA) practices should be followed to ensure reproducibility. A strict numerical code verification is mandatory to ensure correctness in the coding of the models. Numerical calculation verification is mandatory to ensure a correct spatial discretisation of the problem.
2. Execute a sensitivity analysis in the operation range: A non-linear Sensitivity Analysis (SA) on the operation range of the cases should be executed to (a) understand the impact of each input in the QoIs, (b) Safely reduce the variables for the UQ.
3. Proceed to the uncertainty quantification in multiple validation points: The extreme cases and a middle point should be investigated to ensure a credible solution. A comparison of the QoIs distribution is required including multiple metrics that allow comparing the results with other similar works and future projects.
4. Provide an overall evaluation of the UQ results: a Final analysis in which range the model is credible is required to a safe use of the model for predictions.

Table 6 ASME V&V40 credibility factors [ASME 2018] analysed on the risk-based assessment. The table shows the maximum possible score ("Max." column), the desired goal ("Goal" column) and the obtained score ("Obt." column). The goal column also includes the description of the activity to achieve that gradation.

Aspect				Max.	Evaluation Goal	Obt.
1. Verification	1.1. Code	1.1.1. software quality assurance (SQA)		C	B [SQA procedures are specified and documented.]	C
		1.1.2. Numerical code verification		D	C [The numerical solution is compared to an exact solution.]	D
	1.2. Calculation	1.2.1. Discretisation error		C	B [Convergence analysis are performed obtaining stable behaviours.]	C
		1.2.2. Numerical solver error		C	B [Solver parameters are based on values from a previously verified model.]	B
		1.2.3. User error		D	B [Key inputs and outputs were verified by the practitioner.]	C
2. Validation	2.1. Computational model	2.1.1. Model form		C	B [Influence of some assumptions is explored.]	B
		2.1.2 Model inputs	2.1.2.1. Quantification of sensitivities	C	B [A SA of the expected key parameters is performed.]	C
			2.1.2.2. Quantification of uncertainties	D	B [UQ is executed on expected key inputs but not propagated to the QoIs.]	D
	2.2. Comparator	2.2.1. Test samples	2.2.1.1. Quantity of test samples	C	A [A single sample is used.]	A
			2.2.1.2. Range of characteristics of test samples	D	A [A single test condition is examined.]	A
			2.2.1.3. Measurements of test samples	C	B [One or more key characteristic are measured.]	C
			2.2.1.4. Uncertainty of test samples measurements	C	A [Characteristics uncertainty is not addressed.]	A
		2.2.2. Test conditions	2.2.2.1. Quantity of test conditions	B	B [Multiple test conditions.]	B
			2.2.2.2. Range of test conditions	D	B [Test conditions representing a range of conditions near nominal range are examined.]	C
			2.2.2.3. Measurements of test conditions	C	B [One or more key test conditions are measured.]	B
			2.2.2.4. Uncertainty of test conditions measurements	C	B [UQ of the test conditions incorporated instrument accuracy only.]	B
	2.3. Assessment	2.3.1. Equivalence		C	B [The types of all inputs are similar, but ranges are not equivalent.]	C
		2.3.2. Output Comparison	2.3.2.1. Quantity	B	B [Multiple outputs were compared.]	B
			2.3.2.2. Equivalency of output parameters	C	B [Most types of outputs are similar.]	C
			2.3.2.3. Rigour of output comparison	C	B [Comparison was performed by arithmetic difference.]	C
			2.3.2.4. Agreement of output comparison	C	B [The level of agreement is satisfactory for some key comparisons.]	B
3. Applicability	3.1. Relevance of the Quantity of interest			C	B [A subset of the QoIs are identical to those for the CoU.]	C
	3.2. Relevance of the validation activities to the CoU			D	B [There is partial overlap between the ranges of the validation points and the CoU.]	C

2.2.2.4 Results

The SA is intended to highlight the input parameters with a considerable impact in the QoIs. To proceed with the latin hypercube sampling (LHS) a uniform distribution is considered for the SA. 500 samples are obtained with LHS and shown in the scatter plot at Figure 17 together with the Pearson's correlation coefficient ρ. Pearson's ρ is a measure of the strength of a linear association between the two variables in each bivariate plot. From a visual analysis of the scatter plot it rises that the data is nonlinear, heteroskedastically distributed, and contains multivariate outlayers, failing 3 of the 7

assumptions required for Pearson's analysis. To overcome this issue Sobol indices are calculated. Sobol indices provide information of the importance of each input taking into account complex factors like nonlinearities, input interactions, and sample dispersion. The Total Sobol index of each input with respect to each QoI are shown at the tornado plot in Figure 17. The larger the index, the more important that input is for the QoI.

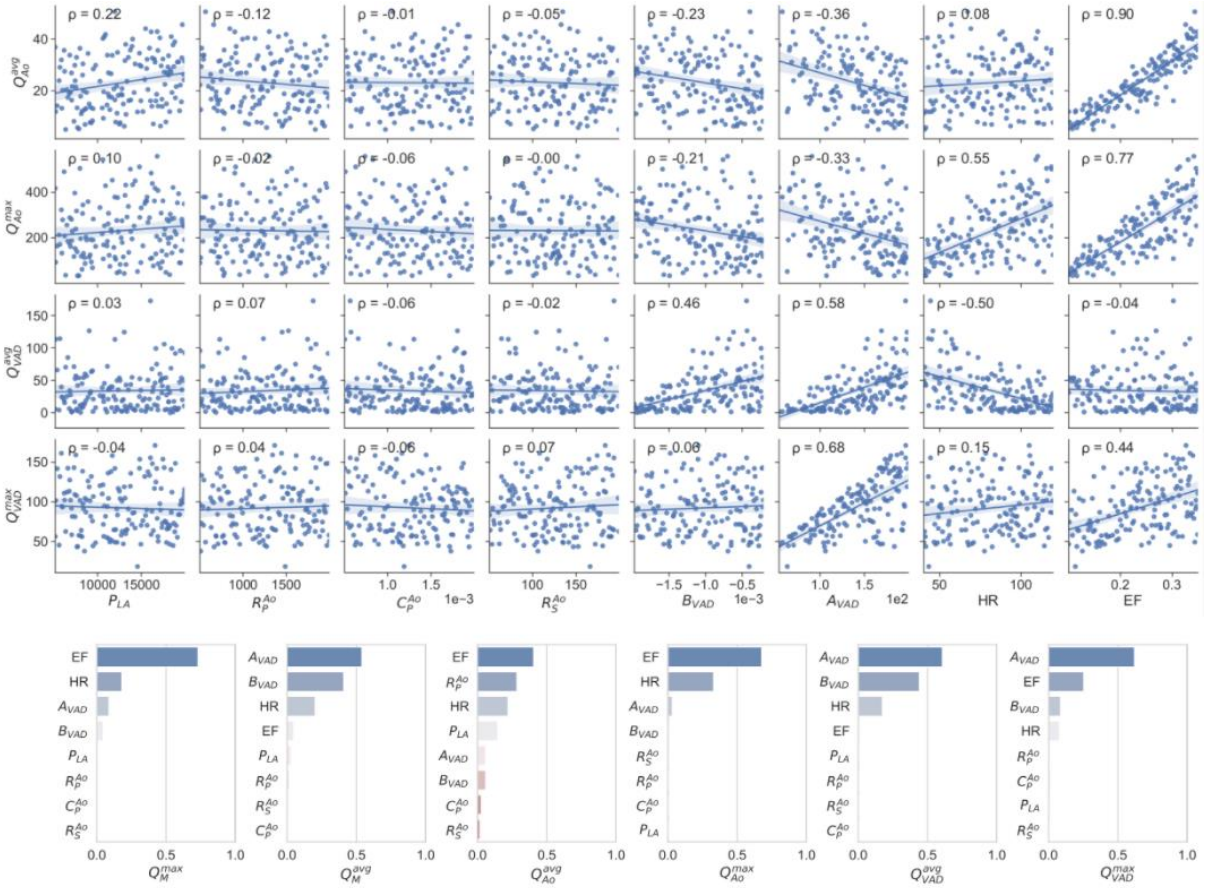
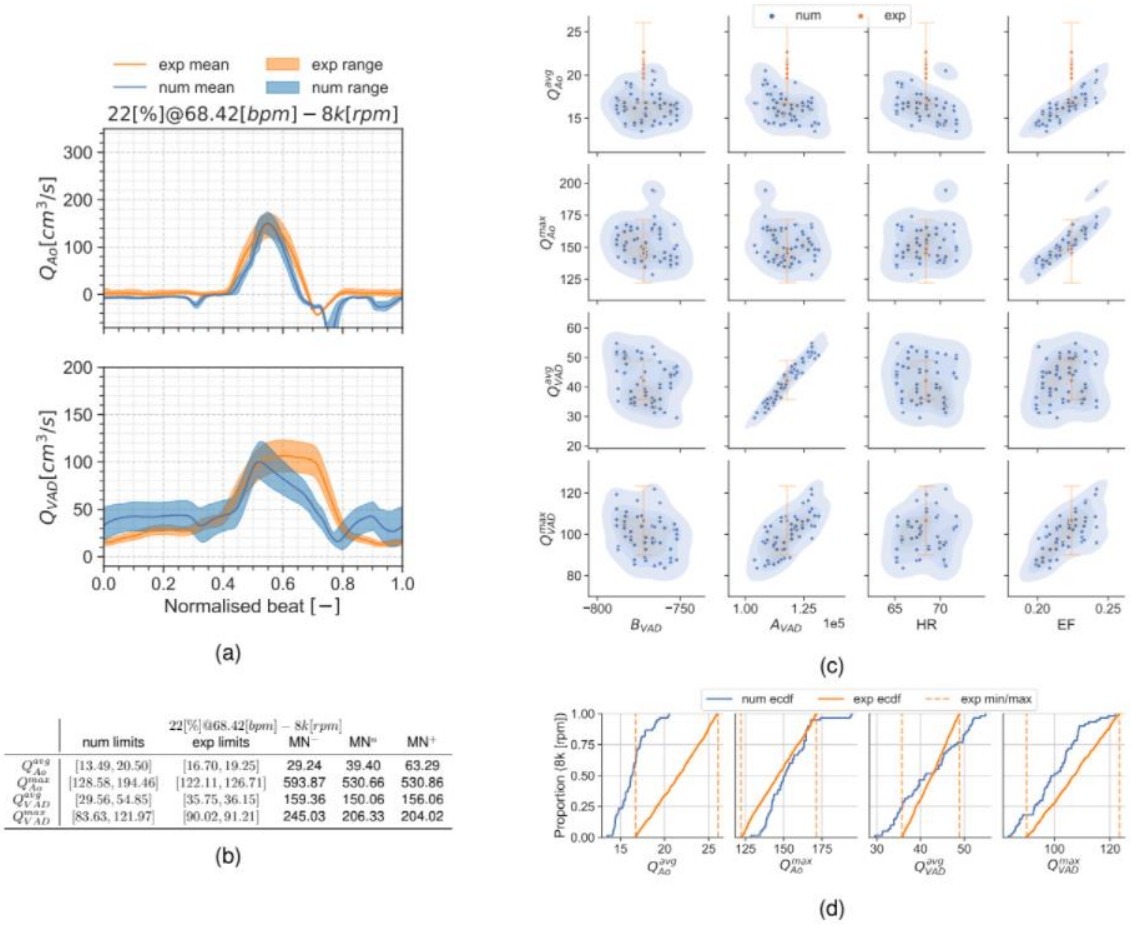


Figure 17 Scatter plots and total Sobol indices tornado plots for the 8 input variables and the 4 QoIs. The scatter plot also shows the Pearson's linear correlation number ρ in the top left corner.

The UQ consists of six validation points. For the sake of brevity, here we show results for a single validation point. As we count with a single execution of the bench experiment, the results are treated with an epistemic error range that is intended to account for the user error. On the contrary, the multiple executions of the numerical experiment let us use the statistical data for the output. Figure 18 shows an example of one validation point (22[%]@68.42[bpm] and 8k[rpm]). Results are analysed through scatter plots, empirical cumulative distribution functions (ECDFs), and multiple validation metrics. The validation metric is computed as in [Wong_2014], the minimum Minkowski L1 norm (MN) is chosen between the experimental range and the numerical distribution.



	num limits	exp limits	MN⁻	MN^	MN⁺
Q_{Ao}^{avg}	[13.49, 20.50]	[16.70, 19.25]	29.24	39.40	63.29
Q_{Ao}^{max}	[128.58, 194.46]	[122.11, 126.71]	593.87	530.66	530.86
Q_{VAD}^{avg}	[29.56, 54.85]	[35.75, 36.15]	159.36	150.06	156.06
Q_{VAD}^{max}	[83.63, 121.97]	[90.02, 91.21]	245.03	206.33	204.02

(b)

Figure 18 Summary for the condition 22[%]@68.42[bpm] and 8k[rpm]. (a): aortic valve and LVAD flows. (b): validation metrics. (c): scatter plot showing the simulation and experimental data. (d): ECDF for the simulation, experimental data limits and the constructed uniform distribution.

2.2.2.5 Conclusion

SA and UQ techniques are mandatory to create credible numerical results, at the cost of a high computational cost. Despite that fact, this work shows that credible simulations for regulatory support are readily available, if the HPC resources are not a technical difficulty.

2.2.3 UrbanAir

The UrbanAir (UA) application is tailored towards assessing and predicting air quality over complex urban areas. It is a multi-scale model that combines two community models – WRF (Weather Research and Forecast model), responsible for weather prediction at mesoscale or regional level, and EULAG – all scale geophysical flow which aims at solving transport of contaminants at city level. Predicting air quality is a challenging problem that requires a trade-off between the accuracy of results and acceptable time-to-solution. UA uses an immersed boundary method to accurately represent complex building geometry, and thus deliver a proper flow around buildings and contaminants transportation. It is therefore a computationally demanding application.

The UA requires accurate weather prediction as input data, as well as accurate emission rates for different pollutants (*e.g.* SO2, NO2, NOx, PM2.5, PM10) and different types of sources. These include points sources – attributed to industrial chimneys, line sources – attributed to road transportation,

and area sources – attributed to heat appliances. The quality of prediction depends on the quality of input data. Taking the prediction of NO2/NOx as an example, which is attributed mainly to road transportation, initial information required includes number of cars passing the street, ratio between diesel and gasoline engines, fuel usage, fuel density, ratio between engine hot and cold start, *etc*. Some of these inputs may be estimated more accurately, *e.g.* number of cars (from statistical data or via additional simulation), while some remain a puzzle, *e.g.* ratio between hot and cold engines. To overcome these shortcomings, non-intrusive UQ has been applied to the EULAG model by means of Stochastic Collocation. The reason is two-folded. First, it allows us to run ensembles – simulations with different set of input parameters – to provide mean results taking into account uncertainty of input parameters of unknown value. Second, it helps to learn which input parameter impacts simulation results by the means of sensitivity analysis.

With VECMAtk we sampled 8 input parameters: total number of cars, diesel to gasoline ratio and fuel usage, density, NO2 index for each type of fuel. For the selected uncertainty quantification method, 256 ensembles were required to be computed. The analysis presented in Table 7, showed four parameters to be the most influential. This analysis allowed us then to limit the number of ensembles from 256 (with 8 input parameters being sampled) to 16 (4 input parameters being sampled) for future runs.

Table 7 Sensitivity analysis of UA input parameters

Parameters	Sobol first index
gas_no2_index	0.14294
gas_density	0.13251
oil_no2_index	0.11094
oil_density	0.10523
no_of_cars	0.0831
gas_usage	0.0592
oil_usage	0.0591
gas_cars_ratio	0.00184

The aforementioned study resulted in 256 ensembles, running on total 6144 CPU cores, consuming ca. 30k CPU core-hours. By limiting the number of runs we lower the CPU core-hours to just 1920. The VECMAtk orchestrates execution of the ensembles on HPC machines via the QCGPJ module – there is no need to populate runs manually over nodes, nor to map application instances to nodes, nor to change the code where different hardware resources are available. Moreover, it allows us to see the differences in results between the ensembles. Last but not least, the ensembles generated and run via the VECMAtk makes it easier to understand differences in results between different runs, and to provide an average result. In this way we can analyze how NO2 concentration is changing with height, as presented in Figure 19. We are also analyzing how contamination concentration is changing through the whole domain at given height, as presented in Figure 20.

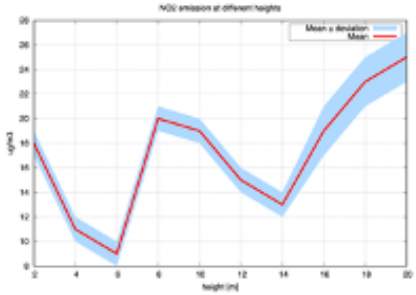


Figure 19 NO2 concentration at different heights for a given point in 2D space.

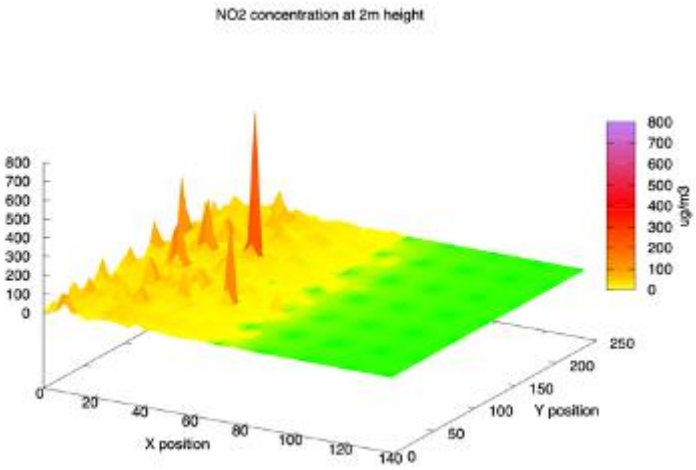


Figure 20 NO2 concentration at 2m height for the whole domain.

See references [Wright_2020], [Groen_2021] and [Suleimenova_2021b].

2.2.5 Coronavirus simulations

The COVID-19 pandemic was not foreseen when the VECMA project was set up, but caused a redirection of some resources once it started. Three such efforts are briefly described below.

2.2.5.1 CovidSim

CovidSim is a well-known and influential epidemiological code, that has been used to advise the UK government on the effects of various non-pharmaceutical interventions, *e.g.* social distancing and school closures. It has a large number of parameters, which prompted us to implement a dimension-adaptive version of Stochastic Collocation in EasyVVUQ. This work was published in Nature Computational Science [Edeling_2021], and is already described in D4.3. Here, we also mentioned the use of deep-active subspaces to quantify the uncertainty in CovidSim due to 19 chosen uncertain inputs. We have since increased this number to 41 without scaling difficulties, and are now planning to execute a 60 dimensional UQ campaign on the PSNC Eagle Supercomputer.

Finally, we also implemented a global derivative-based sensitivity metric into EasySurrogate, which allows us to identify important inputs using the deep-active subspace network. The results for the case of 41 parameters are shown below. The same parameters are identified as important compared to our dimension-adaptive study of [Edeling_2021], which makes us confident in these results. Note that the

majority of inputs have no impact on the output. We think that machine-learning methods as described here therefore may have the potential of weeding out many uninfluential parameters from a large input set, as they scale well with respect to the input dimension. If one is wary of using a machine-learning surrogate, it may be possible to construct a different (`classic') surrogate model on the identified set of important inputs, if this set is small enough.

Quite a number of toolkit components were engaged in the study of 41 parameters. The network was trained using EasySurrogate, on data from a simple Monte Carlo campaign generated by EasyVVUQ, which was submitted to the PSNC Eagle supercomputer using the QCG-PilotJob mechanism, with the overall workflow and data transfer managed by Fabsim3.

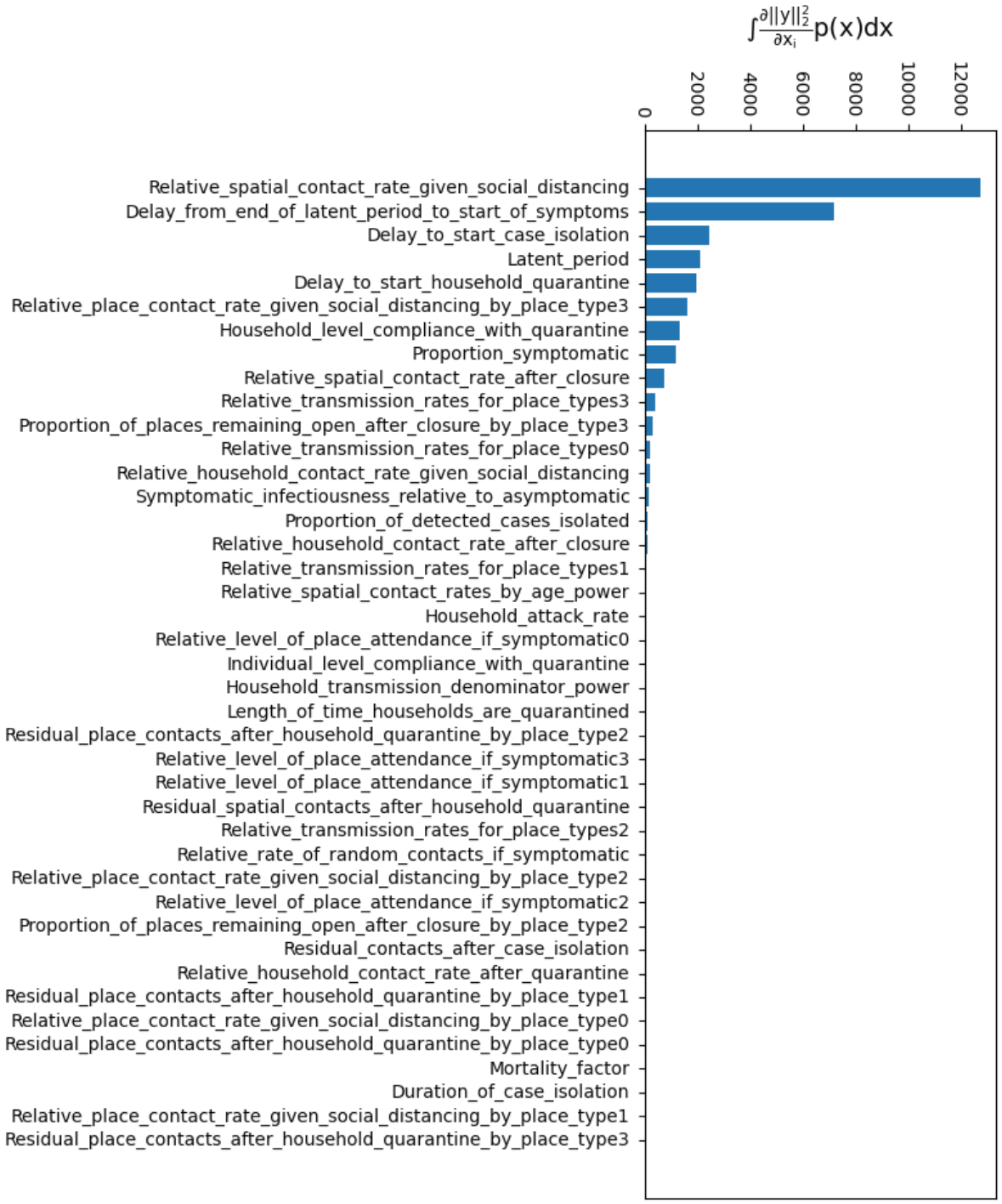


Figure 21 The global, derivative-based sensitivity indices extracted from the deep active subspace network.

2.2.5.2 Flu and Coronavirus Simulator (FACS)

The Flu and Coronavirus Simulator (facs.readthedocs.io) is an agent-based simulation tool that models the spread of COVID-19 in local areas. The tool is in use by a range of health partners, and has for example been applied to forecast COVID-19 spread in several boroughs in West London.

VECMAtk benefitted this application by facilitating ensemble simulation forecasts in a scalable and portable way, by automating a range of pre- and post-processing tasks, and by facilitating the analysis of sensitivity analysis covering a range of a key simulation parameters (*e.g.*, latent period and public health measure uptake rates).

Many of the FACS workflows are defined in FabCovid19, which in turn relies on the FabSim3 automation toolkit. In addition, we use EasyVVUQ to perform parts of our sensitivity analysis for the code, and QCG-PilotJob to more rapidly execute our ensemble- and scenario-based forecasts. A first impression of the work being done in this regard can be found in [Groen_2021].

2.2.5.3 Dutch-Covid Model

This study is already described in detail in D4.3 and was largely completed at that time. It used EasyVVUQ to not only assess the impact of physical parameters, but also examined the effect of random seeds. The only update here is that our article was accepted for computation in PLOS Computational Biology [Gugole_2021], a leading journal in the field.

3 Conclusions

Section 2.1 VECMA exemplar applications showed that all the VECMA applications have made use of the VECMA developed tools to their research areas. The use of these techniques has improved the "actionability" of the results by quantifying the uncertainties in the predictions and addressing verifiability and verification of the workflows.

Section 2.2 External applications showed the results from external partners (or, in some cases, work external to the VECMA project done by VECMA partners in collaboration with external partners) applying VECMA tools to problems in fusion, biotechnology, studies of pollution and epidemiology.

This deliverable has shown the that the approach that VECMA has developed is useful to an extensive range of applications. A range of UQPs have been applied to a set of applications covering fusion, climate, materials science, biology, migration, *etc*. The UQPs have allowed the researchers to understand what impact the uncertain inputs have on the uncertainty of the model outputs, even for problems with many inputs, and to verify and validate these results. The implementation of these ideas into the VVUQ toolkit has allowed the members of the project to perform this work effectively and efficiently and has also been taken up by users outside of the VECMA project to investigate UQ and V&V for their problems, allowing them a deeper understanding of the sources of uncertainty in their results.

4 Annexes

None

5 References

[ASME_2009] American Society of Mechanical Engineers, 2009. “Standard for Verification and Validation in Computational Fluid Dynamics and Heat Transfer: ASME V&V 20”. The American Society of Mechanical Engineers (ASME).

[ASME_2018] American Society of Mechanical Engineers, 2018. “Assessing Credibility of Computational Modeling through Verification and Validation: Application to Medical Devices - V V 40 - 2018”. Asme V&V 40-2018, p. 60.

[Bartoli_2018] Bartoli, C. R., Zhang, D., Kang, J., Hennessy-Strahs, S., Restle, D., Howard, J., Redline, G., Bermudez, C., Atluri, P., and Acker, M. A., 2018. “Clinical and in vitro evidence that subclinical hemolysis contributes to lvad thrombosis”. The Annals of thoracic surgery, 105(3), pp. 807–814.

[Benjamin_2019] Benjamin, E. J., Muntner, P., Alonso, A., Bittencourt, M. S., Callaway, C. W., Carson, A. P., Chamberlain, A. M., Chang, A. R., Cheng, S., Das, S. R., et al., 2019. “Heart disease and stroke statistics—2019 update: a report from the American heart association”. Circulation, 139(10), pp. e56–e528.

[Bosak_2021] Bosak, Bartosz, Tomasz Piontek, Paul Karlshoefer, Erwan Raffin, Jalal Lakhlili, and Piotr Kopta. ‘Verification, Validation and Uncertainty Quantification of Large-Scale Applications with QCG-PilotJob’. In Computational Science – ICCS 2021, edited by Maciej Paszynski, Dieter Kranzlmüller, Valeria V. Krzhizhanovskaya, Jack J. Dongarra, and Peter M.A. Sloot, 12746:495–501. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2021. https://doi.org/10.1007/978-3-030-77977-1_39.

[Chivukula_2019] Chivukula, V. K., Beckman, J. A., Prisco, A. R., Lin, S., Dardas, T. F., Cheng, R. K., Farris, S. D., Smith, J. W., Mokadam, N. A., Mahr, C., et al., 2019. “Small lv size is an independent risk factor for vad thrombosis”. ASAIO journal (American Society for Artificial Internal Organs: 1992), 65(2), p. 152.

[Chivukula_2020] Chivukula, V. K., Beckman, J. A., Li, S., Masri, S. C., Levy, W. C., Lin, S., Cheng, R. K., Farris, S. D., Wood, G., Dardas, T. F., et al., 2020. “Left ventricular assist device inflow cannula insertion depth influences thrombosis risk”. Asaio Journal, 66(7), pp. 766–773.

[Coster_2021] David P Coster, Rainer Fischer, Olivier Hoenen, Jalal Lakhlili, Onnie O Luk, Roland Preuss, Bruce D Scott, and Udo von Toussaint. ‘Building a Turbulence-Transport Workflow Incorporating Uncertainty Quantification for Predicting Core Profiles in a Tokamak Plasma’. Nuclear Fusion, 2 November 2021. https://doi.org/10.1088/1741-4326/ac359f.

[Crommelin_2021] Crommelin, D., & Edeling, W. (2021). Resampling with neural networks for stochastic parameterization in multiscale systems. Physica D: Nonlinear Phenomena, 422, 132894.

[Edeling_2020] Edeling, W., & Crommelin, D. (2020). Reducing data-driven dynamical subgrid scale models by physical constraints. Computers & Fluids, 201, 104470.

[Edeling_2021] Edeling, W., Arabnejad, H., Sinclair, R., Suleimenova, D., Gopalakrishnan, K., Bosak, B., ... & Coveney, P. V. (2021). The impact of uncertainty on predictions of the CovidSim epidemiological code. Nature Computational Science, 1(2), 128-135.

[Fang_2015] Fang, J. C., Ewald, G. A., Allen, L. A., Butler, J., Canary, C. A. W., Colvin-Adams, M., Dickinson, M. G., Levy, P., Stough, W. G., Sweitzer, N. K., et al., 2015. “Advanced (stage d) heart failure: a statement from the heart failure society of america guidelines committee”. Journal of cardiac failure, 21(6), pp. 519–534.

[Garcia_2008] Garcia, M. A. Z., Enriquez, L. A., Dembitsky, W., and May-Newman, K., 2008. “The effect of aortic valve incompetence on the hemodynamics of a continuous flow ventricular assist device in a mock circulation”. ASAIO journal, 54(3), pp. 237–244.

[Groen_2019] Groen, Derek, Robin A. Richardson, David W. Wright, Vytautas Jancauskas, Robert Sinclair, Paul Karlshoefer, Maxime Vassaux, et al. ‘Introducing VECMAtk - Verification, Validation and Uncertainty Quantification for Multiscale and HPC Simulations’. In Computational Science – ICCS 2019, edited by João M. F. Rodrigues, Pedro J. S. Cardoso, Jânio Monteiro, Roberto Lam, Valeria V. Krzhizhanovskaya, Michael H. Lees, Jack J. Dongarra, and Peter M.A. Sloot, 479–92. Cham: Springer International Publishing, 2019.

[Groen_2021] D. Groen, H. Arabnejad, V. Jancauskas, W. N. Edeling, F. Jansson, R. A. Richardson, J. Lakhlili, L. Veen, B. Bosak, P. Kopta, D. W. Wright, N. Monnier, P. Karlshoefer, D. Suleimenova, R. Sinclair, M. Vassaux, A. Nikishova, M. Bieniek, Onnie O. Luk, M. Kulczewski, E. Raffin, D. Crommelin, O. Hoenen, D. P. Coster, T. Piontek and P. V. Coveney, “VECMAtk: a scalable verification, validation and uncertainty quantification toolkit for scientific simulations”, Phil. Trans. R. Soc. A. 379, 20200221 (2021), DOI:10.1098/rsta.2020.0221

[Gugole_2021] Gugole F, Coffeng LE, Edeling W, Sanderse B, de Vlas SJ, Crommelin D (2021) Uncertainty quantification and sensitivity analysis of COVID-19 exit strategies in an individual-based transmission model. PLoS Comput Biol 17(9): e1009355. https://doi.org/10.1371/journal.pcbi.1009355.

[Hindmarsh_2005] Hindmarsh, A. C., Brown, P. N., Grant, K. E., Lee, S. L., Serban, R., Shumaker, D. E., and Woodward, C. S. (2005). SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers. ACM Transactions on Mathematical Software, 31, 363–396.

[Jahani_2021] A. Jahani, H. Arabnejad, D. Suleimanova, M. Vuckovic, I. Mahmood and D. Groen (2021) “Towards a Coupled Migration and Weather Simulation: South Sudan Conflict”. In: Paszynski M., Kranzlmüller D., Krzhizhanovskaya V.V., Dongarra J.J., Sloot P.M. (eds) Computational Science – ICCS 2021. Lecture Notes in Computer Science, vol 12746. Springer, Cham. DOI: 10.1007/978-3-030-77977-1_40

[Jansson_2021] Jansson, F., Edeling, W., Attema, J., & Crommelin, D. (2021). Assessing uncertainties from physical parameters and modelling choices in an atmospheric large eddy simulation model. Philosophical Transactions of the Royal Society A, 379(2197), 20200073.

[Jorde_2014] Jorde, U. P., Kushwaha, S. S., Tatooles, A. J., Naka, Y., Bhat, G., Long, J. W., Horstmanshof, D. A., Kormos, R. L., Teuteberg, J. J., Slaughter, M. S., et al., 2014. “Results of the destination therapy post-food and drug administration approval study with a continuous flow left ventricular assist device: a prospective study using the intermacs registry (interagency registry for mechanically assisted circulatory support)”. Journal of the American College of Cardiology, 63(17), pp. 1751–1757.

[Lakhlili_2020] Lakhlili, Jalal, Olivier Hoenen, Onnie O. Luk, and David P. Coster. ‘Uncertainty Quantification for Multiscale Fusion Plasma Simulations with VECMA Toolkit’. In Computational Science – ICCS 2020, edited by Valeria V. Krzhizhanovskaya, Gábor Závodszky, Michael H. Lees, Jack J. Dongarra, Peter M. A. Sloot, Sérgio Brissos, and João Teixeira, 719–30.

Cham: Springer International Publishing, 2020. https://doi.org/10.1007/978-3-030-50436-6_53.

[Liao_2018] Liao, S., Neidlin, M., Li, Z., Simpson, B., and Gregory, S. D., 2018. “Ventricular flow dynamics with varying lvad inflow cannula lengths: In-silico evaluation in a multiscale model”. Journal of biomechanics, 72, pp. 106–115.

[Lowe_2003] Lowe, G. D., 2003. “Virchow’s triad revisited: abnormal flow”. Pathophysiology of haemostasis and thrombosis, 33(5-6), pp. 455–457.

[Luk_2018] Luk, OO, O Hoenen, A Bottino, BD Scott, and DP Coster. ‘Optimization of Multiscale Fusion Plasma Simulations within the ComPat Framework’. In 45th EPS Conference on Plasma Physics. European Physical Society, 2018.

[Luk_2019a] Luk, O. O., O. Hoenen, O. Perks, K. Brabazon, T. Piontek, P. Kopta, B. Bosak, A. Bottino, B. D. Scott, and D. P. Coster. ‘Application of the Extreme Scaling Computing Pattern on Multiscale Fusion Plasma Modelling’. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 377, no. 2142 (8 April 2019): 20180152. https://doi.org/10.1098/rsta.2018.0152.

[Luk_2019b] Luk, O.O., O. Hoenen, A. Bottino, B.D. Scott, and D.P. Coster. ‘ComPat Framework for Multiscale Simulations Applied to Fusion Plasmas’. Computer Physics Communications 239 (June 2019): 126–33. https://doi.org/10.1016/j.cpc.2018.12.021.

[Luk_2021a] Luk, O. O., J. Lakhlili, O. Hoenen, U. von Toussaint, B. D. Scott, and D. P. Coster. ‘Towards Validated Multiscale Simulations for Fusion’. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 379, no. 2197 (17 May 2021): rsta.2020.0074, 20200074. https://doi.org/10.1098/rsta.2020.0074.

[Luk_2021b] Luk, OO, J Lakhlili, U von Toussaint, LE Veen, Y Yudin, O Hoenen, BD Scott, and DP Coster. ‘Performing Uncertainty Quantification on Multiscale Fusion Plasma Simulations with the VECMA Toolkit’. In 47th EPS Conference on Plasma Physics, 2021. http://ocs.ciemat.es/EPS2021PAP/pdf/P2.1075.pdf.

[Neidlin_2021] Neidlin, M., Liao, S., Li, Z., Simpson, B., Kaye, D. M., Steinseifer, U., and Gregory, S., 2021. “Understanding the influence of left ventricular assist device inflow cannula alignment and the risk of intraventricular thrombosis”. Biomedical engineering online, 20(1), pp. 1–14.

[Nikishova_2019] Nikishova, A., et al. "Semi-intrusive multiscale metamodelling uncertainty quantification with application to a model of in-stent restenosis." Philosophical Transactions of the Royal Society A 377.2142 (2019): 20180154.

[Ong_2013] Ong, C., Dokos, S., Chan, B., Lim, E., Al Abed, A., Osman, N., Kadiman, S., and Lovell, N. H., 2013. “Numerical investigation of the effect of cannula placement on thrombosis”. Theoretical Biology and Medical

[Prisco_2017] Prisco, A. R., Aliseda, A., Beckman, J. A., Mokadam, N. A., Mahr, C., and Garcia, G. J., 2017. “Impact of lvad implantation site on ventricular blood stagnation”. ASAIO journal (American Society for Artificial Internal Organs: 1992), 63(4), p. 392.

[Richardson_2020] Richardson, Robin A., David W. Wright, Wouter Edeling, Vytautas Jancauskas, Jalal Lakhlili, and Peter V. Coveney. ‘EasyVVUQ: A Library for Verification, Validation and Uncertainty Quantification in High Performance Computing’. Journal of Open Research Software 8 (29 April 2020): 11. https://doi.org/10.5334/jors.303.

[Rossini_2020] Rossini, L., Braun, O. O., Brambatti, M., Benito, Y., Mizeracki, A., Miramontes, M., Nguyen, C., Martinez-Legazpi, P., Almeida, S., Kraushaar, M., et al., 2020. “Intraventricular flow patterns in patients treated with left ventricular assist devices”. ASAIO Journal.

[Suleimenova_2017] D. Suleimenova, D. Bell, and D. Groen, (2017) “A generalized simulation development approach for predicting refugee destinations”, Scientific Reports, 7:13377. DOI: 10.1038/s41598-017-13828-9

[Suleimenova_2021a] D. Suleimenova, H. Arabnejad, W. N. Edeling, and D. Groen, (2021) “Sensitivity-driven Simulation Development: A case study in forced migration”, Philosophical Transactions of the Royal Society A, 379 (20200077). DOI: 10.1098/rsta.2020.0077

[Suleimenova_2021b] D. Suleimenova, H. Arabnejad, W. N. Edeling, D. Coster, O. O. Luk, J. Lakhlili, V. Jancaskas, M Kulczewski, L. Veen, D. Ye, P. Zun, V. Krzhizhanovskaya, A. Hoekstra, D. Crommelin, P. V. Coveney and D. Groen, (2021) “Tutorial applications for Verification, Validation and Uncertainty Quantification using VECMA toolkit”, Journal of Computational Science, 53 (101402). DOI: 10.1016/j.jocs.2021.101402

[Suter_2020] Suter, James L., Robert C. Sinclair, and Peter V. Coveney. "Principles governing control of aggregation and dispersion of graphene and graphene oxide in polymer melts." Advanced Materials 32.36 (2020): 2003213.

[Tan_2003] Tan, K. T., and Lip, G. Y., 2003. “Red vs white thrombi: treating the right clot is crucial”. Archives of internal medicine, 163(20), pp. 2534–2535.

[Varga-Szabo_2008] Varga-Szabo, D., Pleines, I., and Nieswandt, B., 2008. “Cell adhesion mechanisms in platelets”. Arteriosclerosis, thrombosis, and vascular biology, 28(3), pp. 403–412.

[Vassaux_2021] Vassaux, M., Wan, S., Edeling, W. N., & Coveney, P. V. (2021). Ensembles Are Required to Handle Aleatoric and Parametric Uncertainty in Molecular Dynamics Simulation. Journal of Chemistry Theory and Computation, in press, doi: 10.1021/acs.jctc.1c00526

[Wong_2014] Wong, K., Samaroo, G., Ling, I., Dembitsky, W., Adamson, R., Del Alamo, J., and May-Newman, K., 2014. “Intraventricular flow patterns and stasis in the lvad-assisted heart”. Journal of biomechanics, 47(6), pp. 1485–1494.

[Wright_2020] D. W. Wright, R. A. Richardson, W. Edeling, J. Lakhlili, 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. Weglarz, D. Crommelin, D. Groen, and P. V. Coveney, “Building confidence in simulation: Applications of EasyVVUQ”, Advanced Theory and Simulations, 1900246 (2020) DOI:10.1002/adts.201900246

[Ye_2021] Ye, Dongwei, et al. "Non-intrusive and semi-intrusive uncertainty quantification of a multiscale in-stent restenosis model." Reliability Engineering & System Safety 214 (2021): 107734.

[Zhao_2008] Zhao, R., Marhefka, J. N., Shu, F., Hund, S. J., Kameneva, M. V., and Antaki, J. F., 2008. “Micro-flow visualization of red blood cell-enhanced platelet concentration at sudden expansion”. Annals of biomedical engineering, 36(7), p. 1130.