9th International Conference on Sensitivity Analysis of Model Output

28th - 30th of October 2019

Castelldefels (Barcelona), Spain

Table of Contents

Keynotes

‘‘Randomized quasi-Monte Carlo methods in global sensitivity analysis’’ - Giray Ökten

‘‘What do We Want from Sensitivity Analysis?’’ - Emanuele Borgonovo

‘‘Recent advances in sensitivity analysis and screening with dependence measures’’ - Amandine Marrel

“Uncertainty Quantification and Bayesian Model Calibration Applied to Stochastic Systems” - David Higdon

“Active Subspaces: Some Ideas for Off-Axis Sensitivity Analysis” - Paul Constantine

Session 1: Meta-modelling in sensitivity analysis I

Global sensitivity analysis for stochastic simulators based on generalized lambda surrogate models

Global sensitivity analysis aims at quantifying the impact of input variables onto the variation of the response of a computational model. Classically, such models (also called simulators) are deterministic, in the sense that repeated runs provide the same output quantity of interest. In contrast, stochastic simulators return different results when run twice with the same input values due to additional sources of stochasticity in the code itself. In other words, the output of a stochastic simulator is a random variable for a given vector of input parameters.

Many sensitivity measures, such as the Sobol' indices and Borgonovo indices [1], have been developed in the context of deterministic simulators. They can be directly extended to stochastic simulators [2,3], despite the additional randomness of the latter. The calculation of such measures can be carried out through Monte Carlo simulation, which would require many model evaluations though. However, high-fidelity models are often time-consuming: a single model run may require hours or even days. In consequence, direct application of Monte Carlo simulations to calculate sensitivity measures becomes intractable.

To alleviate the computational burden, surrogate models are constructed so as to mimic the original numerical model at a smaller computational cost though. For deterministic simulators, surrogate models have been successfully developed over the last decade, e.g. polynomial chaos expansions [4]. However, the question of appropriate surrogate modelling for stochastic simulators arose only recently in engineering.

In this study, we propose to use generalized lambda distributions to flexibly approximate the response of a stochastic simulator. Under this setting, the parameters of the generalized lambda distribution become deterministic functions of the input variables. In this contribution, we use sparse polynomial chaos expansions to represent the latter. To construct such a sparse generalized lambda model, we develop an algorithm that combines feasible generalized least-squares with stepwise regression. This method does not require repeated model evaluations for the same input parameters to account for the random nature of the output, and thus it reduces the total number of model runs drastically.

Once the stochastic emulator is constructed, one can easily evaluate the conditional mean and variance, which is needed for the Sobol' indices calculation. Because the generalized lambda distribution parametrizes the output quantile function, the surrogate model is expressed as a deterministic function of the input variables and a latent uniform random variable that represents the randomness of the output. As a result, instead of calculating the Sobol' indices for each input variable through sampling, we can derive analytically the Sobol' indices by some suitable post-processing. Moreover, the generalized lambda model provides the conditional distribution of the output given any input parameters. Therefore, distribution-based sensitivity measures, such as Borgonovo indices, can also be calculated straightforwardly.

References

[1] E. Borgonovo, A new uncertainty importance measure. Reliab. Eng. Sys. Safety, 92:771-784, 2007.

[2] A. Marrel, B. Iooss, S. Da Veiga, and M. Ribatet, Global sensitivity analysis of stochastic computer models with joint metamodels, Stat. Comput., 22:833-847, 2012.
[3] M. N. Jimenez, O. P. Le Maître, and O. M. Knio, Nonintrusive polynomial chaos expansions for sensitivity analysis in stochastic differential equations, SIAM/ASA J. Uncer. Quant., 5:378-402, 2017
[4] B. Sudret, Global sensitivity analysis using polynomial chaos expansions, Reliab. Eng. Sys. Safety, 93:964-979, 2008

Xujia Zhu
zhu@ibk.baug.ethz.ch
Chair of Risk, Safety and Uncertainty Quantification
ETH Zurich, Switzerland

Bruno Sudret
sudret@ibk.baug.ethz.ch
Chair of Risk, Safety and Uncertainty Quantification
ETH Zurich, Switzerland

Efficient sensitivity analysis using Gaussian Process-based model order reduction

To optimise or assess hazards in engineering systems one must understand and predict their behaviour across a range of conditions. Where hazard assessment is a concern, the range of potential conditions is necessarily broad and therefore, even with efficient models, the necessary sensitivity analysis requires considerable computational effort. To mitigate this, direct interrogation of the underlying model may be replaced by a reduced model encapsulating the system behaviour in a cheaper, simpler framework. This model order reduction is typically achieved by developing surrogate models, ideally with a reduced input space.

A common choice of surrogate is the Gaussian Process , which is computationally efficient and facilitates further analysis through the analytic tractability of its form. This work employs Gaussian Processes with Automatic Relevance Detection (ARD) kernels (Wipf and Nagarajan, 2007), which conveniently mimic the Gaussian form with a different length-scale for each input dimension. The similarity between (vector) inputs x0 and x1 is measured by the kernel as:

k(x0, x1) = exp(-x0^T L^(-2) x1)

for diagonal length-scale matrix L. Very long length-scales characterise irrelevant input dimensions which may safely be ignored: a further step towards model order reduction. Also, importantly, this kernel enables semi-analytic calculation of Sobol' indices for the Gaussian Process surrogate. The Sobol' indices measure the proportion of output variance ascribable to each combination of input dimensions; in other words, they measure the relevance of various subspaces of the input space to the output y.

Recently ARD kernels have (Milton and Brown, 2019) been used to find an optimal rotation of the input basis, maximising the Sobol' indices successively in the subspaces spanned by the first input dimension, followed by the first two input dimensions, and so on. This greatly increases the effectiveness of model order reduction, allowing inputs to linearly combine instead of just re-ordering them.

This work applies these successive methods of model order reduction and sensitivity analysis to the simulation of thermal runaway in Li-ion batteries. The temperature-time data was obtained from an oven simulation (Bugryniec, Davidson and Brown, 2018; Bugryniec et al., 2019) of Li-ion cells at 218C with randomly sampled values (based on the literature) for heat transfer properties.

References

[1]Bugryniec, P. J. et al. (2019) ‘Pursuing safer batteries: Thermal abuse of LiFePO 4 cells’, Journal of Power Sources, 414, pp. 557–568. doi: 10.1016/j.jpowsour.2019.01.013.
[2]Bugryniec, P. J., Davidson, J. N. and Brown, S. F. (2018) ‘Assessment of thermal runaway in commercial lithium iron phosphate cells due to overheating in an oven test’, Energy Procedia. Elsevier B.V., 151, pp. 74–78. doi: 10.1016/j.egypro.2018.09.030.
[3]Marrel, A. et al. (2009) ‘Calculations of Sobol indices for the Gaussian process metamodel’, Reliability Engineering and System Safety, 94(3), pp. 742–751. doi: 10.1016/j.ress.2008.07.008.

[4]Milton, R. A. and Brown, S. F. (2019) ‘Optimal Reduced Order Modelling’, In preparation. pp. 1–8.
[5]Wipf, D. P. and Nagarajan, S. (2007) ‘A New View of Automatic Relevance Determination’, in Proceedings of the 20th International Conference on Neural Information Processing Systems, NIPS’07. Curran Associates Inc., pp. 1625–1632. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.143.8009&rep=rep1&type=pdf.
[6]Wu, Z. et al. (2016) ‘Global sensitivity analysis using a Gaussian Radial Basis Function metamodel’, Reliability Engineering and System Safety. Elsevier, 154, pp. 171–179. doi: 10.1016/j.ress.2016.06.006.

Aaron S. Yeardley
asyeardley1@sheffield.ac.uk
Department of Chemical and Biological Engineering
University of Sheffield
Sheffield, S1 3JD
United Kingdom

Robert A. Milton
r.a.milton@sheffield.ac.uk
Department of Chemical and Biological Engineering
University of Sheffield
Sheffield, S1 3JD
United Kingdom

Peter J. Bugryniec
pbugryniec1@sheffield.ac.uk
Department of Chemical and Biological Engineering
University of Sheffield
Sheffield, S1 3JD
United Kingdom

Solomon F. Brown
s.f.brown@sheffield.ac.uk
Department of Chemical and Biological Engineering
University of Sheffield
Sheffield, S1 3JD
United Kingdom

Facing High-Dimensional Simulators: Faster Kriging?

The continuous growth in computing capabilities and the advancement in data driven techniques make simulation experiments increasingly relevant for enterprises and decision makers. However, 'simulation models are often tedious to build, need substantial data for input modeling, and require significant time to run' . This might be an obstacle to exploit the simulator capabilities, especially if the simulator is used for decision-making in real time. One main strategy is to substitute the time-consuming simulator with a fast-running emulator over a region of interest (Kleijnen and Sargent 2000; Ankenman, Nelson, and Staum 2010; Rosenbaum and Staum 2017)

Kriging is among the most well-known and applied emulation methods in simulation (Kleijnen 2017; Ankenman, Nelson, and Staum 2010; Chen, Ankenman, and Nelson 2013). The merits of Kriging range from its analytical tractability to the fact that it yields not only a point estimate of the prediction, but also a corresponding error quantification (T.J. Santner, Williams, and Notz 2003). Forecast accuracy, the ability of using all available information, and speed in training and prediction are of primary interest to the analyst in practical applications. Accuracy is essential for obvious reasons. However, the ability to use all available information often generates a trade-off between accuracy and speed. In training, memory requirements associated with matrix inversion limit the size of the dataset that can be used. Similarly, the manipulation of large matrices increases the execution time in prediction, reducing the convenience in using the emulator.

This work addresses this issue, proposing a new implementation that enables the application of Kriging to simulators of larger dimensions and datasets of larger size than in current practice. The algorithm is a hybrid between the Gaussian process regression algorithms of the statistical machine learning literature and the typical algorithmic implementation of Kriging in simulation. The calculation of the Kriging mean and variance borrows from the work of Rudi, Camoriano, and Rosasco (2015) on Nyström regularization, with a modification to take into account the constraint of positive variance, which is often neglected in machine learning studies. We provide theoretical foundations, proving the analytical expressions of the estimation error for both the mean and variance estimators. The computational analysis shows that the algorithm requires a total cost of order O(nm(d+m)) in time, O(m^2 + n) in space, where d is the number of inputs, n the sample size, and m< n the size of the Nyström centers. This promises to be a notable saving compared to universal Kriging algorithms, whose requirements are essentially of O(n^3) in time and O(n^2) in space.

The work then reports results of extensive numerical experiments to compare the performance of the proposed algorithm against current Kriging implementations (see the recent analysis in (Erickson, Ankenman, and Sanchez 2018) on simulators of increasing dimensionality. Findings show notable savings in time and memory requirements that allow one to breach the 10,000 simulator-input barriers. Moreover, we challenge the fast Kriging emulator in the estimation of complex functionals of the simulator output distribution, global sensitivity measures, that require several thousands of prediction runs, serving as a test of the time advantage promised by the theoretical analysis of the algorithm.

References

Ankenman, Bruce, Barry L. Nelson, and Jeremy Staum. 2010. "Stochastic Kriging for Simulation Metamodeling." Operations Research Publication 58 (2):371–82.

Xuefei Lu
xuefei.lu@polimi.it
Politecnico di Milano
Bocconi University
Milan, Italy

Alessandro Rudi
alessandro.rudi@gmail.com
INRIA - École Normale Supérieure
Paris, France

Emanuele Borgonovo
emanuele.borgonovo@unibocconi.it
Bocconi University
Milan, Italy

Lorenzo Andrea Rosasco
lrosasco@mit.edu
DIBRIS, Università degli Studi di Genova, Genova, Italy
LCSL, Massachusetts Institute of Technology & Istituto Italiano di Tecnologia,
Cambridge, USA

Session 2: Meta-modelling in sensitivity analysis II

Bayesian model averaging of sparse polynomial chaos expansions

We consider a model response y=f(x) function of d independent random variables x = . Polynomial chaos expansion (PCE) is a powerful method for high-dimensional representation of input- output relationship when model inputs are independent of each other [1, 2]. The Bayesian sparse PCE of [3] proved itself to be very efficient to perform this task. The latter relies on a sophisticated algorithm that finds the best sparsest polynomial chaos expansion possible. Let us denote by M_n(x) a given PCE representation.

By best possible it is meant that a trade-off is to find between the number of terms in the expansion and the variance error (both to be as small as possible). The performance of a given PCE candidate is measured through its associated likelihood, denoted p(yIM_n), also called model evidence. Indeed, evidence is the key quantity to compare competing models.

In [3], the best PCE is identified by stepwise regression during which the current subset is enriched with a new element. Then, the evidence associated with the enriched PCE is computed and compared to the previous one. If the new evidence is better than the previous one, then the enriched subset is kept, otherwise, the previous subset is selected. Notably, the enrichment process obeys to an algorithm that circumvents the curse of dimensionality. The iteration automatically stops when no more relevant terms are added to the subset.

Model evidence in the procédure of Shao et al. [3] is approximated by exp(-0.5*KIC_n) where KIC_n is the Kashyap information criterion associted with M_n. While the cited authors investigate the best sparse PCE representation (the one with highest evidence), in the present work, a collection of competing Bayesian sparse PC representations (of various cardi- nality) are investigated. This allows accounting for metamodel errors in the calculation of sensitivity indices. A comparison between the Sobol' indices obtained with the original approach of [3] and those computed with the Bayesian model averaging approach are undertaken onto different benchmark functions.

References

[1] B. Sudret. Global sensitivity analysis using polynomial chaos expansions. Reliability Engineering and System Safety, 93:964–979, 2008.
[2] G. Blatman and B. Sudret. Adaptive sparse polynomial chaos expansion based on least angle regression. Journal of Computational Physics, 230(6):2345–2367, 2011.
[3] Q. Shao, A. Younes, M. Fahs, and T. A. Mara. Bayesian sparse polynomial chaos expansion for global sensitivity analysis. Computer Methods in Applied Mechanics & Engineering, 318:474–496, 2017.
[4] R. L. Kashyap. Optimal choice of AR and MA parts in autoregressive moving average models. IEEE Trans. Pattern Anal. Machine Intell., 4(2):99–104, 1982.

Thierry Mara
mara@univ-reunion.fr
University of Reunion

Variogram-based Sensitivity Analysis: Bridging Derivative-based and Variance-based Approaches?

Sensitivity Analysis, as a science, has evolved significantly over the past few decades. There now exists a range of approaches and methods for SA, based on different theories and philosophical characterizations of sensitivity of system/model responses to controlling factors. The "derivative-based" and "variance-based" approaches are two general, perhaps the most commonly used approaches to SA. The former globally aggregates the measures of local sensitivity (i.e., partial derivatives) across the factor space, while the latter decomposes the total variance of a model output and attributes its parts to different factors that are allowed to vary in the factor space. The theoretical differences between different approaches often make them result in different, sometimes conflicting sensitivity assessments, when applied to the same problem.

This presentation introduces a recently developed "variogram-based" theory for SA. This theory, called Variogram Analysis of Response Surfaces (VARS), utilizes the concepts of anisotropic variograms and covariograms to characterize the sensitivity of model response to controlling factors across the full range of factor "perturbation scales". Therefore, VARS directly recognizes that there is a spatially-ordered variance and covariance structure in the response surface, and hence in its derivatives, and uses such information to generate a comprehensive set of indices for global sensitivity. VARS may be viewed as a "unifying theory", as it is theoretically shown that the derivative-based and variance-based approaches are special cases of VARS, and their sensitivity indices (e.g., variance-based total-order effects and elementary effects) can be estimated by VARS as well. It is further shown that VARS is highly computationally efficient and statistically robust because, in parts, it is based on pairs of points in the factor space rather than individual points.

References

[1] Razavi, S., and H. V. Gupta, (2015), What do we mean by sensitivity analysis? The need for comprehensive characterization of ''global'' sensitivity in Earth and Environmental systems models, Water Resources Research, 51, 3070–3092, doi:10.1002/2014WR016527.
[2] Razavi, S., and Gupta, H. V., (2016), A new framework for comprehensive, robust, and efficient global sensitivity analysis: I. Theory, Water Resources Research, 51, doi:10.1002/2015WR017558.
[3] Razavi, S., and Gupta, H. V., (2016), A new framework for comprehensive, robust, and efficient global sensitivity analysis: II. Application, Water Resources Research, 51, doi:10.1002/2015WR017559.

Saman Razavi
saman.razavi@usask.ca
University of Saskatchewan

Hoshin Gupta
saman.razavi@usask.ca
University of Saskatchewan

Razi Sheikholeslami
razi.sheikholeslami@usask.ca
University of Saskatchewan

Nhu Cuong Do

nhu.do@usask.ca
Postdoctoral Fellow
Global Institute for Water Security
University of Saskatchewan
Saskatchewan, Canada

Amin Haghnegahdar

Sensitivity analysis with generalized chaos expansions

Thanks to the functional ANOVA decomposition, the Sobol' indices give, for a square integrable non-linear model and independent input variables, the parts of the output variance due to each input and to each interaction between inputs ([7], [2]). In addition, the total Sobol' index provides the overall contribution of each input [3], including interactions with other inputs. More generally, the interaction magnitudes can be measured by the corresponding indices in the functional ANOVA decomposition.

Many methods exist to efficiently estimate the first-order Sobol' indices. One of the most popular and powerful one is polynomial chaos (PC) expansion. It consists in approximating the response onto the specific basis made by the orthonormal polynomials built on the input distributions. Indeed, its strength stands on the fact that, once the expansion computed, it easily gives all the ANOVA indices (in particular the total Sobol' indices) [8]. Notice that, as the PC expansion must be truncated in practice, the estimated ANOVA indices provide (up to estimation error) lower bounds of the true ones.

In general, the estimation of the total Sobol' indices (and other ANOVA indices) suffers from the curse of dimensionality (number of inputs) and can be too costly in terms of the number of necessary model evaluations [5]. Low-cost computations of upper and lower bounds for total Sobol' indices are then very useful. DGSM (Derivative-based Global Sensitivity Measures), computed from the derivatives of the model output, give such economical upper and lower bounds (see [4] for a review). For upper bounds, general optimal results are obtained in [6]. For lower bounds, only special cases (uniform, Normal and Gamma) have been investigated in [9, 4]. The bounds given in [4] are quite rough as they are smaller than the first-order Sobol indices. In our work, we follow the tracks opened by [9] using PC expansions, but for both more general distributions and expansions. Recall that it is desirable to consider a wide range of input distributions. Indeed, transporting input distributions to uniform ones induces undesirable properties in the modified model.

The aim of our work is to investigate extensions of PC expansions for sensitivity analysis, with a focus on DGSM. Hence, we consider expansions obtained by tensorized Hilbert basis of functions, that we call generalized chaos expansion. We first prove that, as for PC expansions [8], any ANOVA index can be obtained by simple algebraic operations on the decomposition coefficients. Lower bounds are obtained by truncation of the expansion. Then, we consider the special case of a Hilbert basis given by the eigenfunctions of an elliptic differential operator (DO) associated to the input distribution. This DO is naturally associated to a Poincaré inequality [1]. By analogy to PC expansions, the corresponding Hilbert basis will be called DO expansions. We show that DO expansion can be rewritten with derivatives, and thus provides accurate lower bounds for DGSM.

There is a closed connection between DO and PC expansion. When using specific weights in Poincaré inequalities, DO expansions correspond to Legendre polynomials used in PC expansions for the uniform distribution, and to Hermite, Laguerre and Jacobi polynomials, used for the Normal, Gamma and Beta distributions [1]. However, except for the Normal distribution, the DO expansions associated to the usual unweighted Poincaré inequality give other expansions, involving non-polynomial functions. In particular, for the uniform distribution, it coincides with the Fourier expansion.

There are several advantages for using DO expansions. Firstly, as shown above, it can be viewed as an extension of PC expansions, which is already a powerful technique for sensitivity

analysis. Secondly, computations can be done for a large range of admissible probability distributions, e.g. those supported by a compact interval and satisfying a Poincaré inequality. Thirdly, DO expansions can be immediately used to approximate DGSM. In particular, the computations involve the same numerical method used to compute DGSM upper bounds, which reduce the global computation cost for computing (upper/lower) bounds on DGSM.

References

[1] D. Bakry, I. Gentil, and M. Ledoux. Analysis and geometry of Markov diffusion operators, volume 348 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer, Cham, 2014.
[2] W. Hoeffding. A class of statistics with asymptotically normal distribution. Ann. Math. Statist., 19(3):293–325, 09 1948.
[3] T. Homma and A. Saltelli. Importance measures in global sensitivity analysis of non linear models. Reliability Engineering and System Safety, 52:1–17, 1996.
[4] S. Kucherenko and B. Iooss. Derivative-based global sensitivity measures. In R. Ghanem, D. Higdon, and H. Owhadi, editors, Springer Handbook on Uncertainty Quantification, pages 1241–1263. Springer, 2017.
[5] C. Prieur and S. Tarantola. Variance-based sensitivity analysis: Theory and estimation algorithms. In R. Ghanem, D. Higdon, and H. Owhadi, editors, Springer Handbook on Uncertainty Quantification, pages 1217–1239. Springer, 2017.
[6] O. Roustant, F. Barthe, and B. Iooss. Poincaré inequalities on intervals - application to sensitivity analysis. Electron. J. Statist., 11(2):3081–3119, 2017.
[7] I. Sobol. Sensitivity estimates for non linear mathematical models. Mathematical Modelling and Computational Experiments, 1:407–414, 1993.
[8] B. Sudret. Global sensitivity analysis using polynomial chaos expansion. Reliability Engineering and System Safety, 93:964–979, 2008.
[9] B. Sudret and C. V. Mai. Computing derivative-based global sensitivity measures using polynomial chaos expansions. Reliability Engineering & System Safety, 134:241–250, 2015.

Olivier Roustant
roustant@emse.fr
Mines Saint-Etienne,
France

Fabrice Gamboa
fabrice.gamboa@math.univ-toulouse.fr
Institut de Mathématiques de Toulouse,
France

Bertrand Iooss
bertrand.iooss@edf.fr
Electricité de France R&D,
France

Poster Session 1: Meta-modelling in sensitivity analysis and applications to engineering/reliability issues

Cross-validation based adaptive sampling for Gaussian process models

In many real-world applications, we are interested in approximating black-box, costly functions as accurately as possible with the smallest number of function evaluations. A complex computer code is an example of such a function. In this work, Gaussian processes (GPs) are used to approximate the output of complex computer code. This surrogate model is known as an emulator. We consider the problem of extending an initial experiment (set of model runs) sequentially to improve the emulator.

A sequential sampling approach based on leave-one-out (LOO) cross-validation is proposed that can be easily extended to a batch mode, where at each iteration a set of input points is selected for evaluation. This is a desirable property since it saves the user time when parallel computing is available. After fitting a GP to training data points, the mean squared LOO error is calculated at each design point. The mean squared LOO error is used as a measure to identify important data points. More precisely, when this quantity is large at a point it means that the quality of prediction depends a great deal on that point and adding more samples in the nearby region could improve the accuracy of the GP model. As a result, it is reasonable to select the next query sample where the mean squared LOO error is maximum. However, the errors are only available at the data points and need to be estimated at unobserved points. To do this, a second GP is fitted to the mean squared LOO errors and the point where the maximum of the modified expected improvement (EI) criterion occurs is chosen as the next query location.

EI is a popular acquisition function in Bayesian optimisation and is used to trade-off between local and global search. However, it has tendency towards exploitation, meaning that its maximum is close to the sample with maximum value. To avoid clustering, a modified version of EI, called pseudo expected improvement (PEI), is employed which is more explorative than EI and allows to discover unexplored regions. PEI is obtained by multiplying EI by an influence function which is a measure of distance between the training data based on the metric of the (second) GP covariance kernel. The results show that the proposed sampling method is promising.

Hossein Mohammadi
h.mohammadi@exeter.ac.uk
University of Exeter,
United Kingdom

Peter Challenor
p.g.challenor@exeter.ac.uk
University of Exeter,
United Kingdom

Application of global sensitivity analysis in the optimization design of aeronautical hydraulic pipelines

This work investigates the design of constraint hoops in the aeronautical hydraulic pipeline system. The non-probabilistic sensitivity analysis is used to screen out the hoops which are both insensitive to the maximum stress response, the maximum displacement response as well as the first order natural frequency. The analysis result can give the guidance to reduce the size and the weight of the pipeline system. Based on the pretreatment analysis, the position coordinates of the remaining constraint hoops are further optimized. The comparison before and after optimization reveals that the dynamic performances of the pipeline system are improved significantly. This study indicates that the proposed method can provide an effective solution to the design of the aeronautical hydraulic pipeline system.

In aeronautical hydraulic systems, the constraint hoops are widely used for fixing and supporting the pipeline system to reduce the impact of the terrible working environment. However in engineering practice, the parameters and locations of the constraint hoops are mostly uncertain. These uncertainties can easily cause local stress concentration which becomes the failure source of the pipeline structure and weakens the reliability of the aeronautical hydraulic system seriously. And apart from this, the positions of the constraint hoops of the hydraulic pipeline system are mostly determined based on the designer's experiences. Hence it is significant to propose an optimization design method of the constraint hoops. In this study, non-probabilistic sensitivity analysis is used as the key part of proposed method.

Sensitivity analysis is a method for analyzing how the uncertainty of model input affects the uncertainty of the output response. However, in engineering practice, it is a great challenge to get the sufficient experimental data. Therefore, the probabilistic model is no longer applicable and the non-probabilistic model such as interval model is an ideal choice. The non-probabilistic method is introduced while only the upper and lower bounds of the uncertainties are easy to get other than the precise random distribution information. For the pipeline system studied in this work, the interval model is used to describe the uncertainty of the spring stiffness and the non-probabilistic global sensitivity indices are introduced to measure the contribution of input variables to the dynamic response of this pipeline system. According to the result of the analysis, the constraint hoops which have no or little effect to the output response are screened out and can be neglected in engineering design. After the screening, the positions of the remaining constraint hoops need to be optimized in order to further improve the dynamic performances of the hydraulic pipeline system.

The finite element model of the complex hydraulic pipeline system which will be optimized in this study is built in ANSYS at first. In this model, there are more than twenty hoops which can be simplified as the spring elements to constrain the radial displacement of the pipeline system. And those spring stiffness of the constraint hoops are considered as the input variables. Basing on the non-probabilistic global sensitivity index, the hoops which have none or less effects on three dynamic responses are screened out and can be neglected in order to reduce the weight in the design. According to the above non-probabilistic global sensitivity analysis, there are three constraint hoop stiffness having little effect on the three responses. The exact positions of remaining hoops need to be determined by adapting the reliability optimization method. Considering the dynamic reliability based on the First-passage theory, the optimization model of this hydraulic pipeline system is established. The objective function is the maximum standard deviation of stress response of the pipeline system, the design variables are the coordinates of the constraint hoops position and the three constraint

functions are the maximum standard deviation of displacement response constraint of the system, the dynamic reliability constraint of the system and the first order natural frequency constraint of the system, respectively. In this study, the optimization method Sequential Quadratic Programming (SQP) is employed to solve the optimization problem. The optimization results show that within about 500 times of the objective function evaluations, the maximum standard deviation of stress response is reduced by about 44%. And apart from this, the three constraint conditions which include the maximum standard deviation of displacement response, the dynamic reliability and the first order natural frequency are satisfied. This indicates that the proposed optimization design method not only reduces weight but also improves performance of the pipeline system. The proposed method provides a reasonable suggestion for the design of the hydraulic pipeline system.

Changcong Zhou
changcongzhou@nwpu.edu.cn
Northwestern Polytechnical University

Zheng Zhang
zhangzheng@mail.nwpu.edu.cn
Northwestern Polytechnical University

Haodong Zhao
hdzhao0816@163.com
Northwestern Polytechnical University

Mengyao Ji
jimengyao@mail.nwpu.edu.cn
Northwestern Polytechnical University

Multi-fidelity Radial Basis Function Metamodel for Global Sensitivity Analysis

Global sensitivity analysis aims at quantifying the relative importance of input variables in determining the value of an assigned model output. Over the last several decades, GSA has received much attention and many different GSA methods have been well developed. Among these approaches, Sobol index has been widely used in the practice application since it can provide some valuable information of model and be realized easily. The Sobol index is often computed by Monte Carlo technique or Quasi Monte Carlo. However, the high computational cost of high dimensional integrals makes it practically unsuitable for most computationally intensive model analysis.

To address this issue and reduce the computational burden, a multi-fidelity metamodel estimator based on Gaussian radial basis function (RBF) is proposed. Since the accuracy of metamodel is to some extent determined by computer experiment, the optimal Latin hypercube design is employed to generate a limited number of design points, which are used to build the Gaussian RBF high and low fidelity metamodels. To allow a high accuracy prediction to the physical model response, the multi-fidelity RBF metamodel is constructed by linear regression mapping of low fidelity RBF metamodel to the high fidelity one. Subsequently, an analytical expression is derived from the multi-fidelity RBF metamodel to estimate Sobol index. Several numerical experiments are performed to compare the performance of the proposed method with single fidelity RBF method, and the results demonstrate that the proposed method is an efficient approach.

Xiaobing Shang
shangxiaobing1992@126.com
Control and Simulation Center
Harbin Institute of Technology
Harbin, China

Application of Sensitivity Analysis Methods for Investigation on Pneumatic End-Position Cushioning

Pneumatic drives find a widespread application in the modern industry, not only as a cheap and reliable alternative to electric drives, but also as a front runner in some special fields. Pneumatic drives are essential in food, medical and diverse explosion-hazardous technologies, where they have gained a reputation of hygienic, safe and reliable systems.

Within the last decades pneumatic technology has been progressing under a motto ‘‘strengthen the strengths and weaken the weaknesses’’. The first part of the expression relates to a further increase in system flexibility, reliability and ease of use of pneumatic applications, which all are mainly brought about by integrated electronic control systems. This have made pneumatic drives smarter, safer and simpler in both operation and design. However, despite of the technological progress in the field of electronics and communication technologies, time-costly manual works are still needed when setting up or adjusting some basic pneumatic components, as throttles or damping devises. This lengthens the duration of start-up and sometimes of regular maintenance procedures.

Especially pneumatic end-position cushioning systems, used in most pneumatic cylinders to absorb an excess of kinetic energy of the moving mass by the end of the stroke, are very sensitive to changes in operational conditions, setup and environment. Whenever one of those parameters has been changed, manual readjustment is needed. The latter has usually a trial-and-error nature. In parallel, a maximum quantity of energy, absorbed by pneumatic end-cushioning, often appears as a limitation factor when sizing the pneumatic cylinder. An engineer would prefer rather a large-scaled cylinder with a proportionally higher damping capability to bring the moving mass softly and safely to the end position. Large-scaling of pneumatic drives, however, contradicts the second part of the development trend: to weaken the weaknesses. High energy consumption is the most aching weakness of pneumatics, and an increase in cylinder diameter results directly in an increase in energy consumption whilst performing the same net work. Summing up, the pneumatic end-cushioning needs to be redesigned to achieve a lower sensitivity against the changes from the outside, thus increasing the robustness of the drive and replicability of the technological processes, as well as to enhance energy absorption without enlarging the system or pushing up the costs.

A common tool used to solve similar engineering tasks is the optimization of parameters having a major influence on the system’s behavior. In most problems in the field of hydraulics and pneumatics these parameters are already well-known and visible. The number of optimized variables seldom exceeds 6…8 degrees of freedom. However, a damping process in pneumatic end-cushioning involves simultaneously different physical domains, which are fluid mechanics, heat transfer and tribology. And there is a great number of factors that may inflict the damping behavior, making optimization difficult.

Thus, to investigate the strategies for performance improvement of pneumatic end-cushioning, sensitivity analysis is used within this study. The prior goal is to find out the design factors with a major influence on the ability of the end-cushioning to absorb kinetic energy. These high-importance factors will be further used for robust optimization, to reduce the above-mentioned sensitivity of the system against the operational and environmental factors. In this paper methods and settings for sensitivity analysis and their application for a model of pneumatic end-cushioning systems are discussed.

In the first step, a lumped-parameters simulation model of pneumatic end-position cushioning is built in ESI ITI SimulationX software. The model inputs are represented by factors that can be affected in the design process. They correspond to geometrical dimensions and material properties of cylinder parts. These inputs are varied between their physically reasonable limits. As model outputs, whose sensitivity against the inputs should be determined, energy losses during the damping process, peak deceleration and damping power are considered. Further, the model is experimentally validated to provide adequate data for a friction model and flow characteristic of the damping throttle.

In the next step sensitivity analysis is performed. Due to nonlinearity and nonmonotonicity of the model and large variations of the input parameters, only global sensitivity methods are considered to be suitable. The sensitivity analysis is performed in the R programming environment by means of Sensitivity Package. Matlab is used for data exchange between the simulation software and R. Ten variance-based methods are applied for analysis of the model with different number of inputs (8 and 20 variables). Their results for various sampling strategies and sample sizes are compared for three model outputs. As a result of the study following questions are answered:
* which methods and sampling strategies provide the lowest dispersion and standard error of total and first-order sensitivity indices for the same simulation costs (number of model runs) for a model with 8 inputs;
* do these methods and sampling strategies perform better for the model with 20 inputs as well;
* how different the dispersions of all three outputs can be by all other conditions being equal;
* can the needed sample size for a many-inputs model be predicted by evaluation of the data gained for a few-inputs model, in order to meet the desired dispersion and standard error of the sensitivity indices for the lowest simulation costs;
* how do the results of sensitivity analysis depend on the applied method and sampling strategy.

This study allows not only to reach the prior goal, i.e. to find out the input factors with the highest influence on the outputs, but also to collect the information about how the sensitivity analysis should be performed in the engineering tasks similar to the presented one in terms of number of variables, and system properties as nonlinearity and nonmonotonicity. This information may contribute to increase the incidence of use of sensitivity analysis as a powerful and efficient tool in industrial engineering in general.

Fedor Nazarov
fedor.nazarov@tu-dresden.de
Chair of Fluid-Mechatronic Systems (Fluidtronics)
TU Dresden, Germany

Ralf Tautenhahn
ralf.tautenhahn@tu-dresden.de
Chair of Fluid-Mechatronic Systems (Fluidtronics)
TU Dresden, Germany

Jürgen Weber
fluidtronik@mailbox.tu-dresden.de
Chair of Fluid-Mechatronic Systems (Fluidtronics)
TU Dresden, Germany

Uncertainty quantification and Sensitivity Analysis in Composite Materials using Sampling based Non-Intrusive Approaches

Composite materials have been increasingly used in aerospace, construction, sports, and transportation industries due to their superior properties. As the process of manufacturing and testing of these composite materials is time-consuming and sometimes unfeasible, modeling and simulation approaches are more suitable and imperative. These materials possess hierarchical structures which require the use of multiscale modeling. This implies the need for the development and integration of models for the reliable prediction of materials at different scales.

Uncertainties in the microstructure of composite materials are inherent to the processes used to manufacture these materials. The effect of these uncertainties may cause uncertainties in the evolution of damage processes. To analyze the effect of these uncertainties in composite materials, the simulation process must be integrated with uncertainty quantification and uncertainty propagation (UP) methods. However, due to the complexities in modeling processes and significant computational costs in uncertainty analysis, very limited advancement has been seen in this direction.

The work proposed in this paper takes a few steps in the direction of an improved material modeling which includes uncertainty and sensitivity analysis in order to produce a more realistic representation of composite materials. In this work, several non-intrusive (sampling-based) approaches and sampling techniques are proposed for multiscale UQ and UP to address these limitations. We apply these approaches to quantify the effect of uncertainties in a composite leaf spring test case where Epoxy resin 914C and Carbon Fiber T300 are the base material. For the structural analysis of the deterministic test case, the commercial solver ABAQUS is used. The uncertainties are introduced in the volume fraction and physical properties of the base materials. This work is conducted in the framework of the European Project COMPOSELECTOR (Multi-Scale Composite Material Selection Platform with a Seamless Integration of Materials Models and Multidisciplinary Design Framework) H2020. The ultimate aim of this project is to develop a Business Decision Support System (BDSS), which integrates materials modeling, business tools, and databases into a single workflow to support the complex decision process involved in the selection and design of composite materials.

Acknowledgment: The authors would like to acknowledge the support of EU H2020 COMPOSELECTOR project (GA N°: 721105).

References

[1] Belouettar S. et al., Integration of material and process modelling in a business decision support system: Case of COMPOSELECTOR H2020 project, Composite Structures, Vol. 204, Pages 778-790, 2018.
[2] Ghanem R. et al., Handbook of Uncertainty Quantification, Springer, ISBN 978-3-319-12384-4
[3] Sobol I. M., On sensitivity estimation for nonlinear mathematical models, Matem. Mod, Vol 2, Issue 1, Pages 112—118, 1990.
[4] Sudret B., Global sensitivity analysis using polynomial chaos expansions, Reliability Engineering & System Safety, Vol. 93, Issue 7, Pages 964-979, 2008.

Dinesh Kumar
dinesh.kumar@list.lu
Materials Research & Technology Department, Luxembourg Institute of Science & Technology
Luxembourg

Yao Kouts
Yao.koutsawa@list.lu
Materials Research & Technology Department, Luxembourg Institute of Science & Technology
Luxembourg

Ahmed Makradi
Ahmed.Makradi@list.lu
Materials Research & Technology Department, Luxembourg Institute of Science & Technology
Luxembourg

Carlos Kavka
kavka@esteco.com
ESTECO SpA
Trieste, Italy

Salim Belouettar
Salim.Belouettar@list.lu
Materials Research & Technology Department, Luxembourg Institute of Science & Technology
Luxembourg

Global Sensitivity Analysis of ballistic free-flight experiments: does the model preserve the behaviour of the system?

Problem statement

Knowledge of the projectile's in-flight behaviour enables its development as a munition. Necessary tasks—such as: autopilot design, navigation algorithms implementation, or trajectory prediction—can be undertaken only if the behaviour of a projectile is correctly characterised and known. Following the importance of the characterisation task, the aerospace and ballistic research community has developed several approaches to successfully identify the projectile aerodynamics. Principally, they can be divided into wind tunnel tests, computational fluid dynamics simulations, and free-flight experiments. The last can be considered as the most realistic method, since the projectile's behaviour is observed—and then used for a data-based identification procedure—under real flight conditions at a ballistic test range.

However, realistic test conditions introduce several sources of uncertainty to the mathematical model of a projectile, which is a nonlinear structure with time-varying output signals. Principally, the initial flight conditions are unknown, meaning that the initial state vector may be considered only as a vector of nominal values with uncertainty around them. Similarly, the influence of the aerodynamics of the projectile is known only approximately, based on wind tunnel or CFD analysis, and remains to be verified during free-flight experiments.

In effect, free-flight experiments can be considered as a setting for a global sensitivity analysis (GSA), with unknown test conditions corresponding to the sources of uncertainty. The benefit of conducting GSA in this case lies in the ability to test if the model behaves in the physical manner, as well as to visualise the influence of uncertain parameters on the time-varying measurements, such as accelerometers and magnetometers. Additionally, according to authors' best knowledge, the utilisation of GSA in the field of ballistics is a novel application, since currently only the local sensitivity analysis has been applied in the domain.

Approach

Our approach takes advantage of two premises in order to efficiently calculate time-varying Sobol' [1] indices:

* In order to significantly reduce the computation time required for each simulation, the analysis is performed on a specially developed model structure: a quasi-LPV (linear parameter-varying) one, obtained by a reference-frame transformation, which reduces the numerical complexity of solving the ordinary differential equation, while being analytically equivalent to the original, nonlinear model [2].
* In order to reduce the amount of simulations necessary, a permutation-based approach of Mara [3] is applied: the model is simulated first for a sampling matrix X1; then for a sampling matrix X2, where X2 is a permuted matrix X1.

Analysis is performed for a quasi-LPV model of a projectile in flight that consists of 12 state equations and 9 output signals. The output signals correspond to three-dimensional accelerometers, magnetometers, and gyroscopes. Uncertain inputs for GSA consist of 8 parameters: the unknown initial translational velocity, the unknown initial pitch and yaw rates, the unknown mass, the unknown initial pitch and yaw angles, as well as unknown normal force coefficient and unknown axial force coefficient. For the unknown parameters, uniform distribution is assumed, and their nominal values are selected in such a way, that they

correspond to testing conditions at the ballistic testing ground in Baldesheim, France [4]. The total sample size is N = 40 000.

Results

Results of the GSA confirm the predictions of the domain knowledge of ballistics and aerodynamics:

* The initial Euler angles influence only the magnetometer measurements, and only for the first few seconds of a flight.
* The effect of axial force coefficient uncertainty is visible especially in the axial accelerometer measurements.
* The effect of normal force coefficient uncertainty is visible in the remaining two accelerometer measurements. Additionally, its effect on magnetometer measurements is visible, and it increases over time, while the influence of initial Euler angles uncertainty diminishes.
* Small mass uncertainty does not influence the sensor measurements.

The observed effects are especially interesting in the context of identification of unknown parameters: knowing that non-influential parameters are non-identifiable [5], one can easily confirm that some of the model parameters might be identifiable using only some of the sensors, and only in some periods of time. For example, any attempt to identify initial Euler angles using magnetometer measurements after a few seconds of flight would be futile. Moreover, the conformity of the results with the domain knowledge suggests that the quasi-LPV model has been correctly developed and effectively preserves the real behaviour of a projectile.

Conclusions

GSA can be considered an effective tool to visualise physical knowledge in ballistic systems and to detect potentially identifiable parameters. In the context of the present work, it can be also concluded that the model structure used in analysis correctly represents the physical behaviour of a studied system: a projectile in flight.

Acknowledgements

This work is in part financed by the French Defence Procurement Agency (DGA, Direction Générale de l'Armement) of the French Ministry of the Armed Forces.

References

[1] I. M. Sobol, ‘‘Sensitivity estimates for nonlinear mathematical models,’’ Math. Model. Comput. Exp., vol. 1, no. 4, pp. 407–414, 1993.
[2] D. Machala, S. Dobre, M. Albisser, M. Gilson, and F. Collin, ‘‘The Analysis of Vehicle’s In-Flight Behaviour Using Quasi-LPV and Nonlinear Models,’’ in AIAA Scitech 2019 Forum, 2019.
[3] T. Mara and O. Rakoto, ‘‘Comparison of some efficient methods to evaluate the main effect of computer model factors,’’ J. Stat. Comput. Simul., vol. 78, no. 2, pp. 167–178, 2008.
[4] M. Albisser and S. Dobre, ‘‘Sensitivity Analysis for Global Parameter Identification. Application to Aerodynamic Coefficients,’’ IFAC-PapersOnLine, vol. 51, no. 15, pp. 963–968, 2018.
[5] S. Dobre, T. Bastogne, C. Profeta, M. Barberi-Heyob, and A. Richard, ‘‘Limits of variance-based sensitivity analysis for non-identifiability testing in high dimensional dynamic models,’’ Automatica, vol. 48, no. 11, pp. 2740–2749, 2012.

Dawid Machala
machala.dawid@gmail.com
French-German Research Institute of Saint-Louis

Collin Floriane
floriane.collin@univ-lorraine.fr
Université de Lorraine
Nancy, France

Marie Albisser
marie.albisser@isl.eu
French-German Research Institute of Saint-Louis

Simona Dobre
simona.dobre@isl.eu
French-German Research Institute of Saint-Louis

Marion Gilson
marion.gilson@univ-lorraine.fr
Université de Lorraine
Nancy, France

Contrast based global sensitivity analysis of failure probability

The failure probability of a structural steel member is investigated using a global sensitivity analysis subordinated to contrast. The study deals with the failure of a load-bearing steel bridge beam under bending. The failure is caused by brittle fracture caused by many repeated stresses. The computational model of fatigue crack progression is based on linear fracture mechanics. The main objective of probability-oriented sensitivity analysis is structural reliability.

The fundamental question in terms of structural reliability is how significant is the effect of input random variables on the on the ultimate limit state of load-bearing steel beam. The sensitivity indices are estimated using double-nested-loop simulation of the Latin Hypercube Sampling method. The equivalent stress range is identified as the dominant variable most influenced the failure probability.

The second dominant variable is the initial edge crack. Both variables exhibit strong interaction effects, mutually and also with the other variables. New findings indicate that interaction effects among input variables strongly influence the failure probability.

The results are important for the lifetime reliability analysis of steel bridges having reached or exceeded expected life. Fatigue is one of the more critical forms of damage in steel bridges. Prediction and accurate assessment of the state of fatigue damage and the remaining fatigue life of steel bridges remains a challenging and unresolved task.

Global sensitivity analysis, which is capable of identifying the most critical and essential contributors of input variables to uncertainty of failure probability, is one of valuable techniques in solving the problem of structural reliability. The results of global sensitivity analysis present a significant contribution to the extension of knowledge in the field of the reliability theory of building structures.

Zdenek Kala
kala.z@fce.vutbr.cz
Faculty of Civil Engineering
Brno University of Technology

Variational Inference for sparse recovery of Polynomial Chaos expansions

We present a novel approach for the construction of Polynomial Chaos expansions based on hierarchical Bayesian models, The approach attempts to estimate posterior densities of the coefficients of Polynonial Chaos expansions using Variational Inference , that is by approximating the posterior distribution of the parameters of interest via minimization of the relative entropy between the true posterior and its approximation.

Additionally, motivated by the Dictionary Learning (DL) literature, the hierarchical Bayesian model under consideration enforces sparsity in the chaos series expansion with the presense of a Bernoull prior on the multicative constants on the coefficients, whose success probabilities are jointly inferred within the VI framework. The efficiency of the approach is demonstrated through several numerical examples.

Panagiotis Tsilifis
tsilifis@protonmail.com
Institute of Mathematics
École Polytechnique Fédérale de Lausanne

A Reliability Analysis with an Active-learning Metamodel for the Reconstruction of a Dissected Aorta Cross-section

Aortic Dissection is a disease that affects the aorta. It develops with the formation of a secondary volume in which blood is collected. This condition is often undiagnosed because of its rapid development and degeneration. Patients with AD often die between one day and one week. In mathematical and computational models of this disease, the level of uncertainty is extremely high. One of the leading causes of this uncertainty is the lack of useful datasets and experiments that allow a better understanding of the genetic and development of AD.

Uncertainties reside mainly in the model's input parameters. They result from errors in the measurement, or precisely, from a lack of information necessary for the development of the model. Therefore, assumptions and hypotheses come to our aid, but they also bring with them a price to pay: uncertainty.

The primary task of Uncertainty Quantification (UQ) is allocating uncertainty, and understanding how and which one is meant to be reduced. This tool is also able to modify models toward a patient-specific approach, that is, the ability to extract the right information from a few selected inputs that will be provided to the model.

This project aims to use the structural reliability analysis technique for the reconstruction of a dissected aorta cross-section. The objective of structural reliability analysis is to determine the probability of failure of a system concerning a performance criterion. This criterion is influenced by the uncertainty in the input parameters to the model. Given a function g(x) that represents the model, the limit state surface can be determined as g(x) = 0, which divides the output domain between success and failure response with respect to a limit-state parameter. The probability of failure is therefore defined as the integral of the joint input PDF with respect to the input parameters, on the input domain that produces a failure status.

Some of the techniques most used to determine this integral are Monte Carlo Simulation, which, however, require a large number of simulations to obtain adequate accuracy; FORM and SORM methods based on the approximation of the model near the limit-state function; or the use of surrogate models such as Polynomial-Chaos Kriging (PCK).

The goal is to use reliability analysis to identify the points to be added to the best-fit ellipse of the dissected cross-section, which will recreate a reasonable healthy aorta cross-section. In this project, the existing algorithm developed by Echard et al. in 2011 is revisited. It can actively enrich a reduced set of model evaluations in the vicinity of the limit-state function. Furthermore, the use of Kriging methods will provide information about the variance of approximation of the model, which could be used as a form of local error. The purpose of the enrichment algorithm of the initial experimental design (ED) is to provide, together with the PCK metamodel, information on the reliability of this reconstruction by using the statistic of significant interest.

The adaptive algorithm is structured as follows. Initially, a random sampling technique generates a small ED of the input parameter, and the corresponding response is computed. A metamodel is then calibrated from this dataset. A large sampling is created, and the responses are calculated from the metamodel. The limit state surface is estimated based on the metamodel. An enrichment criterion of the initial ED based on a misclassification function will define the best candidate in the input space that has the lowest value of misclassification function; its response is also added to the initial ED. The algorithm starts again but with a new

initial dataset from which the PCK is rebuilt. The stopping criterion terminates the algorithm based on the convergence of the statistic of interest of the metamodel.

The reference cross-section of the aorta is collected from a patient who was diagnosed with AD. The assumption is that the aorta best-fit geometry is an ellipse. The area and the tilt of the reference ellipse are then used as geometric criterions to set the reliability analysis, so to locate the safe and failure reconstruction responses. Physiological properties of the aorta in its longitudinal direction fix both geometric constraints. A set of points initially define the dissected aorta cross-section, which represents its segmentation. By adding one random point in space to its best-fit ellipse, its shape and orientation will change. This change will produce a new ellipse with a different area and tilt. Then, the limit-state parameter on the two geometric properties will restrict the selection of the new random point to a safe response status, avoiding the failure domain. The quantile estimation is computed after the division of the domain in safe and failure conditions and by setting the limit-state parameter equal to the quantile of interest.

In conclusion, the usage of the adaptive algorithm together with the metamodel technique of Polynomial-Chaos Kriging will define the areas in the cross-section plane in which a point can be used to enrich the dissected segmentation for its reconstruction.

Gian Marco Melito
gmelito@tugraz.at
Institute of Mechanics
Graz University of Technology
Austria

Katrin Ellermann
ellermann@tugraz.at
Institute of Mechanics,
Graz University of Technology
Austria

Improving computing time of the Dempster-Shafer theory through a Bootstrap method for applications in vehicle safety

Car companies aim to build safe cars to prevent occupant injuries in the event of a crash. The department of vehicle safety therefore has to assess --- or rather forecast --- the crashworthiness and intervene if the car design will not be able to achieve certain targets. The National Highway Traffic Safety Administration --- the United States federal civilian agency for road and vehicle safety --- performs crash tests to rate the safety of the vehicle as part of the New Car Assessment Program star rating. This star rating is based on the vehicle safety score that results from a frontal, side and rollover test. Five stars stand for the best crashworthiness and zero stars for the worst one in the relevant vehicle category.

The vehicle safety score is calculated from the star rating of the individual crash tests. For the frontal crash, the star rating is determined separately for the passenger and driver. Both depend on the probability of injuries to the considered occupant with respect to a baseline value. This so-called relative risk is combined from the probabilities of injury for different body parts --- head, chest, neck and femur. Sensors integrated in the driver and the passenger dummy measure injury criteria values from which the required probabilities are computed through risk curves.

The car company itself performs these crash tests during development to predict the rating awarded by the administration. Identical constellations of the same crash test usually lead to non-identical results due to aleatoric uncertainty and the usual epistemic uncertainty inherent in the development process. One can model the resulting uncertainty of the injury criteria by introducing separate intervals for each criterion.

We propose to then use the Dempster-Shafer theory (DST) to forecast the outcome of the NCAP rating. The DST propagates uncertainties and yields the best and worst possible rating arising from these uncertainties. The plausability curve represents the best possible result whereas the belief curve characterizes the worst. Both curves have the form of a cumulative distribution function. For the calculation, one has to construct all DST interval cells, C_i for $i\in\{1,...,n\} \subset \N$, and calculate their corresponding basic probability assignments, p_i.

A performance bottleneck of the DST is the high number of optimization problems to be solved --- $2n$ in total. The minimum and maximum values resulting from the optimization are needed to obtain the plausability and belief curve, respectively. To reduce the number of optimization problems to $2n_\text{s}, n_\text{s} < n$, we suggest to use a Bootstrap method for the minimum and maximum values as well as for the corresponding probabilities. For this purpose, we randomly select n_s interval cells along with their probabilities and calculate the respective minimum and maximum values. The remaining cells and their probabilities --- altogether $n - n_\text{s}$ cells --- are then approximated by a Bootstrap method.

By randomly choosing the n_s interval cells, it is likely that the cells with the highest probabilities are not included. As a consequence, the approximated curves are significantly distorted since the largest jumps of the exact curves are missing. For this reason, we propose to take the $n_\text{b} < n_\text{s}$ biggest probabilities together with their optimized values into account. Only after that, we choose $n_\text{r} = n_\text{s} - n_\text{b}$ random cells from the not yet selected $n - n_\text{b}$ ones as before, optimize them and apply the Bootstrap method.

We present an adaptive approach to choose n_s, n_b, and n_r that the optimization problems are reduced and the approximated curves reflect the exact DST. The exact DST plausability and belief curve are not known. Therefore, we need to specify when the bootstrapped curves approximate the exact curves closely. This consideration is also implemented in the approach. The NCAP star rating is used as an example to demonstrate our approach. Instead of optimizing $n = 1296$ interval cells, the approach can closely approximate the exact DST curves with only optimizing $n_s = 250$ cells at most.

Jonas Jehle
jonas.jehle@bmw.de
BMW Group

Volker Lange
volker.vl.lange@bmw.de
BMW Group

Matthias Gerdts
matthias.gerdts@unibw.de
Universität der Bundeswehr München

Session 3: Sensitivity analysis and engineering issues

On the convergence of quantile perturbed law-based indices in robustness analysis

Quantile-oriented PLI

Perturbed law-based indices aim at assessing the impact of a perturbation on a probabilistic quantity of interest (QoI). They were first introduced in a structural reliability framework for the study of a failure probability. In more recent works, they were adapted to the study of output quantiles when perturbing the distributions of model inputs. In this last case, they allowed the analysis of epistemic uncertainties which are physical parameters of thermal hydraulic simulation code. In practice, they are suitable for robustness analysis in any kind of uncertainty propagation approach.

We denote q the quantile at a given level of the output Y=g(X) of a given model g with uncertain inputs X=(X1,…,Xd), and q' being the same quantile for the output distribution of g when a perturbation is introduced on a parameter of the i-th input Xi. When applied to the study of output quantiles of a computer code, a perturbed law-based index is defined as a signed ratio of the quantile of the perturbed distribution over the quantile of the initial distributions [1,2]. More precisely the PLI is q'/q -1 when q'>q and 1-q/q' else. Basically, the PLI measures the percentage of increase or decrease of the quantile value when, for example, the mean or standard deviation of Xi are moved of a certain quantity (see, for example, [3], section 3.3 : Methodologies of input perturbation).

Estimation of quantile-PLI

The PLI is generally estimated using a plugin estimator based on a couple of estimators of the quantile q and q', as done by the PLIquantile function of the 'sensitivity' R package [4]. This last method uses a common empirical estimator for the initial quantile based on the empirical distribution function associated with a given output sample. The perturbed quantile is estimated through the same set of inpuput-output values. For this purpose, we transform the empirical distribution function into the perturbed distribution function. This one is obtained by weighting each term of the sum of the empirical distribution function by a likelihood ratio to handle the perturbation. This method relies on the same principle as used for importance sampling Monte-Carlo algorithms.

Asymptotic law of the quantile-PLI estimator

In spite of their advantages and use in an industrial context, no theorem was proved to guarantee the convergence of the quantile-oriented version of PLI as exists for failure probabilities. The convergence of the estimators of q and q' has been long established and CLT for both estimators are mentioned in many papers ([5] for example). However, proving such a theorem for the estimator of the PLI requires to study the asymptotic law of the couple of the two estimators which is not straightforward when the estimators are correlated. In particular, when both initial and perturbed estimators come from the same sample.

We have proven recently that, under the existence of a third moment of the likelihood ratio, a bivariate CTL exists for the joint distribution of the two estimators. We also give the covariance matrix of the bivariate Gaussian distribution asymptotically reached. This allows to obtain a CLT for the estimator of the PLI the same way as the one proved in [3] (Appendix D) for failure probabilities-PLI.

Perspectives

This result first of all suggest to compute approximate bounds of confidence intervals (CI) for the estimator of quantile-PLI. But such an asymptotic CI requires to know the density function of the output to be computed. We propose to use kernel estimators for the output density and the maginal perturbed input density, which could also be used for the estimation of the corresponding CDF and allow a better estimation of q and q'. The next steps would then consist in implementing such an estimation method for the PLI and compare it with its ''empirical'' estimator, as well as the estimated bounds of the asymptotic CI provided by the CLT should be compared against the bootstrap bounds currently provided by the PLIquantile function in R.

References

[1] Sensitivity analysis using perturbed-law based indices for quantiles and application to an industrial case, R. Sueur, B. Iooss, T. Delage, 10th International Conference on Mathematical Methods in Reliability, 2017
[2] Robust analysis of epistemic uncertainties propagation studies in LOCA assessment thermal-hydraulic model, T. Delage, R. Sueur, B. Iooss, Best Estimate Plus Uncertainty International Conference, May 13-19, 2018
[3] Paul Lemaître. Sensitivity analysis in structural reliability, PhD Thesis, University of Bordeaux, 2014
[4] Package 'sensitivity' version 1.12.2, September 2018, B. Iooss, A. Janon and G. Pujol
[5] Lihua Sun, L. Jeff Hong. Asymptotic representations for importance-sampling estimators of value-at-risk and conditional value-at-risk. Operations Research Letters 38, 246-251, 2010.
[6] Peter W. Glynn. Importance sampling for Monte Carlo estimation of quantiles. Proceedings of the Second International Workshop on Mathematical Methods in Stochastic Simulation and Experimental Design, 180-185, 1996.

Roman Sueur
roman.sueur@edf.fr
EDF R&D
Institut de Mathématiques de Toulouse

Jérôme Stenger
jerome.stenger@edf.fr
EDF R&D
Institut de Mathématiques de Toulouse

Clément Gauchy
clement.gauchy@edf.fr
EDF R&D
Institut de Mathématiques de Toulouse

An Application Example of the JRC Methodology for Interoperability Testing: Design of Experiments, Uncertainty and Sensitivity Analysis

In this work we apply the methodology for interoperability testing developed at the Smart Grids Interoperability Laboratory of the Joint Research Centre of the European Commission to test the interoperability between a data concentrator and a set of smart meters [1].

Data concentrators and smart meters are fundamental components of a smart electricity grid. Smart meters are simple electronic devices that register real-time consumption and generation of electricity, in a household or an industry, and send the data to the electricity retailer for monitoring and billing.

A data concentrator connects smart meters from the same neighborhood providing communication capability with a large number of them. The data concentrator can be configured to request periodic electricity information from each smart meter.

For the correct functioning of the smart grid, components and systems must be interoperable. Interoperability is, loosely speaking, the ability of such components / systems to communicate appropriately and understand each other so that they can properly perform the function they are supposed to perform.

Design of experiments and the sensitivity / uncertainty analysis are relevant components of the methodology, which can reveal the limits of a system under test and give valuable feedback about the critical conditions which do not guarantee interoperability. The design and analysis of experiments employed in the JRC methodology supply information about the crucial parameters that either lead to an acceptable system performance or to a failure of interoperability.

The system is stress-tested under different conditions by varying three parameters: the rate at which meter data are requested by the data concentrator, the number of smart meters connected to the data concentrator and the amount of information contained in each data packet that is supposed to be transferred from the smart meters to the data concentrators.

For each combination of these three factors (identified by quasi Monte Carlo sampling), the share of readings correctly received by the data concentrator from a set of smart meters is measured in a real experiment. This quantity is taken as a measure of the level of interoperability between the two devices.

Each experiment lasts 3 hours plus the time to set-up the infrastructure. Therefore, only a small amount of experiments can be executed, for a given time budget. For this reason, a restricted number of uncertain factors is usually considered, otherwise the number of experiments required to fill the factors' space would necessarily be too large thus making the analysis unfeasible.

The uncertainty analysis shows that a change of the system's conditions has an effect on the system's overall interoperability and the critical region where the system is not interoperable can be ascertained through Monte Carlo filtering.

The sensitivity analysis works in parallel to the uncertainty analysis helping the experimenter to understand important and non-important factors to facilitate the subsequent

implementation of a meta-model that is needed for the uncertainty analysis and the Monte Carlo filtering.

It was decided to implement a High Dimensional Model Representation (HDMR) meta-model [2] using the INTEROP software developed by Broda Ltd. Starting from the original SobolGSA application, INTEROP allows the automatic implementation of a Python code for the HDMR that can be used to run the meta-model a large number of times thus covering the space of uncertain factors as much as possible and allowing the user to find the critical region(s) where the system is not interoperable.

The meta-model computes the share of readings correctly received by the data concentrator from a set of smart meters, as a quantitative measure of the level of interoperability between the two devices.

The meta-model was run over a fine regular grid in the space of the two factors: "rate at which meter data are requested by the data concentrator" and "number of smart meters connected to the data concentrator", for a given value of the third factor "amount of information contained in each data packet ".

We will show the results of running the HDMR on the space of the three factors, highlighting the critical areas where the share of readings correctly received from the data concentrator is too low to be acceptable.

The HDMR allows the estimation of the sensitivity indices, which are computed to identify the most and least important factors. This information is crucial because, being impossible to display the input – output relationship on a 4D space, the sensitivity analysis helps understanding which minimum set of 3D scatterplots is necessary to display in order to visualize the critical regions in the space of the factors.

References

[1] Papaioannou, I.; Tarantola, S.; Lucas, A.; Kotsakis, E.;Marinopoulos, A.; Ginocchi,M.; Olariaga-Guardiola,M.; Masera,M. Smart Grid Interoperability Testing Methodology; EUR 29416 EN; Publications Office of the European Union: Luxembourg, 2018; ISBN 978-92-79-96855-6.

[2] Li G., Hu J., Wang S., Georgopoulos P., Schoendorf J., Rabitz H. Random Sampling-High Dimensional Model Representation (RS-HDMR) and orthogonality of its different order component functions, Journal of Physical Chemistry A, vol. 110, no. 7, (2006) 2474–2485.

Stefano Tarantola
stef.tara@libero.it
Joint Research Centre - European Commission

Nikoleta Andreadou
nikoleta.andreadou@ec.europa.eu
Joint Research Centre - European Commission

Alexandre Lucas
alexandre.lucas@ec.europa.eu
Joint Research Centre - European Commission

Ioannis Poursanidis
iapoursanidis@gmail.com
Joint Research Centre - European Commission

Study on a gas crisis in the EU, uncertainty and sensitivity analysis

Natural gas plays an important role in the share of energy in the EU. Nevertheless, only a few Member States have significant indigenous resources, which makes the EU very dependent on third countries such as Russia, Algeria, Norway and Qatar to satisfy its gas demand. In January 2013 a commercial dispute between NAFTOGAZ of Ukraine and GAZPROM produced the largest gas crisis ever experienced by the EU and some neighbouring countries. This event triggered legislative initiatives at a European level starting with Regulation No 994/2010 followed by Regulation (EU) 2017/1938 on "measures to safeguard the security of gas supply". The measures were focusing mainly on infrastructure development, diversification of supply resources and strengthening regional cooperation between countries that have a common supply source and therefore are exposed to the similar risks.

Shortly after the occurrence of the 2009 gas crisis, JRC started developing a new simulation tool: the ‘‘gas emergency flow’’ model (GEMFLOW; Szikszai and Monforti, 2011). The aim was to assess the possible outcomes of a supply disruption beforehand and minimize losses during an emergency by finding the optimal distribution of flows based on Monte-Carlo modelling principles. This new tool has been modified and improved several times over the years in order to have a more enhanced view on the resilience of the natural gas supply system during supply crises (Rodríguez-Gómez et al., 2016). In its present status, the described model cannot serve as a hydraulic software that is currently used by the national system operators, but it is able to draw significant conclusions on the European gas system's capabilities.

GEMFLOW simulates the reaction of an interconnected regional gas system to a gas crisis. Countries are linked to one another by maximum one (virtual) pipeline summing up the capacities of all (real) pipelines connecting them. All storage spaces and capacities as well as production and LNG capacities are supposed to be virtually available to meet gas demand, leading to a further simplification of the system. Besides infrastructure capacities (adjusted to the specific disruption) the inputs that are required for the specific simulation are the demand and the actual stock level (to determine the actual storage withdrawal capacity). A crisis is simulated a user-defined number of times for a specified length by changing the availability of a supply source (either a cross-border point, a supply corridor or the amount of commodity in underground storages) and by randomising, within a set of constraints and rules on sharing remaining resources, the allocation of the remaining supply to all countries. Figure 1 provides an overview of the geographical structure of GEMFLOW. Figure 1.- Geographical structure of GEMFLOW indicating aggregated pipeline connections between countries (EU countries, gas transit countries and gas producer countries) by means of lines.

The analysis of a crisis scenario focuses on the status of the supply system during and at the end of the crisis. Consequently the main results of the simulations are country (or region) level distributions for the unserved gas (the part of demand of a country or region in the examined area that is not met due to the crisis), days of imbalance (supply < demand) in the affected countries and utilisation of infrastructure (percentage of the capacities of pipelines, storages, production and LNG terminals that are used to cope with the crisis).

In this study we analyse a gas crisis (interruption of transit gas across Ukraine to the EU) and pay special attention to the identification of the most important uncertain parameters and to the characterisation of uncertainty in key output variables. To identify the most important input parameters and their contribution to the output uncertainty we use Sobol sensitivity indices computed with the Polynomial Chaos Expansion (PCE) technique (Sudret, 2008, Shao et al., 2017). The PCE technique allows the estimation of the Sobol first order and total order

indices with fewer model runs. It therefore can successfully be applied even when the resources are limited and/or in the case complex models must be analysed, as it is the case of GEMFLOW. We combine these sensitivity indices with graphic techniques like scatterplots and parallel line plots.

Ricardo Bolado
ricardo.bolado-lavin@ec.europa.eu
European Commission, DG-JRC

Nicola Zaccarelli
nicola.zaccarelli@ec.europa.eu
European Commission, DG-JRC

Rossana Rosati
rossana.rosati@ec.europa.eu
European Commission, DG-JRC

Daniel Albrecht
daniel.albrecht@ec.europa.eu
European Commission, DG-JRC

Design for sensitivity analysis and sensitivity auditing on models of the water-energy-food nexus

Quantitatively understanding the water-energy-food nexus is a needful requisite for informing and steering new national and global policies toward sustainability. The EU funded MAGIC project aims at offering to a wide set of actors (policy officers and makers, international agencies, NGOs,…) a rich representation of the nexus key issues, that avoids relying on any closed model and could be tailored to the stakeholder's perspective. This target is pursued through Quantitative Story-Telling, applied to relevant narratives about the nexus, that is backed up by an analysis approach grounded in bio-economics.

Our contribution to the project as applied statisticians is focused on operationalizing the bio-physical, social and quality checks required to sustain any narrative about nexus, by providing a formal representation apt for its quantification. Since the different perceptions, goals and values bound to any stakeholder, we have designed a parameterised approach to a computer model based on a hierarchy of processors, a logical device borrowed from relational analysis. In order to apply this kind of model to a set of MAGIC case studies, a software system has been designed to suitably declare the processors' hierarchies and quantify the knowns (i.e. observations, stemming from official statistics or technical knowledge, where multiple observer could be attached to the same observable, thus allowing different observations) including a solver module to obtain unknown quantities from its graph based representation.

We are fully aware that any model application would be useless without enabling the user to understand its ground (transparency about any assumption and input data), weaknesses (different kind of uncertainties simultaneously affecting its results) and strengths (the key relationships it enables to discover and highlight). Therefore, a strategy to allow sensitivity analysis and eventually enable sensitivity auditing of the results has been explicitly addressed during the development of the model. It had required a twofold translation process, made open to users via a Domain Specific Language used to declare the case study at hand and the input data:
* form narratives expressed as multipurpose grammars to a structured computer model and
* from the formal computer model to a semantically relevant mathematical model suitable to sustain sensitivity as well as scenario analyses.

The mathematical model of the solver is based on the analogy with a multi-commodity network flow problem and is suitable for different parameterization of constraints and objective function, so to accommodate different stakeholders. From the technical standpoint (i.e. in the model-landscape) the solver is based on the application of an heuristic algorithm for optimization (belonging to the class of Mixed Integer Linear Programming, applied to a generalized multi-commodity transportation problem that demonstrated apt to capture the interdependent network flows problem to which the hierarchy of processor could be mapped) coded in Python by leveraging the Gurobi Optimizer libraries.

The application of standard sensitivity analysis methods to this model is not straight, since the solution to the flow problem does not satisfy the uniqueness conditions; moreover, the network model encompass a large set of parameters (further expanded by potential stakeholders choices) that can each take on many different values, making the sensitivity analyst to cope with a vast space of potential parameter combinations. Therefore, a suitable approach should be selected and an apt design identified, then the procedure has to be automated via a further software module.

In our problem the focus is at highlighting the key parameters in the formal representation of a complex system, the coupled human-environment semantically relevant for a specific case study, eventually analyzed at multiple scales and keeping into account its openness. Therefore, we would not perform a Global sensitivity analysis, since we are not interested in the output uncertainty, but concerned with respect to the output itself. The input parameter intervals to be explored does not represent analyst's uncertainty about that parameters but accommodate also for the different stakeholders' perspectives, indeed. Whereas approach for Variance-based sensitivity analysis could summarize the sensitivity by identifying how much of the output variance is due to each contributory source of uncertainty, conversely we aimed at input parameters' importance. Nonetheless, we were able to borrow useful analogies form the approaches for Global sensitivity analyses in the field of transportation networks.

We have tailored a sensitivity analysis strategy allowing us to start with a small set of different parameters combination — aiming at a screening phase; then the results steered toward an increased set of points in the parameter space needed to compute a fully quantitative measure. This approach demonstrated both flexible, accurate and informative, and also achievable at a reasonable computational cost

Acknowledgements:
Authors acknowledge financial support by the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 689669 (MAGIC). This work reflects the authors' view only; the funding agency is not responsible for any use that may be made of the information it contains.

References

[1] Giampietro, M., Aspinall, R. J., Ramos-Martin, J., Bukkens, S. (eds.), (2014), Resource Accounting for Sustainability Assessment – The Nexus between Energy, Food, Water and Land Use, Routledge, London
[2] Matthews, K.B., Blackstock, K.L., Rivington, M., Waylen, K., et al. (2017), Delivering more than the "Sum of the Parts" - Using Quantitative Story Telling to address the challenges of conducting science for policy in the EU land, water and energy nexus. In: MODSIM2017, 22nd International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, December 2017, pp. 15-21.
[3] González-López, R., Giampietro M., (2017), Multi-Scale Integrated Analysis of Charcoal Production in Complex Social-Ecological Systems, Frontiers in Environmental Science, 5, 54.
[4] Staiano, M. , Nebot Medina, R., (2019), Engineering Software Tools for Capturing the Complexity of the Water-Energy-Food Nexus: Lessons from the MAGIC Project, 10th International Multi-Conference on Complexity, Informatics and Cybernetics (IMCIC 2019), 15 March 2019, Orlando (FL), USA.
[5] Staiano, M., Siciliano, R., Pandolfo, G., Nebot Medina, R., (2019), Semantic Model-to-Model Translation to Capture the Nexus Complex Predicate, 10th International Multi-Conference on Complexity, Informatics and Cybernetics (IMCIC 2019), 15 March 2019, Orlando (FL), USA.
[6] Thompson, E. L., Smith, L. A., (2019). Escape from model-land. Economics Discussion Papers, No 2019-23

Michele Staiano
michele.staiano@unina.it
Department of Industrial Engineering
University of Napoli Federico II

Italy

Carmela Iorio
carmela.iorio@unina.it
Department of Industrial Engineering
University of Napoli Federico II

Giuseppe Pandolfo
giuseppe.pandolfo@unina.it
Department of Industrial Engineering
University of Napoli Federico II

Roberta Siciliano
roberta@unina.it
Department of Industrial Engineering
University of Napoli Federico II

Monte Carlo simulations for process innovation and design in process and chemical engineering applications

Normal 0
false false false DA X-NONE X-NONE

In the first part of this contribution, a review of process systems engineering discipline and how they contribute to design of next generation processes in chemical engineering will be given. In particular, development and application of PSE tools and methods to advance process and product development challenges in pharmaceutical and bio-based industries will be presented. In the next part, we present successful applications of monte carlo simulations using case studies.

First, we present recent progress in designing next generation wastewater treatment plants in the era of circular economy and resource recovery paradigm. To this end, a Monte Carlo-based framework for conceptual process synthesis and design, also called SPDLab tool, is presented and its application to address the challenges of wastewater treatment design is discussed. The key feature of the framework is that numerically it enables to work with predictive and complex non-linear mechanistic models of individual technological treatment units. This alone helps engineers to use the tool on industrially relevant problems and scales (as opposed to just working with simple toy problems often encountered in literature). In general for process design and in particular wastewater treatment plants, many uncertainties exists at the design stage, namely due to changing feed flowrate (quantity) and composition (quality) among others.

Therefore the framework also includes uncertainty and sensitivity analyses as well as simulation based optimization techniques to handle solving design problems under uncertainties.Selected application of the framework to address wastewater treatment process synthesis and design is highlighted. As a second domain applications of Monte Carlo engineering, we focus on using Monte Carlo based framework to support process innovation and optimization for fermentation based production in dairy industry.

As a third case study, we present application of the framework for innovative process development applications for oleochemical industry by providing an interface with commercial process simulator. This includes among other development of surrogate models for supporting mixed integer nonlinear programming based process design, sensitivity analysis using polynomial chaos expansion (PCE) based sobol index to study process bottlenecks understanding among others are presented and discussed.

Overall the capabilities of Monte Carlo engineering in complementing simulation based engineering for process innovation and design is clearly evident and emphasized as enabling technology in the era of digitalization for processing industries.

Gurkan Sin
gsi@kt.dtu.dk
Process and Systems Engineering Center (PROSYS)
Department of Chemical and Biochemical Engineering
Technical University of Denmark
Denmark

Session 4: Sensitivity analysis and environmental issues

Why so many published sensitivity analyses are false: a systematic review of sensitivity analysis practices

The importance of sensitivity analysis is widely acknowledged. Sensitivity analysis is prescribed in national and international guidelines in the context of impact assessment . When the output of a model feeds into policy prescription and planning, a sensitivity analysis would appear as an essential element of due diligence.

Despite the clear importance of sensitivity analysis, there are a number of problems observed in practical sensitivity analysis and uncertainty analysis, which can be found in all fields of research. These problems range from confusions in terminology to statistically inaccurate techniques which can underestimate model uncertainty. Specifically:
* While most practitioners of SA distinguish it from UA, modellers overall tend to conflate the two terms, e.g. performing an uncertainty analysis and calling it a sensitivity analysis.
* The sensitivity analysis methodology often relies on so-called local techniques which are invalid for nonlinear models.

Based on a recently published work [4], and on some additional material developed since it was published in early 2019, the talk will back up these assertions with evidence. Demonstrating that there is a systematic problem in practical sensitivity analysis might be a first step towards improving the situation. Some reviews of sensitivity analysis practice do already exist: in [5], an assessment of the state of sensitivity analysis was performed using a bibliometric approach. The authors in [6] review the state of sensitivity analysis (or lack thereof) in hydrological modelling. However, to the authors' knowledge, there is no detailed cross-disciplinary assessment of the state of sensitivity analysis, as practised by modellers.

Accordingly, the talk will address the following points:
* To assess the "state" of sensitivity analysis across a range of academic disciplines. We do this by a systematic review of a large number of highly cited papers in which sensitivity analysis is the focus in some respect.
* To discuss – based on this review - known problems and misinterpretations of sensitivity analysis, why these might occur, and propose some ideas for how these problems might be addressed.

To this effect, an extensive literature review was carried out based on highly cited articles that have a focus on sensitivity analysis. The literature search was conducted on the Scopus database using as search criteria the strings "sensitivity analysis" and "model/modelling", and "uncertainty" in the title, abstract or keywords. The search was restricted to journal articles in English in the period 2012-2017, in order to provide a sample of recent research.

This search resulted in around 6000 articles. After binning these articles in 19 subject areas, from Agricultural and Biological Sciences all the way to Social Sciences, we selected the top twenty most-cited papers from each field. This resulted - after some screening and censoring, into a total of 280 papers which were read by the authors in order to answer the following questions:

* Was an uncertainty analysis performed? If so, was a global or local approach used?
* Was a sensitivity analysis performed? If so, was a global or local approach used?

* Was the paper primarily focused on the method of sensitivity analysis, or on the model (application)?
* Was the model used linear, nonlinear, or was it unclear?

The review indicated that a significant fraction of papers investigated use sensitivity analysis approaches which fail elementary considerations of experimental design and do not properly explore the space of the input factors, with the result that uncertainty is generally underestimated and sensitivity is wrongly estimated. Up to 65% of the reviewed (highly cited) papers are based on inadequate methods (i.e. varying one input factor at a time). Even in the most generous interpretation, where all models with unclear linearity are assumed linear, still over 20% of papers contain inadequate methodology. Further, a significant number of papers confuse sensitivity and uncertainty analysis, which is likely to exacerbate the problem with spreading good practice.

In the talk we investigate the reasons for such trends. It is observed that the problem with sensitivity analysis is partly attributable to the fact that mathematical modelling is not a discipline in its own right, and every branch of science and technology approaches modelling following its own culture and practice. Uncertainty and sensitivity analyses are likewise orphans of a disciplinary home. The similitudes and differences between mathematical modelling – the subject of the present review – and statistical modelling are also discussed in relation to the present debate among statistician about misuse or abuse of statistics such as the p-test. The talk will conclude with suggest good practices for both teachers and practitioners.

References

[1] European Commission, ‘‘European Commission IMPACT ASSESSMENT GUIDELINES,’’ 2009.
[2] Office of Management and Budget, ‘‘Proposed Risk Assessment Bulletin,’’ 2006.
[3] U.S. Environmental Protection Agency (EPA), ‘‘Guidance on the Development, Evaluation, and Application of Environmental Models,’’ 2009.
[4] A. Saltelli, K. Aleksankina, W. Becker, P. Fennell, F. Ferretti, N. Holst, S. Li, and Q. Wu, ‘‘Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices,’’ Environ. Model. Softw., vol. 114, pp. 29–39, Apr. 2019.
[5] F. Ferretti, A. Saltelli, and S. Tarantola, ‘‘Trends in sensitivity analysis practice in the last decade,’’ Sci. Total Environ., vol. 568, pp. 666–670, Oct. 2016.
[6] M.-J. Shin, J. H. A. Guillaume, B. F. W. Croke, and A. J. Jakeman, ‘‘Addressing ten questions about conceptual rainfall–runoff models with global sensitivity analyses in R,’’ J. Hydrol., vol. 503, pp. 135–152, Oct. 2013.

Andrea Saltelli
andrea.saltelli@gmail.com
University of Bergen, Norway & Universitat Oberta de Catalunya, Spain

Ksenia Aleksankina
s0950577@sms.ed.ac.uk
University of Edinburgh & NERC Centre for Ecology and Hydrology
Edinburgh, UK
William Becker
William.Becker@ec.europa.eu
European Commission, Joint Research Centre, Ispra, IT

Pamela Fennell

pamela.fennell.13@ucl.ac.uk
University College London, UK

Federico Ferretti
federicoferretti@hotmail.it
European Commission, Joint Research Centre, Ispra, IT

Niels Holst
niels.holst@agro.au.dk
Aarhus University
Denmark

Qiongli Wu
qiongli.wu@gmail.com
Wuhan Institute of Physics and Mathematics
Chinese Academy of Sciences
Wuhan, China

Sensitivity analysis of an avalanche flow dynamic model using one sample methods and aggregated indices

Avalanche flow dynamic models depend on poorly known inputs [Naaim et al., 2013]. The outputs of these models are commonly both functional (e.g., the velocity and snow depth of an avalanche in different points of a discretized path) and scalar (e.g., runout distance). These models are employed for land-use planning as far as for the design of defense structures. Thus, it is required to assess the impact of the uncertainty of the input parameters on the outputs, and this is the aim of the sensitivity analysis. Nevertheless, very few sensitivity analysis have been developed so far in the avalanche field.

Classically, if the output of a numerical model is scalar, its total variance can be split into partial variances by using the Hoeffding decomposition. Each of the partial variances is associated to an input or to an interaction between inputs. Then, the Sobol' indices are calculated as the ratio of each partial variance and the total variance. These indices are statistical tools that quantify the influence of each input on the output.

If the model output is functional, it is possible to apply this sheme to each element of the vector but this could lead to redundancies in the results. Also, they could be difficult to interpret if the dimension of the output is high. Therefore, it can be more informative to apply the sensitivity analysis to the whole set of outputs by calculating the aggregated Sobol' indices proposed by [Lamboni et al., 2009, Gamboa et al., 2013]. In addition to the complex output structure of avalanche models, their functional outputs (e.g., velocity and snow depth profile) have a high number of zeros corresponding to the beginning of the starting zone and the end of the runout zone. Thus, in our application, only the non-null values are used in the computation of sensitivity indices, in other words, we are interested in computing the indices of conditional random variables. The samples are constructed from acceptance-rejection sampling. Due to our particular conditional output, we cannot use traditional Sobol' indices estimation methods (e.g., the Fast Amplitude Sensitivity Test (FAST), Random Balanche Design, polynomial chaos expansion) because they require a classical sampling scheme. That's why this application represents a novelty in sensitivity analysis. We tackle our problem using methods that do not require any specific sampling design, as for example, the one-sample or given data methods [Plischke et al., 2013, Strong and Oakley, 2013] and the effective algorithm for computing global sensitivity indice (EASI) method developed by [Plischke, 2010].

Application is made in a 2-D avalanche model based on the Saint-Venant equations. Bootstrapping intervals are calculated to analyze the accuracy of the methods. In previous studies of the sensitivity analysis of avalanches models, little attention has been given to the application of a formal statistical method to quantify the uncertainty of the inputs. Also, the complete set of outputs has not been used to quantify the uncertainty. Therefore, these are the two objectives of this work.

In conclusion, both estimation methods lead to similar results. These results show that the static friction coefficient is the most influential but the other inputs are non negligible since they significantly contribute to the total variability, with variations along the avalanche path as function of topography and flow dynamics.

References

[1][Gamboa et al., 2013] Gamboa, F., Janon, A., Klein, T., and Lagnoux, A. (2013). Sensitivity analysis for multidimensional and functional outputs. ArXiv e-prints.

[2][Lamboni et al., 2009] Lamboni, M., Makowski, D., Lehuger, S., Gabrielle, B., and Monod, H. (2009). Multivariate global sensitivity analysis for dynamic crop models. Field Crops Research, 113(3):312 – 320.
[3][Naaim et al., 2013] Naaim, M., Durand, Y., Eckert, N., and Chambon, G. (2013). Dense avalanche friction coefficients: influence of physical properties of snow. Journal of Glaciology, 59(216):771–782.
[4][Plischke, 2010] Plischke, E. (2010). An effective algorithm for computing global sensitivity indices (EASI). Reliability Engineering & System Safety, 95(4):354 - 360.
[5][Plischke et al., 2013] Plischke, E., Borgonovo, E., and Smith, C. L. (2013). Global sensitivity measures from given data. European Journal of Operational Research, 226(3):536–550.[Strong and Oakley, 2013] Strong, M. and Oakley, J. E. (2013). An efficient method for computing single-parameter partial expected value of perfect information. Medical Decision Making, 33(6):755–766. PMID: 23275450.

Heredia Maria Belen
maria-belen.heredia@irstea.fr
Univ. Grenoble Alpes
Grenoble, France

Clémentine Prieur
clementine.prieur@univ-grenoble-alpes.fr
Institute of Engineering Univ. Grenoble Alpes
Univ. Grenoble
Grenoble, France

Nicolas Eckert
nicolas.eckert@irstea.fr
Univ. Grenoble Alpes
Grenoble, France

A New Approach for Global Sensitivity Analysis on any Given Data: Data-driven Variogram Analysis of Response Surfaces

Global Sensitivity Analysis can be a powerful tool for understanding how various processes in complex systems interact and how the system behavior varies by perturbing these processes. Most of the existing GSA techniques are sampling-based, requiring a ‘‘model’’ to determine the influence of input variables on the response variables. However, in a burgeoning era of big data with the exponential growth in data availability and accessibility, there is an increasingly pressing need to develop efficient GSA techniques that can extract sensitivity-related information from any available data on a process or set of processes, independent of having input–output functional relationships (i.e., a model).

To overcome this challenge, we extend the theory of Variogram Analysis of Response Surfaces (VARS), a recently developed approach to GSA, to enable its application to any collected dataset. This extended version, called Data-driven VARS (D-VARS) works on any given data to approximate the anisotropic variance and covariance structure in the underlying (but unknown) response surface, across the full range of ‘‘perturbation scales’’, and uses that information to estimate the variogram-based, variance-based, and derivative-based indices of global sensitivity. D-VARS enables the user to efficiently conduct GSA for cases where the properties of input-output distributions and of the underlying response surface are unknown and only a (small) sample of the input-output space is available.

D-VARS is first tested on a synthetic case study with data samples of varying sizes. In this case study, the underlying response surface and the distributional properties of the inputs and outputs are perfectly known and used for the assessment of the efficiency and robustness of D-VARS. Second, we apply D-VARS on a multivariate hydro-climatic dataset collected from the southern boreal forest at Boreal Ecosystem Research and Monitoring Sites (BERMS), Saskatchewan, Canada, to extract information on sensitivities and relationships within the different variables. The results of the two case studies confirm that D-VARS is effective and robust, and can perform well with data samples of small size.

Razi Sheikholeslami
razi.sheikholeslami@usask.ca
University of Saskatchewan

Saman Razavi
saman.razavi@usask.ca
University of Saskatchewan

Session 5: Sensitivity analysis and nuclear issues

Advanced global and target sensitivity analyses using dependence measures and statistical tests - Application to a nuclear safety test-case

Uncertainty quantification [1] is gaining more and more attention in the field of nuclear engineering, not only to assess the reliability and safety of nuclear power plants , but also to provide robust methodological tools which can complete the deterministic (and sometimes over-conservative) analyses already performed, in order to better identify the safety margins. As an example, safety analyses involving thermal-hydraulic issues such as an intermediate-break loss-of-coolant accident (IBLOCA) can be treated using probabilistic and statistical approaches known as the ‘‘best-estimate plus uncertainty’’ (BEPU) methodology in the nuclear engineering community [2]. Such an IBLOCA case may involve around a hundred input variables which have to be handled within the uncertainty modeling and propagation phases. Moreover, the thermal-hydraulic computer model used to simulate the IBLOCA (here, the CATHARE code as in [3]) is complex, strongly nonlinear and costly-to-evaluate. All these constraints make the estimation of the model output, and furthermore, of any quantity of interest (QoI) based on this output, quite difficult. For instance, in the IBLOCA test-case, a typical QoI can be a high-order quantile over the second peak of cladding temperature (PCT).

From this perspective, it is widely recognized that global sensitivity analysis (GSA) plays a key role in the BEPU methodology [4], in particular for ‘‘screening’’ and ‘‘ranking’’ purposes [5]. Screening aims at identifying the significantly influential inputs on the model output. These inputs, called here ‘‘primary influential inputs’’ (PII), constitute the effective input dimension. As for the ranking phase, it aims at providing a quantitative measure of this influence, in order to sort the PII by order of influence. Among several GSA methods available in literature, a recent emphasis has been put over the use of generalized dependence measures to build sensitivity indices. This interest is explained by the relevant theoretical and numerical properties of these indices regarding the two purposes mentioned previously [6]. Among several dependence measures, the use of the Hilbert-Schmidt Independence Criterion (HSIC) led to define a sensitivity index which measures the statistical dependence between the model output and each input under consideration [7]. Formally, the HSIC index relies on the covariance between two functions chosen among a reproducing kernel Hilbert space (RKHS) equipped with a well-chosen kernel. Moreover, HSIC-based statistical tests for independence can be built for screening purposes, such as proposed in [8, 9].

These statistical procedures can be used to test the hypothesis of dependence between the output and each input so as to discriminate between the significant input variables and the non-significant ones. Depending on the size of the available sample (i.e., the couples gathering the input realizations and their corresponding output values), one can apply either asymptotic tests (if the size is sufficiently large) or permutation-based tests [8].

From the general reliability and risk management perspectives, one can show that GSA applied on the model output may be insufficient to catch the underlying influence on the input uncertainty on a specific reliability-oriented QoI (e.g., a high-order quantile). In the context of dependence measures, dedicated indices known as ‘‘target sensitivity indices’’ have been recently proposed in [10]. This target sensitivity analysis (TSA) is used to efficiently identify

the inputs that could potentially be strongly influential on the reliability-oriented QoI but which could have been ignored by the GSA.

In the present work, a three-step approach is proposed to perform GSA and TSA from a unique Monte Carlo sample of input-output realizations. This approach is developed in the specific context of complex computer code simulation and a high-dimensional input vector. In the first step, GSA using HSIC indices and statistical tests is performed so as to identify the PII. In the second step, TSA using the targeted HSIC (proposed by [10]) together with statistical tests is performed to identify the possible input variables that might be influential from the reliability-oriented point of view (i.e., the variables that most contribute to the crossing of a threshold quantile value for the output). The third step results in an aggregation strategy so as to rank the inputs regarding these two different analyses. This step combines both the values of the indices and the p-values given by the statistical tests.

Finally, this three-step approach is applied to a complex IBLOCA test-case to demonstrate its efficiency and find the set of PII. As a perspective, these results could be used to efficiently build a surrogate model by considering only the PII as the explanatory inputs (e.g., as in [9]). The obtained surrogate model could then be used to go deep inside the analysis of the physical relations between the model output and the PII.

References

[1] De Rocquigny E., Devictor N., Tarantola S. (2008). ‘‘Uncertainty in industrial practice: a guide to quantitative uncertainty management’’. Wiley.
[2] Prosek A., Mavko B. (2007). ‘‘The stats-of-the-art theory and applications of Best-Estimate Plus Uncertainty methods’’, Nuclear Technology, 158.69.
[3] Mazgaj P., Vacher J.-L., Carnevali S. (2016). ‘‘Comparison of CATHARE results with the experimental results of cold leg intermediate break LOCA obtained during ROSA-2/LSTF test 7’’, EPJ Nuclear Science & Technology 2, 1.
[4] De Crécy A. et al. (2008). ‘‘Uncertainty and sensitivity analysis of the LOFT L2-5 test: Results of the BEMUSE programme’’, Nuclear Engineering and Design, 238, pp. 3561-3578.
[5] Saltelli A. et al. (2008). ‘‘Global Sensitivity Analysis. The Primer’’. Wiley.
[6] Gretton A. et al. (2005). ‘‘Measuring statistical dependence with Hilbert-Schmidt norms’’, Proc. of the International Conference on Algorithmic Learning Theory (ALT 2005), pp. 63-77.
[7] Da Veiga S. (2015). ‘‘Global sensitivity analysis with dependence measures’’, Journal of Statistical Computation and Simulation 85.7, pp. 1283-1305.
[8] De Lozzo M., Marrel A. (2016). ‘‘New improvements in the use of dependence measures for sensitivity analysis and screening’’, Journal of Statistical Computation and Simulation 86.15, pp. 3038-3058.
[9] Iooss B., Marrel A. (2019). ‘‘Advanced Methodology for Uncertainty Propagation in Computer Experiments with Large Number of Inputs’’, Nuclear Technology, In Press, DOI:10.1080/00295450.2019.1573617.
[10] Raguet H., Marrel A. (2018). ‘‘Target and conditional sensitivity analysis with emphasis on dependence measures’’, ArXiv e-prints, pp. 1-48, arXiv: 1801.10047v2.

Vincent Chabridon
vincent.chabridon@gmail.com
EDF R&D, Département PRISME
France

Amandine Marrel
amandine.marrel@cea.fr
CEA, DEN, DER, SESI, LEMS
France

Bertrand Iooss
bertrand.iooss@edf.fr
EDF R&D, Département PRISME
France

Estimation of higher-order sensitivity indices of final repository models by means of the RS-HDMR and BSPCE metamodeling approaches

Final repository models can behave in a highly nonlinear, non-monotonic or even discontinuous manner which may cause parameter interactions and many uncertainties. Thus, by better understanding parameter interactions, uncertainties of these models may also be better quantified. Probabilistic methods of higher order sensitivity analysis provide a possibility for identifying parameter interactions. Moreover, in contrast to methods of probabilistic sensitivity analysis that use linearization of the model output, it is said that higher order sensitivity analysis may be better appropriate for detecting non-linear or, especially, non-monotonic influences. From the decomposition of the total unconditional variance of the model output, sensitivity measures are computed in terms of sensitivity indices in this type of analysis. The first-order index (SI1) and the total-order index (SIT) are the two key indices which are used in practice. Whereas SI1 refers to the individual impact (or main effect) of one parameter on the model, SIT represents the total effect of one parameter on the output in interactions of any order with all other parameters. The second-order sensitivity indices (SI2), describe the interactions between two model parameters.

Sensitivity indices can efficiently be estimated by the Random-Sampling High Dimensional Model Representation (RS-HDMR) metamodeling approach. This approach is based on truncating the ANOVA-HDMR expansions up to the second order, while the truncated terms are then approximated by orthonormal polynomials (Zuniga et al [1]). By design SIT in this method is a sum of SI1 plus all corresponding SI2's for a considered parameter. To potentially reduce numerical noise upon the higher order indices with low significance, a good selection of maximal polynomial orders of orthonormal polynomial decomposition is required [2]. RS-HDMR belongs to a wider class of methods known as polynomial chaos expansion (PCE). PCE is a widely used approach in metamodeling [3]. Usually only a few terms are relevant in the PCE structure. The Bayesian Sparse PCE method (BSPCE) makes use of sparse PCE. Selection of the proposed PCE structure is based on a Bayesian approach using the Kashyap information criterion for model selection [4]. Using BSPCE, SI1 and SIT can be estimated. In this work we used the SobolGSA software [5] which contains both the RS-HDMR and BSPCE methods.

In this paper, we have compared the RS-HDMR and BSPCE methods using a model for a generic LILW repository in a salt mine. The model includes a barrier in the near field which is chemically dissolved (corroded) over time by magnesium-containing brine, resulting in a sudden significant change of the model behaviour and usually a rise of the radiation exposure. We considered two different sets of parameters: one with 6 parameters and one with 20 parameters (LILW6 and LILW20 models, respectively). The LILW20 model also considers the 6 parameters of the LILW6 model. For the time-dependent analysis, 31 time points were used.

We compared the results for SI1 obtained with the EASI method [6] with those obtained using both metamodeling approaches. EASI is a Fourier-based technique for the computation of the SI1 indices. It belongs into the same class of algorithms as FAST and Random balance designs (RBD). EASI is a computationally cheap method for which existing data can be used. Unlike the FAST and RBD methods, which use a specifically generated sample set that contains suitable frequency data for the input factors, in EASI these frequencies are introduced by sorting and shuffling the available input samples.

The SI1 time curves of both models and approaches agree with those obtained with the EASI method. Different sets of maximal polynomial orders and different numbers of simulations were examined to find optimal coefficients for the two LILW models using the RS-HDMR

approach. While SI1 indices for both approaches agree well for most of simulations, for some choices of numbers of model runs, there are quite some differences in the SIT indices. For the BSPCE approach the highest accuracy is obtained for both models with the largest number of model runs which is twice of the number of model runs used with the most optimal RS-HDMR metamodel. This indicates that there might be interaction effects higher than second order in the systems or there is numerical noise when the number of model runs used to build the BSPCE metamodel is too high.

Acknowledgements

The work was financed by the German Federal Ministry for Economic Affairs and Energy (BMWi). We would also like to thank Dirk-A. Becker for his constructive feedback.

References

[1] M.M. Zuniga, S. Kucherenko, N. Shah, Metamodelling with independent and dependent inputs. Computer Physics Communications, 184(6): 1570–1580, (2013).
[2] S. M. Spiessl, S. Kucherenko, Noise-adjusted higher-order sensitivity indices of final repository models. EGU General Assembly 2019, Geophysical Research Abstracts Vol. 21, EGU2019-14005, https://meetingorganizer.copernicus.org/EGU2019/EGU2019-14005.pdf(https://meetingorganizer.copernicus.org/EGU2019/EGU2019-14005.pdf), (2019).
[3] B. Sudret, Global sensitivity analysis using polynomial chaos expansions. Reliability Engineering and System Safety, 93: 964–979, (2008).
[4] Q. Shao, A. Younes, M. Fahs, T-A. Mara, Bayesian sparse polynomial chaos expansion for global sensitivity analysis. Computer Methods in Applied Mechanics and Engineering, 318: 474-496 (2017).
[5] S. M. Spiessl, S. Kucherenko, D.-A. Becker, O. Zaccheus, Higher-order sensitivity analysis of a final repository model with discontinuous behaviour. Reliability Engineering and System Safety, doi: https://doi.org/10.1016/j.ress.2018.12.004(https://doi.org/10.1016/j.ress.2018.12.004), (2018).
[6] E. Plischke, An effective algorithm for computing global sensitivity indices (EASI). Reliability Engineering and System Safety, 95: 354–360, (2010).

Sabine Spiessl
sabine.spiessl@grs.de
Gesellschaft für Anlagen- und Reaktorsicherheit (GRS)
Braunschweig, Germany

Sergei Kucherenko
s.kucherenko@imperial.ac.uk
Imperial College London
London, UK

A new intrusive Monte-Carlo solver for the resolution of the uncertain (non)linear Boltzmann equation: application to sensitivity analysis to deal with nuclear waste

Many physical or economical applications rely on Monte-Carlo codes to solve deterministic partial differential equations (PDEs). This is the case for example for (non-exhaustive list) neutronics [1,8], photonics [2], plasma physics [3], social economics [4] etc. In neutronics/photonics for example, the Monte-Carlo resolution implies the sampling of the physical variables: for the position, for the velocity and for the time. The simulations are generally costly but the MC resolution is competitive with respect to other methods due to the high dimensional (x(3)+v(3)+t(1)=7) deterministic problem. The numerical parameter to control the accuracy of the code is Nmc the number of MC particles. The more is important, the more accurate the result is. The convergence rate obeys the central limit theorem: the error is O(1/sqrt(Nmc)).

Obviously, performing a sensitivity analysis with respect to different parameters X in R^d is also of great interest and importance in every of the aforementioned applications (initial/boundary conditions, but above all matter parameters such as cross-sections, reaction rates etc.). In fact, in our physical applications (neutronics, photonics mainly), we even would like to be able to perform systematic sensitivity studies to make sure we focus our future efforts on the reduction of the variability of the relevant input(s) among the components of . As a consequence, we very often face a 7+d dimensional problem.

Non-intrusive methods are usually applied (use of a black box code solving the physics), in particular to estimate Sobol indices with the efficient integration scheme described in [5]. But it demands a high number of code evaluations. In our MC resolution context, each one of them is already costly. One accurate run can take several hours on several hundreds of processors.

When applying the non-intrusive method of [5] to compute the Sobol indices of an output of an MC simulation code, basically, the physical space (x,t,v) and the uncertain space (X) are both explored thanks to two different MC experimental designs. The first one has Nmc particles to explore the space relative to the physical variables (x,t,v) , the second one has N runs for the space relative to the uncertain variable X. The number of MC particle is generally of about Nmc=10^6-10^9 for a tractable accuracy, see [1]. The number of runs is generally N=(d+2)n where n=10^3-10^6, see [5,6]. In this non-intrusive context, the two MC samplings are tensorised in the sense we finally treat and process Nmc N = 10^9-10^{15} particles. A sensitivity study is consequently costly and sometimes not even accurate enough.

The main idea of the present work comes from the fact that MC experimental designs should allow avoiding the tensorisation of the Nmc particles and N runs. For this, we need to be able to sample the whole space relative to (x,t,v,X) with the same MC design. This implies sampling the uncertain parameters X within the code, hence the intrusiveness of the approach. But we also need to make sure the operation we introduce within this code are still consistent with the uncertain PDE we aim at solving. In [7], we show a consistent and convergent modification of the code is possible. The paper presents a new generalised Polynomial Chaos (gPC) based intrusive MC solver to tackle sensitivity analysis problems for the linear Boltzmann equation. The new scheme implies only one run to compute the Sobol indices. The cost of this run is only slightly more costly than a classical one. The new solver shows important gains (factor from 4 to more than 60 in term of restitution times for equivalent accuracies on the Sobol

indices). It even allows performing studies in high dimensions (7+d=7+6=13) which were not achievable non-intrusively.

The aim of the paper/talk is two-fold:
We first present the method described in [7], used to solve a sensitivity analysis problem for the uncertain linear Boltzmann equation. The intrusive method only needs simple and localized modifications of an already existing MC simulation code. It furthermore allows important gains (by avoiding the tensorisation) but remains slightly sensitive to the dimension (as it is still gPC based). The solver now allows performing several sensitivity analysis and tackle robust optimisation problems on the uncertain linear Boltzman equation (see [7]).

Second, we present our first results for the sensitivity analysis of the uncertain Boltzmann equation coupled to the Bateman system. This nonlinear coupling (described in [1]) is intensively used to quantify the amount of nuclear waste (isotopic depletion, see [8] for example). This part put together two recent ingredients described in [1] and [7]. It also ensures a proof of concept the methodology presented in [7] can be applied easily to any MC simulation code (for linear or nonlinear problems).

The conditions for the efficient coupling and development of the solver in an MC code will also be investigated, interpreted and presented.

References

[1] Adrien Bernede and Gaël Poëtte. An Unsplit Monte-Carlo scheme for the resolution of the linear Boltzmann equation equation coupled to (stiff) Bateman equations. Journal of Computational Physics, 354:211 – 241, 2018
[2] Allan B. Wollaber. Four decades of implicit monte carlo. Journal of Computational and Theoretical Transport, 45(1-2):1–70, 2016
[3] C. Benedetti, A. Sgattoni, G. Turchetti, and P. Londrillo. ALaDyn: A High-Accuracy PIC Code for the Maxwell Vlasov Equations. IEEE Transactions on Plasma Science, 36(4):1790–1798, 2008
[4] L. Pareschi, P. Vellucci, and M. Zanella. Kinetic models of collective decision-making in the presence of equality bias, Physica A
[5] Andrea Saltelli. Making best use of model evaluations to compute sensitivity indices. Computer Physics Communications, 145(2):280 – 297, 2002
[6] Bertrand Iooss and Paul Lemaître. A Review on Global Sensitivity Analysis Methods, pages 101–122. Dellino, Gabriella and Meloni, Carlo, Springer US, Boston, MA, 2015
[7] Gaël Poëtte. A gPC-intrusive Monte-Carlo scheme for the resolution of the uncertain linear Boltzmann equation. Journal of Computational Physics, 385:135 – 162, 2019
[8] Jan Dufek, Dan Kotlyar, and Eugene Shwageraus. Numerical stability of the predictor-correct method in Monte Carlo burnup calculations of critical reactors. Annals of Nuclear Energy, 56:3

Gael Poette
gael.poette@gmail.com
CEA

Poster Session 2: Application of sensitivity analysis to environmental and nuclear issues or with temporal and/or spatial inputs/outputs

The Application of Uncertainty and Sensitivity Analysis in Bottom-Up Building Stock Energy Models

As cities world-wide seek to reduce emissions of greenhouse gases to meet international obligations and to mitigate the effects of changing climates to address local concerns about health and well-being, tools are needed to compare the effects of different interventions and assess the impacts of different development scenarios. Building-stock energy models incorporate representations of large numbers of individual buildings in order to create a model of a neighbourhood, entire city or region . In contrast to data-driven approaches which use statistical and machine learning techniques to relate overall energy consumption to characteristics of the building stock, this study focusses on building-stock energy models which are based on the aggregation of simulated energy flows for individual buildings and can be described as bottom-up, building physics based models. Existing discussions of uncertainty and sensitivity analysis in building-stock energy models

Uncertainty (UA) and sensitivity analysis (SA) of models for individual buildings is a well explored topic. In particular, Tian (2013)presents a detailed review of the subject, together with recommendations for appropriate methods for different problem settings. Naber et al. (2017)extend the review to encompass models at an urban or national scale, and note a limited uptake of UA/SA methods. However, Naber et al. do not perform a systematic review nor offer any detailed analysis of the methods used and their appropriateness. Therefore, there is a need for a review of the application of UA/SA to building-stock energy models, together with a need to determine how the form of the model in question – quasi steady state, dynamic, etc. – might affect the application of UA/SA techniques.

Aim of this paper
The aim of our present paper then, is to review the existing literature on the application of UA/SA to building-stock energy models, consider the forms of model encountered and the UA/SA methods used. Based on this, a research programme is then set out to support the application of UA/SA to building-stock energy models.

Method
A review was undertaken of journal and conference publications together with book chapters to explore UA/SA approaches employed to date, using the database Scopus. Given the relatively recent emergence of these models, the period of review was limited to publications from 2010 onwards. A total of 570 publications were identified, since this exceeded the resources available for review, the top 100 publications by citation count were selected for review. After abstract screening 39 publications were retained. A screening of the full texts resulted in the exclusion of a further 15 publications, either because the model described was not a bottom-up, physics-based model, or because, although the abstract discussed application at scale, the case detailed in the publication referred only to a single building.

Results
Of the 24 publications that remained after the screening process, only 7 undertook some form of uncertainty analysis. This means that over two-thirds of do not include an assessment of

the bounds of inference of the model presented. Our findings are therefore in agreement with the more general results of the systematic review undertaken by Saltelli et al.(2019)which found three quarters of publications selected through a search focussed on sensitivity analysis did not contain an uncertainty analysis.

Our corresponding findings for sensitivity analysis are similar to those by Shin et al. (2013)in their review of the application of sensitivity analysis in hydrological modelling who found 36% of the total papers mentioned some form of SA (46% in this review) but only 19% of those discussing SA actually undertook some form of analysis and only half used robust methods. In this study 33% of the papers which met the modelling criteria applied some form of SA of which only a quarter were robust techniques.

Conclusions

We conclude that Uncertainty Analysis (UA) and Sensitivity Analysis (SA) are not common practice in building-stock energy modelling and that if UA and SA are performed, only a few parameters are assessed and that the use of standardized methodologies is underrepresented. To foster UA and SA we outline a dedicated research programme being undertaken under the auspices of the IEA EBC Annex 70. This research programme aims to: (i) highlight the benefits and the costs of different UA and SA approaches, (ii) generate indications which types and sets of variables, input data and modelling approaches are relevant for different types of model output (e.g. energy use and consumption, emissions, costs and benefits), differentiating by use case of the model and spatial and temporal resolution, and (iii) provide guidelines about good and best practice on how to perform UA and SA in different cases. This should raise awareness about the relevance of uncertainties, give modellers and their clients a guideline on how to prioritize UA and SA depending on the use case and enable them to state the impact of uncertainties at least in a rough semi-quantitative manner.

References

[1] Kavgic, M., Mavrogianni, A., Mumovic, D., Summerfield, A., Stevanovic, Z., & Djurovic-Petrovic, M. (2010). A review of bottom-up building stock models for energy consumption in the residential sector. Building and Environment, 45(7), 1683–1697. https://doi.org/10.1016/j.buildenv.2010.01.021

[2] Naber, E., Volk, R., & Schultmann, F. (2017). From the Building Level Energy Performance Assessment to the National Level: How are Uncertainties Handled in Building Stock Models. Procedia Engineering, 180, 1443–1452. https://doi.org/10.1016/j.proeng.2017.04.307

[3] Saltelli, A., Aleksankina, K., Becker, W., Fennell, P., Ferretti, F., Holst, N., … Wu, Q. (2019). Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices. Environmental Modelling & Software. https://doi.org/10.1016/j.envsoft.2019.01.012

[4] Shin, M.-J., Guillaume, J. H. A., Croke, B. F. W., & Jakeman, A. J. (2013). Addressing ten questions about conceptual rainfall–runoff models with global sensitivity analyses in R. Journal of Hydrology, 503, 135–152. https://doi.org/10.1016/j.jhydrol.2013.08.047

[5] Tian, W. (2013). A review of sensitivity analysis methods in building energy analysis. Renewable and Sustainable Energy Reviews, 20, 411–419. https://doi.org/10.1016/j.rser.2012.12.014

Pamela Fennell
pamela.fennell@ucl.ac.uk
University College London

Paul Ruyssevelt
p.ruyssevelt@ucl.ac.uk
University College London

Erika Mata
erika.mata@ivl.se
IVL Swedish Environmental Research Institute

Martin Jakob
martin.jakob@tep-energy.ch
TEP Energy GmbH

Treatment of Uncertainties with Respect to Geomechanical Modelling to Provide Evidence of the Integrity of the Geologic Barrier of a Radioactive Waste Repository in Rock

Within the framework of the plan approval procedure for the closure of the Morsleben Repository , a safety case is currently being prepared. Part of the safety case is the modelling of the geologic barrier. In Germany, it has been required by law since 2010 to carry out a sensitivity and uncertainty analysis of the models used within the safety case for a repository of high-level radioactive waste in deep geologic formations. The same law postulates a shift of the focus of the safety case from containment ("methods or physical structures designed to prevent or control the release and the dispersion of radioactive substances" (IAEA: Specific Safety Guide No. SSG-23, Vienna 2012)) towards the isolation of the radionuclides from the biosphere. Isolation is to be guaranteed by a multi-barrier system consisting of the geologic and of engineered barriers. This is the reason why the models of the geologic barrier are to provide evidence of not only structural stability of the disposal facility but now more importantly to provide evidence of the integrity of the geologic barrier.

In the case of rock salt, the integrity of the geologic barrier is affected, if certain geomechanical loading states and induced mechanisms apply, which leads to damage and loss of the initial tightness of the rock salt. These integrity-relevant processes are:
* the fluid-pressure-driven opening of grain boundaries, if the pressure exceeds the normal stress and adhesive forces at the boundaries, and/or
* generation and growth of (interconnected) cracks due to deviatoric loading.

To assess the geomechanical loading of the geologic barrier and the barrier's short and long-term evolution, numerical calculations are carried out. As a result of these calculations, a statement on the structural stability and barrier integrity can be given, using the conservative fluid-pressure-criterion (minimum stress criterion) aiming at mechanism (1) and the dilatancy criterion aiming at shear stress-induced damage processes (2). These results are influenced by model assumptions, simplifications as well as by data-, parameter-, and model-uncertainties and variabilities.

To treat these uncertainties of geomechanical modelling, a stepwise approach was used to show the range of possible solutions and to identify relevant issues and parameters. First, a catalogue of uncertainties was compiled after systematic screening of the whole modelling sequence from input data to output. The uncertainties and possible relevant issues identified cover categories, i.e. mine and geological/geomechanical model, backfill-planning, numerical code and model used, including its simplifications, boundary- and initial conditions, constitutive models and model parameters. These uncertainties were quantified and bandwidths were assigned. In a second step, the impacts of individual uncertainties on the barrier integrity were analysed by means of bandwidth studies. These studies use several generic models, each representative of specific conditions in different mine areas. The calculations evaluate the long-term evolution of the geomechanical state within the rock salt barrier, e.g., the evolution of the thickness of the intact rock salt barrier. This allows to identify relevant uncertainties and to quantify the impact of each uncertainty on the barrier integrity.

For each generic model, a reference parameter set was established first and a reference run was performed. Based on the catalogue of uncertainties, a different number of variants from the reference run resulted for each generic model. The majority of these variants contain one change in the parameter set compared with the reference parameter set. However, there are

also some variants that contain multiple changes in the parameter set. Time-dependent criteria that represent the geomechanical state are evaluated for each variant and then compared with the reference run.

In a next step, the relevance of these issues will be checked on more realistic, comprehensive location-specific models by additional safety analyses. If the calculations show a relevant impact of uncertainties on the barrier integrity, further iterative steps will become necessary to check if these uncertainty bandwidths can be narrowed down or can be accepted. If necessary, additional measures to narrow down these uncertainties will have to be taken.

Johanna Wolf
johanna.wolf@bge.de
Bundesgesellschaft für Endlagerung (BGE), Peine / Federal Company for Radioactive Waste Disposal (Germany)

A Joint Exercise on Sensitivity Analysis in Repository Performance Assessment

Radioactive wastes have to be effectively isolated from the biosphere for a very long time. It is international consensus that repositories for such wastes should be established in appropriate disposal facilities to minimize the transport of radionuclides to the accessible environment. The safety-related long-term performance of radioactive waste repositories depends on a number of interacting features, events, and processes, including features of the engineered and natural barrier systems; geologic, climatic, and human-intrusion events; and coupled thermal, hydrological, mechanical, and chemical processes. Performance assessment (PA) simulations are used to determine if radionuclide releases will remain within the relevant regulatory limits, which may be defined differently depending on the host country, the repository concept, waste characteristics, site characteristics, etc. Numerical models used in PA usually calculate the release of radionuclides from the wastes, their transport through the system and, sometimes, their radiological effects in the biosphere. Such models are complex, coupled, and may be highly nonlinear. They are subject to aleatory, epistemic, and model uncertainty. The number of uncertain inputs can be in the hundreds, and outputs of interest (often concentration or dose) can span several orders of magnitude. Therefore, uncertainty and sensitivity analysis is a substantial part of repository PA.

Motivated by exchanges between the European Union's Implementing Geological Disposal of radioactive waste Technology Platform (IGD-TP) and NEA's Integration Group for the Safety Case (IGSC), since 2015 a collaboration of several waste management, consultant and research organisations has evolved, addressing several issues related to the treatment of uncertainties. In 2017, a sensitivity analysis (SA) exercise was initiated, aimed at identifying advantages and drawbacks of various SA methods vis-à-vis the characteristics of the model(s) under scrutiny and at exploring the possibilities for joint or shared approaches to SA dependent on the models' specifics. The cooperation started with participants from USA and Germany, and was later joined by groups from Belgium, Finland, Russia, and Switzerland.

Global sensitivity analysis has been a widely applied tool in repository performance assessment for several decades. Methods commonly applied are based on linear and rank regression and (partial) correlation [1]. Recent advances in computer hardware and SA methodology have made feasible more computationally expensive methods (e.g., Sobol variance decomposition) that do not presuppose linearity or monotonicity, and some application to repository PA has been made (e.g., [2, 3]). A primary challenge is choosing SA methods that capture the behaviour of a complex, coupled repository system without requiring a prohibitive number of evaluations of computationally expensive numerical models.

In this exercise, we apply a variety of "common" and "new" SA methods to repository system models selected from among the models in use by participants. For this initial phase of the comparison, we limit ourselves to analysing the sensitivity of outputs to epistemic uncertain inputs, and we consider the models themselves as "black boxes", mathematically representing some mapping from the input variable space to output variable space. Application of a variety of SA methods allows us to address questions regarding the contribution (or lack thereof) of each input to the output uncertainty, interactions between input variables and effects on output uncertainty, and general model behaviour (linearity or lack thereof, monotonicity etc.), as well as to assess which SA methods alone or in combination return the most information about which models.

As a first step, a questionnaire was developed in order to characterise the features of potential test cases. Based on the responses, it was then decided to address a series of models with increasing complexity with regard to the number of inputs, dependency of inputs and model behaviour:

* A generic repository for spent nuclear fuel (SNF) and high level waste in clay (GRS),
* A generic repository for commercial SNF in shale (SNL),
* A near-surface repository for low and intermediate level waste at Dessel/ Belgium(SCK CEN),
* A generic repository for low and intermediate level waste in a salt mine (GRS),
* A mechanistic fuel matrix degradation model (SNL),
* A generic repository for commercial SNF in fractured crystalline rock (SNL).

Most of the models generate time-dependent outputs. The sensitivity analysis methods applied comprise

* Scatterplots,
* CUSUNORO curves,
* Standardized Regression Coefficients (SRC),
* Standardized Rank Regression Coefficients (SRRC),
* SRC with bijective transformation,
* Pearson's Correlation Coefficients (CC),
* Spearman' Rank Correlation Coefficients (RCC),
* Partial Correlation Coefficients (PCC),
* Partial Rank Correlation Coefficients (PRCC),
* 1st order Sensitivity Index (SI1), calculated using PCE, EASI, CSN or RS-HDMR,
* 2nd order Sensitivity Index (SI2), calculated using Harmonic Regression or RS-HDMR,
* Total order Sensitivity Index (SIT), calculated using PCE or RS-HDMR.

The on-going exercise is expected to deepen understanding of when and how to use different SA methods most effectively and to yield a step towards a common approach to sensitivity analysis in repository PA. Although it is too early to evaluate the exercise as a whole, we present a few general and specific findings.

References

[1] Helton, J. C. and F. J. Davis, Sampling-Based Methods. In A. Saltelli, K. Chan, & E. M. Scott (Eds.), Sensitivity Analysis(pp. 102-153). Chichester, U.K.: Wiley (2000).
[2] Spiessl, S. and D.-A. Becker, Investigation of Modern Methods of Probabilistic Sensitivity Analysis of Final Repository Performance Assessment Models (MOSEL). GRS-412, Gesellschaft für Anlagen- und Reaktorsicherheit (GRS) gGmbH, Köln (2017).
[3] Cormenzana, J. L., Probabilistic Sensitivity Analysis for the "Initial Defect in the Canister" Reference Model. Working Report 2013-25. Posiva-Oy, Finland (2013).
[4] Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525. This paper is Sandia publication SAND2019-3872 A.

Dirk-Alexander Becker
Dirk-Alexander.Becker@grs.de
Gesellschaft für Anlagen- und Reaktorsicherheit (GRS)

Dusty Brooks
dbrooks@sandia.gov
Sandia National Laboratories (SNL)

Joan Govaerts
Joan.Govaerts@sckcen.be
Studiecentrum voor Kernenergie – Centre d'Etude de l'Energie Nucléaire (SCK-CEN)

Lasse Koskinen
Lasse.Koskinen@posiva.fi
Posiva Oy

Sergei Kucherenko
s.kucherenko@imperial.ac.uk
Imperial College, London

Pekka Kupianen
Pekka.Kupiainen@posiva.fi
Posiva Oy

Paul Mariner
Paul.Mariner@sandia.gov
Sandia National Laboratories (SNL)

Barbara Pastina
Barbara.Pastina@posiva.fi
Posiva Oy

Elmar Plischke
Elmar.Plischke@tu-clausthal.de
Technische Universität Clausthal (TUC)

Using Bayesian networks to represent distributions of successful model parameter combinations: Simulation of diatom growth in the Elbe River

Normal 0
21
false false false DE X-NONE X-NONE

Ecosystem models differ in complexity and the number of free parameters that are involved. Given a limited number of observations, it is usually impossible to strictly calibrate the full set of model parameters. This problem of model over-parametrization is well-known and has been identified in many studies. From the mathematical point of view, it might call for a reduction of the number of model parameters. Either some details addressed in the model could be discarded or model parameters could be combined into a smaller set of more abstract parameters. The severe disadvantage of parameter aggregation is that compound parameters may lose their process oriented interpretation. Even if the model successfully predicts variables of interest, it may be unable to explain why certain events are going to happen. A process-based interpretation, however, is important for any informed management action.

The method we propose in this study keeps the full set of model parameters, each of which has a clear meaning in the context of biological processes. However, instead of providing for each of these parameters its most probable value complemented by an error bar, the approach adheres to a description of the joint distribution of all parameters. Bayesian network (BN) technology is employed to describe inter-relationships between parameter values that enable a reasonable reproduction of existing observations. A large ensemble of successful simulations underlying the BN was generated based on Markov Chain Monte Carlo (MCMC) technique.

The example model we use to illustrate the approach predicts chlorophyll a concentrations at station Geesthacht Weir on the Elbe River in Germany. Concentrations at different instants of time are simulated independently from each other, extracting each chlorophyll concentration value at the endpoint of a Lagrangian trajectory of a water body that travelled down the river. Smoothness of data simulated at Weir Geesthacht results exclusively from the fact that the environmental conditions governing algae growth tend to persist for some time. According to the model, the key effects influencing algae (diatom) growth are those of changing light conditions, changing temperature and, during some periods, also lack of silica.
Although the model equations are simple - many processes like grazing by zooplankton, for instance, are not explicitly described - simulations agree surprisingly well with observations. Its success and its clear structure make the model a good candidate for demonstrating the approach we propose. Model simulations can be compared with data from five different years (1997-2001). An important question is to which extent model parameters might be controlled differently by observations in different years, leading to different interpretations of observed variability.

Seven parameters representing different aspects of algae growth were selected for the study. The model is fast enough to enable generation of 106 effective parameter combinations using the MCMC approach. This data set was then used to train the BN. We argue that the BN summarizes the outcome of model calibration in a most comprehensive way. Interrelationships between parameters implicitly represent effects of over-parametrisation. The interactive BN allows users to see how choosing a certain value for one specific parameter narrows or shifts the ranges of acceptable values for all other parameters. Our study provides

also an example of bimodal marginal distributions. Due to parameter interrelationships documented in the BN, this indicates existence of different concepts (e.g. dominant processes) that are able to explain available observations. The BN also documents which kind of additional knowledge would resolve an existing ambiguity.

Ulrich Callies
ulrich.callies@hzg.de
Helmholtz-Zentrum Geesthacht,
Geesthacht, Germany

Housam Dibeh
housam.dibeh@hzg.de
Helmholtz-Zentrum Geesthacht,
Geesthacht, Germany

Posterior Parameter Inference for Personalising Models of Individual Cardiac Myocytes

Cardiac function is determined by the integrated response of millions of myocytes working together. Myocyte contraction is triggered by an electrically-evoked calcium-transient. Since calcium ions are not normally transmitted between myocytes, each cell is individually responsible for its calcium handling. The ability to describe calcium dynamics at a single cell level is thus increasingly recognised as important for understanding cardiac physiology, pathology and response to treatment [1].

Myocyte calcium dynamics can be characterized in terms of the activity of five proteins that describe the movement of calcium into and out of the cell and within intracellular compartments. Variation and covariation of these key proteins account for the biological variability inherently present among myocytes and likely represent physiologically important common regulatory networks and substrates, respectively. Increases in variation in cardiac myocyte proteins, as occurs with age, can increase the risk of arrhythmia. Decreased covariation may result from a degradation in key regulatory networks. Variation and covariation represent novel, and potentially important, measures of cardiac function. Yet no method has been developed to systematically record them.

Mathematical models have been used to encode the biophysics of the proteins that regulate calcium dynamics and how these proteins interact to form the myocyte calcium handling system. These models depend on a number of parameters. If there is no uncertainty in the parameter values, then once the initial conditions are set, the model output is deterministic parameter uncertainty quantification directly relates to the experimental error and the variation and covariation in the activities of the proteins. However, conventional methods for interrogating cardiac physiology have focused on describing the mean behaviour of a 'representative cell'. This is obtained by fitting model parameters to mean experimental values, using different sets of cells for studying each protein. The canonical approach is justified by the experimental challenge associated with making multiple measurements in single myocytes, but it is crucially unable to measure the covariance between parameters.

The contribution of this work is the development of a statistical methodology that fully quantifies uncertainty in the parameters of the Hinch biophysical model [2] of calcium dynamics in single cardiac myocytes. The model consists of a set of coupled ordinary differential equations , with high-dimensional parameter vector (order of 40 parameters).

We work under the Bayesian inference paradigm, that treats the parameters as random variables, and aims at obtaining samples from their joint posterior distribution, given the data and prior information. The benefits deriving from performing Bayesian inference are two-fold. First, not only the parameter posterior variance but also their posterior covariance is obtained, overcoming the limitation of the existing methods; moreover, non-identifiability of some of the parameters from the data naturally results in high posterior variance, thus automatically conveying which part of the parameter vector is informed by the data.

There is not conjugacy between likelihood and prior, despite choosing a Gaussian distribution for both, because the ODEs solution is non-linear in the parameters; as a consequence, the posterior cannot be sampled from using direct techniques. Conventional Bayesian sampling algorithms, such as Markov chain Monte Carlo (MCMC), are inefficient because the likelihood appears to be very informative about low-dimensional aspects of the high-

dimensional parameter vector. This results in slow mixing of the Markov Chains, and impractically large values of the autocorrelation function.

When the likelihood is Gaussian, the gradients of the posterior are analytically expressed in terms of the sensitivities of the ODE solution to the parameter, which are, in turn, numerically obtained as solutions to an adjoint set of ODEs. Some improvement in the posterior sampling schemes is thus obtained by incorporating geometric information about the posterior, using gradient-based algorithms such as the Metropolis-Hastings Adjusted Langevin Algorithm (MALA) [3]. However, the outcome of the posterior samplers still appears to be dominated by the likelihood.

In order to increase the effectiveness of the sampling schemes, we thus perform dimensionality reduction of the parameter space, in a way that is informed by the likelihood, since this is the cause of poor mixing.

We follow the work of [4], that identifies a projection of the parameters on a low dimensional space, based on the (average) Fisher information matrix (FIM); this is also analytically expressed in terms of the model sensitivities when the likelihood is Gaussian. In particular, given the projection matrix, the posterior samplers now target a subset of the coefficients of the linear transformation of the parameter vector that have smaller variability.

The 'effective dimensionality' of the problem based on this method appears to be of the order of 10-15 parameters. This procedure does introduce some bias, because the new target is a reduced posterior distribution; however, MCMC and MALA samplers have better mixing properties, and the bias is compensated by lower variance when re-projecting on the full parameter space.

The outcomes of our research are (i) a robust and effective method for approximate sampling from the posterior in the challenging high-dimensional Hinch model and (ii) the scientific conclusions that this enables us to make about calcium-dynamics at a single cell level. This talk focuses on the first aspect, and, more broadly, on how our investigation sheds light on the important question of how to exploit dimension reduction methods in Bayesian computation.

References

[1] Altschuler, SJ and Wu, LF. Cellular heterogeneity: do differences make a difference? Cell, 141(4):559–563, 2010.
[2] Hinch, R, Greenstein, JL, Tanskanen, AJ, Xu, L and Winslow, RL. A simplified local control model of calcium-induced calcium release in cardiac ventricular myocytes. Biophysical Journal, 87(6):3723–3736, 2004
[3] Girolami, M and Calderhead, B. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. Journal of the Royal Statistical Society: SeriesB (Statistical Methodology), 73(2):123–214, 2011.
[4] Cui, T, Martin, J , Marzouk, YM, Solonen, A, and Spantini, A . Likelihood-informed dimension reduction for nonlinear inverse problems.Inverse Problems, 30(11):114015, 2014.

Marina Riabiz
marina.riabiz@kcl.ac.uk
King's College London
Alan Turing Institute
Jon Cockayne
J.Cockayne@warwick.ac.uk

University of Warwick
Alan Turing Institute

Pawel Swietach
pawel.swietach@dpag.ox.ac.uk
Oxford University

Mark Girolami
mag92@eng.cam.ac.uk
Cambridge University
Alan Turing Institute

Steven Niederer
steven.niederer@kcl.ac.uk
King's College London

Chris Oates
chris.oates@newcastle.ac.uk
Newcastle University
Alan Turing Institute

Sensitivity and Uncertainty Analysis for Geochemical Models for Repository Safety assessment

A key component of safety assessment for radioactive waste repositories in deep geological formations is the simulation of potential radionuclide release scenarios and the transport of radionuclides through the repository system. The realistic modelling of geochemical processes is of high relevance for assessing the migration of radionuclides in groundwater systems. One important retardation process for radionuclides to be considered is sorption onto mineral surfaces of the host rock/sediments. Due to the heterogeneities of the natural system and the high complexity of the geochemical models with uncertain and varying input parameters, it is important to understand which of the input parameter uncertainties have considerable influences on the model output uncertainty. This in turn allows model reduction required to upscale from the molecular to the plug scale. Consequently, we performed sensitivity and uncertainty analysis (SA/UA) to investigate and understand our geochemical model.

For the quantification of the contaminant retention in groundwater, the solid/liquid distribution coefficients (Kd-values) calculated for a given groundwater/rock system are traditionally used. Most often conventional concepts with constant Kd-values are applied in reactive transport simulations. Such an approach has the advantage to be simple and computationally fast but cannot reflect changes in geochemical conditions that will occur during the evolution of the repository system, e.g. due to climatic changes. Due to the German safety criteria with an assessment period of about 1 million years it is necessary to consider the impact of such geochemical changes on the radionuclide transport and retardation.

For this, we developed a new approach, where the smart Kd-concept (www.smartkd-concept.de) [1] is modified in complex geochemical models including mechanistic sorption models, and implemented it in reactive transport calculations [2, 3]. Possible migration scenarios for repository-relevant radionuclides (isotopes of Am, Cm, Cs, Ni, Np, Pu, Ra, Se, Th and U) through a typical sedimentary rock system covering potential repository host rocks, namely salt and clay formations in Northern Germany as natural geological barrier, were developed in the first stage.

Smart Kd-values and their associated sensitivities and uncertainties were computed for a wide range of important geochemical input parameters/boundary conditions such as pH value, ionic strength, concentration of competing cations and complexing ligands, e.g. calcium and dissolved inorganic carbon. Our toolbox coupled the geochemical speciation code PHREEQC [4] with the numeric tool UCODE [5] and SimLab2.2/4 [6]. SimLab has the advantage to permit a simultaneous variation of all input parameters according to their probability density functions (pdfs) and mutual correlations. It provides (in contrast to UCODE, which also incorporates some simple SA/UA algorithms) methods for Global SA. Comparable SA/UA have been done with a new software package RepoSUN, which is based on SimLab [7]. For the varying parameters uncertainty intervals were defined from field investigations and log-uniform distributions were assumed for all parameters, except for the pH value, which was assumed to be uniformly distributed. The results, i. e. the smart Kd-values, were analyzed in the following way:

* As the output distribution covers several orders of magnitude, a log-transformation was performed on the output data. Then the computed a-priory multidimensional smart Kd-matrices are accessible for interpolations during subsequent transport simulations.
* A histogram of the model output visualizes the distribution.

* Some statistical measures were calculated to characterize the distribution (minimum, maximum, mean, standard deviation).
* Sensitivity measures were calculated for each of the input parameters: the standardized regression coefficient (SRC), the standardized rank regression coefficient (SRRC) and the variance-based first-order sensitivity index (SI1).

On the basis of the results it could be shown that the smart Kd-approach goes considerably beyond the conventional concepts. We can demonstrate that constant Kd-values used in previous transport simulations [8] are a crude assumption, as in reality they rather range over several orders of magnitude. Moreover, with the results from the SA, the most important input parameters influencing the radionuclide retardation can be identified (key parameters of the model). The calculated sensitivity indices allowed us to assess the most and less sensitive parameters. SA is a useful means for reducing the complexity of a geochemical model by focusing on the most important input parameters.

Acknowledgement:
This work was funded by the German Federal Ministry for Economic Affairs and Energy (BMWi) under grant Nos. 02 E 10518 and 02 E 10528 as well as grant Nos. 02 E 11072A and 02 E 11072B.

References

[1] Stockmann, M., Schikora, J., Becker, D.-A., Flügge, J., Noseck, U., Brendler, V. (2017). Chemosphere, 187: 277-285.
[2] Noseck, U., Brendler, V., Flügge, J., Stockmann, M., Britz, S., Lampe, M., Schikora, J., Schneider, A. (2012). Report GRS-297, Braunschweig.
[3] Noseck, U., Brendler, V., Britz, S., Stockmann, M., Fricke, J., Richter, C., Lampe, M., Gehrke, A., Flügge, J. (2018). Report GRS-500, Braunschweig.
[4] Parkhurst, D.L., Appelo, C.A.J. (2013). U.S. Geological Survey Techniques and Methods, Book 6, Chap. A43, Amsterdam.
[5] Poeter, E.P., Hill, M.C., Lu, D., Tiedemann, C.R., Mehl, S. (2014). U.S. Geological Survey and Methods, Report GWMI2014-02.

Madlen Stockmann
m.stockmann@hzdr.de
Helmholtz-Zentrum Dresden-Rossendorf e.V. (HZDR)
Institute of Resource Ecology
Germany

Dirk-Alexander Becker
Dirk-Alexander.Becker@grs.de
Gesellschaft für Anlagen- und Reaktorsicherheit (GRS) gGmbH
Braunschweig, Germany

Ulrich Noseck
Ulrich.Noseck@grs.de
Gesellschaft für Anlagen- und Reaktorsicherheit (GRS) gGmbH
Braunschweig, Germany

Vinzenz Brendler
v.brendler@hzdr.de

Helmholtz-Zentrum Dresden-Rossendorf e.V. (HZDR)
Institute of Resource Ecology
Germany

The principle and potential of ensemble weather forecasting through different decision support tools used in agronomy. Sensitivity analysis results of a crop and grape worm simulation model to its internal and external parameters

Numerical modelling is very useful in agricultural decision-making. However model approximations and uncertainties in model parameters and external forcings can have a significant impact on the output, which is usually a processing date. Reliable information on the treatment date is important for the agricultural sector in order to maintain good pesticide consumption practice. Avoiding early treatment is necessary to reduce crop protection products while late treatments will cause damage and loss in field crops. For crop diseases and pests, many decision support systems are proposed. They aim to take into account the meteorological conditions in the modeling of epidemic dynamics to assist farmers in their decision-making. For example, in France, the Septo-LIS® tool is proposed for positioning the first treatment on wheat septoria and the EVA model (adapted from Chavent, 1983) can be used to fix the date of treatment on the grape worm.

Crop management is highly dependent on weather conditions. In addition to past and actual data, these models use more and more weather forecasts. Nevertheless, the atmospheric flow is chaotic and weather forecasts remain imperfect and uncertain. In recent years, operational weather prediction centers, including French meteorological service (Météo-France), have implemented probabilistic prediction systems that provide an estimate of the uncertainty of the weather forecasts. Probabilistic prediction consists in a set of perturbed forecasts, called ensemble forecasts, that account for the different sources of uncertainty. For example, three ensemble prediction systems (EPS) that cover different time and length scales are operationally used by Météo-France, providing forecasts from hours to several weeks. To our knowledge, ensemble weather forecasts are not yet used in agricultural applications.

The uncertainty of weather forecasts is a major source of variation in the agronomy model outputs. Sensitivity analysis is then used to identify the most influential external meteorological parameters as well as internal influential parameters that can influence the farmer's decision. The purpose of this study is to present the sensitivity analysis results of a crop and grape worm simulation model to its internal and external parameters. The aim is to quantify the relative impact of the three main sources of uncertainties: agronomic model parameters, past weather observations and future forecasts. Since model parameters are numerous the most influental one are identified with standard sensitivity analysis methods such as the Morris method for global sensitivity analysis. Once the most influential parameters are selected, we quantified (within a Sobol' scheme) the impact of these parameters on the processing date by taking into account different past weather scenarios. Then, we do sensitivity analysis studies on future weather forecasts and quantify the impact over the agronomy model output.

This communication will present the principle of ensemble weather forecasting and the potential of using these probabilistic forecasts through different decision support tools used in an operational way for issues in wine and wheat. In the second part we will present the sensitivity analysis results that quantify the relative impact of the three type of uncertainties.

Ivana Aleksovska
ivana.aleksovska@acta.asso.fr
INRA, Météo-France, ACTA

Laure Raynaud
laure.raynaud@meteo.fr
INRA, Météo-France, ACTA

Robert Faivre
robert.faivre@inra.fr
INRA, Météo-France, ACTA

François Brun
Francois.Brun@Acta.Asso.Fr
INRA, Météo-France, ACTA

Flow prediction in ungauged catchments using probabilistic random forests regionalization and new statistical adequacy tests

Flow predictions in ungauged basins (PUBs) remains an elusive challenge in hydrological sciences and engineering, even with the advances achieved during the ‘‘PUB decade’’. Meeting this challenge is made difficult by the uncertainty in the ‘‘regionalization’’ model used to transpose hydrological data (e.g., flow indices) from gauged to ungauged basins, and by the uncertainty in the hydrological model used to predict streamflow in the ungauged basin.

In this work we attack the PUB problem by exploiting recent advances in flow index selection and Bayesian / machine-learning regionalization methods, and by proposing two statistical metrics, DistanceTest and InfoTest, to assess the adequacy of a model before estimating its parameters. DistanceTest quantifies whether a model (hydrological or regionalization) is likely to reproduce the available hydrological information in a catchment. InfoTest is based on Bayes Factors and quantifies the information added by a model (hydrological or regionalization) over prior knowledge about the available hydrological information in a catchment). Our study demonstrates that the proposed model adequacy tests can be treated as a prerequisite for a model (hydrological or regionalization) being considered capable of providing meaningful and high quality flow time series predictions in ungauged catchments. If a model is found inadequate a priori and rejected, the modeler is spared the effort in conditioning the model parameters, which can be a substantial saving when the conditioning is implemented using computationally expensive Monte Carlo techniques. The ability to diagnose model adequacy a priori can help to identify the main sources of predictive error.

First, an extensive set of available flow indices is reduced using principal component (PC) analysis to a compact orthogonal set of ‘‘flow index PCs’’. These flow index PCs are regionalized under minimal assumptions using Random Forests regression augmented with a residual error model, and used to condition hydrological model parameters using a Bayesian scheme. Second, ‘‘adequacy’’ tests are proposed to evaluate a priori the hydrological and regionalization model performance in the space of flow index PCs.

We report an empirical case study based on 92 catchments in northern Spain. Results show that: (1) a small number of principal components capture approximately 87% of variability in the flow indices; (2) adequacy tests with respect to regionalized information are indicative of (but do not guarantee) the ability of a hydrological model to predict flow time series in ungauged catchments. The adequacy tests identify the regionalization of flow index PCs as adequate in 12 of 16 catchments, but the hydrological model as adequate in only 1 of 16 catchments.

Overall, the case study results suggest that the hydrological model is the main source of uncertainty in comparison to the regionalization model, and hence that priority should be given to the improvement of the hydrological model structure and input data (the effects of which are not disaggregated in this work).

Cristina Prieto Serra
prietoc@unican.es
Environmental Hydraulics Institute ‘‘IHCantabria’’
Universidad de Cantabria
Santander, Spain

Nataliya Le Vine
Nataliya_LeVine@swissre.com

Swiss Re, Armonk
NY, USA

Dmitri Kavetski
dmitri.kavetski@adelaide.edu.au
School of Civil, Environmental and Mining Engineering
University of Adelaide
Australia

Eduardo García
edelwar@gmail.com
Environmental Hydraulics Institute "IHCantabria"
Universidad de Cantabria
Santander, Spain

César Alvarez
alvarezc@unican.es
Environmental Hydraulics Institute "IHCantabria"
Universidad de Cantabria
Santander, Spain

Raúl Medina
medinar@unican.es
Environmental Hydraulics Institute "IHCantabria"
Universidad de Cantabria
Santander, Spain

Better understanding the dynamics of a green roof model using GSA
Context and problem Statement

Within the last ten years, green roofs have become a very important component of sustainable urban development. Growing environmental awareness as well as striking economical and ecological advantages are the main driving forces for their increasing success. Green roof means that building is covered with vegetation and a growing medium, both typically planted over a waterproofing membrane. It may also include additional layers such as a root barrier and drainage and irrigation systems. Green roofs have many benefits mainly to the stormwater runoff reduction but also to the life expectancy and energy balance of a building. However, as pointed out in [1], a green roof hydrologic performance is difficult to predict and control, due to a lack of methodologies to perform quantitative analyses. Models describing the water balance, based on unsaturated flow in porous media, have been proposed in the literature [2]. According to them, the water retention capacity of a green roof depends on several factors as the physical properties of the growth media, the initial water content, the vegetation types, the weather condition as the temperature, the relative humidity, the solar radiation, etc. These factors are uncertain due to measurements, expert judgements or simply insufficient knowledge.

Some studies have been carried out to evaluate the impact of the different factors on the water retention capacity. In general, the studies focus on one particular factor at a time, such as the precipitation, the initial water content, the depth of media or the plant species. They consist in testing different scenarios for the factor and evaluating the corresponding water retention capacity. A review of the results can be found in [2]. However, because of the wide ranging results, it is not possible to specify an optimal green roof build-up that will maximise green roof hydrologic benefits in any country or climate zone. Research is therefore necessary to help select green roof materials in different areas. Sensitivity Analysis (SA) can serve in better understanding of the dynamic behavior of green roofs, and in facilitating design processes.

Regarding the green roof model, sensitivity analyses have been performed but only for thermal performance purposes [3, 4]. The aim of this study is to perform global sensitivity analysis of a green roof model, and to assess the most influential factors on water retention capacity of a green roof located in a place presenting a meso-climate, semi-oceanic with a continental degraded marked influence. The results of this study can help better understanding the dynamics of the moisture content (model output), taking into account uncertainties of 6 parameters characterizing the growing medium: the residual moisture content, the moisture content and the hydraulic conductivity at saturation, the porosity.

Green roof model

The moisture content is described using the Richard's equation, partial derivative equation combined with the Mualem model and Van Genuchten relation [2]. The Richard's equation is simulated using Hydrus-1D software, a finite-element modeling environment simulating water flows in variably saturated porous media. This work is based on an in situ experimental green roof plot settled in Tomblaine, north-east of France, 4840ʾN 613ʾE.

Strategy for Sensitivity analysis of the green roof model

In the cases of computationally expensive models, the metamodelling technique which maps inputs and outputs is a very useful and practical way of making computations tractable.

To compute the Sobol' sensitivity indices, SobolGSA software, a general purpose metamodeling software [5, 6, 7, 8], has been used.The procedure applied to compute the sensitivity indices is the following:

* Generate the samples for each parameter using Quasi Monte Carlo sampling implemented in SobolGSA software. The 6 uncertain parameters follow a uniform distribution.
* Simulate the model running Hydrus-1D for the different samples in order to generate the corresponding moisture content.
* Compute the sensitivity indices using SobolGSA software based on Random Sampling-High dimensional model representation (RS-HDMR) method.
* Tests have been made with different sample sizes (128, 512 and 1024 samples) in order to analyse the convergence of the sensitivity indices. It appears that 512 samples are sufficient.

To conclude, the moisture content at saturated conditions has been found to have the highest sensitivity index, explaining almost 78% of the moisture content variance. It is followed by some characteristic parameter of the medium. This is due to an important macro porosity of the growing medium, facilitating the water drainage. The sensitivity indices of the other parameters are small showing that they are not influential and can be neglected.

References

[1] Y. Li and Jr. RW Babcock. Green roof hydrologic performance and modeling: a review. Water Science and Technology, 69(4):727-38, 2014.
[2] A. E. Taigbenu. The green element method. Springer Boston, MA, 2007.
[3] H. T. Rakotondramiarana, T. F. Ranaivoarisoa, D. Morau. Dynamic Simulation of the Green Roofs Impact on Building Energy Performance, Case Study of Antananarivo, Madagascar. Buildings, 5:497-520, 2015.
[4] A. Pianella, L. Aye, Z. Chen, N. Williams. Effects of design and operating parameters on green roof thermal performance in Melbourne. Proceedings of the 5th International Conference on Zero Energy Mass Customised Housing, Kuala Lumpur, 2016.
[5] SobolGSA software, download at http://www.imperial.ac.uk/process-systems-engineering/research/free-software/sobolgsa-oftware/(http://www.imperial.ac.uk/process-systems-engineering/research/free-software/sobolgsa-software/)
[6] I. Sobol', S. Kucherenko. Global Sensitivity Indices for Nonlinear Mathematical Models. Review, Wilmott, 1 (2005) 56-61.
[7] I. Sobol', S. Kucherenko, Derivative based Global Sensitivity Measures and their link with global sensitivity indices, Mathematics and Computers in Simulation, V 79, Issue 10, (2009) 3009-3017
[8] I. Sobol', S. Kucherenko. A new derivative based importance criterion for groups of variables and its link with the global sensitivity index Computer Physics Communications, V. 181, Issue 7, (2010) 1212-1217.

Collin Floriane
floriane.collin@univ-lorraine.fr
University of Lorraine
Sergei Kucherenko
s.kucherenko@imperial.ac.uk
Imperial College of London

Claverie Rémy
remy.claverie@cerema.fr
CEREMA

Global sensitivity analysis based on covariance decomposition: new approach and results

Models with dynamic outputs are widely used in practical engineering. Global sensitivity analysis of these models can provide useful information for risk assessment and decision making. Recently, many studies on dynamic systems use a multivariate GSA method based on covariance decomposition .

In this method, a scalarization of the equation of covariance decomposition is made by taking the trace of each term to obtain a set of useful sensitivity indices. However, there also exist other approaches to make the scalarization and get useful sensitivity indices. In this work, we propose to make the scalarization by taking the summation of all the elements in each term of the equation of covariance decomposition.

It shows that the corresponding sensitivity indices are equivalent to the traditional variance-based sensitivity indices on the summation of dynamic outputs. Actually, the summation of dynamic outputs can be considered as the zero frequency component (obtained by discrete Fourier transform) of the dynamic outputs, and the higher frequency components can be obtained easily through a weighted summation of dynamic outputs. Thus, the proposed approach can easily provide a global sensitivity analysis on all the frequency components of the dynamic outputs.

Therefore, the proposed approach makes a connection between the global sensitivity analysis of model output and the global sensitivity analysis of the frequency components of model output. In the field of failure detection of systems (such as water distribution system), the frequency components of model output can be very useful features. Therefore, the proposed method can be helpful for failure detection of systems.

Finally, the proposed approach is applied to a water distribution system to find the uncertain parameters having a significant effect on the frequency components of the model output.

Sinan Xiao
sinan.xiao@iws.uni-stuttgart.de
University of Stuttgart

Gödze Köse
goezde.koese@iws.uni-stuttgart.de
University of Stuttgart

Sergei Oladyshkin
sergey.oladyshkin@iws.uni-stuttgart.de
University of Stuttgart

Wolfgang Novak
wolfgang.nowak@iws.uni-stuttgart.de
University of Stuttgart

Parallel Session 6: Sensitivity analysis and algorithms Optimization

A cost-saving strategy for computing the Sobol' indices based on replicated Latin Hypercube Designs

A cost-saving strategy for computing Sobol' indices to conduct sensitivity analyses of computer models is presented. It relies on a sampling structure of the input variables referred to as replicated Latin Hypercube Designs . As recently pointed out, two of those designs are enough to compute estimates of the whole set of first order Sobol' indices. However, this is only true if a particular class of estimators is used, but by which small indices are poorly estimated. In this work, new estimators that take advantage of the permutation structure of replicated LHDs are proposed so that the precision of small indices can be improved. These estimators are shown to be particularly relevant when being implemented as part of a framework consisting of two stages.

The first stage is to compute all the first order Sobol' indices using two rLHDs only. Then, the precision of small indices can be reevaluated by the new estimators and, as a bonus, the associated total Sobol' indices can be estimated as well. Some academic examples are performed to show the superiority of the new estimators for computing small indices.

Lastly, a nuclear application concerning research reactors is carried out to demonstrate the interest of such a two-stage strategy to save simulations when the computer model is time consuming.

Damblin Guillaume
guillaume.damblin@cea.fr
CEA DEN

Alberto Ghione
alberto.ghione@cea.fr
CEA DEN

Sensitivity Analysis in Systematic and Representative Benchmarking of Optimization Algorithm Performance

Introduction

Besides the conventional and established applications of sensitivity analysis in the field of optimization [1], sensitivity analysis-based approaches can also be used in research on systematic performance analysis of meta-heuristic optimization algorithms, and the construction of synthetic industrially relevant benchmark problems.

The No Free Lunch Theorems for search and optimization [2] imply that there is no universal best algorithm for all problem types. The challenges in the field of non-convex optimization are thus moving towards the quests to identify and develop efficient algorithms for particular problem types. Industrially relevant product and process engineering problems often involve computationally expensive, highly non-linear, non-convex simulations responses. The industrial problems for which performance matters the most, are also those for which it is the most difficult to identify and develop efficient optimization approaches.

Therefore often artificial test functions and benchmark problems are used in the optimization community. For most of the commonly used artificial benchmark problems for meta-heuristic optimization algorithms, it is, however, difficult to relate relative algorithm performance on one artificial problem, to any real-world problems, simulation-based problems or other artificial benchmark problems. In order to increase the value of such benchmarks from a set of many isolated statistical tests to scientific tests with some generalization value, benchmark functions with a more systematic structure are required.

Various forms of variable screening and sensitivity analysis [3] can aid the construction of such improved benchmark problems [4]. Besides more systematic comparative assessments of optimization algorithms, it is also important to find ways to relate algorithm performance on artificial benchmark tests to real-world and industrial optimization problems [5]. Also here sensitivity analysis could aid to characterize the response behavior of real-world and simulation-based problems in order to match the problem features with particular synthetic benchmarks.

By means of academic examples as well as industrial case studies we demonstrate innovative approaches which set steps towards systematic and representative benchmarking and empirical optimization algorithm performance analysis.

The Characterization of Optimization Problems

A first step towards obtaining generalizable information from optimization benchmarks is to answer questions such as: What will happen to the optimization algorithm performance when a characteristic feature of the benchmark problem would be changed?

Problem changes of practical interest could for example involve modifications of the problem dimension, the distribution of the first and higher order sensitivity indices of the model output w.r.t the model input, the smoothness of the objective function, or its multi-modality.To investigate the answers to questions of the previously stated kind, benchmark functions parameterized w.r.t several function features were developed. By means of such parameterized test problems the performance of meta-heuristic algorithms is systematically investigated, showing clear trends and generalizable quantitative performance characteristics of optimization algorithms w.r.t. particular problem features.

Most multidisciplinary engineering optimization problems are not only composed of an objective function but also contain one or more constraints. The objective and all the constraint functions can be dependent on the results of several computationally expensive simulations with non-linear responses. Not only the structure of the individual simulations responses is important for the performance of optimization algorithms, but also the structure and dependencies between the simulation responses. Design variables which are important for the objective function, are not necessarily also those variables which are important for those of the non-linear constraints. In order to develop synthetic benchmark problems which are representative for particular real-world or industrial problems, it is necessary to be able to construct benchmark problems with similar response structures, and similar relations between the objective and constraints. To address this challenge, various forms of sensitivity analysis could be used to characterize the structure of the real-world and industrially relevant optimization problems. These characteristics could then be used to construct computationally affordable benchmark functions, on which the performance of different algorithms and algorithm settings can be investigated. Such an approach could be used to select efficient algorithms for particular optimization problems, and to tune or optimize the optimization algorithm parameters to problems with particular features. An approach along the lines of these ideas has been developed and tested on a relatively complex industrial engineering problem, namely the optimization of car body structures involving crashworthiness, vibrational comfort, and weight requirements.

Discussion and Outlook

Although the results of the presented ideas and approaches are promising, the area of research of using sensitivity analysis methods for systematic insight in optimization algorithm performance is still very young, and not widespread. There is much potential for improvement, and there are still many open challenges: Which variable screening and or sensitivity analysis methods are the most suitable for such applications? How to effectively balance the trade-off between function evaluation cost and the nature and quality of the obtained sensitivity measures? Also, from the computational simulation perspective, various approaches to increase the efficiency of the response characterization need to be investigated [6]. Sensitivity analysis-based approaches can contribute towards systematic optimization algorithm assessments, with more generalizable results. It can also seems an essential ingredient to extricate the problem of matching and developing efficient algorithms for complex computationally expensive industrial optimization problems.

References

[1] Castillo, E., Mínguez, R., & Castillo, C. . Sensitivity analysis in optimization and reliability problems. Reliability Engineering & System Safety, 93(12), 1788-1800.
[2] Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. IEEE transactions on evolutionary computation, 1(1), 67-82.
[3] Iooss, B., & Lemaître, P. (2015). A review on global sensitivity analysis methods. In Uncertainty management in simulation-optimization of complex systems (pp. 101-122). Springer, Boston, MA.
[4] Sala, R., Baldanzini, N., & Pierini, M. (2017). Global optimization test problems based on random field composition. Optimization Letters, 11(4), 699-713.
[5] Sala, R., Baldanzini, N., & Pierini, M. (2016). Representative surrogate problems as test functions for expensive simulators in multidisciplinary design optimization of vehicle structures. Structural and Multidisciplinary Optimization, 54(3), 449-468.

[6] Biehler, J., Gee, M. W., & Wall, W. A. (2015). Towards efficient uncertainty quantification in complex and large-scale biomechanical problems based on a Bayesian multi-fidelity scheme. Biomechanics and modeling in mechanobiology, 14(3), 489-513.

Ramses Sala
sala@rhrk.uni-kl.de
Institute of Applied Mechanics
Department of Mechanical and Process Engineering TU
Kaiserslautern, Germany

Ralf Müller
ram@rhrk.uni-kl.de
Institute of Applied Mechanics
Department of Mechanical and Process Engineering TU
Kaiserslautern, Germany

Niccoló Baldanzini
niccolo.baldanzini@unifi.it
Dipartimento di Ingegneria Industriale
Università degli Studi di Firenze
Italy

Marco Pierini
marco.pierini@unifi.it
Dipartimento di Ingegneria Industriale
Università degli Studi di Firenze
Italy

How sensitivity analysis can help AI optimization algorithms

Metaheuristics are strategies that guide a global optimization process of non-linear objective functions. The goal is to efficiently explore the search space in order to find near–optimal solutions. Metaheuristics deal with a dichotomy between exploration, discover unexplored areas, and exploitation, to dig around a good solution. Population based metaheuristics ensure a wide exploration of the search space.

Starting from a set of initial points, the metaheuristic iterations randomly explore the neighborhood of each point. A neighbor is generated by offsetting a set of variables of the current point. The variables and the offset are randomly chosen. We propose to use the information gathered during the iterations for guiding the algorithm toward cleverer choices. The idea is to use the points evaluations to characterize the variables behavior (relevance and shape) thanks to a sensitivity analysis (SA) method.

Sensitivity analysis method are widely used to eliminate non influential variables before the optimization process. In this paper, we present an another way to use the information given by a SA method for improving the convergence of the metaheuristic. We assume that proceeding in two steps may not be suitable. Removing variables definitively can be damaging to the optimization process. Irrelevant variables at the beginning of the algorithm may become relevant further in the search process and could discriminate the points (during the exploitation process). Moreover, integrate a sensitivity analysis method directly in a metaheuristic saves evaluations of the objective function. The goal is not to compute accurate sensibility analysis indices, but to obtain enough information on the variable behavior in order to guide the optimization process.

According to their search process, metaheuristics can be split into three families: those which search along one direction (variable), those which select a subset of variables and those which search along all directions. For the two first families, SA would help to focus on the most influential variables. In all cases the information about variable behavior (monotony, non-linearity, ...) would help adapting the offset when generating a new neighbor.

In a previous work, Morrisʾ one-at-time method, has been integrated in Artificial Bee Colony algorithm. ABC algorithm integrates fairly well Morrisʾ method, because of its one-direction neighborhood search process. They both offset a point according to a single variable at time and analyze the impact on the objective function output.

Among all search processes implemented in metaheuristics, ABC neighborhood search is a particular case. Many metaheuristics algorithms search a neighbor in a hyper-sphere, offsetting various variables at a time (Simulated Annealing, Differential Evolution, Particle Swarm Optimization).

We generalized this approach. We defined a naive sensitivity analysis method able to:
* identify non influential variables;
* rank the variables;
* determine the behavior of each input variable,in a multidimensional neighborhood context.

The sensibility analysis is defined according to the following constraints:
* the sensitivity analysis method has to cope with complex functions since it will be used in the context of hard optimization;

* the sensitivity analysis method must not have experimental design constraint since the evaluation points are given by the initialization and the iterations of the algorithm
* the sensitivity analysis has to deal with a reasonable number of evaluations since metaheuristics usually have a limited number of function evaluations
* the sensitivity indices must be very fast to calculate in order to be integrated into the optimization algorithm

None of the most famous sensitivity methods entirely achieve these objectives. The covering of the search space with the evaluation points leads to a poor estimation of the Sobol' indices. Moreover, this method does not make difference between monotonic or non-linear influential variables. Morris' method requires less points and the pair (μ^*,s^2) allows to distinguish non-linear behaviors from monotonic behaviors. But Morris method is constrained by an experimental design (trajectories by moving one variable at time). With the same computational cost than Morris' method, linear and rank correlation coefficients are not constrained by an experimental design. However, this method assumes a linear or monotonic model.

The simple idea is to replace the elementary effects in the Morris' method by the linear (or rank) correlation coefficients computed in a local neighborhood of a set of evaluation points. As done in Morris' method, the normalized mean (mj) and standard deviation (sj) of the local coefficients are computed, by variable. A large value of mj indicates an influential variable with a monotonic effect, and a large value of sj indicates an influential variable with a non-linear effect. We propose to integrate this knowledge about the variables behavior to adapt the offset value for each variable. We focus on two key behaviors of metaheuristic search process:
* the shift of a current point: Non-linearity leads to a small shift and, on the contrary, monotony leads to a large shift.
* the variable selection to be shifted: replace the random selection by a selection of the most influential variables.

This highlights how a sensitivity analysis computed during the optimization process, can improve the algorithm focusing on the most influential variables and adapting the hypervolume search according to the behavior of each variable.

Astrid Jourdan
aj@eisti.eu
EISTI, Quartz Laboratory

Peio Loubiere
plo@eisti.eu
EISTI, Quartz Laboratory

Patrick Siarry
siarry@u-pec.fr
LISSI EA 3956 UPEC

Rachid Chelouah
rc@eisti.eu
EISTI, Quartz Laboratory

Parallel session 7: Sensitivity analysis and environmental issues II

Global sensitivity analysis of fishery management simulation models: Efficient conditioning and robust combination of the Morris and Sobol methods

Fishery systems are complex systems that need to be managed to ensure the sustainable exploitation of fishing resources, the economic efficiency of the fishery and the social wellbeing of fishing communities. One of the main problems in the management of fisheries is the inherent uncertainty which arises principally from the natural variability of the system, the errors made in the decision-making process and the problems in the implementation of the management actions. Management strategy evaluation approach arose as a procedure to formalize the incorporation of uncertainty in the fishery management decision-making process. Its objective is to evaluate the performance of fishery management strategies, by means of simulation, before they are put in place. The main difference between MSE and other simulation approaches is that, as well as the natural fishery system, it also simulates the entire management process, from data collection to the generation of the management advice.

Although global sensitivity analysis (GSA) seems a natural complement of MSE, the two approaches are rarely combined. The infrequent use of GSA for MSE models is mainly due to i) the complexity of the models, which inhibits the use of the available GSA software, ii) the high computational cost, and iii) the lack of practical applications in the field. To promote the application of GSA in MSE simulation models we define a scientifically sound and computationally efficient methodology to apply GSA to complex fishery MSE models. First, we propose a set of guidelines to condition the model efficiently so the correlation in the input factors is avoided and the number of input factors is reduced. Second, we introduce new selection and convergence criteria to combine the Sobol and Morris methods robustly. Furthermore, these criteria are applicable to any other simulation model.

We illustrate the approach using a complex implementation of a bio-economic MSE model where, initially, the number of parameters was greater than 1000 and many were correlated. In general, the parameters of the fleets' short-term dynamics were the input factors with the highest impact on the output indicators. Regarding the observation and estimation errors, their impact was bigger in the economic performance of the fishery than in the biological indicators of the fish species.

Dorleta Garcia
dgarcia@azti.es
AZTI

Inmaculada Arostegui
inmaculada.arostegui@ehu.es
UPV-EHU

Raúl Prellezo
rprellezo@azti.es
AZTI

Sensitivity Analysis of an atmospheric dispersion model – First order index using the IA-estimator

The Joint Research Centre of the European Commission has recently developed a software tool, named Accident Damage Analysis Module , to assist the regulatory authorities in the European Union (EU) for assessing the consequences of potential industrial accidents in Seveso sites. In particular, the atmospheric dispersion module of the tool focusses on the calculus of concentration of a toxic or a flammable substance following the unintended release of such a substance. Details on this tool and its validation are reported elsewhere (Fabbri et al., 2017, Fabbri et al., 2018).

In the current work, a variance-based sensitivity analysis has been carried out on the dispersion model implemented in ADAM, by using a reference accident scenario consisting of a downward release of a pressurised-liquefied toxic (chlorine) from a 6-inches bottom hole of a horizontal cylinder. The authors propose a Monte Carlo estimation of the Sobol' sensitivity indices, in particular first order ones, which makes use of the Improved Algorithm-IA (Azzini&Rosati, ongoing submission). The new formula shows a higher level of convergence if compared with similar benchmark estimators and allows the detection of not significant factors with very few model runs.

The sensitivity analysis was conducted by fixing the geometry of the vessel, and the release hole and by varying all other relevant factors in a quite wide range (substance quantity, storage/meteorological/environmental conditions, and model parameters). Two model outputs were selected i.e., the downwind distances at which the toxic cloud reaches a concentration level of 15000ppm and 1500ppm, respectively.

In this initial study, 13 variables were identified and included in the analysis. Uncertainty ranges were identified to account of possible variability in the input model, and uniform distributions were assumed. Because of the exploratory nature of the study, some range of uncertainty were initially considered very large, as for instance the tank filling level, which was set to represent all conditions between 'almost full' and 'almost empty'. Meteorological and environmental factors as well as some specific model parameters play also a relevant role in the model output, and they were therefore included in this study.

The estimation of the Sobol' first order indices has been carried out using the new improved algorithm-IA. This estimator is based on the the evaluation of two symmetric integrals to compute the sensitivity effects and dynamic output variances, which are calculated and applied to each investigated input. Therefore, the computation of the indices requires 2N(k+1) model evaluations, where N is the sample size and k is the number of inputs. The results of the global sensitivity analysis show a significant convergence with both MC and QMC sequences.

Referring to the case of 15000ppm and QMC technique, at a size of N=50 (total cost of 1400 runs), six out of thirteen inputs (i.e., substrate roughness in proximity of the release, substrate thermal conductivity and diffusivity, ambient temperature, and humidity) are already detected as negligible, with both sensitivity indices below 0.01. The input ranking identified at this stage is confirmed by the evaluations at larger samples. Apart from one factor (difference around 0.07 for both main and total order index), all sensitivity indices are correctly computed at N=100, and no relevant modification appears starting from N=200.

References

[1]Ermak, D.L., 1990: User's manual for SLAB: an atmospheric dispersion model for denser-than-air releases, Lawrence Livermore National Laboratory, Livermore, California, UCRL-MA-105607.
[2] Fabbri, L., M. Binda, and Y. Bruinen de Bruin, 2017: Accident Damage Analysis Module (ADAM) – Technical Guidance, Publications Office of the European Union, Luxembourg, EUR 28732 EN, 2017, ISBN 978-92-79-71879-3, doi: 10.2760/719457.
[3]Fabbri, L., M. Binda, and M.H. Wood, 2018: Evaluation of the Accident Damage Analysis Module (ADAM) tool - Verification and Validation of the implemented models in ADAM for Consequence Analysis, Publications Office of the European Union, Luxembourg, EUR 29363 EN, ISBN 978-92-79-94668-4, doi:10.2760/582513.
[4]Lilburne L., and S. Tarantola, 2009: Sensitivity analysis of spatial models, International Journal of Geographical Information Science, 23(2), 151-168.
[5]Saltelli, A., P. Annoni, I. Azzini, F. Campolongo, M. Ratto and S. Tarantola, 2010: Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index, Compute Physics Communications, 181(2), 259-270.
[6]Saltelli, A., M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana and S.
[7]Tarantola, 2008: Global Sensitivity Analysis. The Primer, Chichester: Probability and Statistics, John Wiley and Sons.
[8]Sobol, I. M., 1993: Sensitivity analysis for non-linear mathematical models, Mathematical Modelling and Computational Experiment, 1, 407-414.

Luciano Fabbri
luciano.fabbri@ec.europa.eu
European Commission, Joint Research Centre (JRC)
Ispra, Italy

Ivano Azzini
ivano.azzini@ec.europa.eu
European Commission, Joint Research Centre (JRC)
Ispra, Italy

Rossana Rossati
rossana.rosati@ec.europa.eu
European Commission, Joint Research Centre (JRC)
Ispra, Italy

Sensitivity analysis for flood evacuation model

The frequency and severity of natural disasters have been significantly increasing throughout the world in the last decades due to several influencing factors including the impact of climate change combined with an ever-growing population in areas at risk. Initiatives to develop decision support systems for mass evacuation in emergency situations are beginning to emerge and these make use of recent progress made in the development of powerful simulation tools for evacuation modelling. While complex scenarios can now be simulated in a qualitatively realistic way, quantitative predictions remain elusive. This is due to the many sources of uncertainty in the simulations. In addition to uncertainties in the input data , the simulated system itself presents inherent challenges linked to the accurate description of human behaviour under stress conditions.

Data61, the digital innovation arm of CSIRO, has developed a new simulation package for evacuation modelling which enables each individual to be described with specific characteristics including their response to stress and onset of panic, social interactions with other agents, among others. In addition, the agents can navigate a complex three-dimensional environment and respond to external stimuli. A flood event has been used as an example of an external stimulus in the present research. The extent of the flood was simulated using CSIRO's owned package SWIFT and imported as a geometrically evolving constraint in the evacuation model. The agents have an ability to sense their environment and potentially react to avoid the areas that have been or are at risk of being flooded. The model may then be used to study the possible outcomes of the evacuation depending on decisions made such as instructing the evacuees to leave at a particular time, towards a specific evacuation area, etc. This provides useful information to test possible evacuation plans and strategies. Yet the outcome is strongly dependent on a number of parameters in the evacuation model, on the agents and on the event that are only partially known. Some a priori values for these parameters are given in the literature but they may not be optimal for simulation and thus lead to evacuation times that are not consistent with observations. The deterministic numerical approach thus appears limited; these limitations can be overcome with an ensemble-based approach making statistical assumptions on the model parameters. The stochastic framework allows to develop reliable methods to reduce the uncertainties present in the model parameters and its outputs. By capitalizing on CERFACS's expertise in uncertainty quantification (UQ), we have started to develop these approaches and identify the most important parameters that may affect the outcome of an evacuation. This is the first step towards developing more accurate models. The computation of sensitivity analysis indices allows to identify predominant parameters on the quantity of interest. The methods for sensitivity analysis that have been implemented so far at CERFACS are non-intrusive and rely on Monte-Carlo based sampling of the input space and propagation with the direct solver or with a reduced-cost surrogate.

In a preliminary study, we considered the evacuation of a room with a single exit. This example has been widely studied in the literature, particularly in relation to the social force model proposed by Helbing et al. (1995, 2000). A given number of agents are placed randomly inside a room from which they attempt to evacuate at a predefined desired speed. The social force model describes the characteristics of the agents, the social interactions between them through a repulsive interaction force as well as additional granular-like forces that capture particular effects in panicking crowds. This model is widely popular in the scientific literature, partly due to its ability to capture many emergent macroscopic phenomena of crowd motion including the so-called faster-is-slower effect. This describes the observation that under

certain conditions of evacuation increasing the speed of the agents actually leads to longer egress times.

On this simplified model, a uniform assumption of the probability density function of the social force model was made. A learning data base was built from an optimized design of experiment in the input space and integration of the evacuation model. The Sobol' sensitivity indices have been computed as well as the probability density function for the total evacuation time. It was shown that the predominant parameters to explain the variance of the evacuation time, are related to the intensity of interactions as well as compression and friction force between agents. Uncertainty quantification and sensitivity analysis from this proof of concept paves the way towards reduction of uncertainties with data assimilation algorithms.

Sophie Ricci
ricci@cerfacs.fr
CECI, UMR5318/CERFACS

Vincent Lemiale
Vincent.Lemiale@data61.csiro.au
CSIRO-Data61

Nicolau Andres Thio
Nico.Andres@data61.csiro.au
CSIRO-data61

Pamphile Roy
roy@cerfacs.fr
CERFACS

Parallel session 8: Sensitivity analysis in spatio-temporal modelling I

Impact of common inhabitants' habits on the risk of overheating in dwellings. Application of temporal sensitivity analysis

To predict building thermal behaviour, many inputs are requested as climatic data, occupant's behaviour scenarios as well as multiple inputs related to building envelope properties, building geometry and HVAC systems' characteristics. In the purpose of reducing energy consumption in a building as building energy use currently accounts for over 40% of total primary energy consumption in the U.S. and E.U. [Cao and al., 2016] sensitivity analyses are applied in several practical issues in building energy analysis [Tian, 2013]. Among the issues, the study of the impact of occupants' behaviours on dynamic thermal response of a building remains a challenging task.

This study seeks to estimate the robustness of a building in summer period to occupant's behaviours as climate change will induce stronger and more frequent heat-waves period and consequently a potential increase in air-conditioning energy use in buildings. To do so, this study uses sensitivity analysis in the aim of answering three questions: How intensely might an occupant's action affect indoor temperature? In which way? And for how long? In particular, this study deals with the problematic of summer comfort in low-energy residential building. The intrinsic characteristics of these buildings as a high-insulated and an airtight enveloped coupled with a bioclimatic design make them more sensitive to overheating.

The inhabitants' actions scrutinise in this study reflect common inhabitants' habits as: the opening of windows for a short period in early morning, at noon and in the evening because it affects the heat balance of the building due to the introduction of an external air-flow; the opening of blinds in the morning, at lunch time, in the afternoon and late afternoon because it controls the introduction of solar heat gains; the electric appliances usage in early morning, at noon, in the evening and in stand-by mode as it corresponds to sensible internal heat gains and finally, the interlock of the night mechanical ventilation system that also affects the introduction of external air-flow. In total, 12 inhabitant's actions are investigated. Moreover, a low-energy house is the case study and climatic data corresponding to the heat-wave period of summer 2018 in Paris area is used. That year, a 16-day period of heat-wave has been recorded in France from the 24th July to 8th August. It is the second warmest summer, just after summer 2003 according to the French weather agency MétéoFrance.

To understand the intensity and the persistence of inhabitants' actions on the building, temporal sensitivity analysis is conducted on the output of our model that is the hourly indoor temperature i.e. sensitivity analysis is performed at an hourly time step to obtain the temporal function of sensibility of each action. For that, here, the RDB-FAST method is used (Random Balance Design – Fourier Amplitude Sensitivity Test). It is a robust and accurate global sensitivity analysis method based on variance decomposition [Tarantola and al., 2006] and is available now on SALib library in Python (Sensitivity Analysis Library) and it is implemented with a LHS (Latin Hypercube Sampling) sampling.

Prior to manipulate the sensitivity analysis results, uncertainty analysis is performed as LHS sampling allows to properly explore the input space. This uncertainty analysis provides two different major information: firstly, it allows to identify period in time when there is a risk of thermal discomfort i.e. for which most of the 1,000 combinations of actions propagated in the

model lead to indoor temperature higher than 27°C that is a common overheating threshold. Secondly, the analysis of the evolution of the variability of the indoor temperature over time allows to distinguish over the previous selected period, potential-free moment in time i.e. when the variability is too low that there is no actions better than others and other moment in time with higher variability for which actions are better than others.

The temporal sensitivity analysis is performed on the indoor air temperature at each hourly timestep. Partial variance results are preferred to sensitivity index in order to work with quantitative results due to the temporal variation of the variance.

Either a positive or a negative sign is assigned to the partial variance of each action and at each time step according to the observed tendency i.e. the sign of regression coefficient value between the 1,000 computed indoor temperatures values and the 1,000 values associated at the action in order to identify actions that lead to higher indoor temperature from actions leading to lower indoor temperature.

By observing the maximum duration in time when partial variance of an action is above a lower limit of variance, this methodology shows, for example than the impact on the indoor temperature of the opening of windows for one hour can last for 23 hours. Results also show that the opening of blind for 4 hours in the afternoon has less negative impact, in intensity, than the opening of windows for one hour at noon.

With this generic methodology it is possible to conclude on the intensity and the persistence of the impact of inhabitants' actions on indoor temperature. Two prospects are: to identify risk and possible improvements during a heat-wave period and if improvements exist, to provide a ''user-guide'' to the inhabitants.

References

[1]Cao, Xiaodong, Xilei Dai, and Junjie Liu. 2016. ''Building Energy-Consumption Status Worldwide and the State-of-the-Art Technologies for Zero-Energy Buildings during the Past Decade.'' Energy and Buildings 128 (September): 198–213. doi:10.1016/j.enbuild.2016.06.089.
[2]Tarantola, S., D. Gatelli, and T. A. Mara. 2006. ''Random Balance Designs for the Estimation of First Order Global Sensitivity Indices.'' Reliability Engineering & System Safety 91 (6): 717–727. doi:10.1016/j.ress.2005.06.003.
[3]Tian, Wei. 2013. ''A Review of Sensitivity Analysis Methods in Building Energy Analysis.'' Renewable and Sustainable Energy Reviews 20 (April): 411–419. doi:10.1016/j.rser.2012.12.014.

Léa Gondian
lea.gondian@univ-smb.fr
Université Savoie Mont Blanc

Jeanne Goffart
jeanne.goffart@gmail.com
Université Savoie Mont Blanc

Coupling of sensitivity analysis and model calibration in radioactive waste disposal safety assessment

Prediction of radionuclides transport in the environment is essential for safety assessment of long-term disposal of radioactive waste. Groundwater flow and transport modelling is a commonly used approach to estimate the distribution migration of the contamination in geological media. This kind of modelling is associated with high level of uncertainties of diverse origins. And almost always the most significant contribution to the uncertainty of the modelling result is associated with the conceptual uncertainty due to lack of data or understanding of the underlying processes.

As an example one can consider a structural geological model – it is absolutely impossible to construct an accurate one, especially for highly heterogeneous media . The variety of geological structural models compliant with the experimental data could be developed both manually and by means of stochastic modelling. The other example deals with the chemical description of sorption processes – the set of reactions, types of sorption centers, etc [1]. A coupled utilization of sensitivity analysis and automatic model parameter calibration is an adequate approach to this problem.

For sensitivity analysis we use the variance-based method – Sobol indices. This method allows evaluating the first order sensitivity indices responsible for the single parameter contribution to the resulting uncertainty, their confidence intervals and the total sensitivity indices that take into account the parameter interactions. Some results of sensitivity analysis indicate problems in the conceptual model, among them such as:
* the model shows low sensitivities to all parameters;
* the model shows low sensitivities to some parameters, and modelers -experts consider it as strongly non-adequate.

The automatic calibration procedure relies on the optimization methods that search through parameter combinations to achieve the best fit of model results and observations [2]. Nowadays heuristic optimization methods become more and more widely applied for complex optimization problems (that means the minimization of the nondifferentiable functions with multiple local minima) in the different scientific fields. The main idea of it is that the behavior of many natural systems has successfully solved their optimization issues and could be efficiently reproduced for other problems.

In this work as the optimization methods we use a modified Particle swarm optimization approach and the heuristic hybrid Cuckoo Search and Harmony Search algorithm was utilized. The hybridization allows gaining advantage from Cuckoo search's ability to efficiently search the whole parametric space and the efficiency of Harmony search to refine the found near-best solution.

This step may lead to one of the «unfortunate» results that almost without exception lead to the necessity of conceptual model revision:
* Optimal values of several parameters come close to the limits of the predefined diapason that could indicate possibly inadequate values.
* Optimization via different algorithms and with different initial guesses fails to converge.
So, in practical experience uncertainty analysis procedure consists of at least two different parameter optimization steps and the sensitivity analysis before each of them.

The described approach was applied to two problems: the choices of the structural geological model and of the set of chemical reactions describing the sorption of radionuclides on goethite. The heterogeneous media is represented by several types of crystalline rock intruded by dikes and having different fracture indices.Two alternative geological cross-sections were developed in accordance with the data from 4 exploratory boreholes. These two models have apparent contradictions in the general orientation of the dikes that are sufficient because their intersections with the gneiss are considered as the potential path of contaminant transport.

In addition to the descriptive data from the exploration wells, the results of pumping tests are available. The groundwater flow model is being developed on the base of both structural models and calibrated to fit the head observations from these pumping tests.

Regarding the modelling of the groundwater flow in the complex heterogeneous media, the necessary to deal with uncertainties comprised:
* The uncertainty of structural elements within intervals of measurements.
* The number of parameters that are significant for the current groundwater flow model.
* The uncertainty of the groundwater flow model parameters that can provide consistency with the observations.

Varied parameters were filtration properties of different structural elements and the uncertain border conditions. The fitness function being minimized was defined as a weighted sum of squared residuals between modelled and experimental water heads. The calibration was preceded by sensitivity analysis, which allowed excluding insufficient for the calibration parameters (filtration coefficients of crushing zone and of moderately cracked dikes). The results clearly indicate a greater likelihood of one of the geological structural models.

The development of the chemical model describing sorption of radionuclides (Eu, Np, U) on goethite required 4 iterations to obtain good coincidence with the experiment. Varied parameters were sorption equilibrium constants and the density of strong and weak sorption sites. The goal of parameter optimization was to fit simulated and experimental isotherms and pH-dependencies. Each iteration included sensitivity analysis and 3 of them included both sensitivity analysis and calibration. The evolution stages of the conceptual model were:
* The selection of description of electrostatic features.
* The definition of the number of sorption centers.
* The importance of the hydrolysis reactions.

The examples given clearly show that sensitivity analysis and parameter optimization steps are not the final stages of the computational model development, but an inseparable part of the modelling process.

Normal 0
false false false RU X-NONE X-NONE

References

[1] Payne, Tim E., et al. "Guidelines for thermodynamic sorption modelling in the context of radioactive waste disposal." Environmental modelling & software 42 (2013): 143-156
[2] Hill, Mary C., and Claire R. Tiedeman. Effective groundwater model calibration: with analysis of data, sensitivities, predictions, and uncertainty. John Wiley & Sons, 2006

Elena Savaleva

esav@ibrae.ac.ru
Nuclear Safety Institute Russian Academy of sciences
Moscow, Russia

Valentina Svitelman
Svitelman@ibrae.ac.ru
Nuclear Safety Institute Russian Academy of sciences
Moscow, Russia

Petr Blinov
blinov@ibrae.ac.ru
Nuclear Safety Institute Russian Academy of sciences
Moscow, Russia

Dmitri Valetov
valetovdk@ibrae.ac.ru
Nuclear Safety Institute Russian Academy of sciences
Moscow, Russia

Georgy Neyzhaev
neyvazhaev@ibrae.ac.ru
Nuclear Safety Institute Russian Academy of sciences
Moscow, Russia

Assessing the Effectiveness of Visualization Techniques for Spatial Sensitivity Analysis

Visualization of sensitivity analysis output in spatially explicit models can potentially enhance the interpretation of uncertainty in model inputs. Yet, it is not obvious which visualization techniques are effective in the context of spatial and spatio-temporal models as the difficulty of selecting an appropriate visualization increases with model complexity and the number of input-output linkages.

This study examines the effectiveness of adjacent and coincident maps depicting the results of global sensitivity analysis applied to a widely used spatio-temporal model of urban growth called SLEUTH. The efficacy of both types of maps is tested for expert and novice end-users with a web-based survey to understand which of the maps is more effective in supporting the comprehension of global sensitivity analysis results. Additionally, the survey results are analyzed to determine which map type is more effective in gaining user's confidence to discern influential model inputs form non-influential ones, and correctly interpret sensitive locations in a model's study area.

Sedap Sala Ayca
sedasalap@gmail.com
San Diego State University

Piotr Jankowski
pjankows@sdsu.edu
San Diego State University

Keith C. Clarke
kclarke@geog.ucsb.edu
University of California
Santa Barbara

Atsushi Nara
anara@sdsu.edu
San Diego State University

Parallel Session 9: Sensitivity analysis and reliability

Variance-based reliability sensitivity analysis and the FORM alpha factors

In reliability assessments, it is useful to compute importance measures that provide information on the influence of the random variables on the probability of failure. Classical importance measures are the alpha factors, which are obtained as a by-product of the first-order reliability method . These factors are the directional cosines of the most probable failure point in an underlying independent standard normal random variable space. They represent the contribution of each random variable on the variance of the linearized limit-state function in this underlying space.

Alternatively, one might assess sensitivity by a variance decomposition of the indicator function, i.e. the function that indicates membership of the random variables to the failure domain [L.Y. Li, Z.Z. Lu, J. Feng, B.T. Wang, Moment-independent importance measure of basic variable and its state dependent parameter solution, Struct. Saf. 38 (2012) 40–47]. This presentation discusses the relation of the latter variance-based sensitivity measures to the FORM alpha factors and analytically shows that there exist one-to-one relationships between them for linear limit-state functions of normal random variables.

We also demonstrate that these relationships enable a good approximation of variance-based sensitivities for general reliability problems. The derived relationships shed light on the behavior of first-order and total-effect indices of the failure event in engineering reliability problems.

Iason Papaioannou
iason.papaioannou@tum.de
Engineering Risk Analysis Group
Technische Universität München

Daniel Straub
straub@tum.de
Engineering Risk Analysis Group
Technische Universität München

Reliability sensitivity analysis with value of information

Reliability assessment and robust design provide decision support for engineering systems under uncertainty. Sensitivity analysis is a key element of such assessments. In many instances, it is relevant to understand if and how further efforts should be undertaken to reduce uncertainty, e.g. by collecting additional data or by improving models. The value of information concept provides a decision-theoretic measure to quantify the benefits of such uncertainty reduction, and has therefore been proposed as a basis for sensitivity analysis [1]. In particular, the expected value of partial perfect information and the expected value of sample information (EVSI) have been proposed as sensitivity measures [1-2]. This talk focuses on the definition and computation of EVPPI for reliability assessment of systems represented through physics-based models, in contrast to [3], which considers Boolean system models. Because the value of information can only be calculated in the context of a decision analysis, we propose and investigate general formulations for such decision analyses, which are representative for a wide range of applications. On this basis, we derive expressions for the EVPPI. Thereafter, we discuss three different computational strategies that can be employed to evaluate the EVPPI for systems that are represented by numerical models. Based on the results of a reliability analysis, these strategies work without additional runs of the (potentially costly) numerical model. Concepts and implementation are demonstrated on some realistic engineering problems. Finally, we draw some conclusions on the relationship between the EVPPI and common sensitivity measures in reliability assessments.

References

[1] Felli, J. C., & Hazen, G. B. (1998). Sensitivity Analysis and the Expected Value of Perfect Information. Medical Decision Making, 18(1), 95–109.
[2] Oakley, J. E. (2009). Decision-Theoretic Sensitivity Analysis for Complex Computer Models. Technometrics, 51(2), 121–129.
[3] Borgonovo, E., & Cillo, A. (2017). Deciding with Thresholds: Importance Measures and Value of Information. Risk Analysis, 37(10), 1828–1848.

Daniel Straub
straub@tum.de
Engineering Risk Analysis Group
Technische Universität München

Max Ehre
max.ehre@tum.de
Technische Universität München

Iason Papaioannou
iason.papaioannou@tum.de
Engineering Risk Analysis Group
Technische Universität München

The role of Sensitivity Analysis in Imprecise Stochastic Simulation

Simulation models have been widely applied in various fields so as to learn and simulate the behaviors of physical systems as well as the performances of engineering structures. However, due to the ubiquitous uncertainties presented in the input parameters and the errors of the simulation models, the practitioners always find it difficult to catch the real system behaviors. Generally, there are two kinds of uncertainties, i.e., the aleatory uncertainty and epistemic uncertainty, where the former one is due to the intrinsic random properties of parameters, and is irreducible, while the later one is caused by the imperfect knowledge of parameters , and can be reduced by learning more knowledge.

Current researches have shown that the error caused by ignoring the epistemic uncertainty can be much more prohibitive than the one caused by the error of simulation models, especially when the response tail distribution is concerned. Thus, properly characterizing the different kinds of uncertainties, and effectively propagating them through the simulation models have become two challenging tasks. The imprecise probability models, such as evidence theory, probability-box model and fuzzy probability models, have been developed and regarded as the most appealing mathematical models for characterizing the two kinds of uncertainties since each of them makes a separation between the two kinds of uncertainties in a unified framework.

However, the efficient and accurate propagation of the imprecise probability models through any simulation models (linear or nonlinear, convex or non-convex) has long been a difficult task, and many efforts have been devoted by researchers in this aspect. Imprecise stochastic simulation developed by the authors is an appealing general non-intrusive framework for effectively propagating the imprecise probability models, and further for uncertainty quantification of model responses, and for structural reliability analysis, under mixed uncertainty environment.

The most appealing characters of this method are that the computational cost is the same with the utilized precise stochastic simulation procedure, the optimization procedure directly performed on the model response functions can be avoided, and the estimation errors are properly addressed. And also, the implementation of this method is almost as simple as the utilized advanced precise stochastic simulation procedure. In this presentation, we show that the sensitivity indices, which are byproducts of the imprecise stochastic simulation and are estimated with trivial efforts, can play important roles in the estimation process.

Firstly, they can be utilized for recognizing the important components in the final estimations. Secondly, they can be severed for measuring the truncation errors of the final estimations. Thirdly, the sensitivity indices can inform the analysts the optimal way for further allocating information so as to reducing the epistemic uncertainty presented in the estimation. Several test examples are introduced to illustrate the theoretical developments.

Pengfei Wei
pengfeiwei@nwpu.edu.cn
School of Mechanics, Civil Engineering and Architecture
Northwestern Polytechnical University
Institute for Risk and Reliability
Leibniz Universität Hannover

Jingweng Song
jwsong0879@163.com
School of Aeronautics
Northwestern Polytechnical University
Institute for Risk and Reliability
Leibniz Universität Hannover

Sifeng Bi
sifeng.bi@irz.uni-hannover.de
Institute for Risk and Reliability
Leibniz Universität Hannover

Matteo Broggi
broggi@irz.uni-hannover.de
Institute for Risk and Reliability
Leibniz Universität Hannover

Michael Beer
beer@irz.uni-hannover.de
Institute for Risk and Reliability
Leibniz Universität Hannover

Session 10: New methods in sensitivity analysis I

Improved Algorithms-IA for main and total order indices estimation

Traditional Sobol' variance-based methods are considered a very popular way to conduct global sensitivity analysis , primarily for their versatility and easiness of coding (Saltelli et al, 2010; Kucherenko&Song, 2017). Regrettably, the number of model evaluations required for an appropriate estimation that strictly depends on the number of inputs represents a severe limit. Therefore, the need for more performing designs to compute Sobol' indices, and in particular the main order index, is still valid

The Monte Carlo estimations of the Sobol' indices can be based on two independent matrices and , and the derived matrices Abi, which are equal to A but the i-th column which is replaced with the i-th column of B (Saltelli et al., 2010).Owen (2013) introduces a formula for the main indices based on three independent points instead of two, using the third point only for replacement purposes, with very good results. A further alternative for Sobol' indices computation is discussed by Tarantola (Tarantola et al., 2006; Lilbure&Tarantola, 2009), which takes into account two dependent matrices Abi and Bai (Bai is equal to B but the i-th column from A) and considers a dynamic variance depending on factor i for each individual estimation of Si / STi.

In this work, improved algorithms-IAs to estimate Sobol' indices are proposed (Azzini&Rosati, ongoing submission). Starting from two independent points x(y,z) and x'(y',z'), the computation of the first-order effect is obtained by the the evaluation of two symmetric integrals (Sobol', 1993), and applying the relation between main and total order indices, also an estimator for the latter is defined. Like in Tarantola, with the aim of capturing all the available information in the most efficient way, four input matrices are considered, and the corresponding four output vectors used in the estimation. To calculate sensitivity indices, dynamic output variances are computed and applied to each investigated input. The total cost of the estimation is therefore 2N(k+1), where N is the sample size and k is the number of inputs.

The new algorithms show a surprising stability, similar to the Owen formula, even at small sampling (N<100) and random numbers, and also good results are achieved both in case of functions with important low-order interaction terms and functions with important high-order interaction terms.

As from tests carried out on a broad set of functions, the rate of convergence is so significant to make the IA competitive compared to similar methods based on only three matrices, with a lower dependence on model dimensionality. Therefore, the improved algorithms-IAs appear to be a valid alternative to the actual best benchmarks using both Monte Carlo numbers and Quasi-Monte Carlo sequences.

References

[1] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto and S. Tarantola, "Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index," Compute Physics Communications, vol. 181, no. 2, pp. 259-270, 2010.
[2] S. Kucherenko and S. Song, "Different numerical estimators for main effect global sensitivity indices," Reliability Engineering and System Safety, vol. 165, pp. 222-238, 2017.

[3] A. B. Owen, "Better estimation of small Sobol' sensititvity indices," ACM Transactions on Modeling and Computer Simulation, vol. 23(11), no. 1-17, 2013.
[4] S. Tarantola, M. Nardo, M. Saisana and D. Gatelli, "A new estimator for sensitivity analysis of model output: An application to the e-business readiness composite indicator," Reliability Engineering and System Safety, vol. 91, pp. 1135-1141, 2006.
[5] L. Lilburne and S. Tarantola, "Sensitivity analysis of spatial models," International Journal of Geographical Information Science, vol. 23, no. 2, pp. 151-168, 2009.
[6] I. M. Sobol', "Sensitivity analysis for non-linear mathematical models.," Mathematical Modelling and Computational Experiment, vol. 1, pp. 407-414, 1993.

Ivano Azzini
ivano.azzini@ec.europa.eu
European Commission, Joint Research Centre (JRC)
Ispra, Italy

Rossana Rosati
rossana.rosati@ec.europa.eu
European Commission, Joint Research Centre (JRC)
Ispra, Italy

Why the Janon Estimator is the Best Thing since Sliced Bread

Sensitivity analysis of large and complex models, which have a high run-time, may often face the limitation that the budget of model runs available is too low to give precise estimates of sensitivity measures. Additionally, the number of model inputs may be high. In such cases, ‘‘screening’’ tools may be used to provide a rough ranking of model inputs or to sort them into influential and non-influential groups. In fact, screening can be viewed as sensitivity analysis at low sample sizes, and the tools available for this task are broadly similar to those used in sensitivity analysis in a more general sense. They can be roughly divided into ‘‘metamodeling’’ approaches, which use a data modelling approach, and ‘‘sampling based’’ approaches, which estimate sensitivity directly from the model samples.

This study offers a comprehensive comparison of ten sensitivity analysis approaches in the screening setting, ranging from recent metamodelling approaches , canonical low-rank approximations and PCE-kriging) to elementary effects, derivative-based sensitivity measures, variogram methods and Monte Carlo estimators. In order to relax the dependency of the results on the test function(s) used, a novel `metafunction' approach is adopted which randomly generates test functions of varying dimensionality (up to 100 inputs) and functional form using a random combination of plausible basis functions and coefficients from a Gaussian mixture distribution, and a range of sample sizes (up to 1000 model runs). This allows a very large range of sensitivity analysis problems to be investigated - in the present study, around 10,000 unique and randomly-generated sensitivity analysis problems were analysed by each of the ten methods.

The results give a detailed comparison of the approaches as ‘‘maps’’ of sample size and dimensionality, and with respect to particular basis functions. They show that in the context of particularly low sample sizes, Monte Carlo estimators, particularly the Janon estimator, outperform metamodels and more commonly-used screening tools. Indicatively and on average, metamodels become competitive at around 20-30 runs per model input, but at lower ratios, sampling-based approaches are considerably more effective as a screening tool.

A discussion on the (perhaps surprising) success of sampling-based approaches is given, along with some practical advice for modellers with a very low budget of model runs. The choice of sensitivity analysis method is important to know in advance because it dictates the sampling strategy.

William Becker
William.Becker@ec.europa.eu
European Commission - Joint Research Centre
Ispra, Italy

Has the Spell Been Broken? Estimating Variance-Based Sensitivity via Nearest Neighbors

In the framework of a factor fixing setting, the analyst tries to identify non-influential input parameters in a simulation model with uncertain inputs. The method usually considered to be fit for this task is the computation of variance-based total effects. An estimator of this quantity from a given sample has sometimes been considered as the holy grail of global sensitivity analysis.

In a recent paper, Devroye, Györfi, Lugosi and Walk show how to estimate the residual variance E[V[Y| Z]] using a nearest neighbor approach.The consequences for obtaining variance-based sensitivity effects from a given sample are immediate. One can simply include or exclude input dimensions from the nearest neighbor search: single dimensions will give first order effects and selecting all but one dimension gives total effects (other combination give subset or superset importance).

Properties and variants of this estimation approach are investigated. Especially in higher dimensions the estimates show a large bias. Bootstrapping techniques together with monotonicity properties are employed to reduce the bias.

Elmar Plischke
elmar.plischke@tu-clausthal.de

Session 11: New methods in sensitivity analysis II

Sensitivity analysis based dimension reduction of multiscale models

A multiscale model couples several single scale models which simulate processes at different temporal and spatial scales [1-3]. In order to develop reliable and trustworthy computational models, usually, Uncertainty Quantification must be performed [4-6]. Often UQ methods require extended computational resources, especially, when the model has a large number of uncertain inputs. However, the coupled structure of multiscale models can provide advantages in UQ. Sensitivity analysis (SA) recognizes the effects of uncertainty in the model inputs to the output parameters [7]. Whenever the variance is a representative measure of model uncertainty, Sobol variance-based method is the preferred approach to identify the main sources of uncertainty [8, 9]. An application of SA to identify ineffectual inputs parameters in order to decrease model dimensionality by equating such parameters to their mean values was demonstrated in [10, 11]. A method to decrease the dimensionality of uncertain inputs of a multiscale model based on SA of its single scale components is discussed in this talk. We demonstrate that in some cases SA of a single scale model provides information on the sensitivity of the final multiscale model output. This then can be employed to reduce the dimensionality of the multiscale model input.

However, the sensitivity of a single scale model response does not always bound the sensitivity of the multiscale model output, which will be demonstrated in two counterexamples. Hence, an analysis of the function defining the relation between single scale components is required to understand whether single scale SA can be used to reduce the dimensionality of the overall multiscale model input space.

Clearly, being the extent of the problem so wide, all possible cases are not considered in this work. However, we argue that in particular cases the method can reduce sufficiently the computational time of UQ by eliminating unimportant parameter equating their values to their means. We will give two examples of the application of this method: a reaction model and the standard Ornstein-Uhlenbeck process.

References

[1]B. Chopard, J.-L. Falcone, A. G. Hoekstra, J. Borgdor , A framework for multiscale and multiscience modeling and numerical simulations, in: Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2011, pp. 2-8. doi:10.1007/978-3-642-21341-0_2.
[2]J. Borgdor , J.-L. Falcone, E. Lorenz, C. Bona-Casas, B. Chopard, A. G. Hoekstra, Foundations of distributed multiscale computing: Formalization, speci cation, and analysis, Journal of Parallel and Distributed Computing 73 (4) (2013) 465-483. doi:10.1016/j.jpdc.2012.12.011.
[3]B. Chopard, J.-L. Falcone, P. Kunzli, L. Veen, A. Hoekstra, Multiscale modeling: recent progress and open questions, Multiscale and Multidisciplinary Modeling, Experiments and Design 1 (1) (2018) 57{68. doi: 10.1007/s41939-017-0006-4.
[4]C. J. Roy, W. L. Oberkampf, A comprehensive framework for veri cation, validation, and uncertainty quanti cation in scienti c computing, Computer Methods in Applied Mechanics and Engineering 200 (25) (2011) 2131-2144. doi: 10.1016/j.cma.2011.03.016. URL http://www.sciencedirect.com/science/article/pii/S0045782511001290(http://www.sciencedirect.com/science/article/pii/S0045782511001290)
[5] L. Uusitalo, A. Lehikoinen, I. Helle, K. Myrberg, An overview of methods to evaluate uncertainty of deterministic models in decision support, Environmental Modelling & Software

63 (2015) 24 { 31. doi:https://doi.org/10.1016/j.envsoft.2014.09.017. URL http://www.sciencedirect.com/science/article/pii/ S1364815214002813
[6] C. Soize, Uncertainty Quanti cation, Springer International Publishing, 2017. doi:10.1007/978-3-319-54339-0.
[7] I. M. Sobol, On sensitivity estimation for nonlinear mathematical models, Matematicheskoe Modelirovanie 2 (1) (1990) 112-118. URL http://www.mathnet.ru/links/85fd3163f401599fc45a5ef55b48a1f5/ mm2320.pdf
[8] A. Saltelli, K. Aleksankina, W. Becker, P. Fennell, F. Ferretti, N. Holst, S. Li, Q. Wu, Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices, Environmental Modelling & Software 114 (2019) 29-39. doi: 10.1016/j.envsoft.2019.01.012. URL http://www.sciencedirect.com/science/article/pii/ S1364815218302822
[9] E. Borgonovo, E. Plischke, Sensitivity analysis: A review of recent advances, European Journal of Operational Research 248 (3) (2016) 869-887. doi: 10.1016/j.ejor.2015.06.032. URL http://www.sciencedirect.com/science/article/pii/ S0377221715005469
[10] I. M. Sobol, Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, Mathematics and computers in simulation 55 (1) (2001) 271-280. doi:10.1016/S0378-4754(00)00270-6.
[11] I. Sobol, S. Tarantola, D. Gatelli, S. Kucherenko, W. Mauntz, Estimating the approximation error when xing unessential factors in global sensitivity analysis, Reliability Engineering & System Safety 92 (7) (2007) 957-960. doi: 10.1016/j.ress.2006.07.001. URL http://www.sciencedirect.com/science/article/pii/ S0951832006001499

Anna Nikishova
a.nikishova@uva.nl
Computational Science Lab, Institute for Informatics
Faculty of Science
University of Amsterdam

Giovanni Eugenio Comi
giovanni.comi@sns.it
Scuola Normale Superiore di Pisa

Alfons G. Hoekstra
A.G.Hoekstra@uva.nl
Computational Science Lab, Institute for Informatics
Faculty of Science
University of Amsterdam

Optimal balance between explorations and repetitions in sensitivity analysis

Sensitivity analysis is well developed for deterministic computer models. When the computer model is stochastic, however, it is less clear what its performance is, or even what it means. In particular, the computer model is repeated several times, say m, at each one of the explorations of the input space. A sensitivity analysis is typically performed with mn(p+1) runs, where n stands for the number of explorations in the input space and p stands for its dimension. Since the computing budget, denoted by T, is limited, we have T=mn(p+1). This is called the budget equation. But there are several couples that satisfy the budget equation. What should we choose? Indeed, it is expected that the performance of the sensitivity analysis depends on this choice. That is, the statistical efficiency of the estimators of the sensitivity indices depends on the choice one makes for the couple (m,n) that satisfies the budget equation.

Our contribution is twofold. First, we show that there is a number of repetitions optimal in a certain sense. Second, we build a method to estimate two kinds of sensitivity indices in the stochastic context. The method is a two-step procedure in which the first step estimates the optimal number of repetitions and the second step estimates the sensitivity indices based on the optimal number of repetitions found in the first step.

We show that this procedure is optimal asymptotically, in the sense that it does arbitrarily as good as the procedure in which the optimal number of repetitions is known. Numerical illustrations are provided to illustrate our theoretical findings.

Gildas Mazo
gildas.mazo@inra.fr
INRA & Université Paris-Saclay

Is it possible to improve existing sample-based algorithm to compute the total sensitivity index?

Variance-based sensitivity indices have established themselves as a reference among practitioners of sensitivity analysis of model output. It is not unusual to consider a variance-based sensitivity analysis as informative if it produces at least the first order sensitivity indices S_j and the so-called total-effect sensitivity indices T_j for all the uncertain factors of the mathematical model under analysis.

Computational economy is critical in sensitivity analysis. It depends mostly upon the number of model evaluations needed to obtain stable values of the estimates. While for the first order indices efficient estimation procedures are available which are independent from the number of factors under analysis, this is less the case for the total sensitivity indices.

When estimating the T_j, one can either use a sample-based approach, whose computational cost depends from the number of factors, or approaches based on meta-modelling/emulators, e.g. based on Gaussian processes. The present work focuses on sample-based estimation procedures for T_j and tries different avenues to achieve an algorithmic improvement over the designs proposed in the existing best practices. We conclude that some proposed sample-based improvements found in the literature do not work as claimed, and that improving on the existing best practice is indeed fraught with difficulties. We motivate our conclusions introducing the concepts of explorativity and efficiency of the design.

Samuele Lo Piano
s.lopiano@gmail.com
Universitat Oberta de Catalunya

Federico Ferretti
federicoferretti@hotmail.it
Joint Research Centre

Arnald Puy
arnald.puy@gmail.com
Université Libre de Bruxelles

Daniel Albrecht
daniel.albrecht@ec.europa.eu
Joint Research Centre

Stefano Tarantola
stefano.tarantola@ec.europa.eu
Joint Research Centre

Andrea Saltelli
Andrea.Saltelli@uib.no
University of Bergen

Poster Session 3: New methods, correlated inputs and new applications of sensitivity analysis

Web-Application for Sensitivity Indices Estimate

The European Commission Competence Centre on Modelling has developed an online application to allow the computation of variance-based sensitivity indices, the so called Sobol' indices (Sobol', 1993), from given Monte Carlo samples.

This Web-App is able to carry out Global Sensitivity Analysis (GSA), that is to compute and apportion the outcome uncertainty to the different model inputs (Saltelli et al., 2008), simply starting from a sample of size N of model factors X (under the assumption of independence of the x- variables) and the corresponding output Y (obtained by the N model executions on the inputs).

The measures of the factor relative importance– main sensitivity effects - are computed by decomposing the output variance (ANOVA decomposition - Sobol', 1993), in terms of partial contributions given by each investigated inputs. These values are normalized, between 0 and 1, obtaining unidimensional measures: the sensitivity indices.

Steps for using the Web-App

Step 1: uploading the data

The user should connect himself to the remote application. From the welcome page it is possible to upload the data file (clicking on BROWSE), that must be a 'csv' file or any ASCII format one, like 'txt'. The file is substantially a matrix (d+1)xN, where d is the number of inputs and (d+1) columns are given by the inputs plus the output. N corresponds to the sample size, which is the number of available cases (rows of the matrix). On the left-hand side of the page, some closed options define the format of the data to be uploaded i.e. which kind of separator has been used (comma, semicolon, tabulation), or the presence of a header line. We strongly recommend using the first row as a header to indicate the name of the variables. In case of a missing header, the variables are denominated as generic xi (e.g. x1;x2;….;xd). The last variable must be the output, indicated automatically with y if no header is used.

When data are successfully uploaded, the first rows/columns of the file are displayed. The number of treated inputs is shown as well as the sample size. Currently, the Web-App can carry out SA with a maximum of 50 independent inputs and one output at a time. The sample cannot exceed the 2^10 rows. Inputs can be either continuous or discrete variables while the output has to be continuous (or at least treatable as a continuous variable).

If the upload fails, the program is not able to recognize the number of variables, a warning message is shown and the rows/columns are not displayed properly.Following a correct uploading, the program is run by clicking on EXECUTE.

Step 2: Execution of the sensitivity analysis

While the program is being executed, the message 'Calculation in Progress. Please wait!' is displayed in red. At the end of this phase, a further message appears to warn the user.
If the program cannot be run correctly, the execution is stopped and an error message is displayed.

Step 3: Analysis of the results

The Web-App provides three types of results which correspond to different tabs. Under the first tab (SA), both sensitivity indices are shown (the main effect index and total order index) for each investigated inputs. In addition, a summary is given about the polynomial chaos expansion (PCE) execution, i.e. the Npce (Number of terms in the expansion) and the Qe2 (amount of variance unexplained by the PCE approximation, a value < 0.10 is recommended). Finally, the output variance decomposition is proposed with the indication of the relevant inputs and interactions.

Clicking on the second tab (SI plot), the first-order sensitivity index (SI) and total order index (ST) values are plotted in a unique graphics. In a second graphic the same values are shown using histograms (in red the SIs and in blue –incrementally – the ST values).
The last Tab (Main Effects) shows the marginal effect of each input (univariate effect) versus the model response. Normally, this allows the investigation of the trends of the effects onto the outcome y.

The EC Web-App appears as a really versatile and user-friendly tool. Sensitivity indices, and corresponding graphics, can be obtained with a few runs of the model. The execution of the program takes only a short time and can be carried out by people without any experience in programming simply uploading the data file.

References

[1] M. Sobol, "Sensitivity estimates for nonlinear mathematical models," Mathematical modelling and computational experiments, vol. 1, no. 4, pp. 407-414., 1993.
[2] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana and S. Tarantola, Global Sensitivity Analysis. The Primer., Chichester: Probability and Statistics, John Wiley and Sons, 2008.
[3] B. Sudret, "Global sensitivity analysis using polynomial chaos expansions," Reliability engineering & system safety , vol. 93, no. 7, pp. 964-979, 2008.
[4] Q. Shao, A. Younes, M. Fahs and T. A. Mara, "Bayesian sparse polynomial chaos expansion for global sensitivity analysis," Computer Methods in Applied Mechanics and Engineering, vol. 318, pp. 474-496, 2017.

Rossana Rosati
rossana.rosati@ec.europa.eu
European Commission, Joint Research Centre (JRC)
Ispra, Italy

Thierry Mara
thierry.mara@univ-reunion.fr
University of Reunion Island
Saint-Denis, France

Ivano Azzini
ivano.azzini@ec.europa.eu

European Commission, Joint Research Centre (JRC)
Ispra, Italy

The variable importance measure index system based on random forest

Random forest is an ensemble learning method for classification, regression and other tanks that operate by constructing a multitude of decision trees. Random forest can be used to rank the importance of variables in a natural way. Variable Importance Measure(VIM) can be implemented to rank or select the predictor variables. Random forest provides two kinds of VIM, i.e., Mean Degrease Gini(MDG) and Mean Decrease Accuracy(MDA)[1]. MDG is the average of the decrease impurity level of samples based on the Gini splitting index of random forest decision tree. Gini splitting index is sensitive to unbalanced samples, and its robustness is poor when data are interrelated[2]. MDA is the average reduction of classification accuracy before and after slight perturbation of variable on the basic of out-of-bag data. The ranking based on MDA has been widely used in many scientific areas. However, it is of interest to select grouped variables rather than to select them individually in many situations. Grouped-variable importance measure (GVIM) based on random forest is presented to enhance the VIM system. After normalizing the MDG, MDA and GVIM indices, the variable importance measure index system based on random forest are established.

The procedure of GVIM can be summarized as follows:
Step 1: Bootstrap sampling from observations to obtain the training data, other observations are "out-of-bag" data.
Step 2: Using the training data to establish a decision tree, the out-of-bag data are used to estimated Mean Squared Error(MSE).
Step 3: Randomly permute the selected grouped-variable predictor from out-of-bag data and keep other predictor values fixed.
Step 4: Estimate the predicted values of the modified out-of-bag data and the corresponding MSE. The decline of the MSE compared with Step 3 is the grouped-variable importance measure.

There are two simulation studies to illustrate the necessary of the GVIM index.
Test 1: A model contains twelve predictor variables, Y=sum(a_i*X_i). all variables X_i are uniformly distributed on the interval [0,1]. The linear correlation coefficient of X_1, X_2, X_3, X_4 are all equal to 0.9 and other variables are irrelevant[3]. The rankings illustrate that the GVIM indices are precise and robust.
Test 2: A model contains ten predictor variables Y=sum(a_i*X_i*X_i+1), where X_i are mutual independent and uniformly distributed on the interval [0,1]. The rankings based on GVIM are related with the coefficients a_i of the model.

Keywords: Variable importance measure; Random Forest; Mean decrease accuracy; Grouped-variable; Gini splitting index

References

[1] Calle M L, Urrea V. Letter to the Editor: Stability of Random Forest importance measures[J]. Briefings in Bioinformatics, 2011, 12(1): 86.
[2] Strobl C, Boulesteix AL, Zeileis A, et al. Bias in random forest variable importance measures: illustrations, sources and a solution. BMC Bioinformatics 2007,8: 25.
[3] Nicodemus K K, Malley J D, Strobl C, et al. The behaviour of random forest permutation-based variable importance measures under predictor correlation[J]. Bmc Bioinformatics, 2010, 11(1): 110.

Ruyang He
he_ruyang@mail.nwpu.edu.cn
Northwestern Polytechnical University
China

Shufang Song
shufangsong@nwpu.edu.cn
Northwestern Polytechnical University
China

Lu Wang
lu.wang@mail.nwpu.edu.cn
Northwestern Polytechnical University
China

A new sensitivity analysis method based on correlation for dependent inputs and outputs

Sensitivity analysis evaluates how the variations in the model output can be apportioned to variations in model inputs. After a long period of development, sensitivity analysis for models with single output has been developed very well, with that for models multiple outputs receiving increasing attention in recent years. For models with multiple outputs, there is not only uncertainty but also correlation among the outputs. Similar to output uncertainty, output correlation is also influenced by uncertainty in the input. If inputs are dependent, correlation between inputs will also contribute to correlation between outputs.

We introduce a new sensitivity analysis technique for models with correlated inputs to measure contributions of inputs to output correlation. The new method allows us to quantitatively distinguish the effects of the correlated and uncorrelated variations of the model inputs on the correlation between outputs. This is achieved by performing off-diagonal elements of covariance decomposition for the correlation contribution of the inputs after decoupling the correlated and uncorrelated parts of the component functions in the high dimension model representation of the outputs. Two numerical examples and a structure model of a rectangular cantilever beam are employed to illustrate the validity and potential benefits of the new indices.

Keywords: Sensitivity analysis; Correlated outputs; Correlation coefficient; Correlated inputs

Liyang Xu
xly@mail.nwpu.edu.cn
School of Aeronautics
Northwestern Polytechnical University

Zhenzhou Lu
zhenzhoulu@nwpu.edu.cn
School of Aeronautics
Northwestern Polytechnical University

Generalized sequential importance sampling for structural reliability analysis of multiple stochastic responses

Estimating the failure probabilities of multiple failure modes is a challenging task in structural reliability analysis. The traditional sequential importance sampling fails to estimate the failure probabilities in a single run, leading to high computational expense and low efficiency. To address this issue, this study develops a generalized Sequence Important Sampling method, in which a unified intermediate event is defined to construct a sequence of intermediate distributions. It iteratively generates samples and makes them gradually approach each failure domain, thus the failure probabilities of multiple failure modes could be estimated simultaneously in a single run.

The simulation techniques are usually applied for structural reliability analysis due to their robustness in handling complex mechanical systems. Among the simulation methods, the crude Monte Carlo Simulation (MCS) is the most basic method for reliability analysis, and independent on the complicacy and dimension of mechanical systems. However, in a context of high level of system safety, the estimation of low failure probability requires a huge demand of samples to achieve an enough accuracy, which is usually intractable, especially for time-demanding models. For these reasons, several variance reduction techniques have been proposed to handle this problem: important sampling (IS) method, directional simulation (DS), line sampling (LS) method, subset simulation (SS) and sequential importance sampling (SIS). Based on the preselected importance sampling density (ISD), IS methods shift the sampling center to the most probable point (MPP) to generate more samples in the failure region, and significantly improve the efficiency for estimating failure probability. However, in high dimensions, IS methods do generally not work due to the fact that samples generated by a fixed ISD impossibly cover the important area of failure domain because this important area is generally unknown and very complex. SS pioneered by Au and Beck has been regarded as one of the most practical reliability method in the last decades. By introducing a sequence of intermediate events, a small failure probability is converted into a product of large conditional failure probabilities, which could be accurately assessed with less computational cost. Iason et al. introduced the SIS that gradually transforms samples generated from the initial distribution to samples exploring the failure region through a sequence of intermediate distributions. This method is originally developed in the context of Bayesian analysis, further applied to structural reliability analysis and reliability-based optimization. Based on the principles of SS and SIS, one can find that the SS could be understood as a special case of SIS if the intermediate distributions are chosen as the optimal ISD of the conditional distributions.

When applied to single failure mode, the traditional sequence important sampling method constructs a series of intermediate sampling density functions based on the limit state function, and utilizes the Markov Chain Monte Carlo method to sample in a step-wise manner, so that the samples gradually approach the failure domain. When there are enough samples in the failure domain, the failure probability can be obtained with high precision. For the multiple failure modes, the intermediate sampling density functions constructed based on each limit state function are different with each other. And the traditional sequence important sampling method unable to select the appropriate intermediate sampling density function to sample to make the obtained samples gradually approach each failure domain. To address this issue, a unified intermediate event concerned all the limit state functions is constructed in the proposed GSIS method, based on which the intermediate sampling density function is constructed. Then, the sampling is iteratively performed with the Markov Chain Monte Carlo method, and the constantly updated samples gradually approach each failure domain. The

failure probabilities of multiple failure modes are obtained simultaneously by a single run of GSIS.

The unified intermediate event constructed in the proposed GSIS method indicates that if any intermediate event for a single limit state function fails, the unified intermediate event must be failed, i.e., the failure domain of the unified intermediate event can be regarded as the union of all the intermediate events. Based on this unified intermediate event, a sequence of intermediate distributions is constructed to sample iteratively, and to drive the samples to simultaneously approach all the failure domains of the failure modes. Hence, all the failure probabilities of the multiple stochastic responses can be obtained in a single run of GSIS. It should be mentioned that if the failure probability of one failure mode is obtained, the limit state function of this failure mode will be removed from the subsequent intermediate distributions. Then, the sampling is iteratively performed based on the new intermediate distributions. The procedure is terminated when the failure probabilities of all the failure modes concerned are obtained.

The efficiency, accuracy and robustness of the proposed GSIS approach have been illustrated with several engineering applications. Compared with the traditional SIS, which has to repetitively perform estimations for each failure mode, and the crude MCS, the proposed GSIS produces proper estimates of the failure probabilities for all the failure modes simultaneously and save a large amount of computational expense. One thing should be mentioned is the improvement on computational efficiency of the proposed GSIS is related to the shared area of the unified intermediate events. The larger the shared area is, the more efficient the GSIS is.

Fuchao Liu
nwpulfc@mail.nwpu.edu.cn
School of Mechanics, Civil Engineering and Architecture
Northwestern Polytechnical University, Xi’an
China

Pan Wang
panwang@nwpu.edu.cn
School of Mechanics, Civil Engineering and Architecture
Northwestern Polytechnical University
China

Improving the estimates of the relative sensitivities of different factors by focusing on parameter sets achieving a good performance

Normal 0
21
false false false DE X-NONE X-NONE

Sensitivity analyses investigate how changes in the model inputs or parameters influence the model outputs. A difficulty when carrying out sensitivity analyses for the model parameters is that the results will depend on the considered parameter ranges. It is can be expected, for example, that larger parameter ranges allow for more variability in the model results, leading to an overestimation of the importance of parameters with larger variation ranges at the expense of parameters with smaller variation ranges. While there are some guidelines about the errors that can be expected for the discharge or precipitation measurement, there is no information about the variability of model parameters that can be considered adequate or reasonable. It is thus difficult to assess the relative importance of model parameters, or between different factors.

This study presents an improvement of the present situation. It investigates the sensitivity of a hydrological rainfall-runoff model with respect to the model parameters, model inputs, observed discharge, the evaluation metrics and the considered period. The analysis is carried out for the 2000 model runs that had the best performance out of a set of 200.000 Monte Carlo runs. As the study set up is based on the observed parameter values, they can be regarded as are representative and adequate for the present purpose. Using these parameter sets, the influence of individual parameters and also the importance of the parameters in comparison to other sources on uncertainty, is assessed.

Another advantage of the approach presented here is that it allows expressing the impact of changes in the different factors in terms of a reduction in performance, unlike most other sensitivity analyses, that need to express their results as a function of model variability as they do not consider the best model parameters.

Carolina Massmann
carolina.massmann@boku.ac.at
Universtity of Life Sciences and Natural Resources
Vienna

Sensitivity analysis for stochastic simulators using differential entropy

This abstract focuses on the sensitivity analysis for a particular kind of computational models called stochastic simulators. Unlike the deterministic ones, stochastic simulators yield a random output for each input because such simulators inherently contain random seeds that make the output random. The stochastic simulator is modeled as a stochastic process M(x,w) where x represents the controllable inputs and w the random seed. This work addresses the global sensitivity analysis on such simulators.

Quantifying the sensitivity on stochastic simulators was addressed in [1] by considering the mean and the variance functions, then performing a GSA on both functions by the mean of the Sobol indices [2]. In this work, the GSA is performed on the differential entropy of the random output. Following the runs of the stochastic simulator (M different points and N realizations on each point), the differential entropy [3] is evaluated on the M points. The GSA is then performed on the differential entropy of the random output, using Sobol indices [2] for instance.

This measure, originally called 'uncertainty' by Shannon [3], infers how confined or spread the random variable is. In addition to the meaning that the entropy held, it enables to get rid of the randomness of the output. Let h be the differential entropy function. The approach consists of performing the GSA on the composed model hoM(x,w).

The proposed method is illustrated through a case study where the aim is to study the exposure of a population to sources of the electromagnetic field (e.g antennas) in virtual cities. In this respect stochastic geometry and advanced computational propagation tools were used to assess the human exposure.

The stochastic city generator generates numerous city realizations based on the same city parameters (street width, anisotropy of the street network, building height). On each realization the exposure of the population to the sources of electromagnetic field is statistically assessed. Hence the exposure, for a given city, can be modeled as a random variable. The aim is to evaluate the sensibility of the exposure (output) to the city parameters (input). One run of the stochastic city generator takes around 3 hours. In this case Sobol indices cannot be derived using Monte Carlo simulations [2]. A polynomial chaos metamodel is rather used and the Sobol indices are derived analytically from the polynomial expansion [4].

Following the simulations on the stochastic city generator over the M points, the differential entropy is evaluated on each point based on the N realizations. A polynomial chaos surrogate model of the entropy is built and the Sobol indices are deduced. Results show that the most influential input is the building height (Sobol index = 0.63).

References

[1] Marrel, A., B. Iooss, S. Da Veiga, and M. Ribatet (2012). Global sensitivity analysis of stochastic computer models with joint metamodels. Stat. Comput. 22, 833–847.
[2] Sobol', I. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Mathematics and Computers in Simulation 55(1-3), 271–280.
[3] Shannon, CE. (1948). A mathematical theory of communication. Bell system technical journal. 27, 379-423 and 623-656.

[4] Sudret, B. (2008). Global sensitivity analysis using polynomial chaos expansions. Reliab. Eng. Sys. Safety 93, 964–979.

Soumaya Azzi
soumaya.azzi@telecom-paristech.fr
LTCI, Télécom ParisTech, Chair C2M

Bruno Sudret
sudret@ibk.baug.ethz.ch
Chair of Risk, Safety and Uncertainty Quantification
ETH Zürich

Joe Wiart
joe.wiart@telecom-paristech.fr
LTCI, Télécom ParisTech, Chair C2M

Variance based global sensitivity analysis of a pharmacokinetic mechanistic model for inhaled compounds in rats

Physiologically based pharmacokinetic models are mechanistic mathematical models used to describe the processes of absorption, distribution, metabolism and excretion (ADME) of a drug in the body. In recent years, the number of publications and submitted new drug applications containing PBPK analyses grew significantly [1,2]. Given their structure, these models are characterized by the presence of uncertainty and variability in the input parameters, and thus, in the model output predictions [3]. For these reasons, there is a strong interest from researchers and regulatory agencies, such as EMA and FDA, in the use of sensitivity analysis to evaluate PBPK models in pharmaceutical research and drug development [4,5].

In this context, we performed a variance based global sensitivity analysis (GSA) on an in-house developed physiologically based model, describing the ADME of inhaled compounds in rats [6]. The model is composed by a part describing the drug pulmonary absorption and another part describing the drug distribution, metabolism and elimination in the body. In this work, we performed two types of GSA, that differ in the considered parameters variability and thus, in the aims: inter-compounds and intra-compound. The aim of inter-compounds GSA is to understand what are the parameters that mostly influence the variability of the model predictions between different drugs [7]. Instead, the aim of the intra-compound GSA is to understand how much the uncertainty associated with the parameters of a given drug impacts the model output uncertainty.

The inter-compounds GSA was performed on the pulmonary absorption model decoupled from the distribution PBPK. In this analysis, the variable parameters are related with the drug dissolution and distribution into the lungs, the drug permeability across the lung tissue, the binding into the cells and systems characteristics, such as lung volume. All the drug related parameters were considered uniformly distributed into a range of values that includes all the internal compound of interest. The considered model outputs were the fraction of the drug that gets absorbed (fa) and the lung-tissue area under the curve (AUC). In this case, we performed one GSA for highly soluble and another one for poorly soluble compounds. The criterion used to divide the compounds into these two classes resembles the well-known one for orally administered compounds [8]. A dose number for inhaled compounds was defined (D0,inh) and it was used to derive the solubility.

We performed the intra-compound GSA on the whole body PBPK model, for three internal compounds of interest (namely A, B and C) that have different molecular characteristics. Here, the parameters that were considered variable are all those of the inter-compounds GSA, plus all the parameters used to describe the processes of distribution, retention in the tissues, hepatic clearance and glomerular filtration. The parameter ranges of variation were defined equal to +-30% of the mean values, except for the active and passive permeabilities across lung tissues and for the blood to plasma ratio (B:P), that were considered equal to +-70% and +-10%, respectively. The considered outputs of interest were the lung tissue and plasma drug concentration AUC.

Concerning the inter-compounds GSA, for highly soluble compounds, the parameter that mostly explain the fa variability is D0,inh, with a total effect (ST) close to 0.9, while for poorly soluble compounds are the mass median aerodynamic diameter (MMAD) and D0,inh (ST $\asymp$ 0.33 and 0.40). For lung AUC, the most important parameters for highly soluble compounds

are the lung tissue binding (LTB) and both passive and active permeabilities, that are parameters that determine the drug retention in the lungs (ST $\asymp$ 0.6, 0.6 and 0.2). For poorly soluble compounds, in addition to the parameters that control the drug retention in the lungs, D0,inh and MMAD (ST $\asymp$ 0.45 and 0.2) are also important. In the intra-compound GSA, for all the considered compounds, the lung AUC variance is mainly explained by the passive permeability variation (ST > 0.5). The most important parameter of plasma AUC for compound A, is the extraction ratio (ST $\asymp$ 0.5), for B are the B:P, the dose and the rat weight (all ST $\asymp$ 0.25) and for C are the dose, the rat weight and B:P (ST $\asymp$ 0.3, 0.2 and 0.25).

In conclusion, in this work we performed two types of GSA that differ in the aim and in the considered parameters variability. The inter-compounds GSA helps in understanding the model general behaviour in the whole space of the parameters and thus, helps to control if the model behaves as expected. Intra-compound GSA, instead, helps in understanding what parameters should be more precisely known (e.g., from in-vitro experiments) to reduce the model output uncertainty.

References

[1] M. Grimstein et al., «Physiologically Based Pharmacokinetic Modeling in Regulatory Science: An Update From the U.S. Food and Drug Administration's Office of Clinical Pharmacology», Journal of Pharmaceutical Sciences, vol. 108, n. 1, pagg. 21–25, gen. 2019.
[2] J. E. Sager, J. Yu, I. Ragueneau-Majlessi, e N. Isoherranen, «Physiologically Based Pharmacokinetic (PBPK) Modeling and Simulation Approaches: A Systematic Review of Published Models, Applications, and Model Verification», Drug Metab Dispos, vol. 43, n. 11, pagg. 1823–1837, nov. 2015.
[3] A. Saltelli et al., Global Sensitivity Analysis. The Primer. John Wiley & Sons, Ltd, 2008.
[4] CHMP (EMA), «Guideline on the reporting of physiologically based pharmacokinetic (PBPK) modelling and simulation», Committee for Medicinal Products for Human Use (CHMP), European Medicines Agency (EMA), London, UK, EMA/CHMP/458101/2016, dic. 2018.
[5] CDER (FDA), «Physiologically Based Pharmacokinetic analyses - format and content: guidance for industry, draft», Center for Drug Evaluation and Research (CDER), U.S. Food and Drug Administration (FDA), Rockville, MD, dic. 2016.
[6] PAGE 27 (2018) Abstr 8549 [www.page-meeting.org/?abstract=8549]
[7] N. Melillo, L. Aarons, P. Magni, e A. S. Darwich, «Variance based global sensitivity analysis of physiologically based pharmacokinetic absorption models for BCS I–IV drugs», J Pharmacokinet Pharmacodyn, vol. 46, n. 1, pagg. 27–42, feb. 2019.
[8] J. E. Hastedt et al., «Scope and relevance of a pulmonary biopharmaceutical classification system AAPS/FDA/USP Workshop March 16-17th, 2015 in Baltimore, MD», AAPS Open, vol. 2, n. 1, pag. 1, gen. 2016.

Nicola Melillo
nicola.melillo01@universitadipavia.it
Department of Electrical, Computer and Biomedical Engineering
Università degli Studi di Pavia
Pavia, Italy

Silvia Grandoni
silvia.grandoni01@universitadipavia.it
Department of Electrical, Computer and Biomedical Engineering
Università degli Studi di Pavia
Pavia, Italy

Nicola Cesari
n.cesari@chiesi.com
Biochemistry and Metabolism Department
Chiesi Farmaceutici S.p.A, Pharmacokinetic
Parma, Italy

Giandomenico Brogin
g.brogin@chiesi.com
Biochemistry and Metabolism Department
Chiesi Farmaceutici S.p.A, Pharmacokinetic
Parma, Italy

Paola Puccini
p.puccini@chiesi.com
Biochemistry and Metabolism Department
Chiesi Farmaceutici S.p.A, Pharmacokinetic
Parma, Italy

Paolo Magni
paolo.magni@unipv.it
Department of Electrical, Computer and Biomedical Engineering
Università degli Studi di Pavia
Pavia, Italy

Global sensitivity based estimability analysis in the parameter optimization of thermodynamic model

In process engineering, different models are developed for prediction, design, optimization or control of a single unit operation or a set of connected unit operations (i.e. plant). These models usually involve several unknown parameters that are very often deduced from experimental measurements by means of a parameter identification process. The latter usually assumes that the available experimental measurements contain the necessary information to accurately identify all the unknown parameters. However, it is well admitted that it is not always the case and the underlying question is to know which parameters are estimable. The objective is therefore to determine the most estimable parameters from the available measurements and possibly to design appropriate experiments to determine the less estimable ones.

In the present paper, the objective is to develop an estimability analysis approach based on global sensitivities and independent of the initial values of the parameters. The case study deals with thermodynamic models used in the predictions of physical properties and liquid-vapor equilibria of concentrated sulfuric acid solutions. More specifically, among the studied models in literature, the Pitzer electrolytic model [1] is chosen.

The local sensitivity approach is first presented, used and discussed. We will then show how its limitations have prompted us to move towards the global approach based on the method developed by Sobol [2] and used to compute the total sensitivity indices. However, one the limitations of this method is that it leads to negative values for first order sensitivity indices especially when their values are very small. For this reason, random balance design [3] method is used to overcome this problem. Global sensitivity indices are then used within an estimability analysis method to rank the parameters from the most estimable to the least estimable.

The most estimable parameters are then identified from the available experimental measurements of water activity and ionic concentrations in the concentrated sulfuric acid solutions. The values of the least estimable parameters are taken from literature or previuous studies. The optimized values of the estimable parameters are then used to compare the model predictions and the experimental measurements. The results show that the estimability approach based on global sensitivities leads to better predictions of experimental measurements using Pitzer model.

References

[1] Pitzer, K. S. (2018). Activity Coefficients in Electrolyte Solutions: 0. CRC press.
[2] Sobol, I. M. (1993). Sensitivity estimates for nonlinear mathematical models. Mathematical modelling and computational experiments, 1(4), 407-414.
[3] Tarantola, S., Gatelli, D., & Mara, T. A. (2006). Random balance designs for the estimation of first order global sensitivity indices. Reliability Engineering & System Safety, 91(6), 717-727.

Ilias Bouchkira
ilyas.bouchkira@um6p.ma
Laboratoire de Modélisation Simulation et Analyse de données
Université Mohammed VI polytechnique
Morocco

Abderrazak M. Latifi
abderrazak.latifi@univ-lorraine.fr
Laboratoire Réactions et Génie des Procédés, CNRS-ENSIC
Université de Lorraine
Nancy Cedex, France

Lhachmi Khamar
lhachmi.khamar@um6p.ma
Laboratoire de Modélisation Simulation et Analyse de données
Université Mohammed VI polytechnique
Morocco

Saad Benjelloun
Saad.BENJELLOUN@um6p.ma
Laboratoire de Modélisation Simulation et Analyse de données
Université Mohammed VI polytechnique
Morocco

Biological diagnosis using global Sensitivity analysis for plant models with correlated parameters

The main aim of plant modeling is to understand the gene-environment interaction so that to aid plant breeding research. To disentangle the genotypic and environmental effects, in plant models each genotype is characterized by a specific vector of model parameters and ideally, these parameters are stable for a certain range of environments. Each parameter can be seen as resulting from the influences of several genes, while the same gene can influence several parameters (pleiotropy). As a consequence, in a family of genotypes (for a same species), there may be strong correlations between parameters.

Moreover, in order to be able to use plant models to characterize and differentiate genotypes, the parametric estimation methods should be precise enough so that the estimation uncertainty remains small and so that statistically significant differences can be detected between the parameters of different genotypes. For this purpose, sensitivity analysis was shown to be a very helpful method in the estimation process, specifically for screening non influential parameters that can be fixed to some nominal values (this is common to all genotypes, the model parameter is thus considered as non-genotypic parameter).

The objective of this paper is to introduce a methodology of global sensitivity analysis (GSA) for plant models with correlated inputs and try to reveal some insights of GSA for biological diagnosis. A genotypic sunflower model called SUNFLO [1][2]is presented for this purpose. The model parameters for a family of 20 genotypes have been estimated or measures with very heavy experimental work and this set of parameter vectors is used to compute the statistical distribution in the parameters space.

In order to distinguish the epistasis and pleiotropy effect, the ideal method should be able to quantitatively compute 3 sensitivity indices for one parameter: 1.The total correlated sensitivity index. It can explain the pleiotropy effect. 2. The main sensitivity index including the correlated and uncorrelated part. It can explain the epistasis effect. 3. The total uncorrelated sensitivity index. It indicates gene-environment interaction.

These 3 indices are not complementary or we cannot say they are a kind of decomposition of the total effect of one parameter, but they somehow overlap to each other. It is reasonable considering the biological explanation we are aiming to reach by sensitivity analysis.

Sobol's indices with dependent inputs are suitable in our case. We referred [3][4] for the computation of these indices. However, in our case the computation cost of the model evaluation is time consuming, and the algorithm for computing the indices involving large amount of model evaluations. We then turn to a new Morris factor screening algorithm with correlated inputs as the first step of analysis. The algorithm was performed as 3 steps: 1. Using orthogonal transformation to transform the factor into independent distribution. 2. Performing elementary effect computing within the optimized trajectory 3.Transform the elementary effect matrix in step2 by the de-orthogonal transformation, then compute the mean value and standard derivation of the elementary effect. The mean value and standard derivation is the final Morris result. We performed analytical test for our new Morris method and compared it with the one in [5]. Our method produces 2 indices for one parameter which remains the basic frame of original Morris method and is convenient for factor screening by two dimensional mapping. While in [5], the number of indices is 4, which means we need to process the four dimensional mapping to decide the screening. The computing cost of our

Morris method is N(k+1), while in [5], it is N(3k+1) due to the sampling strategy. As such, our method is more efficient and effective.

The SUNFLO model output is the biomass production for the crop sunflower by inputting environmental information, mainly temperature and precipitation. It simulated the plantʾs phenology, the leaveʾs development, the accumulation and distribution of biomass, and the production of seeds. For the aim of simplification, we donʾt take into account the water budget and nitrogen constraints in our model. And the temperature was optimized for the calculation of RUE(Radiation Use Efficiency).

We took out 12 genotypic parameters from SUNFLO model. Parameter distribution characteristic and correlation matrix was obtained from the 20 values of each parameter from 20 genotypes. In the first step, we detect 7 parameters to be screened. Then in the second step, we get 3 biological meaningful indices for the 5 parameter remained. Two genetic parameters keep high indices in both Morris and Sobolʾs analysis:M3, thermal date at seed physiological maturity and HIseed, harvest index of the seed. By comparing the 3 indices, we found that the sensitivity contribution of M3 is mainly caused by the main effect and its correlation with the other parameters, which means M3 has strong pleiotropy effect but rarely gene-environment interaction. Biologically, we can find the QTL(quantitative trait locus) of M3, and consider it as the none gene-environment sensitive trait. It is a heuristics step forward of connecting the gene data to plant models at the organ level. And it could enlighten the application of plant models for virtual breeding.

References

[1] J. Lecoeur, R. et al. (2011) ‘‘Quantifying physiological determinants of genetic variation for yield potential in sunflower. SUNFLO: a model-based analysis’’. Functional Plant Biology, 38:246–259.
[2] Casadebaig, P., et al. (2011). "SUNFLO, a model to simulate genotype-specific performance of the sunflower crop in contrasting environments." Agricultural and Forest Meteorology 151(2): 163-178.
[3] Kucherenko, S., et al. (2012). "Estimation of global sensitivity indices for models with dependent variables." Computer Physics Communications 183(4): 937-946.
[4] Zhang, K., et al. (2015). "A new framework of variance based global sensitivity analysis for models with correlated inputs." Structural Safety 55: 1-9.
[5] Ge, Q. and M. Menendez (2017). "Extending Morris method for qualitative global sensitivity analysis of models with dependent inputs." Reliability Engineering & System Safety 162: 28-39.

Zhongming Xu
xuzhongming17@mails.ucas.ac.cn
Wuhan Institute of Physics and Mathematics
Chinese Academy of Sciences

Qiongli Wu
wuqiongli@wipm.ac.cn
Wuhan Institute of Physics and Mathematics
Chinese Academy of Sciences

A software design for flexible model composition with global sensitivity analysis

The implementation of simulation models composed of many interacting sub-models poses a challenge for the modeller during both model formulation and software construction. I present a solution to this problem in the form of Universal Simulator, an open-source software and programming framework). In this design, model building blocks are defined as C++ classes all derived from a common base class supplied by the framework. A specific model is composed from these classes as a hierarchy of objects connected by inputs and outputs. A simple domain-specific language ‘BoxScript’ is used to compose models, including parameter values and the specifics of sensitivity analysis, such as parameter distributions and sets of alternative sub-models. Post-processing of simulation outputs, including graphics and sensitivity analysis, is integrated with R. I will present the design principles behind and the use of Universal Simulator applied to a simple food web model and a complex model of the indoors climate in a commercial greenhouse.

Niels Holst
niels.holst@agro.au.dk
Department of Agroecology
Aarhus University
Denmark

Parallel Session 12 Sensitivity analysis in spatio-temporal modelling II

Shapley effects for sensitivity analysis with correlated inputs

The global sensitivity analysis methods [3] allow to determine which input variables of a numerical model contribute the most to the variability of the model outputs, or on the contrary which are not important and possibly which variables interact with each other. The standard quantitative methods compute the variance-based sensitivity measures also called Sobol' indices. In the case of independent scalar inputs $T \in Rd$), the interpretation of the Sobol' indices is simple because the variance decomposition of the model output $Y = f(X)$ is unique [10]. As the sum of all Sobol' indices is equal to one, the indices are interpreted as proportions of explained variance. However, in many applications, it is common that the input variables have a statistical dependence structure imposed by a probabilistic dependence function [5] (e.g., a copula function) or by physical constraints upon the input or the output space [4]. As discussed for example in [8, 2], estimating and interpreting Sobol' indices with correlated inputs is not trivial .

The Shapley value, introduced in [9], is a solution concept in cooperative game theory and a powerful tool in economical modeling. It consists in fairly distributing both gains and costs to several actors working in coalition. The Shapley value applies primarily in situations when the contributions of each actor are unequal. The Shapley value ensures each actor gains as much or more as they would have from acting independently. In [6], Art Owen has proposed to use that concept to measure variable importance in sensitivity analysis of model outputs. Then, the so-called Shapley effects can be interpreted as an importance measure of model input if the actors are identified with a set of random model inputs and the value assigned to each coalition is identified to the explanatory power of the subset of model inputs composing the coalition. In the framework of dependent input variables, the Shapley effects allow namely to bypass the intricate issue of variance decomposition [11, 7, 2].

The two main properties and advantages of the Shapley values in the frame- work of variable importance measures are that they cannot be negative and they sum-up to the total output variance. The equitable principle driving the allocation rule states that an interaction effect is equally apportioned to each input involved in the interaction.

From the conceptual point of view, the major issue then remains to under- stand the effect of the dependence between inputs on the variance-based Shapley values. The aim of the work presented in this talk is to propose a thorough in- vestigation of several particular cases, simple enough so that we can provide some guidance to their interpretation. For the sake of practical applications, [11] has proposed two estimation algorithms of the Shapley values. In this com- munication, we will numerically study the convergence of these algorithms for estimating Shapley effects. As for the Sobol' indices, one important issue in practice is the numerical cost in terms of number of model evaluations required to estimate the Shapley effects. To ease the computational burden, a classical solution is to use a metamodel which is a mathematical approximation of the numerical model from an initial and limited set of runs. Therefore in this com- munication, we will also synthesize and pursue the efforts of [2, 1] about the metamodel-based algorithms for estimating the Shapley values.

References

[1] N. Benoumechiara and K. Elie-Dit-Cosaque. Shapley effects for sensitivity analysis with dependent inputs: bootstrap and kriging-based algorithms. ESAIM: Proceedings and Surveys, 2019, To appear.
[2] B. Iooss and C. Prieur. Shapley effects for sensitivity analysis with cor- related inputs: comparisons with sobol' indices, numerical estimation and applications. International Journal for Uncertainty Quantification, 2019, To appear.
[3] B. Iooss and A. Saltelli. Introduction: Sensitivity analysis. In R. Ghanem, D. Higdon, and H. Owhadi, editors, Springer Handbook on Uncertainty Quantification, pages 1103–1122. Springer, 2017.
[4] S. Kucherenko, O.V. Klymenko, and N. Shah. Sobol' indices for problems defined in non-rectangular domains. Reliability Engineering and System Safety, 167:218–231, 2017.
[5] D. Kurowicka and R. Cooke. Uncertainty analysis with high dimensional dependence modelling. Wiley, 2006.
[6] A.B. Owen. Sobol' indices and Shapley value. SIAM/ASA Journal on Uncertainty Quantification, 2:245–251, 2014.
[7] A.B. Owen and C. Prieur. On Shapley value for measuring importance of dependent inputs. SIAM/ASA Journal on Uncertainty Quantification, 5:986–1002, 2017.
[8] A. Saltelli and S. Tarantola. On the relative importance of input factors in mathematical models: Safety assessment for nuclear waste disposal. Jour- nal of American Statistical Association, 97:702–709, 2002.
[9] L.S. Shapley. A value for n-persons game. In H.W. Kuhn and A.W. Tucker, editors, Contributions to the theory of games II, Annals of mathematic studies. Princeton University Press, Princeton, NJ, 1953
[10] I.M. Sobol. Sensitivity estimates for non linear mathematical models. Math- ematical Modelling and Computational Experiments, 1:407–414, 1993.
[11] E. Song, B.L. Nelson, and J. Staum. Shapley effects for global sensitivity analysis: Theory and computation. SIAM/ASA Journal on Uncertainty Quantification, 4:1060–1083, 2016

Clémentine Prieur
clementine.prieur@univ-grenoble-alpes.fr
Grenoble Alpes University

Bertrand Iooss
bertrand.iooss@edf.fr
Electricité de France R&D

Shapley values for sensitivity analysis: some recent advances

In this work we present recent results on the use of Shapley values as a tool for sensitivity analysis of computer experiments and machine learning methods.We investigate the Shapley values generated from moment-independent and value-of.information based sensitivity measures. Then, we introduce and study the Shapley-Owen effects for investigating interactions and their nature, providing theoretical connections to the superset importance indices and to the moments of the dimension distribution.

In the last part of the work, we propose a theoretical framework for the use of Shapley values in neural networks models.

Giovanni Rabitti
giovanni.rabitti@phd.unibocconi.it
Bocconi University

Emanuele Borgonovo
emanuele.borgonovo@unibocconi.it
Bocconi University

Cascade Sensitivity Measures

Sensitivity measures quantify the extent to which the distribution of a model output is affected by changes in individual (random) input factors. Sensitivity measures are thus often constructed via partial derivatives either of outputs with respect to inputs ("local" sensitivity measures) or of an output summary in the direction of random inputs. One drawback of such sensitivity measures is that they do not fully account for interactions among or statistical dependence between input factors and extensions have so far focused on higher order derivatives [1]. However, dependence structures between input factors might substantially impact the sensitivity to an input factor and we argue that for statistically dependent input factors, a stress on one input should also precipitate stresses in other input factors. We propose a novel sensitivity measure, termed cascade sensitivity, which explicitly accounts for the direct and indirect dependence between input factors. The proposed cascade sensitivity is defined as the partial derivative of a distortion risk measure applied to the output, in the direction of a stressed random input factor; hence it is closely related to the approaches of [2,3]. Distortion risk measures are quantile based measures of risk subsuming expected utility and the two most common risk measures in financial risk management, the Value-at-Risk and the Tail-Value-at-Risk. Even though the cascade sensitivity is defined through a directional derivative, the derivative is taken after a transformation of the random vector of inputs, that enables a stress on one input factor to propagate through the entire input vector, changing all its components according to the input vector's dependence structure. Specifically, our cascade sensitivity framework is underpinned by a variation of the inverse Rosenblatt transform, which is applied to the random input vector.This transformation of the input vector allows the cascade sensitivity to fully capture the impact of dependence between input factors. Thus, stressing an input impacts on the output risk measure both directly and indirectly, via the generated cascade of stresses on other (dependent) inputs. In particular, the cascade sensitivity of an input factor decomposes into components, each reflecting the direct or indirect contribution of another input factor to the sensitivity of the output. Thus, the decomposition of the cascade sensitivity provides a quantification of the contribution of each individual input factors' indirect effects – stemming from the dependence with the one being stressed – to the cascade sensitivity.

A sensitivity measure that fully reflects the dependence of the random vector of inputs is useful in applications where the dependence structure of the inputs is of particular interest, such as but not limited to risk management. In these applications, inputs are often high dimensional and their aggregation highly non-linear. Moreover, the calculation of the distribution of the output proceeds by Monte Carlo simulation, which requires costly evaluation of the model's aggregation function at different simulated scenarios. We provide explicit analytical representations of the cascade sensitivity, which do not require calculation of the gradient of the aggregation function and allow for a straightforward implementation on a single Monte Carlo sample, thus avoiding multiple evaluations of the model's aggregation function. These alternative representations of the cascade sensitivity are derived for two types of stress on an input factor: a) a perturbation of the distribution of an input factor, such that the stressed input factor follows a mixture distribution, and b) an additive random shock applied to the (tail of the) input factor itself. Thus, our proposed cascade sensitivity framework is practically useful, as we illustrate through an application to a commercially used London Insurance Market portfolio model.

References

[1] Borgonovo, Emanuele, and Elmar Plischke. "Sensitivity analysis: a review of recent advances." European Journal of Operational Research 248.3 (2016): 869-887.
[2] Hong, L. Jeff. "Estimating quantile sensitivities." Operations Research 57.1 (2009): 118-130.
[3] Tsanakas, Andreas, and Pietro Millossovich. "Sensitivity analysis using risk measures." Risk Analysis 36.1 (2016): 30-48.

Silvana Pesenti
silvana_pesenti@hotmail.com
University of Toronto

Pietro Millossovich
Pietro.Millossovich.1@city.ac.uk
Cass Business School

Andreas Tsanakas
A.Tsanakas.1@city.ac.uk
Cass Business School

Parallel Session 13: New methods in sensitivity analysis III

Accounting for Correlation Effects in Global Sensitivity Analysis through Variogram Analysis of Response Surfaces

Global Sensitivity Analysis is useful in characterizing the behavior of advanced environmental models, which involve a considerable number of parameters representing different components of natural systems. These parameters are often correlated and their joint distribution may follow a variety of forms. Most GSA applications, however, assume that the parameters are independent, for mathematical and technical convenience. Methods to handle correlated parameter in GSA are limited, in general.

This presentation extends a recently developed ‘‘variogram-based’’ theory for GSA, called Variogram Analysis of Response Surfaces (VARS), to properly account for correlation effects in the model input/parameter space. This extension results in a new implementation, which generalizes the star-based sampling (STAR) technique to allow sampling from multivariate distributions. The new implementation called gSTAR-VARS involves four steps: (1) applying inverse Nataf transformation to sample ‘‘star centers’’ based on the multivariate probability distribution function, (2) sampling ‘‘star points’’ from conditional distributions based on the star centers, (3) extracting pairs and clustering them into different ‘‘perturbation scales’’, and (4) approximating directional variograms and covariogram for the generation of VARS indices for global sensitivity; gSTAR-VARS simultaneously generates a range of sensitivity indices, including total-variogram effects, total-order effects, and elementary effects.

The gSTAR-VARS is first validated by comparing its sensitivity results to those of other existing methods developed to deal with correlated parameters across multiple test functions. These comparisons show very good agreement between gSTAR-VARS and other methods. Second, gSTAR-VARS is applied to a real case study that involves an integrated water management model for the Saskatchewan River Basin in Canada. The model includes a large number of inputs, some of which are highly correlated. The results show the effectiveness and efficiency of gSTAR-VARS in properly accounting for correlation effects in GSA.

Nhu Cuong Do
nhu.do@usask.ca
Postdoctoral Fellow
Global Institute for Water Security
University of Saskatchewan
Saskatchewan, Canada

Saman Razavi
saman.razavi@usask.ca
Assistant Professor
Global Institute for Water Security & School of Environment and Sustainability
University of Saskatchewan
Saskatchewan, Canada

Sensitivity Analysis: The Cut-HDMR method

The mathematical model of a physical system generally provides one or more output response(s), the result(s) of a mathematical function, depending on the values of input variables, sometimes many input variables. This multivariate representation may suffer from the curse of dimensionality as the number of sample points needed to build the mathematical model function with a sufficient accuracy increases exponentially with the dimensionality. Although most physical systems model outputs generally do not show particular difficulties from the mathematical point of view, efficient high dimensional model representations methods are indispensable when we want to reduce the computational effort needed in representing the input‐output relationships of a physical system.
Various forms of HDMR can be constructed for different objectives. Among them, two types of HDMR are considered and used in the present paper.

First, the ANOVA-HDMR - the same as the analysis of variance (ANOVA) decomposition used in statistics - is a representation of a mathematical function where the multivariate terms (component functions) - often made of orthonormal polynomials - are summed to approximate the original function [y = S f(xi) + S f(xi,j) + …]. Moreover, this addition allows looking at the model output variance as a sum of the variances of the single or interacting inputs [s2 = Ssi2 + Ssi,j2 + …]. The decomposition thus helps depicting, and understanding the relationships between the inputs and the output of the original mathematical function.

Second, the cut-HDMR is a slight variation of the HDMR and is being known[1] for more than twenty years. In particular, a study[2] made in 2012 showed that cut-HDMR is computationally more efficient than the ANOVA decomposition. The cut-HDMR terms (component functions) are defined along cut lines, planes, hyperplanes, etc. across a reference point x in the input space.
The cut-HDMR approach can either approximate the original function by an addition (cut‐HDMR‐additive), or a multiplication (cut‐HDMR‐multiplicative) of terms (component functions). In particular, the cut‐HDMR‐multiplicative technique is attractive as its separable nature makes it able to approximate the original multi‐dimensional integration as a product of a series of one-dimensional integrals.

Three reference functions were proposed for the HDMR based analysis. One additive function: namely a mere sum of weighted variables; one multiplicative function: a weighted product of variables showing thus possible interaction(s); and eventually one third function, a mix of the first two functions with weighted terms that make the function featuring any desired proportion of additive and interactive characteristics.

The three reference functions differently combine the xi variables whose values are in the range [0.0, 1.0] and are generated by a quasi-random generator. A quasi-random generator, provides regularly spaced sample points, avoids clusters of points, gaps, and guarantees the sampling of the whole input space.

The three reference functions were decomposed by the HDMR, the cut-HDMR-additive and the cut‐HDMR-multiplicative representation techniques. The respective HDMR derived terms (component functions) were used to approximate the original mathematical function, or model function, to calculate the variances, and the sensitivity indices against sample size. The obtained values of the approximated functions, the convergences of the variances and the sensitivity indices values are compared and evaluated against theoretical values and sample size.

Comparison of the results obtained with the three different HDMR techniques confirmed that cut‐HDMR is able to approximate a mathematical model function (additivity, interactions, …).

Moreover, from a comparison of the behaviour of an unknown mathematical model output with the three reference functions outputs characteristics, some valuable information may come out helping to draw at a glance conclusions about the additive vs interaction properties of the unknown model and this just after a sole analysis of its output values, with little or no extra computational cost.

Keywords:

Model, modelling, model run time, global sensitivity analysis, variance, sensitivity indices, high dimensional model representations, HDMR, cut-HDMR, Monte‐Carlo simulations, additive models, interactions

References

[1] Global uncertainty assessments by high dimensional model representation (HDMR) by Genyuan Li, Sheng-Wei Wand, Herschel Rabitz, Sookyun Wang, Peter Jaffé, 2002,
[2] Xufang Zhang, Department of Civil and Environmental Engineering, University of Waterloo, Waterloo, ON, Canada, N2L 3G1, 2012,

Daniel Albrecht
daniel.albrecht@ec.europa.eu
EU Commission, Joint Research Center
DG.JRC.I.2-CC-MOD-SAMO

Stefano Tarantola
stefano.tarantola@ec.europa.eu
EU Commission, Joint Research Center
DG.JRC.C.3

An efficient approach for estimating quantile-based global sensitivity measures

Global sensitivity analysis is a very useful tool to evaluate the influence of input variables in the whole distribution range. However, conventional global sensitivity measures can't focus on some certain regions of output value. Kucherenko et al. proposed two global sensitivity measures based on quantiles of the output [1] and used two diff erent Monte Carlo estimators to estimate the measures. The measures can be seen as expectations over the range of x_i of L1 norm and square of L2 norm on a set of alpha-th quantile of the output CDF and alpha-th quantile of conditional output CDF on the input x_i being fixed at X_i, correspondingly. However, when the quantile level alpha is very small or high, i.e., alpha 0.95, the required sample will be quite large by using aforementioned methods. Thus, we propose a new approach to estimate quantile based global sensitivity measures efficiently by utilizing importance sampling based subset simulation(Subsim-IS) and surrogate models.

Importance sampling based subset simulation are presented by Song [2], whose aim is to deal with small failure probability problems. Introducing a set of intermediate failure events, Subsim separates the original probability space into a sequence of subsets, and then the small failure probability can be expressed as a product of larger conditional failure probabilities. Since the quantile level alpha is also probability, similar strategy can be adopted as well.

Assume alpha= which follows descending sequence and Y=g(x), and we can define a set of regions Q_i={x: g(x)<=q_Y(alpha_k)} (k=1,2, …, m). According to the multiplication theorem and the definition of conditional probability in the probability theory, the following equation holds: alpha_m=alpha_1*{alpha_2/alpha_1}*…*{alpha_m/alpha_(m-1)} =P(Q_1)*{P(Q_2|Q_1)}*…*{P(Q_m|Q_(m-1))}. Define conditional probabilities p_1=alpha_1 and p_k=alpha_k/ alpha_{k-1}, and alpha_k can be expressed by alpha_k=p_1*p_2*…*p_k. The detailed procedures of Subsim-IS for estimating quantiles are as follows:
* Generate N_1 samples according to the joint PDF of inputs.
* Compute the corresponding response values, and reorder these values in ascending order. Calculate the alpha_1-th quantile and define the first subset Q_1={x: g(x)<=q_Y(alpha_1)}.
* From these {p_k-1}*{N_k} (k=2,3, …, m) conditional samples which lie in the Q_{k-1}, choose the point with maximum joint PDF values as the sampling center of IS density function.
* Generate N_k (k=2,3, …, m)samples by IS density function and find M_k points that lie in Q_{k-1}. Then find the p_k-th conditional quantile, which is also global alpha_k-th quantile.
* Repeat step 3 and step 4 until all required quantiles have been solved.
The procedures of estimating conditional quantiles are similar, but we need use the strategy of double loop reordering (DLR) approach in [1] to generate the conditional distribution samples. Finally, the quantile-based sensitivity measures can be estimated after solving all quantiles.

It is noted that the proposed procedures are used in the small quantile level case, and little modifications are needed when dealing with large quantile level problems. Also, when model input-output relationship is rather complex or implicit, surrogate model such as adaptive kriging can be used to replace the real model, which contribute to reducing computational workload further.

Keywords: Global sensitivity analysis, Quantile based global sensitivity measure, Importance sampling based subset simulation, Surrogate model

References

[1] Kucherenko S, Song SF, Wang L. Quantile based global sensitivity measures. Reliab Eng Syst Saf 2019; 185: 35–48.
[2] Song SF, Lu ZZ, Qiao HW. Subset simulation for structural reliability sensitivity analysis. Reliab Eng Syst Saf 2009;94(2):658–65.

Zhiwei Bai
bzw1020@163.com
Northwestern Polytechnical University
China

Shufang Song
shufangsong@nwpu.edu.cn
Northwestern Polytechnical University
China

Lu Wang Wang
lu.wang@mail.nwpu.edu.cn
Northwestern Polytechnical University
China

Sergei Kucherenko
s.kucherenko@imperial.ac.uk
Imperial College London
UK

Parallel Session 14: New methods in sensitivity analysis and correlated inputs I

New indices for Multivariate Sensitivity Analysis using fuzzy clustering

Many simulation models produce temporal or spatial data. Sensitivity analyses of such models require the use of the so-called Multivariate Sensitivity Analysis (MSA) methods that are often based on a dimension reduction principle. Model outputs are typically projected onto predefined or data-driven orthogonal bases such as polynomial or eigen vectors [1,2]. Classical scalar sensitivity analysis methods applied on the coefficients of the different components may give some rich information on the impact of model input factors on the variation of the shape of dynamic or spatial model outputs. Generalized Sensitivity Indices (GSIs) ([2,3]) can also be used to rank the overall contribution of input factors on the variability of these model outputs. The use of projection-based methods is however so far limited by the selection of the associated bases which is constrained by orthogonality requirements. Indeed, these bases does not always allow to extract relevant and interpretable information on structural properties of multivariate outputs. More applicable MSA methods are thus expected to be developed [4].

Clustering techniques have been designed to identify groups of similar objects in multivariate data sets. They may thus be particularly adapted to capture the variability of behaviors of model multivariate outputs. While binary clustering has been extensively used in scalar SA to assess the importance of factors leading to region of interest [5], there is still a need of quantitative sensitivity analysis methods taking benefit of multivariate outputs clustering with any number of clusters.

In this work, we propose to make use of a fuzzy clustering procedure to enhance results of MSA on model with multivariate outputs. The main idea relies on the extensive use of the output of a fuzzy clustering method: the so-called membership functions (MF, valued in [0,1]), which quantify for any model response the degree of membership to each cluster. Membership function exactly correspond to posterior probability of membership produced by mixture-based clustering methods.

Our approach is based on the analogy between MF and classical MSA projections. Indeed, MF provide a kind of decomposition, as they sum up to one. However, they are different to basis decomposition as they do not provide an additive decomposition among orthogonal components. Nevertheless, this analogy allows to extend MSA indices to clustered outputs. More precisely, we introduce:
* MF-Sensitivity Indices (MF-SI) : SI on a MF of a given cluster. They allow to answer the question ‘which parameters influence the membership to a given cluster?’
* dMF-Sensitivity Indices (dMF-SI) : SI on pair-wise MF differences. They allow to answer the question ‘which parameters drive the output from one cluster to another?
* MF-Generalized sensitivity indices (MF-GSI) : GSI computed on a MF vector to answer the question ‘‘What is the overall contribution of the parameters wrt change of behavior / clusters.’

We present the computation of these indices on a dedicated toy model producing temporal signals with one or two maxima in response to five parameters and show that the model behavior can be efficiently reported by the newly proposed indices.

References

[1] Campbell, K., McKay, M. D., & Williams, B. J. (2006). Sensitivity analysis when model outputs are functions. Reliability Engineering & System Safety, 91(10-11), 1468-1472.
[2] Lamboni, M., Monod, H., & Makowski, D. (2011). Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models. Reliability Engineering & System Safety, 96(4), 450-459.
[3] Lamboni M., (2018). Multivariate sensitivity analysis: minimum variance unbiased estimators of the first-order and total-effect covariance matrices. Reliability Engineering & System Safety 06/2018;, DOI:10.1016/j.ress.2018.06.004.
[4] P. Wei, Z. Lu, J. Song, Variable importance analysis: a comprehensive review, Reliability Engineering & System Safety 142 (2015) 399–432.
[5] Raguet, H., & Marrel, A. (2018). Target and Conditional Sensitivity Analysis with Emphasis on Dependence Measures. arXiv preprint arXiv:1801.10047.

Sebastien Roux
sebastien.roux@inra.fr
INRA (French Institute for Agricultural Research)

Matieyendou Lamboni
matieyendou.lamboni@univ-guyane.fr
Université de Guyane
France

Samuel Buis

Sensitivity indices and their upper bounds for the models with correlated input factors

Variance-based sensitivity analysis (Vb-SA) [1-3] and Multivariate Sensitivity Analysis (MSA) [4-5] provide the first-order and total sensitivity indices for single and multivariate response models, respectively. In the presence of dependent or correlated input factors, some mathematical methodologies have been developed to deal with that issues for single response models [6-7]. However, the proposed methodologies distinguish two total indices for a given input factor, with a possible inconsistency when ranking input factors. In this abstract, we propose a new methodology for Vb-SA for models with correlated input factors. Our methodology allows for defining only one first-order sensitivity index and one total sensitivity index for each input, with the former less than the latter. We also provide the upper bound of the total index, which should be used for a screening purpose.

Main results

Our methodology relies on the ability to model dependent input factors (X) using equations. For d correlated input factors (X), we show that there exists a function r, a vector of d-1 independent and normally distributed variables Z such that

* Xj and Z are independent;
* $X_{\sim j} = r(X_j, Z)$,

where $X_{\sim j}$ stands for all inputs except Xj.

Point ii) expresses X as a function of independent variables, and it can be used to obtain the derivatives of each component of $X_{\sim j}$ with respect to Xj. These derivatives are interesting for defining the derivative global sensitivity measure (DGSM) [8-9] for correlated input factors and for proposing an upper bound of the total index latter.

Namely, let M stands for a response model. Our model with correlated input factors can be written as

$Y = M(X)$ subject to $X_{\sim j} = r(X_j, Z)$,
or equivalently
$Y = M(X_j, r(X_j, Z))$.

The model $Y = M(X_j, r(X_j, Z))$ includes only independent input factors. Therefore, we use the classical Vb-SA to define the first-order and total sensitivity indices (SIs) of Xj and the upper bound of the total index [9-10]. As a matter of fact, we have this well-known relationship $S_j \leq S_{T.j} \leq UB_j$, where Sj is the first-order index, ST.j is the total index and UBj is the upper bound of the total index. For optimizing the number of models runs for estimating all indices, we derive new expressions of the Sis and the upper bound using only the original dependent input factors. The new expressions are obtained by combining the classical definition of SIs with some change of variables.

Conclusion

We propose a new Vb-SA for models with correlated input factors. Our methodology yields to the definitions of only one first-order index, one total index, and one upper bound for each input factor. The proposed indices are coherent with the Sis for independent input factors, and they should be used for ranking input factors without any troublesome. In the next future, it is interesting to i) compare our new indices with those proposed in [6-7], ii) propose efficient estimators of our indices, iii) extend this mathematical methodology to cope with the multivariate response models and functional outputs with dependent input factors.

References

[1] Lamboni M. (2018) Uncertainty quantification: A Minimun Variance Unbiased (joint) Estimator of the non-normalized Sobol's indices. Stat. Papers.
[2] Sobol, I. M. (1993) Sensitivity estimates for nonlinear mathematical models', MMCE 1, 407–414.
[3] Saltelli A., Ratto M., Andres T., Campolongo F., Cariboni J., Gatelli D., Salsana M., Tarantola S. (2008) Global sensitivity analysis - The primer. Wiley.
[4] Lamboni M., Monod H., Makowski D. (2011) Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models. Reliability Engineering and System Safety 96: 450 - 459.
[5] Lamboni, M. (2018) Multivariate sensitivity analysis: Minimum variance unbiased estimators of the first-order and total-effect covariance matrices. RESS.1016/j.ress.2018.06.004.
[6] Kucherenko S., Tarantola S., Annoni P. (2012) Estimation of global sensitivity indices for models with dependent variables. Comput Phys Commun 183(4):937–946.
[7] Mara T.A., Tarantola S., Annoni P. (2015) Non-parametric methods for global sensitivity analysis of model output with dependent inputs. Environ Model Softw 72:173–183.
[8] Kucherenko S., Rodriguez-Fernandez M., Pantelides C., Shah N. (2009) Monte Carlo evaluation of derivative-based global sensitivity measures. Reliability Engineering and System Safety 94 (2009) 1135– 1148.
[9] Sobol I. M., Kucherenko S., Derivative based global sensitivity measures and the link with global sensitivity indices, Mathematics and Computers in Simulation 79 (2009) 3009-3017.
[10] Lamboni M., Iooss B., Popelin A.-L., Gamboa F., Derivative-based global sensitivity measures: General links with Sobol' indices and numerical tests, Mathematics and Computers in Simulation 87 (0) (2013) 45-54.

Matieyendou Lamboni
matieyendou.lamboni@univ-guyane.fr
Université de Guyane
France

Sergei Kucherenko
s.kucherenko@imperial.ac.uk
Imperial College London
UK

Regional sensitivity analysis for multivariate outputs based on generalized variance

In engineering structural systems, there exist various kinds of uncertainties relating to material performance, external loads, structural dimension parameters and so on. Global sensitivity analysis techniques measure how the whole uncertainties of model input variables can affect the uncertainties of model output, and provide valuable information for structural design, thus playing a significant role in robust optimization and model simplification. Currently, global sensitivity analysis is mainly focused on three aspects: variance-based method proposed by Saltelli et al. (2008) and Sobol (2001), Morrisʾs screening method (Morris 1991) which is suitable for high-dimensional models, and Borgonovoʾs moment-independent method (Borgonovo 2007). A large amount of research has been done on the above-mentioned GSA methods including theory modification and efficient computational methods. However, although GSA can quantify the average effect of model inputs on output performance considering the whole distribution range of input variables, it fails to capture the regional uncertainty information of input variables, which may also be important for engineering decision-makers.

Regional sensitivity technique is then developed for solving this problem. It quantifies the effect of the subranges of model inputs on model output performance and provides detailed information about the links between internal uncertainties of input variables and model output, thus can be regarded as a complementary for GSA technique. Recently, many researchers have done work on theories about regional sensitivity analysis. Sinclair (1993) and Bolado-Lavin et al. (2009) defined contributions to sample mean (CSM) plot for measuring the ratio of output expectation when a single input variable lies on any of its subranges with respect to the original output expectations when all variables varies in their whole distributions. Similarly, Tarantola et al. (2012) developed contributions to sample variance (CSV) plot for measuring the ratio of output variance when an input variable lies on its subranges. Since Tarantolaʾs work fails to show how the output variance change when the uncertainties of input variables reduces to its subranges, Wei et al. (2014; 2015) modified Tarantolaʾs work and proposed to use output expectation after uncertainty reduction of input variables instead of constant one in the formula of regional variance, and defined the revised mean and variance ratio functions. Pannier et al. (2015) not only considered the detailed information of input variables but also the details of model output, and divided the result space into a finite number of subdomains and provided the detailed description of input-output interrelation in subdomains by Sobolʾs partial derivative sensitivity measure (Sobol 2009).

On the other hand, it should be noted that the traditional global and regional sensitivity theory all concern a unitary output. But when the output performance is a function of time, location or other parameters, and in other cases when several different outputs should be taken into consideration at the same time, multivariate methods are in great necessity for those kinds of engineering contexts. There already exist some approaches for global sensitivity analysis with multivariate outputs. Campbell et al. (2006) and Lamboni et al. (2011) decomposed each output variable with classical orthogonal basis function, and defined global sensitivity indices based on output decomposition method. Gamboa et al. (2013) assumed the covariance matrix of outputs can represent the uncertainties of multivariate outputs, and developed generalized Sobol indices based on covariance decomposition. Garcia-Cabrejo and Valocchi (2014) illustrated the equivalence of the above two decomposed methods, and proposed to use polynomial chaos expansion as an efficient computational method. But those methods share the same disadvantage for not involving the correlations of multivariate outputs.

This paper aims at considering this kind of correlation among model outputs and combining it with regional sensitivity techniques. It will provide more detailed information for applicants about how to reduce the total uncertainties of multivariate outputs efficiently. The generalized variance is firstly introduced to represent a scalar form of uncertainties for vector output which has clear geometry interpretation representing the correlation information of the uncertainty of model outputs. Then regional sensitivity index, i.e., the generalized variance ratio function, is defined based on generalized variance. As for computational issues, two methods including single loop Monte Carlo simulation and sparse grid method are developed for efficient estimation of the proposed regional index with multi-outputs. At last, two typical examples are used for explaining how to reduce the uncertainty of outputs with the help of the computational results and illustrating the effectiveness of the proposed method in engineering decision-making.

Jingwen Song
jingwen.song@irz.uni-hannover.de
Northwestern Polytechnical University

Parallel Session 15: New methods in sensitivity analysis and correlated inputs II

How sensible is the PAWN index to its tuning parameters?

Pianosi and Wagener [1 , 2] have recently proposed the density-based index PAWN as a new measure to compute sensitivity indices of skewed model outputs. PAWN relies on Cumulative Distribution Functions to characterize the maximum distance between the unconditional output distribution YU , i.e. obtained by moving all parameters simultaneously, and the conditional output distribution YCij , i.e. obtained by fixing the i -th parameter to j = 1,2,..., n values or intervals within its uncertainty range. The diference between YU and YCij is then assessed via the Kolmogorov-Smirnov test. The final PAWN index for a given parameter (Ti) is obtained by calculating the mean, the median, the maximum or any other summary statistic over all the KS values computed between YU and YCij .

The accuracy of PAWN is therefore conditioned by three tuning parameters that need to be fixed by the practitioner, plus a random factor: the initial sample size for YU (N), the number of conditioning intervals (n), the selection of a summary statistic (theta) and the stochasticity in the sampling of YU (epsilon). This means that, with PAWN, the uncertainties that characterize any given model design extend to the structure of the sensitivity analysis itself. Knowing how the PAWN index is conditioned by the choices taken by the practitioner during the setting of the sensitivity analysis is therefore essential for a proper assessment of its reliability.

Here we will present the results of a study aimed at assessing how sensible is the PAWN index to N, n, epsilon and theta. Our results indicate the following:
* The uncertainty in PAWN indices is largely conveyed by the inclusion or exclusion of max as a possible summary statistic (i.e. max in theta, max not in theta).
* In the max in theta setting, the PAWN indices of very influential model inputs are largely conditioned by the first-order effect of theta. Those of slightly influential model inputs are more determined by the interactions between (N, n, theta). Interactions can be as important as to explain 15-50% of the PAWN indices uncertainty.
* In the max not in theta setting, the PAWN indices of both influential and slightly-influential model inputs are largely conditioned by the simultaneous first-order effect of N , n and theta. Interactions can be as important as to explain 15-30% of the PAWN indices uncertainty.

These results indicate that the computations leading to PAWN indices are significantly non-additive, especially in the case of moderate to slightly-influential model inputs. In the talk we will discuss the implications of this characteristic for a factor-fixing and factor-prioritization setting, and elaborate on the importance of conducting sensitivity analysis of sensitivity indices as a way to test their robustness.

References

[1] F. Pianosi and T. Wagener. "A simple and ecient method for global sensitivity analysis based on cumulative distribution functions". Environmental Modelling and Software 67 (2015), 1-11.
[2] F. Pianosi and T. Wagener. "Distribution-based sensitivity analysis from a generic input-output sample". Environmental Modelling and Software 108 (2018), 197-207.
Arnald Puy
arnald.puy@gmail.com

Université Libre de Bruxelles

Samuele Lo Piano
s.lopiano@gmail.com
Universitat Oberta de Catalunya

Andrea Saltelli
Andrea.Saltelli@uib.no
University of Bergen

Computation of global sensitivity measures for problems with inequality constraints

Mathematical modelling plays crucial role in the design and operation of man-made systems ranging from small appliances to industrial manufacturing and distribution facilities and further to overarching approaches such as Life Cycle Analysis. Many ensuing models encompass numerous physical, chemical, biological, economic, etc. phenomena in multi-scale settings and involve making optimal decisions regarding system configuration and its optimal model-predictive control, which places them in the realm of process systems engineering.

However, optimal design and operation of such process systems is subject to multiple sources of uncertainty the effect of which on model outputs must be quantified in order to ensure robust system design and operational performance. This can be achieved by employing tools for model analysis under uncertainty such as the Global Sensitivity Analysis (GSA) of model outputs with respect to uncertain inputs. There are a number of methods classified under GSA with variance-based approaches being some of the most used owing to their intuitive interpretation and ease of use. In particular, methods based on Sobolʾs seminal contributions [1] have been extended to give access to sensitivity indices of model outputs to (groups of) model inputs even in the case when the latter are statistically dependent [2].

These methods, however, do not cover a wide class of process models involving inequality constraints originating from physical limitations, quality requirements, economic feasibility or regulatory compliance [3, 4]. In this work, we are building on our previous developments on variance-based GSA for models with inequality constraints (which we denote cGSA) [3]. Our original approach extended the definition of Sobolʾs sensitivity indices (SIs) to models defined in non-rectangular (due to constraints) n-dimensional domains Q^n and allowed their evaluation through brute force sampling within an enveloping unit hypercube H^n with the subsequent application of acceptance-rejection formulae for the computation of SIs. Although Sobolʾs low discrepancy sequence was employed for efficient sampling of the input space, the approach proved to be excessively computationally demanding owing to the fact that the infeasible space H^n\ Q^n was sampled with the same resolution as Q^n.

This problem is especially detrimental when model constraints are implicit in nature, i.e. when the model must be solved every time when a constraint must be evaluated (e.g. when the constraints are process KPIs). It should be noted that the number of feasible model realisations contributing to the evaluation of cGSA measures may decrease exponentially with n together with the volume of the feasible space Q^n. This in turn necessitates the sample size to be increased to maintain the accuracy of SI evaluation.

To enhance the computational efficiency of cGSA, we explore here two different approaches. The first one introduces an adaptive sampling method for the identification and exploration of parametric spaces of arbitrary geometric shape. The main feature of the method is that the ratio of the total number of sampled points in the input space to the number of feasible samples tends asymptotically to unity. This leads to minimal sampling of the infeasible space and therefore a reduced number of infeasible model evaluations compared to the previous approaches. In the case of implicitly constrained problems the overall costs of cGSA can thus decrease dramatically. The adaptive sampling approach can be further improved using local interpolation of constraint functions to reduce the need in model evaluation for the verification of constraints. This results in further speedup of cGSA.
The second approach explored in this work relies on the construction of adaptive metamodels of model constraints followed by the construction of a metamodel of the model function itself.

Based on an initial uniform sampling of the enveloping hypercube H^n, metamodels of constraints are used to guide further, adaptive sampling within the feasible domain Q^n accompanied by metamodel updates. Thus, the approximation of the model function can be iteratively improved within the feasible domain while avoiding sampling the infeasible space. cGSA measures can then be evaluated by using computationally cheap metamodels.

The approaches described above are exemplified on a case study involving a multidimensional g-function [3] as well as on a model of the production of a biopharmaceutical drug (monoclonal antibodies) by a mammalian cell culture in a bioreactor under product quality constraints. It is demonstrated that cGSA can successfully identify key influential model inputs with the highest contribution to the output uncertainty. This information can be used to reduce the dimensionality of the input parameter space for further exploration and exploitation of subspaces of inputs, as well as for the design of experiments. In addition, based on the sample constructed within the cGSA procedure an explicit description of the operational Design Space of this process can be constructed as a set of linear inequalities significantly extending the feasible space compared to a typically used approximation with an empirically defined inscribed hyperrectangle [5].

References

[1] Sobol, I.M. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Mathematics and Computers in Simulation, 2001. 55(1-3): p. 271-280.
[2] Kucherenko, S., Tarantola, S., Annoni, P. Estimation of global sensitivity indices for models with dependent variables. Computer Physics Communications, 2012. 183(4): p. 937-946.
[3] Kucherenko, S., Klymenko, O.V., Shah, N. Sobol' indices for problems defined in non-rectangular domains. Reliability Engineering & System Safety, 2017. 167: p. 218-231.
[4] Klymenko, O.V., S. Kucherenko, and N. Shah, Constrained Global Sensitivity Analysis: Sobol' indices for problems in non-rectangular domains. 27th European Symposium on Computer Aided Process Engineering, Pt A, 2017. 40a: p. 151- 156.
[5] Kotidis, P., Demis, P., Goey, C.H., Correa, E., McIntosh, C., Trepekli, S., Shah, N., Klymenko, O.V., Kontoravdi, C. Constrained global sensitivity analysis for bioprocess design space identification, Comput. Chem. Eng., 2019, in press.

Oleksiy Klymenko
o.klymenko@surrey.ac.uk
Department of Chemical and Process Engineering
University of Surrey
Guildford, UK

Panagiotis Demis
p.demis@surrey.ac.uk
Department of Chemical and Process Engineering
University of Surrey
Guildford, UK

Sergei Kucherenko
s.kucherenko@imperial.ac.uk
Department of Chemical Engineering
Imperial College London
London, UK

Application of global sensitivity analysis for constrained models and visualization

Most existing global sensitivity analysis techniques were designed under the hypothesis that model inputs are independent. However, in many cases there are dependencies among inputs, which may have significant impact on the results. Such dependencies in the form of correlations [1] and constraints [2] have been considered in the generalized GSA framework developed by Kucherenko and co-authors.

Consider a model f(x,theta), and the vector of constraint functions g(x,theta), where x is a vector of process parameters defined in the box domain H^d, theta is a vector of uncertain model parameters defined in m-dimensional real space R^m with a given pdf. Define probability p(x) of occurrence of an undesirable (reliability estimation analysis) or desirable (pharmaceutical/ chemical engineering) event as p(x) = P(g(x,theta)>g*). Here g* is a vector of given thresholds. Probability p(x) can be presented in a convenient form as an expectation with respect to theta p(x) =E[I(g(x,theta)>g*)], where I is an indicator function. Define design space (DS) as DS(x, p(x)>p*), where p* is a critical (acceptable) value of probability of meeting constraints. The notion of DS is widely used in pharmaceutical industry [3]. A brute force numerical estimation of p(x) requires (Ntheta x Nx) model runs, where Nx is the number of points in the space of the process parameters x, Ntheta is the number of sampled points in the theta - space. For a typical practical problem this number can be very large which results in costly computations. For such models GSA is an efficient practical tool which can dramatically reduce computational costs and provide useful additional information about importance of parameters and model structure.

Once DS is identified it is necessary to present it graphically. For more than three process parameters visualization of DS is a challenging task. “Guidance for Industry, Q8(R2) Pharmaceutical Development” [3] has the following recommendations: “When multiple parameters are involved, the design space can be presented for two parameters, at different values (e.g., high, middle, low) within the range of the third parameter, the fourth parameter, and so on.” Considering DS in d-dimensional process parameters space, a graphical presentation of DS via 2D plots prescribed in [3] would result in NP = d(d-1)/2 * 3^(d-2) 2D plots.

There are two main parts in the problem of DS identification to which GSA can be applied. The first part is identification of uncertain parameters contributing to the acceptance of imposed constraints with potential further model reduction. This problem is linked to the problem of identification of critical regions of inputs responsible for “extreme” values (values for which constraints are satisfied) of model response. It is known as a factor mapping setting of GSA. It typically requires rather complex techniques such as Monte Carlo Filtering. We propose an efficient GSA formulation using an indicator function I(g(x,theta)>g*) as a “model output”. An application from the chemical industry is used to illustrate how this formulation results in dramatic reduction of the number of required model runs.

The second part is concerned with application of GSA for visualization of DS. This problem belongs to a class of models with constraints. Consider DS in d-dimensional process parameters space and assume that GSA revealed that parameter xi is unimportant. It means that any 2D projections corresponding to different values of xi will be the same. Hence instead of three values (“high”, “middle”, “low”) for this parameter only one is sufficient for the graphical presentation, which would result in the reduction of NP to (d+4)(d-1)/2 * 3^(d-3) 2D

plots. In the case of s- non-important parameters NP will be reduced to [(d-s)(d+5s-1)+9s(s-1)]*3^(d-s-2)/2.

For illustration we consider a 3-dimensional Sobol' "g-function" f(x) in the unit hypercube H^3 with three different sets of parameters ai. For these three sets the corresponding values of Sobol' sensitivity indices SI (main effect) and SIT (total effect), are: (a) a ={0; 1; 2} ; SI ={ 0.706; 0.0818; 0.0234}, SIT ={ 0.856; 0.229; 0.104}; (b) a ={0; 1; 99}; SI ={0.761; 0.0957; 0.0020}, SIT ={0. 887; 0.238; 0.003}; (c) a ={0; 49; 99}; SI ={0.996; 0.0004; 0.0005}, SIT ={0.996; 0.0005; 0.0006}.

A synthetic DS is defined as DS(x, f(x)>1), where f acts as "probability p(x)". Analysis of DS shapes for different sets of parameters show that in the first case (a) all three 2D projections are important and the prescription given above requires to plot NP =d(d-1)/2 * 3^(d-2) = 9 2D projections. GSA reveals that in this case all three input parameters are important. In the second case (b) although all 2D projections onto the plains (x1, x2), (x1, x3), (x2, x3) are important, 2D projections on the plane (x1, x2) are independent of parameter x3. Hence only NP =(d+4)(d-1)/2 * 3^(d-3) = 7 2D projections are required for the visualization of DS. GSA shows that in this case parameter x3 is not important. In the third case (c) only 1D projection to the x1 axis is important while for two other parameters DS shape is defined in the whole domain of their change [0,1] hence no 2D projections are of any interest. GSA shows that in this case only parameter x1 is important. We can summarize that indeed GSA is a valuable practical tool both in identification and visualization of multidimensional DS.

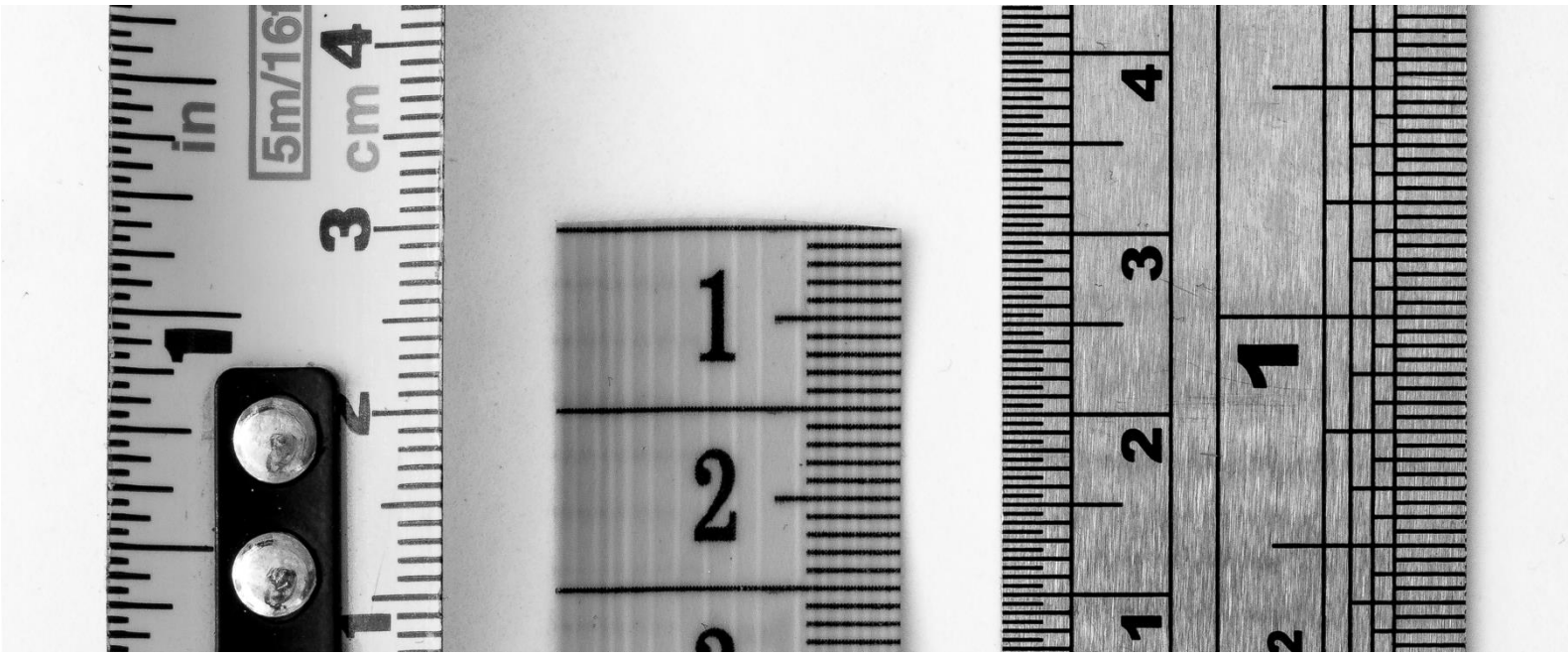
References

[1] Kucherenko, S., Tarantola, S., Annoni, P. 2012. Estimation of global sensitivity indices for models with dependent variables, Computer Physics Communications, 183, 937–946.
[2] Kucherenko, S., Klymenko, O.V., Shah, N. 2017. Sobol' indices for problems defined in non-rectangular domains, Reliability Engineering and System Safety 167, 218–231.
[3] Guidance for Industry, Q8(R2) Pharmaceutical Development, U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER), November 2009, ICH, Revision 2

Sergei Kucherenko
s.kucherenko@imperial.ac.uk
Department of Chemical Engineering
Imperial College London
London, UK

Local Organizing Committee Members

Nuria Febrer
Fakhri Jaan
Samuele Lo Piano
Andrea Saltelli



Sponsors

The event is organized by Open Evidence research at Universitat Oberta de Catalunya (Barcelona, Spain), the Joint Research Centre of the European Commission, the University of Bergen (Bergen, Norway), and the French CERFACS, Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique.

OPEN EVIDENCE

UNIVERSITY OF BERGEN

European Commission

CERFACS