7th International Conference on
Sensitivity Analysis
of Model Output

July 1-4, 2013 — Nice, France
University of Nice, Valrose Campus

www.gdr-mascotnum.fr/mascot13.html

Poster presentations
Proceedings

SAMO 2013 - July 1st-4th 2013

University Nice Sophia Antipolis, Valrose Campus, Nice, France

7th International Conference on Sensitivity Analysis of Model Output

Scope of the conference

Modelling activities are steadily increasing in all scientific disciplines, ranging from financial to environmental assessments. Sensitivity Analysis is crucial both in the modelling phase and in the interpretation of model results. It contributes to model development, model calibration, model validation, reliability and robustness analysis, decision-making under uncertainty, quality-assurance, and model reduction.

SAMO conferences are devoted to advances in research on sensitivity analysis methods and their interdisciplinary applications, they are held every third year. The aim of the conference is to bring together researchers involved in the developments and improvements of methods and strategies and users of sensitivity analysis in all disciplines of science, including physics, operations research, chemistry, biology, nanotechnology, engineering, environmental science, nuclear and industrial safety, economics and finance, etc.

The first day (July 1) is organized jointly with the MASCOT-NUM network and is devoted to presentations by PhD students working on the topics covered by the SAMO conference and MASCOT-NUM (uncertainty in simulation, sensitivity analysis, design and modelling of computer experiments, model validation, optimization under uncertainty, applications, etc.). A submission call has been launched to PhD students. Eight PhD students have been selected for oral presentations, other student submissions being considered for poster communications. A prize of 1000€ will be conferred by the MASCOT-NUM's scientific committee to the best student communication (to be used by the student to go to a meeting).

Program Committee

Paola Annoni (Joint Research Centre, Italy)
Didier Auroux (University of Nice, France)
François Hemez (Los Alamos National Lab., USA)
Sergei Kucherenko (Imperial College, UK)
Bertrand Iooss (EDF R&D, France)
Jeremy Oakley (University of Sheffield, UK)
Alberto Pasanisi (EDF R&D, France)
Clémentine Prieur (Univ. of Grenoble, France)
Luc Pronzato (University of Nice, France)
Andrea Saltelli (Joint Research Centre, Italy)
Bruno Sudret (ETH Zürich, Switzerland)
Stefano Tarantola (Joint Research Centre, Italy)

Organizing Committee

Luc Pronzato (University of Nice, France)
Bertrand Iooss (EDF R&D, France)
Didier Auroux (University of Nice, France)
Corinne Julien (University of Nice, France)
Anne-Laure Popelin (EDF R&D, France)
Joao Rendas (University of Nice, France)
Stefano Tarantola (Joint Res. Centre, Italy)
Irmela Zentner (EDF R&D, France)

Poster presentations

Carlos Lima Azevedo (LNEC Lisbon, Portugal) & Biagio Ciuffo (JRC Ispra, Italy), Global sensitivity analysis of high dimensional traffic micro-simulation models: a multistep approach

Boguslaw Bieda (Univ. Krakow, Poland), Stochastic approach for Life of the Municipal Solid Waste (MSW) Landfill using Monte Carlo simulation

Eduardo Castano (Universidad Autónoma de Querétaro, México), Victor M. Aguirre (Instituto Tecnológico Autónomo de México, México) & Antonio Villeda (Universidad Autónoma de Querétaro, México), A Response Surface Approach for Sensitivity Analysis in Differential Equations

Biagio Ciuffo (JRC Ispra, Italy), Qiao Ge & Monica Menendez (ETH Zurich, Switzerland), Quasi-OTEE versus kriging-based approaches for the sensitivity analysis of computationally expensive traffic simulation models: an exploratory study

Biagio Ciuffo & Serenella Sala (JRC Ispra, Italy), Global sensitivity analysis to identify archetypes for the impact assessment of chemicals

Benoit Delinchant (Univ. Grenoble, France), Reliable Based Design Optimization using k-sigma method and local sensitivity

Ivan T. Dimov & Rayna S. Georgieva (Academy Sciences Sofia, Bulgaria), Global Sensitivity Study of Compact Models in Nanodevices with Correlated Inputs

Fanny Douard & Bertrand Iooss (EDF, France), Dealing with uncertainty in technical & economic studies of investment strategy optimization

Thiago R. Figueiro, Jean-Herve Tortai, Nader Jedidi (Univ. Grenoble, France), Mohamed Saib, Patrick Schiavone (Aselta Nanographics , France), Selection of Test Patterns for Model Calibration based on Sensitivity Analysis

Michael Flechsig, Thomas Nocke, Claus Rachimow (Postdam Institute for Climate Impact Research, Germany), Multi-run simulation environments – requirements, development and application

Allan B. Hedin (Swedish Nuclear Fuel and Waste Management Co., Sweden), Sensitivity analyses of a fast analytical radionuclide transport model

Son Hong Hoang & Remy Baraille (Shom Toulouse, France), Schur vectors for estimating parameters in filtering algorithm for data assimilation

Karin Kandanaond (Rajabhat Univ., Thailand), Robust Assessment of Different Machine Learning Methods for Time Series Forecasting using a Factorial Design of Experiment

Matieyendou Lamboni, Renate Koeble, Adrian Leip (JRC Ispra, Italy), Uncertainty and sensitivity analysis of land shares model over EU 27

Nicolas Marie (Univ. Paris X, France), Rough paths theory applied to computation of sensitivities

George P. Petropoulos (Univ. of Aberystwyth), Hywel Griffiths (University of Aberystwyth, UK) & Stefano Tarantola (JRC Ispra), Towards Operational Products Development from Earth Observation: Exploration of SimSphere Land Surface Process Model Sensitivity using a GSA approach

Claudio M. Rocco (Univ. Central Venezuela, Venezuela), Elvis Hernandez (Univ. Catholica Portugesa, Portugal), Uncertainty and Sensitivity Analysis in Multiple Criteria Decision Problems using Rule Learner Techniques

Nathalie Saint Geours (Irstea, France), Stefano Tarantola (JRC Ispra, Italy), Linda Lilburne (Landcare Research, New Zealand), Computing variance-based sensitivity indices for spatial inputs with “map labelling”: study of sampling size bias

Hana Sulieman (American University of Sharjah, UAE), Profile-based Sensitivity in D-Optimal Designs for Precise Parameter estimation

Qiong-Li Wu, Benoît Bayol, Fenni Kang (Ecole Centrale de Paris, France), Jeremie Lecoeur (3Syngenta Seeds, France), Paul-Henry Cournede (Ecole Centrale de Paris, France), Sensitivity Analysis for Plant Models with Correlated Parameters : Application to the Characterization of Sun over Genotypes

Mauricio Zambrano-Bigliarini, Zuzanna Zajace, Stefano Tarantola (JRC Ispra, Italy), Global sensitivity analysis for the calibration of a fully-distributed hydrological model

7th International Conference on Sensitivity Analysis of Model Output, July 1-4 2013, Nice

Global sensitivity analysis of high dimensional traffic micro-simulation models: a multi-step approach

CARLOS LIMA AZEVEDO
LNEC, National Laboratory of Civil Engineering, 1700 Lisbon, PT

BIAGIO CIUFFO
European Commission Joint Research Centre. 21027 Ispra, IT

Calibration and validation procedures are increasingly ascending transportation modellers' and practitioners' top priorities, as the use of such tools is quickly spreading and its models progressively improved. Traffic simulators are increasingly being applied in many different traffic scenarios and the consistency with the available data needs to be assured. Furthermore, the access to both new and advanced modelling techniques and detailed traffic and behavioural data, is increasing the level of detail of improved traffic simulation models. These challenges have been linked to the need of a consistent understanding of the simulators performance, along with the appropriate calibration and validation procedures to monitor its uncertainty. Sensitivity analysis is the tool used with this aim.

Only a small number of sensitivity analysis practitioners make use of the advanced techniques made available in the recent years. The One-At-Time approach has been applied to microscopic simulation model in order to select the parameters to be calibrated and to get further insight on the meaning of its values after the calibration of car-following models. Analysis of variance based methods (ANOVA) were recently applied to draw inference about the effects of the traffic simulation models' parameters (*1*,*2*). However, almost all these studies only accounted for first order effects, ignoring the interaction of parameters. A recent and more efficient method based on variance decomposition as well has also been used for model a car-following model sensitivity analysis (*3*). All these applications were however applied to specific driver behaviour models. In order to explore the entire set of parameter interactions over the whole parameter space of typical traffic simulators, global SA methods are required. Although global methods may provide far more insight for nonlinear models, they can require large numbers of model runs. In addition, the interpretation of such global sensitivity tests can be difficult because of the number of parameters involved and the potential complexity of the causal relationships.

To overcome these issues, we propose a multi step approach. In general, a first sensitivity analysis is carried out on the sub-models and a second sensitivity analysis on the individual parameters belonging to the subset of the most influential sub-models. At each step, the sensitivity analysis approach used in this study was the variance-based approach based on Sobol's decomposition of variance (*3*). This approach requires the evaluation of the model $N*(k+2)$ times, where k is the number of model parameters (or groups) and N is the dimension of the Monte Carlo experiment.

More specifically, after grouping the parameters by modelling affinity, each parameter combination within the same group was mapped using Sobol's quasi-random sequences (*6*) and extracted from an uniform distribution defined by predefined ranges. The variance-based approach is then applied and the groups with highest share of the model variance are selected. A second sensitivity analysis is then carried out on the individual parameters of the selected groups. If number of parameters is still large, the previous steps are repeated.

The proposed methodology was tested on the advanced driver behaviour model MITSIM, accounting for a total of one hundred model parameters (*4*). This model is of particular interest due to the high interaction of all complex models describing the driver behaviour. The model integrates four levels of decision-making: target lane, gap acceptance, target gap and acceleration, in a latent decision framework based on the concepts of short-term goal and short-term plan (refer to *5* for a complete description). All previous calibrations of MITSIM considered the entire set of demand parameters (traffic input data) and only a small subset of supply parameters (driver behaviour) were generally selected by the modeller. The network chosen for this study was the A44 road in the region of greater

Porto, Portugal. It is a two-lane urban motorway with a total of 3940m and 5 main interchanges. Data of one week combining loop sensor data, video counts samples and vehicle paths count samples was used in this in the simultaneous dynamic estimation of a generic weekday traffic input data.

Several Goodness of Fit (GoF) measures between real and simulated outputs were computed, based on different statistics and different loop detector data, and used in the analysis. The first grouping step was based on the different MITSIM driver behavioural sub-models (Table 1). The sensitive analysis on groups allowed the selection of three most important sub-models, namely the car-following model, the lane utility model and the drivers' heterogeneity model.

Sub-model	Number of Parameters	1st order sensitivity index	Total sensitivity index
Driver Reaction Time Model (G1)	4	0.07	0.10
Car Following Acceleration Model (G2)	11	0.74	0.75
Free Flow Acceleration Model (G3)	7	0.01	0.04
Merging Model (G4)	4	0.01	0.04
Mandatory Lane Change Control (G5)	5	0.01	0.05
Yielding Model (G6)	2	-0.06	0.05
Nosing Model (G7)	6	0.03	0.05
Nosing Control (G8)	4	-0.04	0.05
Courtesy Yielding Model (G9)	4	0.03	0.04
Driver Heterogeneity Parameters (G10)	7	0.23	0.28
Target Gap Acceleration Model (G11)	13	0.00	0.04
Gap Acceptance Model (G12)	9	-0.03	0.05
Lane Utility Model (G13)	16	0.09	0.11
Target Gap Model (G14)	6	0.01	0.04
OD variability (G15)	2	-0.02	0.05

Table 1 Group sensitivity analysis – First order and total order sensitivity indices (N=2048)

The final sensitivity analysis has then been performed with the 34 model parameters from the selected sub-models and allowed individuating a group of 8 parameters accounting for almost the 90% of the output's variance, with a consequent significant simplification of the subsequent model calibration/estimation phase.

In synthesis, the proposed approach allowed identifying in a quantitative and objective way the most important parameters of the MITSIM model. In addition it has allowed choosing among different possible measures of goodness of fit and among different traffic measures. Further research will be carried out also to consider other sensitivity measures like those in (*7*).

Aknowledgements
Research contained within this paper benefited from participation in EU COST Action TU0903 – Methods and tools for supporting the Use caLibration and validaTIon of Traffic simUlation moDEls. This work makes use of results produced with the support of the Portuguese National Grid Initiative (https://wiki.ncg.ingrid.pt).

References:

1. Ciuffo, B., V. Punzo V. and V. Torrieri, V.. A framework for calibrating a microscopic simulation model. *Proceedings of the 86th Annual Meeting of the Transportation Research Board*, Washington, D.C.,
2. Park, B., Qi, H. Development and Evaluation of a Procedure for the Calibration of Simulation Models. *Transportation Research Record*, n. 1934, pp. 208–217. 2005
3. Punzo, V., Ciuffo, B. Sensitivity analysis of car-following models: methodology and application. *Proceedings of the Transportation Research Board 90th Annual Meeting*, Washington D.C., 2011.
4. Toledo, T., Koutsopoulos, H. N. and Ben-Akiva, M.,Integrated driving behavior modelling. *Transportation Research Part C*, Vol. 15, p96-112, 2007.
5. Toledo, T., Integrated Driving Behaviour Modelling. *PhD Thesis*, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 2003.
6. Sobol, I.M. Uniformly distributed sequences with an additional uniform property. USSR *Computational Mathematics and Mathematical Physics*, vol. 16 (5), pp. 236-242, 1976.
7. Plischke E., Borgonovo E. and Smith C.L., Global Sensitivity Measures from Given Data. European Journal of Operational Research, 226:536–550, 2012

[Presenting author's name; affiliation and regular mailing address]
[Biagio Ciuffo, European Commission Joint Research Centre. 21027 Ispra, IT. biagio.ciuffo@ext.jrc.ec.europa.eu]

Stochastic approach for Life of the Municipal Solid Waste (MSW) Landfill using Monte Carlo (MC) Simulation

BOGUSLAW BIEDA

AGH-University of Science and Technology, Al. Mickiewicza 30,30-059 Kraków, POLAND

This paper presents the application and benefits of Monte Carlo (MC) simulation in modeling the transit time contaminants in porous media. The transport of contaminants in a soil is represented by the one-dimensional (1D) form of the advection-dispersion equation given in [1]. For calculation of the transit time to estimate the landfill thickness of compacted clay The has been used the computer program CONTRANS. The output report provided by Crystal Ball® (CB), using with the MC, and presented in the study is the Sensitivity Chart. The present study is based on the data concerns the Mittal Steel Poland in Kraków.

Keywords: Landfill, Monte Carlo Simulation, Crystal Ball®, sensitivity analysis, advection-dispersion equation, Matlab.

Experimental procedures

The program CONTRANS (CONtaminant TRANSit) has been used to estimate the thickness of compacted soil liner of a landfill [2]. The source code of the computer program written in MATLAB based on the flow chart presented in [2] has been done in [4].

The 1D advective-diffusive governing equation for reactive solute transport in saturated soil is given by Acar et al. [1] as follows [2]:

$$R_d\left(\frac{\partial C}{\partial t}\right) = D\left(\frac{\partial^2 C}{\partial z^2}\right) - v_s\left(\frac{\partial C}{\partial z}\right) \quad \textbf{(1)}$$

where R_d is the retardation factor, C is the solute concentration in the direction of transport, D is the diffusion coefficient, v_s is the seepage velocity, z is the depth (distance within the anaerobic granules) and t is the time. To use CB , we must perform the steps given in [3].

Results

The equation (2) used for this study (see below) has been adopted from the (1).

$$Z2 = (HG*HC/POROS)*TH/EDC \quad (\textbf{2})$$

where:
HG = hydraulic gradient, HC = hydraulic conductivity, POROS = porosity, TH = linear thickness, EDC = diffusion coefficient

The statistic data used for contaminant transport simulation model Z2 using CB has been taken from the [4]. Sensitivity Chart, the result of the MC simulation used CB after running 10,000 trials was shown on the Figure 1.

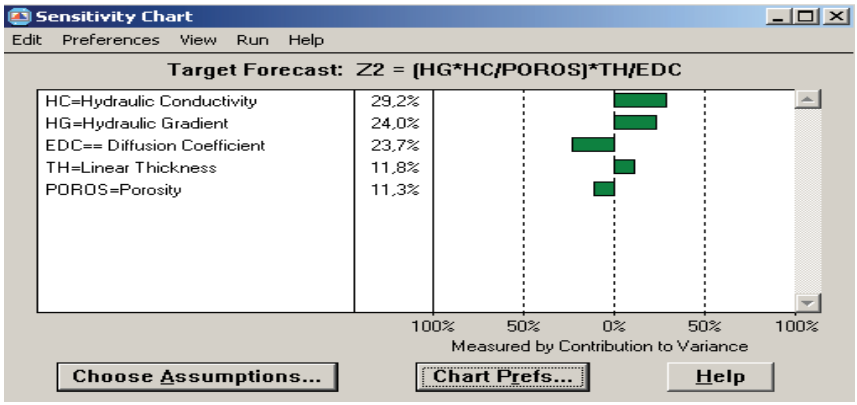

Fig. 1. Sensitivity Chart - Target Forecast Z2

Conclusions

The following important information can be obtained from the Sensitivity Chart:

Sensitivity refers to the amount of uncertainty in a forecast that is caused by the uncertainty of an assumption as well as by the model itself. Positive coefficients indicate that an increase in assumption is associated with an increase in the forecast; negative coefficients imply the reverse. In the Sensitivity Chart (Fig. 2), we can see that Hydraulic Conductivity (HC), Hydraulic Gradient (HG) and Linear Thickness (TH) have a positive influence on Z2 (29.2%, 23.7% and 11.3%, respectively), and Diffusion Coefficient (EDC) as well as Porosity (POROS) have a negative influence on Z2 (24% and 11.84%, respectively).

The simulation results suggest that the MC simulation is efficient tool to use in practice over a large parameter range. In deterministic models significant variations in material properties lead to the radically different results than those predicted by the methods based on the stochastic transit time contaminants models. Hope fully this study will encourage other researchers to consider this approach as well in their works.

Acknowledgement

This study was supported by the science financial grant for 2011-2013. Author used the part of his paper untitled Stochastic approach to municipal solid waste landfill life based on the contaminant transit time modeling using the Monte Carlo (MC) simulation. The Science of the Total Environment, vol. **442**, pp. 489–496, **2013**.

References

1. Acar Y. B., Haider L. Transport of low-concentration contaminants in saturated earthen barriers. Journal of Geotechnical Engineering **116**, **1990**.
2. Cocka E. Report: use of computer program to estimate the thickness of a compacted clay liner in a landfill, Waste Management & Research **17**, pp. 227-230, **1999.**
3. Evans J.R., Olson D.L. Simulation and Risk Analysis. Prentice Hall: Upper Saddle River, New Jersey 07458.1998.
4. Bieda B. Stochastic approach to municipal solid waste landfill life based on the contaminant transit time modeling using the Monte Carlo (MC) simulation. The Science of the Total Environment, vol. **442**, pp. 489–496, **2013**.

7th International Conference on Sensitivity Analysis of Model Output, July 1-4 2013, Nice

A Response Surface Approach for Sensitivity Analysis in Differential Equations.

EDUARDO, CASTAÑO
Universidad Autónoma de Querétaro, Querétaro, México

VÍCTOR, M., AGUIRRE
Instituto Tecnológico Autónomo de México, Ciudad de México, México

ANTONIO, VILLEDA
Universidad Autónoma de Querétaro, Querétaro, México

We show the advantages of using a response surface approach to evaluate the impact of different factors to the problem of estimating the solution and structural parameters of a system of differential equations. The factors considered are: distance of initial point to the true value, noise in the data and amount of smoothing in the estimation method. The method of estimation is based on the statistical analysis of functional data known as profiled estimation and given in the paper [1]. We considered simulating from the predator - prey system of differential equations. As responses of the experiment we used bias and variation of the estimation procedure.

The approach followed considered a replicated full 3^3 factorial experiment. This allowed three levels for each factor and required only 27 runs for each replica. With this design we are able to estimate linear and quadratic main effects and two way interaction terms. Also from the replication we had 27 degrees of freedom to estimate the experimental error term and hence to test the statistical significance of each of the estimated effects.

Our study showed robustness of the estimation method for both responses when the smoothing constant is relatively small, and that this property deteriorates as the smoothing constant increases exponentially. This suggests a strategy of model fitting which consists in estimating the system with small values of the smoothing constant and then increases its value up to a point when the solutions begin to show abnormalities.

The important issues of this approach to study sensitivity are: 1) the reduced number of scenarios to be considered, which in our case were 27 with one full replicate; and 2) the richness of the information obtained from the study. We were able to estimate linear, quadratic and interaction effects of the different factors, discard those that were unimportant using a statistical test, and also to plot the response surface for bias and variation as function of the factors considered.

This approach represents an improvement with respect to the method of trying different initial conditions on the solution system of differential equations, but without a structured strategy. The structure is given by the well-established methodology of statistically designed experiments.

Key words: system of differential equations, parameter estimation, profiled estimation, functional data analysis, inverse problems

References :

[1] Ramsay, J., O., G. Hooker, D. Campbell, and J. Cao, "Parameter estimation for differential equations: a generalized smoothing approach", J. R. Statist. Soc., **B69**, 741-796, 2007.

[Víctor, M. Aguirre; Professor, Statistics Department; Instituto Tecnológico Autónomo de México (ITAM), Río Hondo #1, Col. Progreso Tizapán, C.P. 01080, Del. Álvaro Obregón, México, D.F., México]
[aguirre@itam.mx]

7th International Conference on Sensitivity Analysis of Model Output, July 1-4 2013, Nice

Quasi-OTEE versus kriging-based approaches for the sensitivity analysis of computationally expensive traffic simulation models: an exploratory study

BIAGIO CIUFFO
European Commission Joint Research Centre. 21027 Ispra, IT

QIAO GE
ETH Zurich, Institute for Transport Planning and Systems. 8093 Zurich, CH

MONICA MENENDEZ
ETH Zurich, Institute for Transport Planning and Systems. 8093 Zurich, CH

One area of model calibration in which meaningful contributions are still needed is the sensitivity analysis (SA) of the input parameters for the calibration itself. The SA explores the relationship between the analysis outcome and the parameter assumptions. Due to the limitation of time and other resources, most calibration procedures cannot afford to calibrate all parameters in the model. Thus, calibration is carried out for only a limited number of input parameters. However, there is usually no formal procedure for selecting these parameters, other than choosing the ones that appear to the model user as most likely to have a significant effect on the result (such criteria is often dictated by former experiences). As one could imagine, the selection of an incomplete set of parameters for calibration may lead to multiple issues, including but not limited to, model imprecisions, and unrealistic values for the calibrated parameters. These problems should not be a surprise, as in traffic models there exist many interactions between the different parameters (e.g., many of the car-following parameters also impact the lane-changing model). Hence, focusing on the incorrect set might have a cascading effect. Therefore, a proper SA, including the initial screening of the parameters, can be very valuable for the subsequent calibration process. Moreover, it may actually reduce the total efforts needed during the actual model calibration. A good SA could provide both quantitative and qualitative information regarding the effects of the different model parameters (and their variations) on the simulation results.

Unfortunately, to the authors' knowledge, there are few examples of applying SA in the calibration of microscopic traffic models: the authors of (*1*) used the One-At-a-Time (OAT) method to evaluate the variance of the VISSIM output. The variance-based SA approach was employed in (*2*,*3*) for microscopic traffic simulation models and in (*4*) for two car-following models. Despite these examples, there appears to be no previous research suggesting a standard SA method that is efficient yet accurate to be applied for the calibration of microscopic traffic models.

In this light, this present paper aims at comparing two recently developed SA methods, in order to better understand their advantages and disadvantages. The first model, called quasi-OTEE, was introduced in (*5*). It is a general screening approach based on the Elementary Effects (EE) method (*6*) but with much higher efficiency. It screens the influential parameters through computing the corresponding EE and qualitatively comparing the Sensitivity Indexes. The case study provided in (*5*) demonstrated that this tool can properly identify the most influential parameters from a computationally expensive model, for which other quantitative SA techniques are not feasible to be applied at the beginning. The second method adopts Sobol indices (7) calculated on a kriging approximation of the simulation model. This method has been presented in (*8*) and is based on the recursive use of the DACE tool to obtain a robust kriging emulator. Effectiveness of the method has been proven in (*8*) where the authors show that Sobol indices calculated on the kriging emulator (based on 128 and 512 model evaluations) achieve approximately the same value than those calculated, following the procedure described in (*7*), on almost 40.000 model evaluations.

The benchmarking exercise was carried out on the same case-study presented in (*8*), namely, on the five "toy" networks shown in Figure 1 and simulated using the mesoscopic version of the AIMSUN model. Seven model parameters were considered in the analysis, and, in both methods, 512

model evaluations were used. The SA was then carried out on four different model outputs calculated locally and globally (for a total number of SA ranging from 16 to 52 depending on the network).

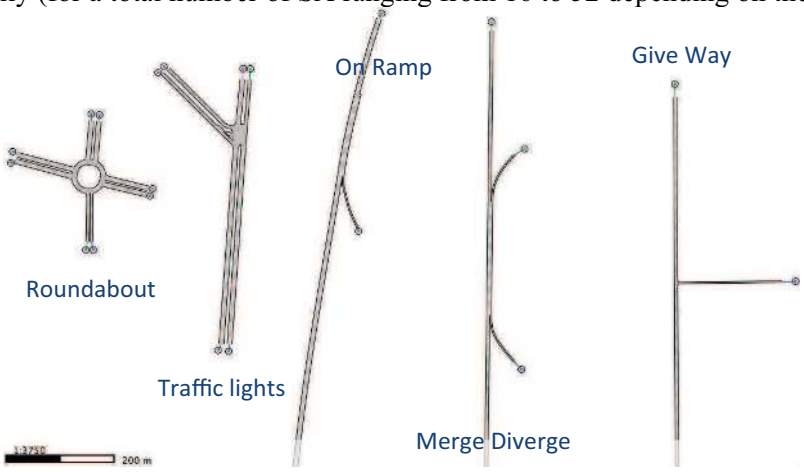

Figure 1. Layout of the five test networks. As the layout changes, the effect of the different parameters on the outputs changes as well.

Preliminary results show that both methods were able to identify, to a good degree, the non-influential parameters. Furthermore, the kriging-based method was also able to provide a reliable estimation of first order and total order sensitivity indices, thus allowing a more powerful insight into the input-output relation of the model. The reliability of a kriging meta-model, however, is expected to suffer from the higher dimensionality of the model itself. The experience carried out therefore suggests the following rule-of-thumb for the sensitivity analysis of computationally expensive traffic simulation models: when the number of parameters is lower than 15-20, kriging-based SA should be preferred. Conversely, when it is higher, the quasi-OTEE method should be adopted since an ill-defined kriging meta-model could lead to misleading results in the analysis.

Acknowledgements
Research contained within this paper benefited from participation in EU COST Action TU0903 – Methods and tools for supporting the Use caLibration and validaTIon of Traffic simUlation moDEls.

References:

1. Kesting, A., and M. Treiber. Calibrating car-following models by using trajectory data. In *Transportation Research Records*, No. 2088, 2008, pp. 148 - 156.
2. Punzo, V., and B. Ciuffo. How parameters of microscopic traffic flow models relate to traffic conditions and implications on model calibration. In *Transportation Research Records*, No. 2124, Transportation Research Board of the National Academies, Washington DC, 2009, pp. 249 - 256.
3. Bartin, B., K. Ozbay, O. Yanmaz-Tuzel, and G. List. Modeling and simulation of unconventional traffic circles. In *Transportation Research Records*, No. 1965, Transportation Research Board of the National Academies, Washington DC, 2006, pp. 201-209.
4. Punzo, V., and Ciuffo, B. Sensitivity Analysis of Car-following Models: Methodology and Application. *Proceedings of the 90th TRB Annual Meeting*, Washington DC, 2011.
5. Ge, Q., and M. Menendez. An improved approach for the sensitivity analysis of computationally expensive microscopic traffic models: a case study of the Zurich network in VISSIM. *Proceedings of the 92th TRB Annual Meeting*, Washington DC, 2013.
6. Morris, M. D. Factorial sampling plans for preliminary computational experiments. *Technometrics*, Vol. 33, No. 2, 1991, pp. 161 - 174.
7. Saltelli, A.; Ratto, M.; Anres, T.; Campolongo, F.; Cariboni, J.; Gatelli, D.; Saisana, M.; Tarantola. S.; *Global Sensitivity Analysis – The Primer*; John Wiley and Sons Ltd: Cheichester, England, 2008.
8. Ciuffo, B., Casas, J., Montanino, M., Perarnau, J., Punzo, V. From Theory to Practice: Gaussian Process Meta-Models for Sensitivity Analysis of Traffic Simulation Models: Case Study of Aimsun Mesoscopic Model. *Proceedings of the 92th TRB Annual Meeting*, Washington DC, 2013.

[Presenting author: Qiao Ge; Address: ETH Zurich IVT, HIL F37.3, Wolfgang-Pauli-Strasse 15, 8093 Zürich, Switzerland]
[qiao.ge@ivt.baug.ethz.ch - http://www.ivt.ethz.ch/people/geqiao]

[Corresponding author: Biagio Ciuffo; Address: European Commission – Joint Research Centre, 21027 Ispra, Italy]
[biagio.ciuffo@ext.jrc.ec.europa.eu]

7th International Conference on Sensitivity Analysis of Model Output, July 1-4 2013, Nice

Global sensitivity analysis to identify archetypes for the impact assessment of chemicals

BIAGIO CIUFFO
European Commission Joint Research Centre. 21027 Ispra, IT

SERENELLA SALA
European Commission Joint Research Centre. 21027 Ispra, IT

Emission of chemicals is increasing over years and the related impacts are greatly influenced by spatial differentiation. Chemicals are usually emitted locally but, due to their physical-chemical properties and persistence, may exert both local and global impact. Besides, variability of environmental parameters of the emission compartment may affect the fate and the exposure up to orders of magnitude of difference. In the scientific community, there is continual debate whether the exclusion of spatial information in applications such as ERA (Health and Environmental Risk Assessment) and LCA (Life Cycle Assessment) may imply misleading results, influencing the decision on products environmental risk and performance. In order to point this aspect out, several spatially resolved multimedia models have been developed at various resolutions for calculating transport and fate of chemicals (*1*,*2*). These models allow assessing the distribution and fate of chemicals in the environment after their emissions, on the basis of chemical (viz. physical chemical) and landscape related properties. Unfortunately, addressing spatial differentiation at high resolution requires significant efforts both in terms of data requirements and of computational capacities (usually not available among practitioners). For this reason, the models adopted so far are mainly simplistic box models in which the concept of spatial differentiation is based on the scale/resolution (cell, country, basin). This approach, however, may reduce the uncertainty of the assessment to the extent in which the "box" is really representative of the removal/transport processes. Another approach is based on archetypes, namely the combination of parameters leading to a certain scenario whose pattern of variability is predictable.

In the present work, we propose a novel approach in which emission archetypes are defined on the basis of the quantitative results achieved by means of a global sensitivity analysis of a complex spatial model. The emission archetypes are built in order to cluster combination of parameters leading to similar behavior in the removal rates of a chemical from the environment.

In particular, global sensitivity analysis techniques have been applied to the Multimedia Assessment of Pollutant Pathways in the Environment model (MAPPE) (*3*) with two aims: to assess the variability in removal rates, focusing on the relative influence of substance properties and of environmental characteristics; to support the development of chemical specific and compartment specific archetypes. Global sensitivity analysis aims at understanding the relative importance of the uncertain inputs in determining the variable of interest. Several techniques can be fruitfully exploited to analyse the sensitivity of a model. In the present paper, we focus our attention on variance-based methods, in which the variance of model outputs is considered as a proxy of their uncertainty. Two different analyses are carried out: (i) sensitivity analysis of environment specific inputs (with the chemical-specific inputs considered as fixed assuming the values corresponding to four different chemicals); and (ii) sensitivity analysis of environment and chemical specific inputs.

Table 1 presents a summary of the results. Chemical-specific properties are predominant with respect to environmental ones. However, once the chemical has been defined, the value of the removal rate is mainly affected by the wind speed, the precipitation and the type of land coverage.

These results suggest the possibility of basing emission archetypes on climatic zones rather than on geo-political characteristics (e.g. continents, countries) (*4*). This hypothesis is tested by evaluating and comparing the distributions of the air removal rate within different climatic zones (considering the Koppen-Geiger Climate Classification, *5*) and within different continents.

Air Compartment		
Chemical category	Key parameters for archetypes (Removal rate)	Key parameters for archetypes (Removal rate order of magnitude)
Overall	Chemical, P, ABL	Chemical, P, U_{10}
Hydrophilic	P, ABL	P, ABL
Lipophilic	U_{10}, P, ABL, OC, Cov	U_{10}, P, ABL, OC, Cov
High volatility	Cov, U_{10}, ABL	U_{10}, Cov, ABL
Multimedia	U_{10}, Cov, ABL, (T)	U_{10}, Cov, ABL, P, (T)

Table 1: Key environmental parameters for different chemical categories as resulting from the sensitivity analysis of the MAPPE model

The comparison is based on the distributions of the kurtosis of the different distributions as shown in Figure 1. From the Figure it is clear that, overall, the climate-based approach is able to provide removal rate estimations with lower uncertainty if compared with the continent-based distributions (the higher the kurtosis the narrower the distribution). Further research is however still necessary as the uncertainty in the removal rate calculation from climate-based archetypes is, in the authors' opinion, still too high.

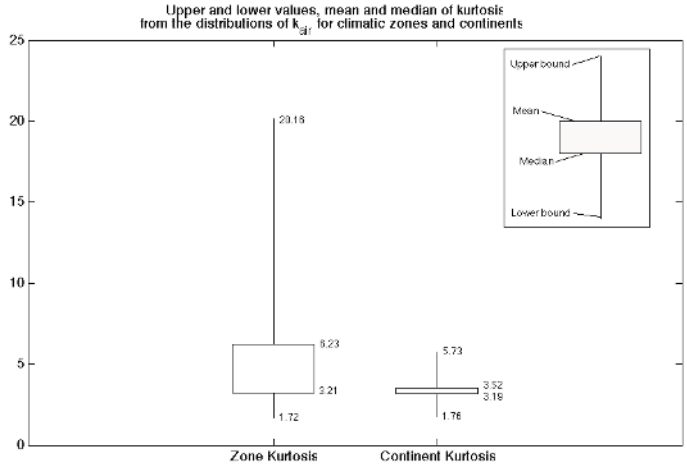

Figure 1: Kurtosis distributions for the distributions of k_{air} from different combinations of chemicals and geographic zones (5)

Aknowledgements

The research was funded by the European Commission under the 7th framework program on environment; ENV.2009.3.3.2.1: LC-IMPACT - Improved Life Cycle Impact Assessment methods (LCIA) for better sustainability assessment of technologies, Grant agreement number 243827

References :

1. Pennington, D.W.; Margni, M.; Ammann, C.; Jolliet, O. Multimedia fate and human intake modelling: Spatial versus non-spatial insights for chemical emissions in Western Europe. *Environ. Sci. Technol.* 2005, 39(4), 1119-1128.
2. Wegener Sleeswijk, A.; Heijungs, R. GLOBOX: A spatially differentiated global fate, intake and effect model for toxicity assessment in LCA. *Sci. Total Environ.* 2010, 408, 2817-2832
3. Pistocchi A, Marinov D, Pontes S, Zulian G, Trombetti M. 2011. Multimedia assessment of pollutant pathways in the environment - Global scale model. Official Publications of the European Communities, Luxembourg EUR 24911 EN.
4. Ciuffo B., Sala S. Global sensitivity analysis of multimedia models: towards specific archetypes for the impact assessment of chemicals 2012 submitted to *Environ. Sci. Technol.*
5. Kottek, M., Grieser, J., Beck, C., Rudolf, B., Rubel, F. World map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift*, 2006, 15 (3), 259-263

[Presenting author's name; affiliation and regular mailing address]
[Biagio Ciuffo, European Commission Joint Research Centre. 21027 Ispra, IT. biagio.ciuffo@ext.jrc.ec.europa.eu]

Reliable Based Design Optimization using k-sigma method and local sensitivity

Benoit DELINCHANT
Electrical Engineering Laboratory, University of Grenoble, FRANCE

INTRODUCTION

This paper presents a method and a tool for solving reliability-based design optimization (RBDO) using local sensitivities, and is applied to a nano magneto-mechanical system. The RBDO problem [1] is formulated as equation (1) where $\{P(H_j(X)\leq 0)\geq P_{Hj}\}$ is a reliability constraint involving the stochastic vector X modeling the uncertainties inherent in the model, and P_{Hj} is the maximal probability allowed. The solution of RBDO problem is not easy, even for simple models, because we must simultaneously solve the optimization problem and evaluate the failure probability. A first approach uses a double loop, one for the constrained optimization problem, and another inside the first one for failure probability using for example Monte-Carlo simulation. This approach is very simple to implement and gives accurate results. But it is not applicable in practice due to high computation time. Other methods approximate the reliability index by using First/Second Order Reliability Method - FORM/SORM [2]. A review of existing methods is proposed in the literature [1].

A DECOUPLED METHOD FOR SOLVING RBDO PROBLEM

In equation (1), design parameters are the mean of stochastic vector **X** (hyperparameters in the case of a normal distribution). These design parameters can vary in a rang fixed by x^{min} and x^{max}. The failure probabilities, or the probabilities of the inequality constraints $H_j(X)$ (*j=1,…,m*) have to be satisfied at the optimal point. In order to simplify this probability evaluation, the decoupled method transforms these constraints from the stochastic domain (equation (1)) to the deterministic one (equation (2)) [1].

$$(1)\ \begin{cases} \text{- minimize} & F(\overline{X}) \\ \text{with} & \overline{X}=\{\overline{x}_1,\overline{x}_2,...\overline{x}_n\}\in \Re^n \\ & x_i^{min}\leq \overline{x}_i\leq x_i^{max} \quad i=1,...,n \\ \text{- subject to} & P(H_j(X)\leq 0)\geq P_{Hj},\ j=1,...,m \end{cases} \qquad (2)\ \begin{cases} \text{- minimize} & F(x) \\ \text{with} & x=\{x_1,x_2,...x_n\}\in \Re^n \\ & x_i^{min}\leq x_i\leq x_i^{max} \quad i=1,...,n \\ \text{- subject to} & H_j(x)+k_\sigma.\sigma_{H_j(x)}\leq 0,\ j=1,...,m \end{cases}$$

$\sigma_{H_j(x)}$ is the standard deviation of the constraint functions $H_j(X)$ and k_σ is the coefficient related to the level of reliability. In this study, we will consider all the inputs independent and governed by a normal distribution. The RBDO problem becomes therefore a standard optimization one, reporting stochastic operations to the evaluation of the standard deviation. Many solutions are available to approximate the standard deviation like Monte Carlo simulation (too time consuming), unscented transform, etc. On the one side, these direct methods require minimizing the number of calls to the model evaluation. On the other side, indirect methods may use Monte Carlo simulation with a very light model based on response surface [5]. In this work, we are investigating efficient direct methods since it is generally very hard to obtain an accurate response surface when parameter space dimension is high.

USE CASE: NANO MAGNETO-MECHANICAL SYSTEM

Our use case is the reliable optimization of a magnetic nanometer switching device which modeling is described in [3]. The design specifications are to find the position and the size (minimal) of magnets subject to constraints on the contact force and the contact length. The positions and the sizes are defined by 4 design parameters considering a total of 19 stochastic variables (19-4=15 with fixed hyperparameters).

GLOBAL SENSIVITY ANALYSIS

In order to estimate the influence of each parameter on the model, several sensitivity studies have been done. First of all, a global study is done in order to check the influence in the overall range of variation of each parameter. The application of the Morris method has led to the identification of 8 influent parameters. In order to dissociate the individual effects from the interaction effects, a Sobol analysis has also been done leading to nearly the same ordering than Morris one.

LOCAL SENSIVITY ANALYSIS

The analysis can be local, around a given set of parameters value. $\sigma_{H_j(x)}$ can be approximated directly from a model linearization. Indeed, Jacobian of our model can be computed by composition of symbolic differentiation and automatic differentiation [4]. Jacobian matrix is of size (n x m) (m is the size of model output vector). Depending on n and m values comparison, a forward or a reverse (respectively, tangent or adjoin) method is preferred. In the general case, m is greater than n, so a reverse approach is preferred. In this case, the Jacobian computation is faster than finite difference approach. This local sensitivity approximation has been compared with classical global sensitivity, leading to the same 8 influent parameters. Furthermore, a global sensitivity can also be performed using this Jacobian [6].

GRADIENT BASED OPTIMIZATION

The sensitivity information allows sorting parameters regarding their influence. But, in our approach, the sensitivity information is also crucial during an optimization process, in order to guide the algorithm to find the optimal reliable solution. Then, we choose to compute the standard deviation σ_y using finite differences in order to keep exact Jacobian computation to the optimization algorithm. From the classical forward finite differences, the partial derivatives of σ_y are given by the following equation (3), in which $\sigma_{y|xi}$ is the standard deviation of the y output, considering only x_i input variation.

$$\frac{\partial \sigma_y}{\partial x_j} = \frac{1}{2\sigma_y} \sum_{i=1}^{n} \left(2\sigma_{y|x_i} \frac{\partial \sigma_{y|x_i}}{\partial x_j} \right) = \frac{1}{\sigma_y} \sum_{i=1}^{n} \left(\sigma_{y|x_i} \frac{\partial \sigma_{y|x_i}}{\partial x_j} \right) \quad (3)$$

We have then performed the optimization using the SQP algorithm in which constraints using k-sigma approach are taken into account. The following table check the results for k-sigma=2 using a Monte Carlo simulation (*10000 runs, 19 parameters with 5% uncertainties*).

	MC μ	**MC σ**	**k_σ=2 (based on MC)**	**k_σ=2 (from optim)**
Contact length [nm] (>=300)	347.6	25.06	[**297.48**, 397.72]	[300, 398.52]
Contact force [10-8N] (>=2)	3.6	0.42	[2.76, 4.44]	[2.87, 4.618]

ACKNOWLEDGMENT

This application study was done during the PhD of Phuong Pham QUANG, 2011.

CONCLUSIONS

In this work, we have proposed an implementation of RBDO based on k-sigma constraints and a local sensitivity approximation. This approximation, which remains valid for reasonable uncertainties (less than 10%), allows to compute the standard deviation as well as its Jacobian to perform gradient based RBDO. A tool implanting this strategy has been developed and is able to add the standard deviation computation as well as its Jacobian from a black box model. Our methodology and tool has been applied on the reliable design of a nanometric magnetic switch.

References :

[1] Y. Tsompanakis and al. “Structural design optimization considering uncertainties”. *Taylor & Francis*, 2008.
[2] A. Saltelli et al.“Sensitivity Analysis”. *Chichester: Wiley Series in Probability and Statistics*, 2000.
[3] P. Pham-Quang, B. Delinchant, J.L. Coulomb and B. du Peloux. “Optimization of a magnetic deformable nano switch based on semi-analytical model”. OIPE 2010, 14-18 Sept 2010, Sofia
[4] P. Pham-Quang, B. Delinchant. “Java Automatic Differentiation Tool Using Virtual Operator Overloading”. *Recent Advances in Algorithmic Differentiation. S. Forth et al. (eds.)*, pp 241-250, Springer; 2012.
[5] Forrester, A. I.J. “Engineering design via surrogate modelling: a practical guide”, *Wiley* 2008
[6] I.M. Sobol, S. Kucherenko “Derivative based global sensitivity measures and their link with global sensitivity indices”, *Mathematics and Computers in Simulation* 79 (2009) 3009–3017

[Benoit DELINCHANT, Grenoble Electrical Engineering Lab., benoit.delinchant@ujf-grenoble.fr - http://g2elab.grenoble-inp.fr]

7th International Conference on Sensitivity Analysis of Model Output, July 1–4 2013, Nice

Global Sensitivity Study of Compact Models in Nanodevices with Correlated Inputs

Ivan, T., Dimov
Rayna, S., Georgieva
Institute of Information and Communication Techmologies, Bulgarian Academy of Sciences (IICT-BAS), Sofia, Bulgaria

Compact models (CMs) are extensively used in transistor-level digital-circuit design and verification. These models are presented by equations based on fundamental solutions of differential equations describing the main physical processes in circuits. Usually these equations have more than one hundred inputs. Compact models are used to simulate circuits with non-identical devices (transistors). This is equivalent to choosing different values for corresponding parameters of each transistor in circuit. It is necessary the distribution of the parameter under consideration to be known in order to choose a particular value. The distribution is determined by fitting against a statistical set of electrical characteristics obtained from physical device simulation. Usually the number of transistors in a microchip is more than 10^6. On the other hand, the smaller transistors become, the more atomic-scale imperfections will affect their performance. That is why the following questions are reasonable: 1. What is the most suitable (and minimal) set of parameters to choose in order to predict the performance of corresponding transistor/circuit/microchip in a most accurate way? 2. How to classify the elements in this set?

The statistical variability in modern Complementary Metal–oxide Semiconductor (CMOS) transistors is introduced by the inevitable discreteness of charge and matter, the atomic scale non-uniformity of the interfaces and the granularity of the materials used in the fabrication of integrated circuits [1]. This statistical variability means that circuits built from billions of transistors with individually-unique properties may not perform as well as expected, despite being manufactured in an identical way. The statistical extraction of compact model parameters done by direct fitting to a statistical set of physically simulated characteristics is presented in various papers [2, 7].

In this work, the threshold-voltage-based BSIM4 compact model (http://www-device.eecs.berkeley.edu/∼bsim3/bsim4.html) is under consideration. An initial sensitivity study of statistical variability with respect to four input parameters has been done: V_{th0} (basic long-channel threshold voltage parameter), U_0 (low-field mobility parameter), R_{dsw} (basic source/drain resistance parameter), D_{sub} - drain-induced barrier-lowering (DIBL) parameter.

Variance-based techniques for global sensitivity analysis have been applied in the particular case study. The input data analised has been obtained during runs of the mathematical model BSIM4. A detailed analysis of metamodeling procedure in this particular case has been done. The Sobol method is one of the most often used variance-based methods [5]. An important advantage of this method is that it allows to compute not only the first-order indices, but also indices of a higher-order in a way similar to the computation of the main effects. The problem of providing global sensitivity analysis applying Sobol approach (and its modifications [4,6]) consists in evaluating total sensitivity indices, and in particular Sobol global sensitivity indices of corresponding order. It represents a problem of multidimensional integration.

As a first step numerical tests to compute sensitivity measures are performed by the sensitivity analysis software SIMLAB v2.2 (http://simlab.jrc.ec.europa.eu/). The results allow to classify input parameters by their importance. The most important parameters are: V_{th0} and U_0. The main effect and the interaction effects has been taken into account. Some unacceptable negative values of sensitivity indices have been obtained (see Figure 1). That is why a subset of three inputs is studied where the input with a negligible influence has been excluded. The results show that the classification of inputs in this case is similar to the first one. Based on the conclusions resulting from the initial sensitivity study of the problem under consideration, two main issues should be taken into account in the next stage: distribution of input parameters and efficient approaches for model sensitivity study with dependent variables. Numerical study has been performed for both cases - uniform and normal distribution of inputs. An approach for estimation variance-based

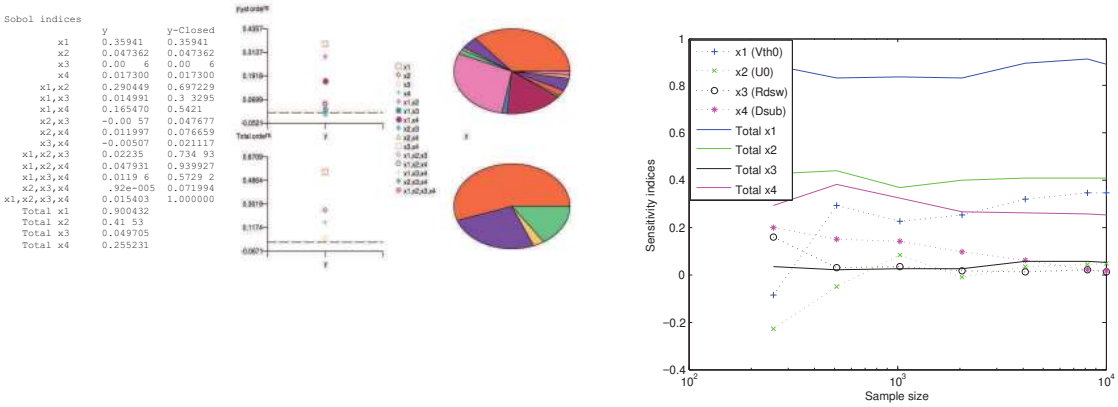

Figure 1: Some results from sensitivity studies of BSIM4 compact model.

sensitivity indices for models with dependent variables presented in [3] has been applied here. Sensitivity indicators have been estimated by Monte Carlo approaches. A comparison of numerical results obtained by sensitivity analysis approaches [4] for both cases (correlated and independent inputs) has been done.

Acknowledgment

The research reported in this work is partly supported by the Bulgarian NSF Grants DMU 03/61/2011, DTK 02/44/2009, and DCVP 02/1.

References:

[1] A. Asenov and B. Cheng, "Modeling and Simulation of Statistical Variability in Nanometer CMOS Technologies", In: *Analog Circuit Design*, H. Casier et al. (eds.), 17–33, 2011.

[2] B. Cheng, D. Dideban, N. Moezi, C.Millar, G. Roy, X. Wang, S. Roy, and A. Asenov, "Statistical Variability Compact Modeling Strategies for BSIM4 and PSP", *IEEE Journal of Design and Test of Computers*, **27**(2):26–35, 2010.

[3] S. Kucherenko, S. Tarantola, and P. Annoni, "Estimation of Global Sensitivity Indices for Models with Dependent Variables", *Computer Physics Communications*, **183**(4):937–946, 2012.

[4] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola, "Global Sensitivity Analysis. The Primer", *John Wiley & Sons Ltd.*, 2008. ISBN: 978-0-470-05997-5.

[5] I. M. Sobol, "Global Sensitivity Indices for Nonlinear Mathematical Models and Their Monte Carlo Estimates", *Mathematics and Computers in Simulation*, **55**(1-3):271–280, 2001.

[6] I. Sobol and E. Myshetskaya, "Monte Carlo Estimators for Small Sensitivity Indices", *Monte Carlo Methods and Applications*, **13**(5-6):455–465, 2007.

[7] K. Takeuchi, M. Hane, "Statistical Compact Model Parameter Extraction by Direct Fitting to Variations", *IEEE Transactions on Electron Devices*, **55**(6):1487–1493, 2008.

[Rayna Georgieva; IICT-BAS, Acad. G. Bonchev Str., Bl. 25A, 1113 Sofia, Bulgaria]
[`rayna@parallel.bas.bg` – `http://parallel.bas.bg/~rayna/`]

Dealing with uncertainty in technical & economic studies of investment strategy optimization

FANNY, DOUARD
EDF R&D, Département MRI, FRANCE

BERTRAND, IOOSS
EDF R&D, Département MRI, FRANCE

Context

To optimize the whole life cost of its nuclear fleet, EDF has developed an asset management methodology. A part of this methodology treats with exceptional maintenance tasks strategies. To help the decision maker to choose the best strategy (how many times do we need to realize exceptional tasks, when,...?), EDF has developed a dedicated tool (described in Figure 1) based on Monte-Carlo simulation to compute many technical economic indicators among which the density function of the Net Present Value (NPV) is the most relevant.

This tool leads to an important simulation time and requires many input data that are surrounded with uncertainty:

- Reliability data : generally there isn't enough (or sometimes not any) feedback data to evaluate reliability model parameters;
- Economic data : economic indicators and duration of Maintenance tasks remain on several hypothesis that can be modified;
- Other data: uncertainty on operating time of power plants, maintenance tasks dates, etc.

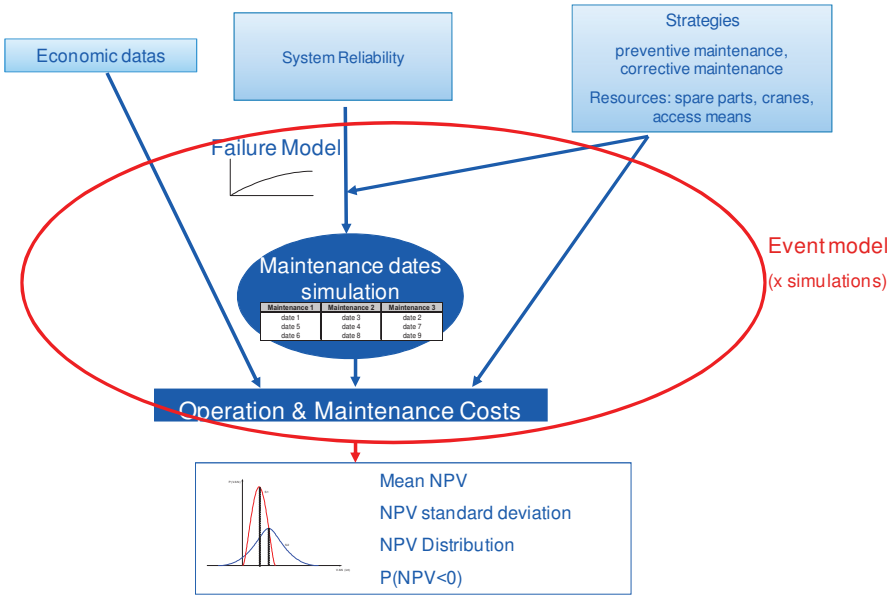

Figure 1 : EDF valorization of maintenance strategies tool description

In order to reduce the output uncertainties, EDF needs to understand the influence of variables and their interactions on the NPV values. The main objectives are then to quantify

- The non-influent variables that can be fixed without consequences on the output uncertainty,
- The most influent input variables: variables to be fixed to obtain the largest reduction of the output uncertainty.

Objective

The problem of sensitivity of NPV is recently discussed in papers in the operations research literature (see [2]). The present work presents the use of factorial fractional designs [1] to treat this problematic, taking into account different sort of output as mean, standard deviation, mode or distribution of NPV. We will briefly introduce factorial fractional designs and their use in a sensitivity analysis context.

The originality of this works stands in the fact that the model output is not a scalar value but a density function. We will discuss which criteria can be used to transform this functional output problem in a scalar output one. We will test different standard distances between two density functions.

We finally will present a numerical application based on a real study that optimizes the replacement of a stator winding of an EDF power plant.

References

[1] D.C. Montgomery, *Design and analysis of experiments,* Wiley, 2008.
[2] M. Baucells and E. Borgonovo, Invariant Probabilistic Sensitivity Analysis, *Management Science*, 2013 (in press).

[Fanny Douard, EDF R&D, 6 Quai Watier, 78401 Chatou, France]
[fanny.douard@edf.fr]

Selection of Test Patterns for Model Calibration based on Sensitivity Analysis

Thiago R. Figueiro
Aselta Nanographics, MINATEC-BHT, 7 Parvis Louis Néel - BP50 - 38040 Grenoble, France
CNRS/UJF-Grenoble1/CEA LTM, 17 Avenue des Martyrs, 38054 Grenoble, France

Jean-Herve Tortai, Nader Jedidi
CNRS/UJF-Grenoble1/CEA LTM, 17 Avenue des Martyrs, 38054 Grenoble, France

Mohamed Saib, Patrick Schiavone
Aselta Nanographics, MINATEC-BHT, 7 Parvis Louis Néel - BP50 - 38040 Grenoble, France

The lithography is one of the key steps in the fabrication of integrated circuits. It consists in transferring the geometric patterns that represents each level of the circuit into a resist on a substrate. As technology advances, the dimensions of these geometric patterns become smaller and smaller, requiring improvements on the precision of the lithography techniques. One possible candidate to address future nodes (below 16nm) is electron beam lithography. If multiple electron beam systems significantly increase the throughput to meet industrial needs, it can be the tool of choice. Nevertheless using a chemically amplified resist (CAR) is mandatory even for systems with a large number of beams. Achieving dense sub 20 nm patterns with CAR is still a challenge as proximity effects degrade the contrast of the aerial image. Bridging, shape rounding or partial development are typical degradation in the desired final pattern shape. Proximity effect correction (PEC) is needed in order to properly delineate dense features as well as meet the required CD uniformity. The correction will even out the non-ideal electron energy deposition using a proper adjustment of the dose assigned to each pattern. Model based PEC requires computing the electron lithography process comprising at least the electron deposition inside the resist as well as the resists behavior under electron exposure. The electron energy deposition is calculated using a convolution approach. The impulse response of the electron beam, which is called PSF (Point Spread Function), is convoluted with the exposed pattern to compute the 2D repartition of electron energy deposited in the resist. Therefore, the quality of a correction is highly dependent on the quality of the PSF model employed and the accuracy of its parameters. The sum of Gaussian functions commonly employed for PSF models is shown in (1).

$$\mathrm{PSF}(x) = \frac{1}{1+\sum \eta_i}\left[\frac{1}{\pi\alpha^2}\exp\left(-x^2/\alpha^2\right) + \sum \frac{\eta_i}{\pi\beta_i^2}\exp\left(-x^2/\beta_i^2\right)\right] \quad (1)$$

There are other empirical models available in the literature, as well as several different approaches to obtain the appropriate value for their parameters [1-2]. However, those techniques are usually limited to certain model characteristics, present some restriction in terms of complexity, or require a prohibitive number of measurements.

One of the most employed strategies to determine the parameters of a PSF model is based on experimental data obtained from measurements on a wafer or mask. The constraint in this approach is to propose a reasonable number of measurements, due to constraints of writing time and metrology time. Figure 1 shows several examples of patterns that may be used.

The selection of the patterns is an important step for the model calibration procedure since it impacts directly the quality of the parameter values extracted during the optimization. This impact is due to the fact that the patterns used must explore the characteristics of the model, being sufficiently comprehensive to include the situations where the model accuracy is required to obtain a precise simulation/correction.

Therefore, it is crucial to select the patterns that are more sensitive to the influence of all parameters on the model. The strategy employed to select those representative patterns is the Sensitivity Analysis. This method indicates the importance of individual parameters' contribution to the model result. In this sense, for each calibration pattern evaluated, the sensitivity of the metrology simulation result over

it is analyzed. The Sensitivity Analysis method used was the calculation of the Global Sensitivity Indices [3,4] both of first order and total order. Global sensitivity indices can be efficiently computed by Variance-based methods [3], implemented using quasi-random sampling algorithms [4,5]. Figure 2 shows one example of evaluation of the first order global sensitivity index for 128 patterns for the three parameters of a two Gaussian PSF. Observe that there are some with high sensitivity to β and several with high sensitivity to η, but none very sensitive to α alone.

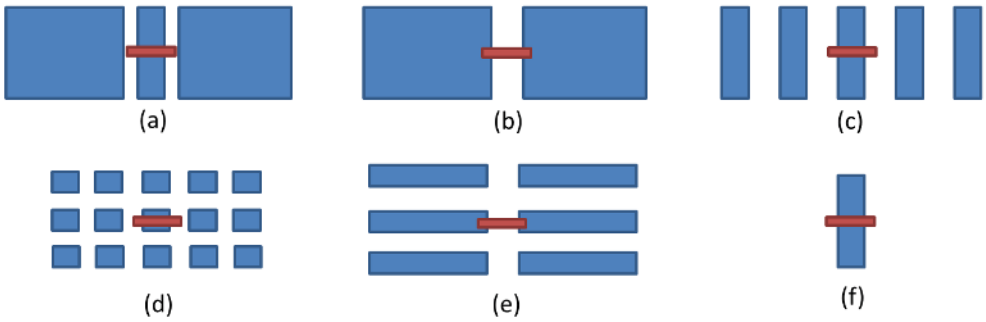

Figure 1. Samples of test patterns (in blue) and the indication where the measurement is taken (in red). (a) line between pads; (b) isolated inverted line; (c) line gratings; (d) contact blocks; (e) tip-to-tip blocks; (f) isolated line.

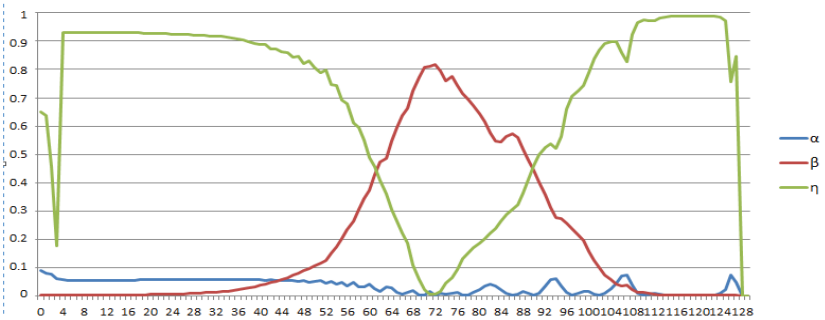

Figure2. First Order Global Sensitivity Index values for 128 different patterns(by changing the dimensions of the shapes presented in figure 1(a) for a two Gaussian PSF model

Another important consideration is the impact of the process variability and the metrology noise over the Sensitivity Analysis. It is known that some patterns may suffer of a higher variability than others (from less than 1nm in the stable cases up to 3 or 4 nm in the most instable ones). Determining the impact of the variation of the measurement over those patterns in the final model calibration is important in order to determine the most robust set of patterns for this purpose. For this purpose, Uncertainty Analysis is employed. The main goal here is to find a compromise between the amount of uncertainty on the parameter choice related to a given pattern and the information it provides to the calibration.

References :

[1] H. Kobinata, Y. Yamada, T. Tamura, K. Fujii, M. Narihiro, Y. Ochiai, "Estimation of Optimum Electron-Beam Projection Lithography Mask Biases Taking Coulomb Beam Blur into Consideration", Jpn. J. Appl. Phys. 42, 3816–3821, (2003)

[2] T. Figueiro, M. Saib, N. Jedidi, P. Schiavone, "Improved electron backscattering representation using a new class of distribution: application to EUV masks", Published in Proceedings SPIE Volume 8323, (2012)

[3] A. Saltelli, S. Tarantola, F. Campolongo, "Sensitivity Analysis as an Ingredient of Modeling", Statistical Science 15, 377-395, (2000)

[4] I.M. Sobol', S.S. Kucherenko, "Global Sensitivity Indices for Nonlinear Mathematical Models. Review, Mathematics and Computers in Simulation", Volume 55, Pages 271 – 280, (2001)

[5] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, S. Tarantola, "Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index", Computer Physics Communications 181(2): 259-270 (2010).

7th International Conference on Sensitivity Analysis of Model Output, July 1-4 2013, Nice

Multi-run simulation environments – requirements, development and application in climate impact research

MICHAEL FLECHSIG
THOMAS NOCKE
CLAUS RACHIMOW
Potsdam Institute for Climate Impact Research, 14476 Potsdam, Germany

A simulation environments is a software tool that supports experimenting with simulation models normally developed outside the environment. Formally, we consider the implementation of a mathematical model M: y = F(x) that transforms a k-dimensional input vector x of factors into an output (vector) y [1]. Often, the input vector x represents parameters, initial or boundary values of the model. For a multi-run experiment a sample is drawn from the factor space in a way that fits to the experiment goal and the model is run for all sample points.

In climate impact research model development and application are the main methodologies besides data analysis and integration. Consequently, evaluation of model results is an important task and uncertainty and sensitivity quantification due to the variability of model factors have been increasingly requested by policy and decision makers in recent years (e.g., [2] and [3]). Typically, simulation models are implemented in programming languages rather than modelling systems, they are non-linear, time consuming and produce large volume of multi-variate and multi-dimensional output. Besides validation tasks, simulation experiments are mainly performed for uncertainty, sensitivity, scenario and re-analyses. For this class of models and tasks methodical challenges arise from (i) integrating and interfacing models implemented in different languages, (ii) sensitivity and uncertainty studies in high-dimensional factor spaces, (iii) comparing experiments performed under different settings, and (iv) computation of experiment-specific measures during experiment analysis.

Besides supplying adequate model-independent / model-free experiment designs for non-additive, non-linear, and non-monotonous models (e.g., qualitative and quantitative, deterministic and probabilistic and Bayesian, local and global designs, [4], [5]) the open architecture of a simulation environment is an important feature for its acceptance. This mainly includes (i) a model interface to import factor values into the model and to export model output to the environment, (ii) load distribution for such multi-run experiments where the factor space can be sampled before simulation and where all single runs are independent from each other, (iii) a post-processor to navigate the coupled factor – state space and derive measures, and (iv) an interface to visual data analysis tools to visualize the measures.

SimEnv [6] is a multi-run simulation environment with the focus on uncertainty, sensitivity and scenario analysis of multi input / output models that meets most of the above criteria: Experiment design is based on pre-defined experiment types for factor spaces of any dimension that only have to be equipped with numerical values. Experiment types cover a variance based technique, a Monte Carlo experiment, local sensitivity analysis, an improved Morris design, (fractional) factorial experiments, Bayesian calibration, and a one-criterial optimization. For probabilistic methods, built-in marginal distributions are supplied for pseudo, stratified and quasi random sampling.
SimEnv comes with a simple model interface that requires only minimal modifications of C/C++, Fortran, Java, Python, Matlab, Mathematica, GAMS or shell script model source code by a SimEnv function call for each factor and each output field under investigation. Multi-variate / -dimensional experiment output is stored in self-describing NetCDF data format.
The environment allows for different flexible load distribution strategies of the individual single runs, supporting multi-core processor machines, experiments in background as well as load distribution to compute clusters, the latter with interfaces to batch queue systems.

In experiment analysis chains of built-in and user-defined operators are interactively specified and applied to experiment output over the factor space to derive secondary experiment output in general and uncertainty / sensitivity measures from secondary output if appropriate. External (reference) data and output from other SimEnv experiments can be embedded for model – data and model or experiment comparisons. Again, analysis output is stored in NetCDF format.
SimEnv is coupled to the visualization system SimEnvVis [7] for interactive explorative visual data analysis. It exploits metadata from the experiment design and the experiment post processor to select appropriate visualization techniques. One of the advantages of SimEnvVis is the ability to cope with multi-run datasets by special visualization techniques like parallel coordinates and graphical tables.

SimEnv has been used in climate impact research for many problem settings, among others, uncertainty and sensitivity analyses [8], [9], and uncertainty quantification in model intercomparisons. [10].

The paper closes with consequences for further investigation into methodology, design and implementation of simulation environments.

References:

[1] Box, G.E.P., Hunter, W.G., Hunter, J.S., "Statistics for experimenters: An introduction to design, data analysis and model building", Wiley, 1978.
[2] Webster, M., "Communicating climate change uncertainty to policy-makers and the public", *Climatic Change*, **61(1-2)**:1-8, 2003.
[3] Schneider von Deimling, T., Held, H., Ganopolski, A., Rahmstorf, S., "Climate sensitivity estimated from ensemble simulations of glacial climate", *Climate Dynamics*, **27**:149-163, 2006.
[4] Saltelli, A., Tarantola, S., Camplongo, F., Ratto, M. "Sensitivity analysis in practice: A guide to assessing scientific models", Wiley, 2004.
[5] Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., "Global sensitivity analysis. The primer", Wiley, 2008.
[6] Flechsig, M., Böhm, U., Nocke, T., Rachimow, C., "The multi-run simulation environment SimEnv. User's Guide", Potsdam Institute for Climate Impact Research, Potsdam, Germany, 2012, http://www.pik-potsdam.de/software/simenv/
[7] Nocke, T., "Visual data mining and visualization design for climate research", PhD Thesis, University of Rostock (in German), 2007.
[8] Knopf, B., Flechsig, M., Zickfeld, K., "Multi-parameter uncertainty analysis of a bifurcation point" *Nonlin. Processes Geophys.*, **13**:531-540, 2006.
[9] Knopf, B., Zickfeld, K., Flechsig, M., Petoukhov, V., "Sensitivity of the Indian monsoon to human activities" *Advances in Atmospheric Sciences*, **25/6**:932-945, 2008.
[10] van Oijen, M. *et al.*, "Bayesian calibration, comparison and averaging of six forest models, using data from Scots pine stands across Europe", *Forest Ecology and Management*, **289**:255–268, 2013.

[flechsig@pik-potsdam.de - http://www.pik-potsdam.de/members/flechsig/]

[nocke@pik-potsdam.de - http://www.pik-potsdam.de/members/nocke/]

[rachimow@pik-potsdam.de - http://www.pik-potsdam.de/members/rachimow/]

7th International Conference on Sensitivity Analysis of Model Output, July 1-4 2013, Nice

Sensitivity analyses of a fast analytical radionuclide transport model

ALLAN B HEDIN
Swedish Nuclear Fuel and Waste Management Co., SKB, Sweden, allan.hedin@skb.se

Introduction

In March 2011, the Swedish Nuclear Fuel and Waste Management Co., SKB, applied for a licence to build a final geological repository for spent nuclear fuel at the Forsmark site, situated around 70 miles north of Stockholm, Sweden. A key component in the licence application is an assessment[1] of the long-term safety of the repository. Probabilistic radionuclide transport and dose calculations are at the core of the analysis. A number of sensitivity analysis methods, e.g. standardised rank regression and cobweb plots, were applied to the calculated dose distributions. The numerical radionuclide transport models used in the assessment can, for several of the analysed scenarios, including the most important ones, be well approximated by simple analytical expressions,[2] that use the same input data distributions as the numerical models. This opens the possibility for additional sensitivity analysis methods that are explored in this paper:

1. The very short execution time of the analytical model facilitates the calculation of variance based sensitivity indices of all orders, for this problem with five input variables.
2. Based on the analytical expressions, a tailored regression model is derived and shown to make good predictions of the results obtained with the numerical transport models, i.e. the analytical model is utilised to derive a sensitivity analysis method for the more complex numerical model.

Variance based sensitivity indices of all orders

The simplified analytical model requires of the order of one microsecond per realisation if only the peak dose, which is the primary entity in the demonstration of compliance with Swedish regulations, is considered. There is thus a good potential for these simplified models to be efficient in the determination of variance based sensitivity indices. The full numerical calculation includes a number of radionculides that contribute insignificantly to the total dose and these and their corresponding uncertain variables are, therefore, not considered in the following. The analytical expression for the dose dominating radionuclide Ra-226 is, somewhat simplified, the following, where the uncertain input variables are D_{Fuel}, $t_{Failure}$, F, K_d^{Ra} and D_e:

$$\mathrm{Dose}^{Ra\,226} = D_{Fuel}(t_{Max} - t_{Failure})\exp\left\{\frac{Pe}{2}\left[1-\sqrt{1+\frac{4t_w\lambda^{Ra\,226}}{Pe}\left(1+\frac{F\sqrt{(\varepsilon_P+(1-\varepsilon_P)K_d^{Ra}\rho)D_e}}{t_w\sqrt{\lambda^{Ra\,226}}}\right)}\right]\right\}$$

For obtaining sensitivity indices, formula (b) in Table 2 of Ref. 3 for first order indices was taken as the starting point. Analogous formulae were derived so that, in addition to the first order indices, also all higher order indices were obtained. Application of these formulae yielded e.g. all sensitivity indices of a 5-dimensional Sobol' g-function with all $a_i = 0$ (see, e.g., Ref. 3 Appendix A) within about one minute of computation time, with input data sampled[4] with a Quasi Monte Carlo scheme. 2^{20} model realisations were executed in this example for each of the 2^5-1 indices, i.e. in total about 30 million realisations. The relative errors were around one percent for all indices except for the 5th order where it was a few percent. The precision was readily increased with additional realisations. Since all sensitivity indices are calculated, total order indices are obtained as sums of relevant first and higher order indices and do not have to be calculated separately. Additional schemes to the mentioned formula in Ref. 3 were tried as starting points, yielding similar or lower efficiency.

The method was applied to the radionuclide transport problem and yielded all sensitivity indices up to and including the 5th order within a few minutes of computation time. The results point to the same

important variables as obtained with e.g. standardised rank regression. The sum of the fist order indices of the three most important parameters, D_{Fuel}, $t_{Failure}$ and F, is 0.46. The sum of the seven 1st, 2nd and 3rd order indices including only these three most important variables was 0.88, i.e. they account for 88 percent of the total variance. The details of this latter information do not emerge if only first and total order indices are determined. It is clear from these results that variable interaction effects give significant contributions to the output variance and this is expected from the nature of the analytical expression above. It is, moreover, not clear that the higher (4th and 5th) order indices provide meaningful information about the real system, considering the simplified nature of the analytical model.

Tailored regression models

The understanding and mathematical formulation of the simplified transport models[2] used in the dose calculations were utilised to construct a number of tailored regression models that include successively more input variables. Such models with two to five variables were constructed for the peak dose of Ra-226, the radionuclide that dominates most of the numerical model realisation. The highest order model yields an R^2-value of 0.99 when regressed on the results of the numerical transport models, see Figure 1, that shows both the regression model expressions as an insert, and the regression results.

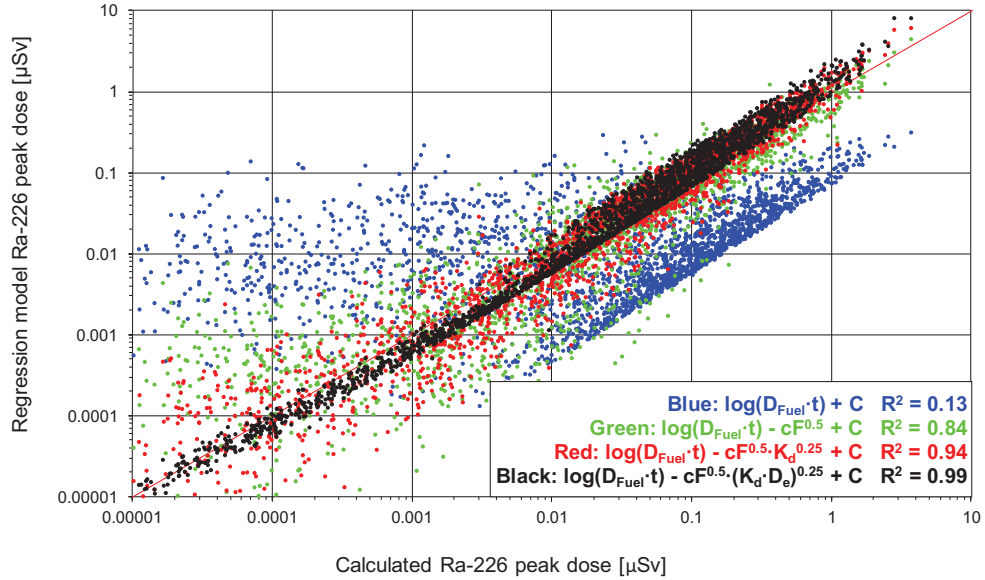

Figure 1. Four tailored regression models, based on successively more variables, for the Ra-226 peak dose. For example, the black dots show the good agreement of the results of a five-parameter tailored regression model with those of the numerical transport model.

Conclusion

A complex model that is well approximated by a simple analytical function has facilitated i) a full exploration of variance based sensitivity indices and ii) the development of tailored regression models that is used for sensitivity analysis of the complex model. Both methods illustrate how input variable uncertainties contribute to the output variance.

References

1. Swedish Nuclear Fuel and Waste Management Co., SKB, "Long-term Safety for the Final Repository for Spent Nuclear Fuel at Forsmark. Main Report of the SR-Site Project," SKB Technical report TR-11-01, 2011. Available at www.skb.se
2. A. Hedin, "Integrated analytic radionuclide transport model for a spent nuclear fuel repository in saturated fractured rock", Nuclear Technology, **138**, pp 179–205, 2002.
3. 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", Computer Physics Communications **181**, pp 259–270, 2010.
4. Freeware for QMC sampling (SobolSeq generator for Microsoft Excel) from www.broda.co.uk

7th International Conference on Sensitivity Analysis of Model Output, July 1-4 2013, Nice

Schur vectors for estimating parameters in filtering algorithm for data assimilation

SON, HONG, HOANG
Res. Eng., SHOM/HOM, Toulouse, France

REMY, BARAILLE
Res. Eng., SHOM/HOM, Toulouse, France

The goal of data assimilation is to correct in an optimal way the system forecast using the given observations. For meteorological or oceanic numerical models, implementing standard filtering algorithms like Kalman filter is an insurmountable task due to very high dimension of the resulting system state (order of 10^6 - 10^7).

We describe in this paper a new approach for the design of a suboptimal filter called as a **Prediction Error Filter** (PEF) [1]. In the PEF, the gain subspace is either constructed in a subspace of leading Schur vectors if the system is of low dimension or is supposed to be given in a prescribed form, with some parameters (vertical error covariance matrix, horizontal correlation length ...) to be estimated from the patterns of the faster growing prediction errors (PE). These patterns, simulated by applying the orthogonal iteration algorithm to the system dynamics, will tend to the real leading **Schur vectors**. A procedure for estimating the filter gain elements from Schur vectors will be presented. By this way the filter is automatically constructed from the numerical model, at low memory and computational cost.

The PE sampling approach is proposed in order to have a possibility to well estimate the PE covariance matrix which is needed to initialize the filter gain. That is the first step in implementing the more advanced algorithm known as an adaptive filtering [2]. In the **adaptive filter** (AF) the optimization of the filter performance is performed by minimizing the mean PE for the system output with the control vector consisting of some pertinent elements of the filter gain. Advantages of such choice of control vector are : (1) Simplicity in selecting parameter intervals ensuring a stability of filter (hence stability of optimization procedure) for unstable dynamical system; (2) Possible choice of control parameters on which the predictor depends linearly; (3) Optimization can be performed by a simultaneous perturbation stochastic approximation algorithm which does not require a development of adjoint code for a linearized system dynamics (i.e., it is possible to perform an optimization by 3 time integrations of numerical model).

A number of numerical examples of filtering problems with dynamical systems, of low and very high dimension (oceanic model), will be given to illustrate the efficiency of the proposed approach. A **sensitivity analysis** of objective functions to control vector in an AF and in a variational data assimilation (VDA) approach will be given from which, for example, one sees that if for a strongly nonlinear Lorenz system, a cost function being a squares of PEs, results in a nearly quadratic function with respect to the proposed parameters in the filter gain, the sensitivity analysis does not lead to a positive conclusion on a possible existence of minimizer in the space of initial system state if we follow the VDA approach.

Performance comparison of the PEF and its adaptive version (optimal in the minimum prediction error sense) with other well known filtering methods, will be also presented.

References :

[1] H.S. Hoang and R. Baraille (2011), Prediction error sampling procedure based on dominant Schur decomposition. Application to state estimation in high dimensional oceanic model, *"Appl. Math. Comput."*, 3689-3709.

[2] Hoang S., Talagrand O. and Baraille R. (2001). On the design of a stable filter for state estimation in high dimensional systems, « *Automatica* », pp. 341-359.

[3] Hoang S., DeMey P., Talagrand O. and Baraille R. (1977). A new ROAF for state estimation in high dimensional systems, 33, « *Automatica* », 1475-1498

[4] Miller R.N., Ghil M. and Gauthier F. (1994). Advanced Data Assimilation in Strongly Nonlinear Dynamical Systems. « J. Atmos. Sci. », pp. 1037-1056.

[5] C.S. Spall C.S. (1998), An overview of the simultaneous perturbation method for efficient optimization, « *Johns Hopkins APL Tech. Digest* », 19 (4) (1998) 482-492.

[Hong Son Hoang; SHOM/HOM, hhoang@shom.fr]

Robust Assessment of Different Machine Learning Methods for Time Series Forecasting using a Factorial Design of Experiment

KARIN KANDANAOND
Rajabhat University Valaya-Alongkorn, Prathumthani, Thailand

The performance of two machine learning methods, artificial neural network (ANN) and support vector machine (SVM), for forecasting time series is considered an open issue for discussions among many authors in the literature. Hence, the purpose of this study is to characterize the capability of these two methods upon the autocorrelation basis. According to the training process, the historical data at time t-50, t-49,…, t-1 were utilized to predict the observation at time t while 50 training vectors were deployed to train the database in each cycle (Fig. 1).

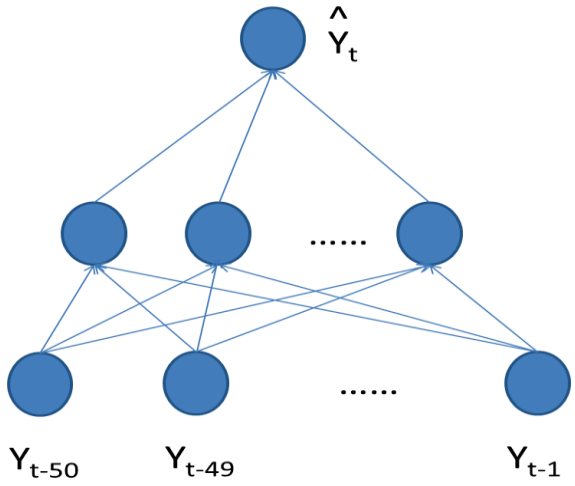

Fig. 1 Prediction of Y at time t

The number of training cycles used was 20,000 cycles and the forecasting error, minimum average percentage error (MAPE), was calculated after each learning cycle. For ANN, the neural network architecture used is the multilayer perceptron (MLP) and the training algorithm of MLP network employed to build models is the Broyden-Fletcher-Goldfarb-Shanno (BFGS). On the other hand, the forecasting model based on the SVM approach is the regression type 1 with C=10.0, epsilon = 0.1 and the kernel is a radial basis function with gamma = 0.1. A statistical package, Statistica version 10, was utilized to facilitate the forecasting by these two methods. For generating data, the numerical simulation in this empirical research is based on the autocorrelated nature (stationary and non-stationary) of the data so the performance of ANN and SVM was compared with respect to the autoregressive integrated moving average (ARIMA) structure. Two classes of ARIMA models, ARMA (1, 1) and IMA (1, 1), as shown in (1) and (2) respectively, were utilized to simulate stationary and non-stationary processes.

$$N_{t+1} = \phi N_t + a_{t+1} - \theta a_t \text{ ; } -1 < \phi < 1, -1 < \theta < 1, \quad (1)$$

$$N_{t+1} = N_t + a_{t+1} - \theta a_t \text{ ; } -1 < \theta < 1, \quad (2)$$

where N_t is the observation at time t
a_t is the random error at time t
ϕ is the autoregressive coefficient
θ is the moving average coefficient.

Afterwards, the experimental results, MAPEs, were compared empirically by using the 2^k factorial design because of its robustness. For stationary processes, three factors, A, B and C, were assigned to three influential factors, AR parameter (ϕ), MA parameter (θ) and methods respectively, while each factor was set at the low and high level as shown in Table 1. Similarly, all factors and their levels for conducting the experimental study of non-stationary processes are shown in Table 2.

Table 1. Factors and levels for stationary processes

Factor	Low	High
A (AR parameter; ϕ)	-0.9	0.9
B (MA parameter; θ)	-0.9	0.9
C (Types of Methods)	ANN	SVM

Table 2. Factors and levels for non-stationary processes

Factor	Low	High
A (MA parameter; θ)	-0.9	0.9
B (Types of Methods)	ANN	SVM

After the experiment, the results were analyzed by a statistical package, Design-Expert version 8, and the most appropriate model was chosen under these two conditions (stationary and non-stationary). The results indicate that these two machine learning methods have different performance under a specific scenario of autocorrelation. Accordingly, when the process is stationary, ANN is a better choice than SVM. However, it turned out to be that SVM had obviously outperformed the ANN for non-stationary case.

[Karin Kandananond; Rajabhat University Valaya-Alongkorn, 1 Moo 20 Paholyothin Rd., Prathumthani, Thailand 13180]
[kandananond@hotmail.com]

7th International Conference on Sensitivity Analysis of Model Output, July 1–4 2013, Nice

Uncertainty and sensitivity analysis of land shares model over EU 27

Matieyendou LAMBONI
EC-Joint Research Centre, Institute for Environment and Sustainability, Italy

Renate KOEBLE
EC-Joint Research Centre, Institute for Environment and Sustainability, Italy

Adrian LEIP
EC-Joint Research Centre, Institute for Environment and Sustainability, Italy

I. Objective: The agricultural sector contributes to $9.3 - 10.6\%$ of total greenhouse gas (GHG) emission, excluding LULUCF, in the EU27 (EEA [1]) and there is a growing interest in modelling environmental impacts of agricultural activities. GHG emission heavily depends on a number of different biophysical characteristics, such as soil, landform, and climate (Lamboni *et al.* [2]; Leip *et al.* [3]) as well as management factors like crop type and fertilizer application. For the quantification of agri-environmental indicators and the estimation of GHG emissions, the distribution of crop shares over the Homogenous Spatial Units (HSU) is required, as many indicators depend on the local combination of land use and environmental conditions. These Spatial Units should express highest possible homogeneity for important factors so that their characterizations and results based on can be regarded as representative. In this paper, we quantify the uncertainty of the prediction of the crop shares for each administrative region (NUTS2) based on land use survey (LUCAS 2009), land use/cover maps (Corine 2006), climate (rain, temperature, vegetation period), soil and topographical variables such as texture, organic content, sand, clay, slope, altitude.

II. Land Shares Model: We used a local multinomial logit model, as a statistical way of downscaling the land use, to estimate the disaggregation function of spatial crops allocation for each administrative region (NUTS2). Cross Validation combined with the F-measure (model effectiveness) [4] were used to select the "optimal" bandwidth during the process of estimation. For each HSU (h) in a NUTS2, characterized by the variables ($\mathbf{x}_h$), the probability to find a land use l is predicted like:

$$\mathbb{P}_r(\widehat{\mathbf{Y}_h = l}) = \frac{\exp^{(\hat{\beta}_l \text{`} \mathbf{x}_h)}}{\sum_{l=1}^{L} \exp^{(\hat{\beta}_l \text{`} \mathbf{x}_h)}}, \tag{1}$$

where, $\hat{\beta}_l$ is the vector of the weighted likelihood estimators for land use l in the NUTS2 and L is the number of land use. $\mathbb{P}_r(\widehat{\mathbf{Y}_h = l})$ is no more than the predicted percentage of land l share in the HSU and the total area of l share in the NUTS2 ($\hat{A}_l$) is:

$$\widehat{A_l} = \frac{1}{H} \sum_{h=1}^{H} \frac{\exp^{(\hat{\beta}_l \text{`} \mathbf{x}_h)}}{\sum_{l=1}^{L} \exp^{(\hat{\beta}_l \text{`} \mathbf{x}_h)}} \times A, \tag{2}$$

where H is the number of HSU in the NUTS2 and A is the area of the NUTS2.

In this model, the uncertain factors are $\hat{\beta}_l,\ l = 1, \ldots, L$ and we assume that the uncertain factors $\hat{\beta}_l$, for each land use, follow a normal distribution with mean (respectively standard deviation) the estimated value of the parameter (resp. standard deviation). This assumption is reasonable for most of the NUTS2 regions because the selected bandwidth leads to more than 200 observations. These distributions of $\hat{\beta}_l$ are used as *a priori* distributions of the parameters of land shares model in Bayesian approach where the unknown distribution of $\widehat{A_l}$ is required.

III. Result: We used the land shares model in (2) to compute the distribution of land shares $\widehat{A_l}$ for each administrative region (NUTS2). First, we use a screening method [5] to select the most influential parameters, and second, we quantify the uncertainty of land shares based only on the most influential parameters. The model output, associated to the mean values of input

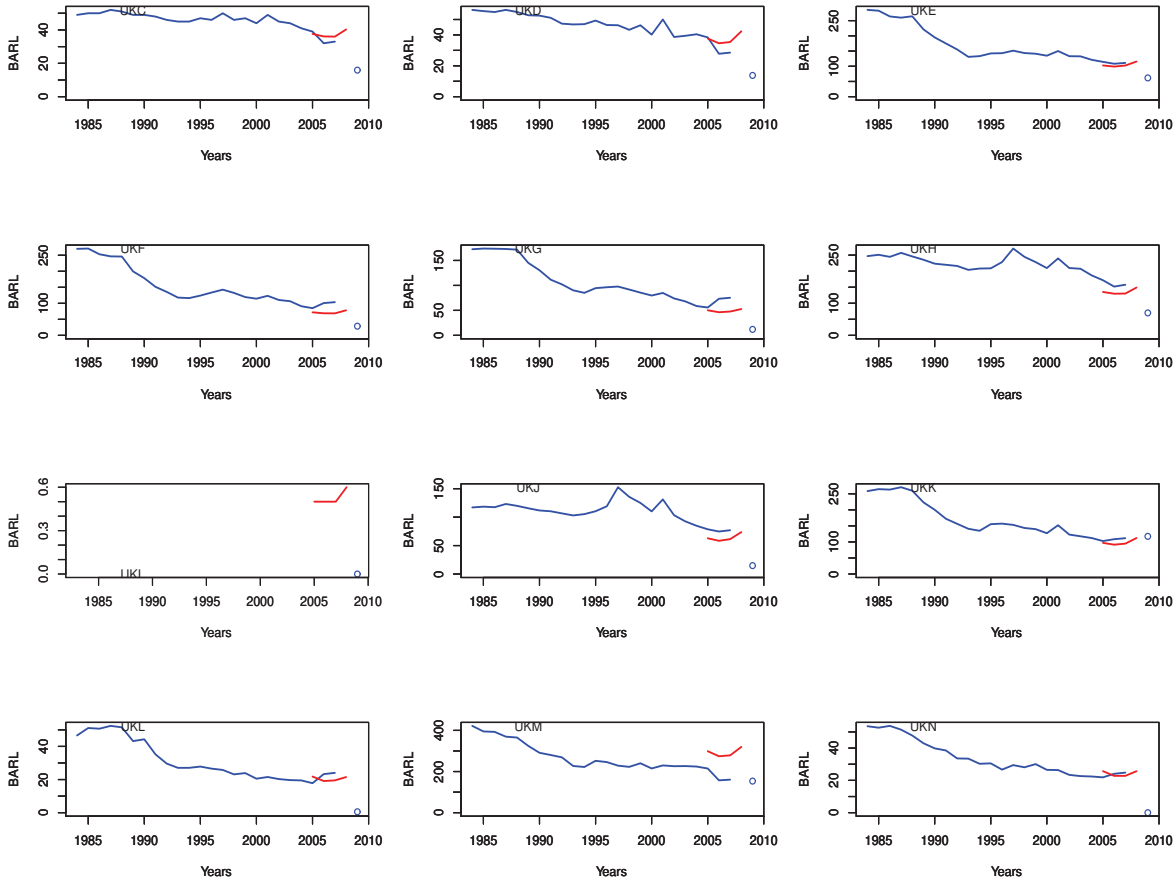

Figure 1: Trends of barley shares from CAPRI (Blue), EUROSTAT (red) and predicted value (point) at NUTS 1 level in UK. We will include information on the probability distribution of estimated crop share.

distributions, is aggregated to a high level (NUTS1 level) to be compared with the trend of land shares from CAPRI data and Eurostat data (Figure 1). In Figure 1, most of predictions follow the trend of crop shares from the statistics.

References: -[1] EEA, 2012. Annual European Community greenhouse gas inventory 1990 - 2010 and inventory report 2012. Submission to the UNFCCC secretariat. European Environment Agency.

-[2] M. Lamboni, D. Makowski, S. Lehuger, B. Gabrielle, H. Monod, Multivariate global sensitivity analysis for dynamic crop models, Field Crops Research 113 (2009), pp. 312-320.

-[3] A. Leip, G. Marchi, R. Koeble, M. Kempen, W. Britz, C. Li, Linking an economic model for european agriculture with a mechanistic model to estimate nitrogen and carbon losses from arable soils in europe, Biogeosciences 5 (1) (2008), pp. 73-94.

-[4] D. Powers, Evaluation: From precision, recall and f-factor to roc, informedness, markedness & correlation, Journal of Machine Learning Technologies 2 (2011), pp. 37-63.

-[5] M. Lamboni, B. Iooss, A.-L. Popelin, F. Gamboa, Derivative-based global sensitivity measures: general links with Sobol' indices and numerical tests, Mathematics and Computers in Simulation, Available online 19/02/2013.

[M. Lamboni; EC-JRC,Ispra, matieyendou.lamboni@jrc.ec.europa.eu]

7th International Conference on Sensitivity Analysis of Model Output, July 1–4 2013, Nice

Rough paths theory applied to computation of sensitivities

Nicolas MARIE
Université Paris X, Nanterre, France

Let S be the prices process of $d \in \mathbf{N}^*$ risky assets, and assume it is the solution of the following stochastic differential equation on $[0,T]$ $(T>0)$:

$$dS_t = b\left(S_t\right) dt + \sigma\left(S_t\right) dW_t \tag{1}$$

with a deterministic initial condition $x \in \mathbf{R}^d$.

In their famous paper [3] published in 1999, by taking equation (1) in the sense of Itô for a d-dimensional Brownian motion W, E. Fournié et al. studied sensitivities with respect to x, b and σ, called *greeks* in finance, of an option's price $\mathbf{E}[F(S_T)]$. By using Malliavin calculus, they proved that *greeks* can be written $\mathbf{E}[F(S_T)\vartheta]$ for F only piecewise continuous and a weight $\vartheta \in L^2$.
About computation of *greeks* in the classical Itô calculus framework, see also E. Gobet and R. Münos [4].

Assume that W is a d-dimensional Gaussian process such that the *enhanced Gaussian process* in the sense of [2], Theorem 15.33 exists over it. Typically, a fractional Brownian motion with Hurst parameter $H > 1/4$ satisfies that assumption.
With bounded coefficients, by taking equation (1) in the sense of rough paths introduced by T. Lyons in 1998 ; the existence of sensitivities with respect to x and σ extends by assuming that F is continuously differentiable such that $\|F(x)\| < C(1+\|x\|)^N$ and $\|DF(x)\| < C(1+\|x\|)^N$ for two given constants $C>0$ and $N \in \mathbf{N}^*$. The regularity of the Itô map for equation (1) together with recent results of T. Cass, C. Litterer and T. Lyons [1] on linear rough differential equations driven by *enhanced Gaussian processes* are crucial.
Under technical assumptions on the Cameron-Martin's space of W, satisfied by fractional Brownian motions for instance ; the sensitivities of $\mathbf{E}[F(S_T)]$ with respect to x and σ can also be written $\mathbf{E}[F(S_T)\vartheta]$ for F only piecewise continuous such that $\|F(x)\| < C(1+\|x\|)^N$, and a weight $\vartheta \in L^2$. Malliavin calculus is crucial as in Itô calculus framework.

For example, the sensitivity of $\mathbf{E}[F(S_T)]$ with respect to the bounded function μ is studied :

$$\begin{aligned} S_t &= \kappa\left(X_t\right) \ ; \kappa : \mathbf{R}^d \to \left(\mathbf{R}_+\right)^d \\ dX_t &= b\left(X_t\right) dt + \sigma\left(Y_t\right) dB_t^{H_1} \\ dY_t &= \mu\left(Y_t\right) dB_t^{H_2} \end{aligned}$$

where, B^{H_1} and B^{H_2} are two independent fractional Brownian motions.

References:

[1] T. Cass, C. Litterer and T. Lyons, Integrability Estimates for Gaussian Rough Differential Equations, *arXiv*, 1104.1813v4, 2011.
[2] P. Friz and N. Victoir, Multidimensional Stochastic Processes as Rough Paths : Theory and Applications, *Cambridge Studied in Applied Mathematics*, **120**, Cambridge University Press, Cambridge, 2010.
[3] E. Fournié, J-M. Larsy, J. Lebuchoux, P-L. Lions and N. Touzi, Applications of Malliavin Calculus to Monte-Carlo Methods in Finance, *Finance Stochast.*, **3**:391–412, 1999.
[4] E. Gobet and R. Münos, Sensitivity Analysis using Itô-Malliavin Calculus and Martingales, and Application to Stochastic Optimal Control, *SIAM J. Control Optim.*, **43**(5):1676–1713, 2005.
[5] N. Marie, Sensitivities via Rough Paths, *arXiv*, 1108.0852, 2011.

[Nicolas MARIE; Laboratoire Modal'X, Université Paris-Ouest Nanterre La Défense, 200 Avenue de la République, 92001 Nanterre Cedex]

7th International Conference on Sensitivity Analysis of Model Output, July 1–4 2013, Nice

[nicolas.marie@math.univ-toulouse.fr – www.math.univ-toulouse.fr/~marie]

7th International Conference on Sensitivity Analysis of Model Output, July 1-4 2013, Nice

Towards Operational Products Development from Earth Observation: Exploration of SimSphere Land Surface Process Model Sensitivity using a GSA approach

GEORGE, P. PETROPOULOS
University of Aberystwyth, Institute of Geography and Earth Sciences, SY23 2EJ, Wales, United Kingdom

HYWEL GRIFFITHS
University of Aberystwyth, Institute of Geography and Earth Sciences, SY23 2EJ, Wales, United Kingdom

STEFANO TARANTOLA
Institute for the Protection and Security of the Citizen, Joint Research Centre, European Commission, Ispra (VA) - Italy

The use of Earth Observation (EO) data combined with land surface process models is at present implemented, or is being explored for obtaining spatio-temporal estimates of parameters characterising land surface interactions such as energy fluxes and surface soil moisture. These techniques aim to improve estimates of these key parameters by combining the horizontal coverage and spectral resolution of remote sensing data with the vertical coverage and fine temporal continuity of simulation process models. In this context, the European Space Agency (ESA) has funded the PROgRESSIon project[1] which explores at a global scale the prototyping of energy fluxes and soil moisture from the synergy of SimSphere model with EO data from ESA-funded/co-funded instruments.

SimSphere was originally developed by Carlson and Boland (1978) and considerably modified to its current state by Gillies et al. (1997) and Petropoulos et al. (2013). It belongs to a special category of deterministic models called Soil Vegetation Atmosphere Transfer (SVAT) models which provide in a detailed description of soil and vegetation canopy processes in a vertical profile representation and at a fine time resolution that is in good agreement with the dynamic timescale of the atmospheric and surface processes. SimSphere is a 1d model with a plant component and its underlying constraint is taken as the balance between all the energy fluxes at the Earth's surface. The model simulates a number of variables characterising the continuous evolving interactions between soil, plant and atmospheric layers over a 24-hour cycle, starting from a set of initial conditions given in the early morning. A review of the model use was provided by Petropoulos et al. (2009a), in which authors underlinedthat very few SA studies have been performed on the model despite the widespread use that is already made of it and its promising potential for much wider assimilation by the community. In response to this requirement, Petropoulos et al. (2009b; 2010) performed a Global Sensitivity Analysis (GSA) on SimSphere based on a Gaussian process emulator (BACCO; Kennedy and O'Hagan, 2001; O'Hagan, 2006). For the first time, their studies made it possible to derive an estimate of the sensitivity of key model outputs characterising land surface interactions to all the model inputs and their interactions. .

Briefly, the BACCO method is based on the use of the Gaussian Emulation Machine for Sensitivity Analysis (GEM-SA) tool, the development of which was funded by the National Environmental Research Council, UK. In the method, the uncertainty of the SA due to the emulator approximation is quantified by computing several statistical metrics. These include the "cross validation root mean square error", the "cross-validation root mean squared relative error" and the "cross-validation root mean squared standardized error". The first measure is the square root of the mean square error of the emulator prediction whereas the second is that value expressed as percentage. The third expresses the residual divided by an estimate of its standard deviation. In addition, the so called "roughness values" and

[1] More information at: http://due.esrin.esa.int/stse/projects/stse_project.php?id=148 [accessed March 1th, 2013]

the "sigma-squared" parameters are estimated. The first describes how rapidly the output responds to changes in each input, whereas the second is the variance of the emulator after standardising the output. The basic SA statistical metrics output from GEM SA include the computation of the relative contributions of the main and joint effects (pairwise contributions only) of the input parameters to the overall output variance, as well as the total effects, definitions of which can be found for example in O'Hagan (2006). Details concerning the BACCO emulation process and the precise method are provided in Kennedy and O'Hagan (2000) and O'Hagan (2006).

The present study aimed to perform a GSA on SimSphere in order to further extend our understanding of the model structure and to establish its coherence, building on the previous analogous works conducted on the model. For consistency and comparability to these previous studies, the GSA implemented herein has also been based on adopting the BACCO method to develop an emulator. Yet, whereas previous SA studies on SimSphere using the BACCO were based on normal probability distribution functions (PDFs) assumption for the model input parameters, in our study a SA is performed assuming uniform PDFs. It also derives PDFs of the most sensitive model inputs directly from satellite imagery acquired from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) imagine radiometer. Our findings, although largely comparable to previous SA studies on the model, showed that the PDF's assumption can influence the absolute SA measures of the model input parameters in respect to the target quantity considered each time. Yet, the nature of the most significant inputs influencing the sensitivity of the examined outputs was not changed. The implications of our main findings are discussed in the context of the future use of the model including its synergy with EO data for deriving operationally key land surface parameters from space.

[George Petropoulos, *University of Aberystwyth, Wales, United Kingdom*]
[george.petropoulos@aber.ac.uk]

References:

1. Carlson, T.N. and F.E., Boland (1978): Analysis of urban-rural canopy using a surface heat flux/temperature model. *J. Appl. Meteorol. 17*, 998-1013.
2. Gillies, R.R.; Carlson, T.N.; Cui, J.; Kustas, W.P. and K.S., Humes (1997): Verification of the "triangle" method for obtaining surface soil water content and energy fluxes from remote measurements of the Normalized Difference Vegetation Index NDVI and surface radiant temperature. *Int. J. Remote Sens.* 18, 3145-3166.
3. Kennedy,M.C. and O'Hagan, A. (2001): Bayesian calibration of computer models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 63 (3), 425–464.
4. O'Hagan, A. (2006): Bayesian analysis of computer code outputs: a tutorial. *Reliab. Eng. Syst. Saf.* 91, 1290–1300.
5. Petropoulos, G., Carlson, T.N. and M. J., Wooster (2009a): An Overview of the Use of the SimSphere Soil Vegetation Atmosphere Transfer (SVAT) Model for the Study of Land-Atmosphere Interactions. *Sensors*, 2009, 9(6), 4286-4308.
6. Petropoulos, G., Wooster, M. J., M. Kennedy, Carlson, T.N., and M. Scholze (2009b): A global sensitivity analysis study of the 1d SimSphere SVAT model using the GEM SA software. *Ecological Modelling*, 220 (19), 2427-2440.
7. Petropoulos, G., M. Ratto and S. Tarantola (2010): A comparative analysis of emulators for the sensitivity analysis of a land surface process model. 6th International Conference on Sensitivity Analysis of Model Output, July 19-22nd, 2010, Milan, Italy, on Procedia - *Social and Behavioral Sciences*, vol. 2 (6), 7716-17.
8. Petropoulos, G.P., I. Konstas and T.N. Carlson (2013): Automation of SimSphere Land Surface Model Use as a Standalone Application and Integration with EO Data for Deriving Key Land Surface Parameters. *European Geosciences Union*, April 7-12th, 2013, Vienna, Austria.

Uncertainty and Sensitivity Analysis in Multiple Criteria Decision Problems using Rule Learner Techniques

Claudio M. Rocco
Universidad Central de Venezuela, Caracas, VENEZUELA

Elvis Hernández
Universidade Católica Portuguesa, Lisboa, PORTUGAL

In many situations, a decision-maker is interested in assessing a set of **m** objects or alternatives Aj characterized simultaneously by **n** criteria or attributes, and defining a ranking able to synthesize the global characteristics of each object, for example, from the best to the worst. This is the case, for example, in the assessment of several projects through attributes such as cost, availability, environmental impact, among others. The behavior of each object, for every criterion, is quantified via "performance values" which can either be numerical or categorical. Several multicriteria decision techniques could be used to this aim [1], like ELECTRE [2], PROMETHEE [3], TOPSIS [4], to name a few. As a result of the assessment, for example, object Ak is ranked as the best.
However, this object ranking, defined as the Base Rank, could be influenced by uncertain factors associated to specific criteria (for example, the criterion "cost of a project" could be affected by variations in the interest rate) or by decision-maker preferences (e.g., criterion weights). In this situation, the decision-maker could be interested in knowing what factors are responsible for a specific model behavior, e.g., what criterion weights affect the position of the object ranked as the best object in the base rank.
Several approaches have been proposed in the literature to evaluate the importance of factors in a model, given a specified behavior, like Monte Carlo Filtering [5], Regional Sensitivity Analysis [6], Generalized Likelihood Uncertainty Estimation [7] and Tree-Structured Density Estimation [8].
The problem addressed in this paper is related to describe the solution space associated to a given model behavior specification. For example, what are the values associated to each criterion that make Ak be ranked as the best object.
The solution space could be a non-convex and/or sparse space. Approaches suggested in [9] or recently in [10-11] are able to extract an approximate hyperbox of the solution space, where factor variations are assigned independently (i.e., factors are "decoupled" [10-11] and the solution space is represented through intervals [x1_inf, x1_sup], [x2_inf, x2_sup], … and [x**l**_inf, x**l**_sup], **l**=**n*****m**). The approach proposed in [9] requires an analytical model f(x1, x2,..,x**l**), while in [10-11] the model is considered as a black-box. In both approaches the widths of the factors could be associated to a sensitivity index.
This paper proposes an approach based on the use of machine learning classification techniques [12] able to provide a description of the solution space, based on a set of "If-Then" rules derived from model samples. Additionally, some data-mining techniques are able to extract the most important factor; others can detect non important factors while others can provide a numerical sensitivity index. However, factors in the solution space are not necessarily decoupled.
An example related to a real decision problem (**m**=20, **n**=4, **l**=80) illustrates the proposed approach. Figure 1 shows the approximate probability distribution for the rank of each object considering the uncertainty of the 80 factors (based on 10000 samples). Objects 15 and 9 have the highest probability to be ranked in the first and second position respectively. This fact means that the ranking of both objects is very robust to

uncertainty effects, considering the performance of the objects (in general, criterion weights (preferences) of specific multicriteria techniques could also be considered).
Thirty one rules were obtained for the condition that objects 15 and 9 are ranked as the first and the second objects. This set of rules is able to correctly explain 96% of the samples. Additionally the techniques detect that factor x75 is the most important factor and there are 62 out of 80 non important factors (i.e., only 18 factors appear in the rules). Figure 2 shows some rules extracted for the "**false**" condition (i.e., objects 15 and 9 are not ranked as the first and the second objects). Similar rules could be derived for the "**true**" condition. In any case, the decision-maker has additional information on what factors are responsible for achieving the desired condition (the typical goal of a Factor Mapping Setting [5]) and what are their possible values.

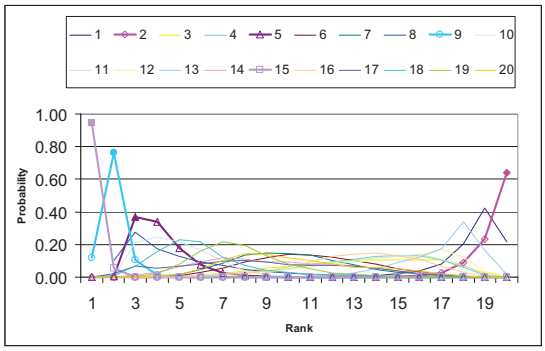

Figure 1: Approximate probability distribution for the rank of each object (10000 samples)

(x69 >= -0.17) and (x75 <= -0.32) and (x74 >= -0.35) => **false**

(x75 <= -0.32) and (x69 >= -0.21) and (x66 >= -0.31) and (x68 <= -0.09) and (x19 >= 1080164) => **false**

(x75 <= -0.3) and (x78 >= -0.34) and (x69 >= -0.17) => **false**

Figure 2: Some rules extracted

References

[1] J. Figuereda, S. Grecco and M. Ehrgott. Multicriteria Analysis: State of Art Surveys. Springer Science, 2005

[2] B. Roy, Revue d'Informatique et de recherché opérationelle 6(8) (1968):57–75

[3] J.P Brans and P.H. Vincke, Manage Sci 31(6) (1985):647–656

[4] C.L Hwang and K.S. Yoon, Multiple attribute decision making: methods and applications. Springer, Berlin, 1981

[5] A Saltelli, S. Tarantola, F. Campolongo and M. Ratto Sensitivity Analysis in Practice. A Guide to Assessing Scientific Models, John Wiley & Sons, Probability and Statistics series, 2004

[6] G.M.Hornberger, R.C.Spear, An approach to the preliminary analysis of environmental systems, J. Environm. Management, 12, 7-18, 1981

[7] K.J.Beven, A.Binley, The Future of Distributed Models: Model Calibration and Uncertainty Prediction, Hydrological Processes, 6, 279-298, 1992

[8] R.C.Spear, T.M. Grieb and Shang N., Factor Uncertainty and Interaction in Complex Environmental Models, Water Resources Research, 30, 3159-3169, 1994

[9] C.M. Rocco, J.A. Moreno and N. Carrasquero: Robust design using a hybrid cellular-evolutionary and interval-arithmetic approach: a reliability application. Reliability Engineering and System Safety, 79:149–159, 2003

[10] M. Zimmermann, J.E. von Hoessle: Computing solution spaces for robust design. International Journal for Numerical Methods in Engineering, DOI: 10.1002/nme.4450, 2012

[11] L. Graff, H. Harbrecht and M. Zimmermann, On the computation of solution spaces in high dimensions, November 2012, Preprint-Number SPP1253-138

[12] I. H.Witten and E. Frank, Data Mining: Practical Machine Learning Tools and Techniques. 2nd Edition, Morgan Kaufmann, San Francisco, 2005

[Claudio M. Rocco; Universidad Central de Venezuela, Facultad de Ingeniería, Apartado Postal 47937, Caracas 1041A, Venezuela]
[croccoucv@gmail.com]

Computing variance-based sensitivity indices for spatial inputs with "map labelling": study of sampling size bias.

NATHALIE SAINT-GEOURS
Irstea – UMR TETIS, Montpellier, France

STEFANO TARANTOLA
Joint Research Centre of the European Commission, Institute for the Protection and Security of the Citizen, Ispra, Italy

LINDA LILBURNE
Landcare Research, Christchurch, New-Zealand

In this paper, we offer to discuss the use of the "map labelling" method to estimate variance-based sensitivity indices for spatially distributed inputs. More precisely, we focus on i) the convergence of indices estimates when the number of randomly generated input maps increases; and on ii) the use of two technical tricks to lower the cost of the analysis in terms of CPU requirements.

Variance-based global sensitivity analysis (VB-GSA) is used to study how the variance of the output of a numerical model can be apportioned to different sources of uncertainty in its inputs [1]. VB-GSA is recognized as an essential component of model building, but it is seldom applied in Earth and Environmental Sciences, partly because most of the numerical models developed in this field include spatially distributed inputs (maps). However, a number of authors have recently introduced methods to compute variance-based sensitivity indices associated to one or several spatially distributed inputs, that possibly exhibit spatial auto-correlation [2,3,4].

The « map labelling » method [2] is one of these methods. It allows calculating first and total-order sensitivity indices associated to an uncertain input map, for the case of a numerical model with a low CPU cost. Let consider a numerical model $\boldsymbol{F}$, with $\boldsymbol{k}$ scalar inputs $\boldsymbol{U_1 \ldots U_k}$ and a single spatially distributed input $\boldsymbol{\{Z(x),\ x\ in\ R^2\}}$. We assume that the uncertainty on $\boldsymbol{Z(x)}$ can be simulated by some stochastic process $\boldsymbol{P}$ (e.g., a geostatistical algorithm), that may account for spatial auto-correlation or any complex structure of variability in $\boldsymbol{Z(x)}$. The steps of the « map labelling » approach are as follows:

i) generate $\boldsymbol{n}$ equiprobable random realizations of the uncertain input map $\boldsymbol{Z(x)}$ and store them in some permanent disk space;
ii) label each realization with a unique integer $\boldsymbol{1 \leq l \leq n}$;
iii) sample random label $\boldsymbol{L}$ along with other scalar inputs $\boldsymbol{U_j}$ to generate an input matrix, following for example the procedure described in [1];
iv) evaluate model $\boldsymbol{F}$ for each line of the input matrix: on line $\boldsymbol{i}$, the sampled value $\boldsymbol{l^{(i)}}$ of random label $\boldsymbol{L}$ indicates that the $\boldsymbol{l^{(i)}}$-th realization of input map $\boldsymbol{Z(x)}$ should be used;
v) from the output vector, calculate the first and total-order sensitivity indices $\boldsymbol{S_L}$ and $\boldsymbol{ST_L}$ of random label $\boldsymbol{L}$. These indices are taken as a measure of the influence of uncertain input map $\boldsymbol{Z(x)}$ on the variance of model output.

In practical cases, the use of the "map labelling" method is limited by the number $\boldsymbol{n}$ of random map realizations that can be generated and stored on the computer. The choice of size $\boldsymbol{n}$ is thus driven by constraints of time – generating random realisations of $\boldsymbol{Z(x)}$ using stochastic process $\boldsymbol{P}$ may be CPU intensive, and constraints of disk space – storing a large number of spatially distributed data may be computationally intractable. The size $\boldsymbol{n}$ will influence the precision and accuracy of sensitivity indices estimates, but also the CPU cost of the analysis.

To investigate this issue, a first part of this paper is devoted to a numerical study of the convergence of sensitivity indices estimates S_L and ST_L when the number n of random input map realizations increases. Results on a simple analytical test case show a bias towards an under-estimation of sensitivity indices S_L and ST_L when n is too small. The efficiency of a $n/(n-1)$ corrective coefficient is tested to correct for this bias.

Next, in a second part of the paper, we discuss the use of two simple technical tricks that can lower the cost of the analysis. The first trick is related to the cost of loading one realisation of $Z(x)$ in the memory of the computer code: if this cost is high, then it is interesting to try and reduce the number of times each realization is loaded in the VB-GSA process. To do so, for each "base-point" of Sobol' sample in the space of the model inputs [1], one single random realization of $Z(x)$ is loaded in the memory of the computer code and it is re-used for all the $(k+1)$ subsequent model runs related to the same base-point. The second trick is related to the cost of storing a large number of random realizations of Z(x) in the disk space. The idea is that these random realisations of $Z(x)$ can be generated "on the fly" for each base-point, then discarded – instead of pre-generating all of them before running sensitivity analysis. Hence, the analysis would not require a large data storage capacity any more. These two tricks make the "map labelling" approach more computationally tractable, and allow using a large number n of random realizations of Z(x).

Both parts of this paper are illustrated on a simple model that computes the total nitrate load in a catchment from spatially distributed input data on climate, irrigation, landuse and soil type.

References :

[1] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana and S. Tarantola, "Global Sensitivity Analysis - The Primer". Wiley, 2008.

[2] L. Lilburne and S. Tarantola, "Sensitivity analysis of spatial models", *Int. J. Geogr. Inf. Sci.*, **23**(2):151–168, 2009.

[3] B. Ioss and M. Ribatet, "Global sensitivity analysis of computer models with functional inputs. *Reliab. Eng. Syst. Saf.*, **94**:1194 – 1204.

[4] N. Saint-Geours, "Sensitivity analysis of spatial models: application to cost-benefit analysis of flood risk management plans", *PhD thesis, Université Montpellier II*, 2012.

Presenting author: Nathalie Saint-Geours; Irstea UMR TETIS Montpellier (F)
nathalie.saint-geours@teledetection.fr

Profile-based Sensitivity in D-Optimal Designs for Precise Parameter Estimation

Hana Sulieman
American University of Sharjah, Sharjah, UAE

For the general form of a single response nonlinear regression model

$$\mathbf{y} = \mathbf{f}(\mathbf{X}, \Theta) + \boldsymbol{\epsilon} \tag{1}$$

where $\mathbf{y}$ is n-element vector of observed response values for particular values of the regressor variables $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$. Θ is a k-element vector of unknown parameters, $\mathbf{f}$ is the vector of predicted values of the response variable for given $\mathbf{X}$ and Θ, and $\boldsymbol{\epsilon}$ is a vector of independent and normally distributed random errors; the local D-optimal design minimizes the determinant of the approximate variance-covariance matrix of the parameter estimates with respect to the experimental settings $\mathbf{X}$. The D-optimality is defined by:

$$max \quad D = |V_0' V_0| \tag{2}$$

where the $n \times k$ matrix V_0 consists of the local sensitivity coefficients, $\frac{\partial \mathbf{f}(\mathbf{X},\Theta)}{\partial \Theta}$ evaluated at an initial parameter estimates Θ_0. Under the linear approximation, the model response surface is replaced by its tangent plane and the usual ellipsoidal joint inference region for Θ is the image in parameter space of a spherical region on the tangent plane. The volume of the ellipsoidal region evaluated at Θ_0 is given by $|V_0' V_0|^{-1/2}$. By maximizing D, this volume is minimized. Thus, for a given nonlinear model response, the D-optimal criterion ensures that the design is such that large regions on the tangent plane map into small regions in the parameter space.When model nonlinearity is pronounced, the *local* D-optimality can produce designs with poor performance and little information about parameters.

To overcome the linearity and peculiarity of the local sensitivity coefficients, Sulieman et.al. (2001, 2004) have developed profile-based sensitivity coefficients which account for model nonlinearity and parameter estimate correlations. For a parameter of interest,θ_i, profile-based sensitivity coefficient is defined by:

$$\mathbf{p}_i = \frac{D\mathbf{f}(\mathbf{X}, \theta_i, \Theta_{-i}(\theta_i))}{D\theta_i} \tag{3}$$

where the operator D means total derivative with respect to θ_i and $\mathbf{p}_i$ is $n \times 1$ vector containing profile-based sensitivity coefficients for θ_i evaluated at the n design points.
In terms of the first and second order derivative information of the model function, $\mathbf{f}(\mathbf{X}, \Theta)$, $\mathbf{p}_i$ is given by:

$$\mathbf{p}_i = \mathbf{v}_i - V_{-i}(V_{-i}' V_{-i} - [e'][V_{-i-i}])^{-1}(V_{-i}' \mathbf{v}_i - V'_{-ii} e) \tag{4}$$

where $\mathbf{v}_i$ is the $n \times 1$ vector of local sensitivity coefficients with respect to θ_i; V_{-i} is an $n \times (k-1)$ matrix consisting of first derivative vectors with respect to Θ_{-i}; V_{-i-i} is the $n \times (k-1) \times (k-1)$ array of the second order derivatives of $\mathbf{f}(\mathbf{X}, \Theta)$ with respect to Θ_{-i}; V_{-ii} is the $n \times (k-1)$ matrix of the second derivatives of with respect to Θ_{-i} and θ_i, and e is the n-element residuals vector.

Sulieman *et. al.* (2009) showed that the profile-based sensitivity coefficients, $\mathbf{p}_i$, measure the influence that θ_i exerts on the predicted response after the removal of its co-dependencies with the remaining parameters. These co-dependencies are measured using second-order derivative information of the model function so as to account for model nonlinearity. Hamilton and Watts (1985)demonstrated that the matrix $[e'][V_{-i-i}]$ is a function of only the intrinsic curvature portion of V_{-i-i} and hence the extent to which profile-based sensitivities differ from the local sensitivities depends on the intrinsic nonlinearity of the model structure and the strength of correlations between parameter estimates.

Using the profile-based sensitivity coefficients defined in equation (4), the profile-based D-optimality can be defined as maximizing:

$$max \quad D_P = |P_0' P_0| \tag{5}$$

where the matrix $P = [\mathbf{p}_1 \mathbf{p}_2 \dots \mathbf{p}_k]$ is evaluated at Θ_0.
With the above interpretation of $\mathbf{p}_i$,by maximizing D_p, the volume of the inference region in a less-correlated parameter space than the original parameters is minimized. Hence, the resulting design produce more precise and less correlated parameter estimates than the corresponding local D-optimal design.

Application of D_p-optimality to model examples will be presented. The applications include both starting designs and sequential designs. The resulting designs will be compared to those obtained by the classical D-optimality. Some computational aspects of the new designs will also be discussed.

References:

[1] D.C. Hamilton and D.G. Watt "A quadratic design criterion for precise estimation in nonlinear regression models", *Technometrics*, **27**(3): 241–250, 1985.

[3] H. Sulieman and I.Kucuk "Global Derivative Based Sensitivity Method For Parameter Estimation",*Journal of the Franklin Institute*, **348**(7): 1556–1573, 2011.

[4] H. Sulieman, I.Kucuk and P.J. McLellan "Parametric sensitivity: A case study comparison". *Computational Statistics & Data Analysis*, **53**(7): 2640–2652, 2009.

[5] H. Sulieman H."Improved local sensitivity measures for regression models with correlated parameters", *Proceedings in Computational Statistics 18th Symposium*, Porto, Portugal, 2008.

[6] H. Sulieman, P.J. McLellan and D.W. Bacon"A Profile-based approach to parametric sensitivity in multiresponse regression models", *Computational Statistics & Data Analysis*, **45**: 721–740, 2004.

[7] H. Sulieman, P.J. McLellan and D.W. Bacon "A Profile-Based Approach to Parametric Sensitivity Analysis of Nonlinear Regression Models", *Technometrics*, **43**(4): 425–33, 2001.

[Hana Sulieman; Deaprtment of Mathematics and Statistics, American University of Sharjah, Sharjah, UAE, 26666]
[`hsulieman@aus.edu` – `https://www2.aus.edu/facultybios/profile.php?faculty=hsulieman`]

7th International Conference on Sensitivity Analysis of Model Output, July 1–4 2013, Nice

Sensitivity Analysis for Plant Models with Correlated Parameters : Application to the Characterization of Sunflower Genotypes

Qiong-Li, WU; Benoît, BAYOL; Fenni, KANG
Ecole Centrale Paris, Lab MAS, 92290,Châtenay-Malabry Cedex, France

Jérémie, LECOEUR
3Syngenta Seeds, Toulouse, France

Paul-Henry, COURNEDE
Ecole Centrale Paris, Lab MAS, 92290,Châtenay-Malabry Cedex, France

The understanding of gene-environment interaction is a crucial issue in plant breeding. For this objective, mechanistic models of plant growth can help disentangle the genotypic and environmental effects [1]. Each genotype is characterized by a specific vector of model parameters [2] and ideally, these parameters are stable for a certain range of environments. As explained in [2], each parameter can be seen as resulting from the influences of several genes (epistasis), 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 (thus common to all genotypes, the model parameter is thus considered as non-genotypic parameter).

Therefore, when performing sensitivity analysis of plant growth models with the specific objective of characterization of genotypes, we need to use a method able to take into account correlated inputs. For example, the classical Sobol method can not be applied straightforwardly. The objective of this paper is thus to introduce a method adapted to the sensitivity analysis for correlated inputs and to apply it to a plant model. The SUNFLO model of sunflower growth [1] is used for this purpose. The model parameters for a family of 20 genotypes have been estimated or measured with very heavy experimental work [1] and this set of parameter vectors is used to compute the statistical distribution in the parameter space.

Sensitivity analysis (SA) for correlated input is discussed in [3][4]. The problem with correlated sample is that the reduction in variance that can be achieved by fixing one factor depends on whether or not other factors have been fixed, and the incremental reduction in variance for each factor depends on the order in which factors are fixed [3].

As in [4], two general settings for SA has been discussed for this issue. Each setting is based on a bet posed on the model $y = f(x_1, x_2, \cdots, x_k)$, where the input can be correlated. In the setting 1, the objective is to find out which parameter would induce the largest reduction in variance if it were fixed to its 'true' value. Because such true value is in general unknown, the bet can be rationally placed by computing the estimates $V_{x_i} = V(E(y|x_i))$, whether or not the input is correlated. Numerically, in the case of correlated inputs, the strategy of sampling and re-sampling matrix in [5] can not be applied because it is based on the assumption that all the factors are independent. Brute-force method can be applied in this case since the computing cost is not so prohibitive in this 'first-order' most important factor searching case.

For the setting 2, the objective is to find out the smallest subset of $\mathbf{x}$ capable of inducing a target reduction in the unconditional variance $V(y)$, as in Sobol's, for the uncorrelated case, a rational selection strategy for the subset of interest in based on the computation of the full set of S_i and ST_i as indicated in [5]. For the correlated case, we follow the work presented in [4]. We set the the target reduction to be the total variance of the model output, so that all the analyzed factors could be selected step by step, which potentially indicates the relative priorities of these factors

for reducing the model variance.

The SUNFLO model estimates the biomass production for the crop sunflower by inputting environmental information, mainly temperature, precipitation, and evaporation reference. It simulates the plant's phenology, the leave's development, the accumulation and distribution of biomass, and the production of seeds. It takes into consideration the plant water budget which determines whether the available water quantity is enough for the plant to grow up in good conditions.

We took out 12 genotypic parameters (as listed in table.1) from SUNFLO model as we know for the analysis. Parameter distribution characteristics and correlation matrix was obtained from the 20 values of each parameter from 20 genotypes. We present here our application of setting 2. In table.1, we can see that the ranking indicated by S_i without input correlation is very different with $V_{x_i}/V = V(E(y|x_i))/V$ considering correlation. It means that Sobol's method can not be applied in our case. We can also see that, though computed with the consideration of correlation, $V_{x_i} = V(E(y|x_i))$ still can not be the value of ranking according to which we select group of parameters that can bring certain percentage reduction of total variance, because fixing one factor depends on whether or not other factors have been fixed as strong correlation exists.

Table 1: Setting 2

Name	$Rank_{V_{x_i}}$	V_{x_i}	V_{x_i}/V	$figuremerit$	S_i
RE	1	3805.64	1.90382	-	0.0159822
$date_{E1}^{TT}$	4	1873.18	0.938112	1.15393	0
HI_{graine}	3	2872.32	1.43812	0.917656	0.629579
RT	2	3620.15	1.81255	0.817215	0.108488
$SFimax$	5	1781.7	0.892071	0.483213	0.0291807
$NFfinal$	7	742.626	0.371507	0.418973	0.000350815
$HI_{max}^{capitule}$	6	1272.91	0.63678	0.369769	0.0559516
$position_{SFiMax}$	12	218.782	0.109527	0.0510487	0.000368093
$coeff_{extinction}$	11	276.681	0.138418	0.0379919	0.0190574
$date_{M0}^{TT}$	9	416.406	0.208339	0.0363175	0.0134685
$date_{M3}^{TT}$	10	370.471	0.185508	0.0176194	0.113339
$date_{F1}^{TT}$	8	584.614	0.292736	0.00782156	0.00206494

References:

[1]J. Lecoeur, R. Poiré-Lassus, A. Christophe, B. Pallas, P. Casadebaig, P. Debaeke, F. Vear, and L. Guilioni. "Quantifying physiological determinants of genetic variation for yield potential in sunflower. SUNFLO: a model-based analysis". *Functional Plant Biology*, 38:246–259, 2011.

[2]V. Letort, P. Mahe, P.-H. Cournède, P. de Reffye, B. Courtois. "Quantitative genetics and functional-structural plant growth models: Simulation of quantitative trait loci detection for model parameters and application to potential yield optimization". *Annals of Botany*, 101(8):1243–1254, 2008.

[3] M.D.McKay, "Variance-based methods for assessing uncertainty importance in NUREG-1150 analyses", *LA-UR-96-2695*, 1–27, 1996.

[4] A.Saltelli, and S.Tarantola, "On the Relative Importance of Input Factors in Mathematical Models: Safety Assessment for Nuclear Waste Disposal", *J. Amer. Statist. Assoc.*, **97**(459):702–709, 2002.

[5] A.Saltelli, "Making best use of model evaluations to compute sensitivity indices", *Comput. Phys. Commun.*, **145**:280–297, 2002.

[Qiong-Li, WU; Ecole Centrale Paris, Lab MAS, 92290,Châtenay-Malabry Cedex, France]
[qiongli.wu@gmail.com –]

7th International Conference on Sensitivity Analysis of Model Output, July 1–4 2013, Nice

Global sensitivity analysis for the calibration of a fully-distributed hydrological model

Mauricio Zambrano-Bigiarini
Water Resources Unit, Institute for Environment and Sustainability, Joint Research Centre, European Commission, Ispra (VA), Italy

Zuzanna Zające
Climate Risk Management Unit, Institute for Environment and Sustainability, Joint Research Centre, European Commission, Ispra (VA), Italy

Stefano Tarantola
Statistical Indicators for Policy Assessment, Joint Research Centre, European Commission, Ispra (VA), Italy

Computer model simulations are widely used in the investigation of complex hydrological systems. In particular, hydrological models are tools that help both to better understand hydrological processes and to predict extreme events such as floods and droughts. Usually, model parameters need to be estimated through calibration, in order to constrain model outputs to observed variables. Relevant model parameters used for calibration are usually selected based on expert knowledge of the modeller or by using a local one-at-a-time (OAT) sensitivity analysis (SA). However, in case of complex models those approaches may not result in proper identification of the most sensitive parameters for model calibration. In particular local OAT SA methods are only effective for assessing the relative importance of input factors when the model is linear, monotonic, and additive, which is rarely the case for complex environmental models. In contrast Global Sensitivity Analysis (GSA) is a formal method for statistical evaluation of relevant parameters that contribute significantly to model performance. GSA techniques explore the entire feasible space of each model parameter, and they do not require any assumptions on the model nature (such as linearity or additivity).

In this work we apply the GSA to LISFLOOD, a fully-distributed hydrological model used for flood forecasting at Pan-European scale within the European Flood Awareness System (EFAS). Two case studies are considered, snowmelt- and evapotranspiration-driven catchments, to identify sensitive parameters for both types of hydrological regimes. Results of the GSA will then be used for selecting parameters that need to be estimated during model calibration. Considering the large number of parameters of a fully-distributed model, a two-step GSA framework is applied. First, we implement the computationally efficient screening method of Morris. This method requires a limited number of simulations and produces a qualitative ranking and selection of important factors. As a second step, we apply the variance-based method of Sobol, only to the subset of factors determined as important during the previous screening. The method of Sobol provides quantitative estimates for first order and total order sensitivity indexes of input factors.

The calibration results after the GSA will be described for both case studies and compared against those obtained by using only prior expert knowledge.

[Mauricio Zambrano-Bigiarini; Water Resources Unit, Institute for Environment and Sustainability, Joint Research Centre, European Commission, Ispra (VA), Italy]
[mauricio.zambrano@jrc.ec.europa.eu – http://www.efas.eu/]