



*7th International Conference on
S*ensitivity* A*nalysis*
of M*odel* O*utput**

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

www.gdr-mascoinum.fr/mascot13.html



**Ph.D. students day
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 Pisanisi (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)

Oral presentations

François Bachoc (CEA/Université Paris VII, France), Maximum Likelihood and Cross Validation for Kriging hyper-parameters estimation

Gaëlle Chastaing (Grenoble Univ., France), Fabrice Gamboa (IMT, France), Clementine Prieur (Grenoble Univ., France), Generalized Sobol sensitivity indices for dependent variables

Clément Chevalier (University Bern, Switzerland), Sequential and batch-sequential Bayesian sampling strategies for the identification of an excursion set

Jana Fruth (University Dortmund, Germany), Olivier Roustant (Ecole des Mines de St-Etienne, France), Sonja Kuhnt (University Dortmund, Germany), Sensitivity Analysis for Functional Inputs in a Sheet Metal Forming Process

Mocmchil Ivanov, Sonja Kuhnt (University Dortmund, Germany), Parallel optimization based on FANOVA graph decomposition

Marjorie Jala (Telecom ParisTech, France), Sequential Design of Experiments for Numerical Dosimetry

Loïc Le Gratiet (CEA/Université Paris VII, France), Asymptotic analysis of the learning curve for Gaussian process regression with noisy observations

Prashant Rai, Mathilde Chevreuil, Anthony Nouy (Ecole Centrale Nantes, France), Regis Lebrun (EADS, France), A regression based method using sparse low rank approximations for uncertainty propagation

PH.D. STUDENT PRESENTATION

Maximum Likelihood and Cross Validation for Kriging hyper-parameters estimation

FRANÇOIS BACHOC
CEA-Saclay, DEN, DM2S, STMF, LGLS, F-91191 Gif-Sur-Yvette, France
Laboratoire de Probabilités et Modèles Aléatoires, Université Paris VII
Site Chevaleret, case 7012, 75205 Paris cedex 13

The choice of the covariance function is an important issue for Kriging models. In most of the cases, this function is selected in a parametric set, generally taking the form $\mathcal{C} = \{\sigma^2 C_\theta, \sigma^2 > 0, \theta \in \Theta\}$. Hence, the selection problem boils down to estimating the hyper-parameters σ^2, θ . The methods we study to address this issue are Maximum Likelihood (ML) and Cross Validation (CV). The CV method consists in choosing θ so as to minimize the empirical mean square Leave-One-Out error, and then choosing σ^2 so that the Kriging predictive variances correspond to the Leave-One-Out errors. The implementation is based on the virtual Leave-One-Out formulas of [1], so that the numerical complexity is the same as ML.

First, we show that CV is a relevant alternative to ML because it is more robust to misspecifications of the parametric set \mathcal{C} [2]. We call misspecification the case in which the true covariance function C_0 of the Gaussian process is not in the set \mathcal{C} . The study is carried-out in a two-step approach. In a first step, we address the case in which $\mathcal{C} = \{\sigma^2 C_{mod}, \sigma^2 > 0\}$ with C_{mod} a given correlation function different from C_0 . This framework makes tractable a closed-form expression of the following quality criterion for an estimator $\hat{\sigma}^2$ of σ^2 :

$$R_{\hat{\sigma}^2, x_0} = \mathbf{E} \left[\left(\mathbf{E} \left[(\hat{y}_0 - y_0)^2 | y \right] - \hat{\sigma}^2 c_{x_0}^2 \right)^2 \right], \quad (1)$$

where x_0 is a prediction point, y is the observation vector, $\mathbf{E} \left[(\hat{y}_0 - y_0)^2 | y \right]$ is the prediction mean-square error conditionally to the observations, and $\hat{\sigma}^2 c_{x_0}^2$ is the estimation of this mean square error. Using the closed form expression of (1), we highlight the fact that ML is more efficient when C_{mod} equals C_0 , but that CV becomes more efficient when C_{mod} becomes different from C_0 . The second step is to study the general case in which correlation hyper-parameters are estimated from data as well. This is carried-out via a numerical study on analytical functions. The results we obtain confirm the ones of the first step.

Second, we address the well-specified case in an increasing-domain asymptotic framework [3]. In this framework, it is known that ML is asymptotically efficient, and hence will be more efficient than CV. Nevertheless, we are interested in quantifying the advantage of ML over CV, as well as in studying the influence of the irregularity of the spatial sampling. The question we address is "is irregular sampling always better than regular sampling for hyper-parameter estimation?". We consider an increasing-domain asymptotic framework, in which the sampling is a random deformation of a periodic grid. The strength ϵ of the random deformation determines the sampling irregularity. Consistency and asymptotic normality are proved for the ML and CV estimators. The asymptotic covariance matrices are deterministic functions of the irregularity parameter ϵ only. By means of an exhaustive study of the asymptotic covariance matrices, it is shown that irregular sampling is generally an advantage to estimation, but we identify cases where it is not. Therefore, a negative answer is given to the claim that irregular sampling is always better for estimation than regular sampling.

References:

- [1] O. Dubrule. "Cross validation of Kriging in a unique neighborhood", *Mathematical Geology*, 15:687–699, 1983.
- [2] F. Bachoc. "Cross Validation and Maximum Likelihood estimation of hyper-parameters of Gaussian processes with model misspecification", *In minor revision for Computational Statistics and Data Analysis*

[3] F. Bachoc. "Asymptotic analysis of the role of the spatial sampling for hyper-parameter estimation of Gaussian processes", *In preparation*

[François Bachoc; CEA-Saclay, DEN, DM2S, STMF, LGLS, F-91191 Gif-Sur-Yvette, France]

[francois.bachoc@cea.fr – <http://www.lpma.math.upmc.fr/pageperso/bachoc/>]

Master: Supélec engineer degree (2007-2010), Applied Mathematics specialization, Ecole Centrale Paris (2010) & Master Mathématiques Vision et Apprentissage, ENS Cachan (2010)

Ph.D. (10/2010 (*starting date*) — 10/2013 (*expected end*)): Parametric estimation of the covariance function in Gaussian process based Kriging models. Application to numerical models validation

Supervisor(s): Josselin Garnier & Jean-Marc Martinez

PH.D. STUDENT PRESENTATION

Generalized Sobol sensitivity indices for dependent variables

GAELE, CHASTAING

Université de Grenoble, LJK/MOISE BP 53, 38041 Grenoble Cedex, France

FABRICE, GAMBOA

Université Paul Sabatier, IMT-EPS, 118, Route de Narbonne, 31062 Toulouse Cedex 9, France

CLÉMENTINE, PRIEUR

Université de Grenoble, LJK/MOISE BP 53, 38041 Grenoble Cedex, France,

Global sensitivity analysis is a stochastic approach whose objective is to identify and to rank the input variables that drive the uncertainty of the model output. To reach this objective, the Sobol indices are widely used as a variance-based method [1].

Let consider the nonlinear regression model $Y = \eta(\mathbf{X})$, where $\mathbf{X} = (X_1, \dots, X_p) \in \mathbb{R}^p$ is a random vector of inputs with a joint distribution denoted by $P_{\mathbf{X}}$. Y is the real-valued response of the model. Under the assumption that inputs \mathbf{X} are independent, that is $P_{\mathbf{X}} = \otimes_{i=1}^p P_{X_i}$, Hoeffding introduces the decomposition of η as a sum of increasing dimension functions [2]. The decomposition is written as

$$\eta(\mathbf{X}) = \eta_{\emptyset} + \sum_{i=1}^p \eta_i(X_i) + \sum_{1 \leq i < j \leq p} \eta_{i,j}(X_i, X_j) + \dots + \eta_{1, \dots, p}(\mathbf{X}) \quad (1)$$

with $\int \eta_u(\mathbf{x}_u) dP_{X_i} = 0 \quad \forall i \in u, \forall u \subseteq \{1, \dots, p\}$ to ensure the uniqueness of (1). The Sobol indices are build on this decomposition, and aims at quantifying the contribution of any subset of inputs in the model. Based on the global variance decomposition of Y , the Sobol index of a set of inputs \mathbf{X}_u is the ratio $S_u = \frac{V(\eta_u)}{V(Y)}$.

However, when inputs are not independent, the uniqueness of the expansion (1) is not true anymore. Thus, the use of Sobol indices may lead to a wrong interpretation because they do not take into account the dependence implicitly included in the Sobol indices. The goal of this work is to provide an exact and unambiguous definition of generalized sensitivity indices adapted to non independent incomes. The second objective is to propose an efficient numerical method able to estimate the new constructed indices.

In this presentation, we propose a generalization of the Hoeffding-Sobol functional ANOVA when the inputs are correlated. Inspired by the work of Stone [3] and Hooker [4], we show that η can be uniquely decomposed as

$$\eta(\mathbf{X}) = \eta_{\emptyset} + \sum_{i=1}^p \eta_i(X_i) + \sum_{1 \leq i < j \leq p} \eta_{i,j}(X_i, X_j) + \dots + \eta_{1, \dots, p}(\mathbf{X}) \quad (2)$$

where each component is constrained by hierarchical orthogonality conditions,

$$\int \eta_u(\mathbf{x}_u) \eta_v(\mathbf{x}_v) dP_{\mathbf{X}} = 0, \quad \forall u \in \{1, \dots, p\}, \forall v \subset u \quad (3)$$

The expansion (2) leads to the construction of generalized Sobol sensitivity indices able to quantify the uncertainty brought by dependent inputs on the model by

$$S_u = \frac{V(\eta_u(\mathbf{X}_u)) + \sum_{u \cap v \neq \emptyset, v} Cov(\eta_u(\mathbf{X}_u), \eta_v(\mathbf{X}_v))}{V(Y)}, \quad \forall u \quad (4)$$

In the second part, we provide a numerical method for the estimation of the generalized sensitivity indices expressed in (4). Considering that each component of the decomposition is subject to orthogonality constraints (3), the procedure consists in constructing directly multivariate basis that satisfy these constraints. Starting from a truncated orthonormal basis, we construct iteratively extended basis satisfying (3). Thus, each component is expanded on suitable basis, and associated coefficients are easily deduced by ordinary least-squares. Nevertheless, the projection of the one-by-one component into finite dimensional subspaces suffers from the curse of dimensionality. To remedy to it, a penalized regression is studied. The aim is to retain a sparse number of informative predictors to fit the model. The greedy approximation [5] deals with the ℓ_0 -penalization. It offers a set of intuitive algorithms able to select a very limited number of predictors. The Lasso regression [6] is a convex relaxation of the ℓ_0 regularization, and offers a statistical stability. These two methods will be compared in our context. Finally, several numerical examples illustrate the interest of our method.

References:

- [1] Sobol, I.M., “Sensitivity estimates for nonlinear mathematical models”, *Mathematical Modeling and Computational Experiment*, **1**(4):407–414, 1993.
- [2] Hoeffding, W., “A class of statistics with asymptotically normal distribution”, *The annals of Mathematical Statistics*, **19**(3):293–325, 1948.
- [3] Stone, C.J., “The use of Polynomial Splines and their tensor products in multivariate function estimation”, *The Annals of Statistics*, **22**(1):118–171, 1994.
- [4] Hooker, G., “Generalized functional ANOVA diagnostics for high-dimensional functions of dependent variables”, *Journal of Computational and Graphical Statistics*, **16**(3):709–732, 2007.
- [5] Temlyakov, V., “Greedy approximation”, *Cambridge University Press*, 2011.
- [6] Tibshirani, R., “Regression shrinkage and selection via the Lasso”, *Journal of the Royal Statistical Society*, **B58**(1):267–288, 1996.

[Ph.D. Gaelle Chastaing; Université de Grenoble, LJK/MOISE]

[gaelle.chastaing@imag.fr – <https://team.inria.fr/moise/gaelle-chastaing/>]

Master: Université Rennes 1, Master Statistique et Econométrie – ENSAI Rennes, Diplôme ingénierie statistique

Ph.D. (10/2010 — 10/2013): The objective of my PhD is the study and the implementation of statistical methods in sensitivity analysis for model with dependent input variables. It concerns the theoretical properties of the developed tools as well as their numerical estimation.

Supervisor(s): F. Gamboa and C. Prieur

PH.D. STUDENT PRESENTATION

Sequential and batch-sequential Bayesian sampling strategies for the identification of an excursion set

CLÉMENT CHEVALIER

University of Bern, Switzerland - IRSN, Fontenay-aux-Roses, France

Abstract: Stepwise Uncertainty Reduction (SUR) strategies aim at constructing a sequence of points for evaluating a function f in such a way that the uncertainty on some quantity of interest decreases to zero. The presentation focuses on SUR strategies in a Gaussian process framework and deals with two different problems. First we consider the problem of identifying the excursion set above a given threshold T of a real-valued function f . Then we study the question of finding the set of “safe controlled configurations”, i.e. the set of *controlled* inputs where the function remains below T , whatever the value of some others *uncontrolled* inputs.

Nuclear criticality safety studies at the IRSN aim at deciding if a system containing fissile material (e.g., Uranium, Plutonium) presents a risk of “criticality accident”, i.e. a risk of self-sustaining nuclear chain reaction. Practitioners dispose of a complex code which can be viewed as a real-valued expensive-to-evaluate function f (possibly observed with some noise) with $d \geq 1$ scalar inputs lying in a compact domain $\mathbb{X} \subset \mathbb{R}^d$. A configuration $\mathbf{x} \in \mathbb{X}$ is considered to be dangerous if $f(\mathbf{x}) \geq T$, with $T = 0.95$, and safe otherwise. Two major problems of interest are formulated.

Problem 1: Identifying the set of all dangerous configurations: $\{x \in \mathbb{X} : f(\mathbf{x}) \geq T\}$.

Problem 2: Identifying the set of *safe controlled configurations*, i.e. the set:

$\{\mathbf{x}_c \in \mathbb{X}_c : \forall \mathbf{x}_u \in \mathbb{X}_u, f(\mathbf{x}_c, \mathbf{x}_u) < T\}$, where some of the d inputs parameters are controlled and lie in a domain \mathbb{X}_c , while others are uncontrolled and lie in \mathbb{X}_u (with $\mathbb{X} = \mathbb{X}_c \times \mathbb{X}_u$).

In our setting, the problems described above need to be solved with a limited evaluation budget of f which justifies the use of a kriging metamodel. From an initial design of experiments of n evaluations we use a sequential sampling strategy allowing to obtain, at each iteration, a batch of $r \geq 1$ points where f is evaluated. This approach is similar to the approach in [1] where an *Expected Improvement* criterion is used to guide the evaluations of f for optimization purposes.

Many sampling criteria based on kriging are already proposed in the literature to deal with the first problem [2,3]. In particular, [2] shows that criteria based on the idea of “Stepwise Uncertainty Reduction” (SUR) have better performances in applications. However, these criteria are themselves expensive to compute. A contribution of this Ph.D. consists in introducing closed-form formulas allowing to drastically reduce the computational effort required to use some SUR strategies for Problem 1. The formulas also allow a fast computation of multi-point criteria, which provide a batch of $r > 1$ points at each iteration. An example is shown on Figure 1, in the case $r = 4$, with a $2d$ -function, and a threshold $T = -10$. From an initial design of 10 evaluations, 5 iterations of the batch-sequential SUR strategies are run, each of them providing 4 points.

The use of SUR strategies can be naturally extended to deal with the difficult Problem 2. However a direct use of the obtained criterion is challenging as it involves, for each $\mathbf{x}_c \in \mathbb{X}_c$, the computation of the conditional exceedance probability of sectional-Gaussian processes, i.e. the probability

$$P \left(\max_{\mathbf{x}_u \in \mathbb{X}_u} \xi(\mathbf{x}_c, \mathbf{x}_u) \geq T \mid n \text{ evaluations} \right), \tag{1}$$

where ξ is our Gaussian process with known (or estimated) covariance parameters. Having no analytical formula to quickly calculate the latter expression, we rely on approximations obtained from the literature [4], on ad-hoc heuristic strategies and closed form formulas extending the results obtained with Problem 1. An example is shown on Figure 2 on the same $2d$ -function, in the case where the first variable is controlled and the second is not. We use again $r = 4$ and 5 iterations.

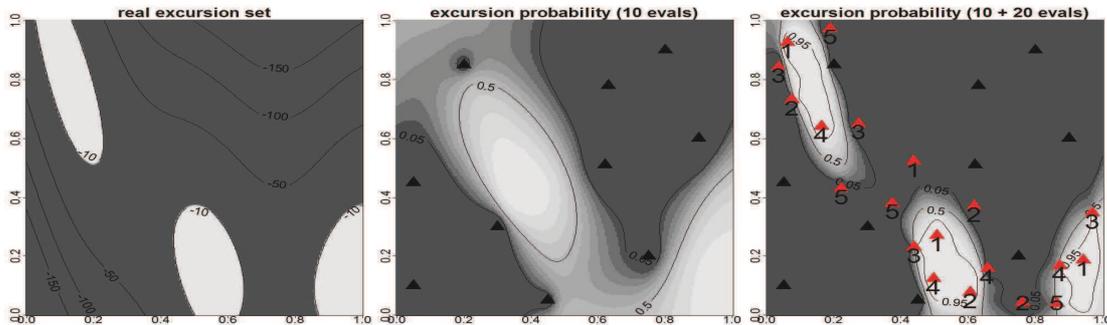


Figure 1: Left: excursion set (in white) of a $2d$ -function f . Middle: probability of excursion calculated from 10 evaluations of f . Right: updated probability of excursion after 5 iterations of a batch-sequential SUR strategy. Numbers correspond to the iteration numbers.

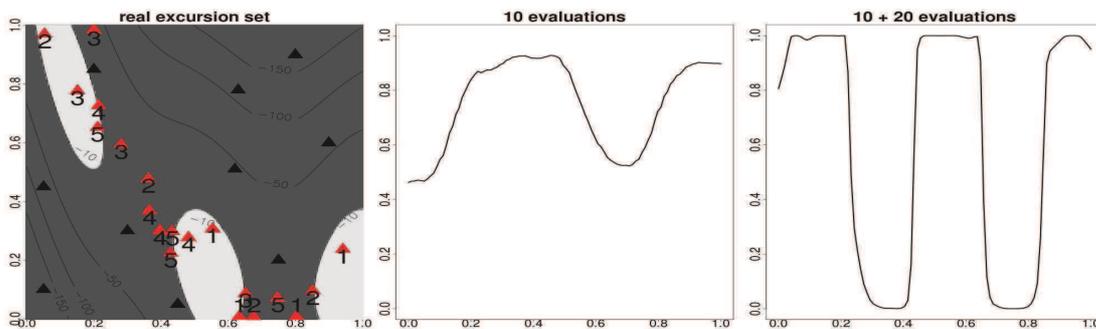


Figure 2: Left: excursion set (in white) of a $2d$ -function with the initial design (black triangles) and the points sampled by one of the proposed batch-sequential SUR strategy for Problem 2 (red). Middle: Sectional excursion probability: $P(\max_{\mathbf{x}_u \in \mathbb{X}_u} \xi(\mathbf{x}_c, \mathbf{x}_u) \geq T \mid n \text{ evaluations})$, as a function of \mathbf{x}_c . Right: updated sectional excursion probability, once 20 new “well-chosen” points are evaluated.

References:

- [1] D. R. Jones, M. Schonlau and J. William. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, **13**(4):455–492, 1998.
- [2] J. Bect, D. Ginsbourger, L. Li, V. Picheny and E. Vazquez. Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing*, **22**(3): 773–793, 2012.
- [3] V. Picheny, D. Ginsbourger, O. Roustant, R. T. Haftka and N.-H. Kim. Adaptive designs of experiments for accurate approximation of target regions. *Journal of Mechanical Design*, **132**(7), 2010.
- [4] R.J. Adler and J.E. Taylor. Random Fields and Geometry. *Springer*

[Clément Chevalier; Alpeneggstrasse 22 CH-3012 Bern]
 [clement.chevalierstat.unibe.ch – www.redice-project.org]

Master: University Pierre et Marie Curie, Paris 6.

Ph.D. (11/2010 — 11/2013): This Ph.D. is funded by the IRSN and the ReDice consortium.

Supervisor(s): David Ginsbourger (UNIBE) and Yann Richet (IRSN)

PH.D. STUDENT PRESENTATION

Sensitivity Analysis for Functional Inputs in a Sheet Metal Forming Process

JANA FRUTH

TU Dortmund University, Dortmund, Germany

OLIVIER ROUSTANT

Ecole Nationale Supérieure des Mines, Saint-Etienne, France

SONJA KUHNT

TU Dortmund University, Dortmund, Germany

In sheet metal forming, springback, the reforming of the sheet metal after the removal of the punch, is one of the major defects (see Figure 1). It is especially present for advanced high strength steels, which are more and more in use in the automotive industry. The forming process can be simulated by complex, but time consuming finite element simulations. By applying effective sensitivity analysis methods to the simulations the relations between process parameters and springback have been successfully analysed [1]. There the process parameters were kept on constant level during the forming process. Now finite element simulation have been developed that allow to change parameters during the forming process in order to further improve the accuracy of the springback analysis. New methods are needed for the analysis of those temporal changeable parameters. This talk presents ideas for an effective sensitivity analysis of functional input including design and graphical representation of the functional influences of the inputs.

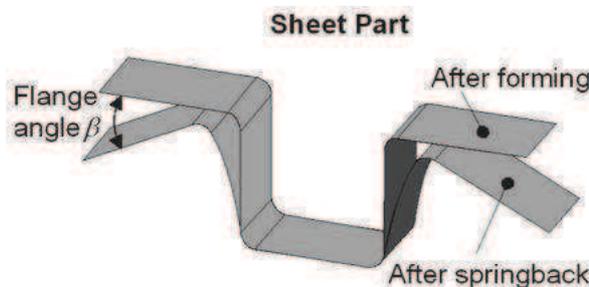


Figure 1: Springback after forming of a sheet metal

We consider an experiment with d_s scalar input variables $x_1, \dots, x_{d_s} \in [0, 1]$ and d_f functional input variables $f_1(t), \dots, f_{d_f}(t) \in [0, 1] \forall t \in [0, 1]$ with scalar response y

$$y = f(x_1, \dots, x_{d_s}, f_{d_s+1}(t), \dots, f_{d_s+d_f}(t)), t \in [0, 1].$$

Our aim is to effectively analyse the variables' influence on y for different periods of the time t . The time is considered to be continuous here. In practice it usually is a discrete set of a high number of equidistant time points at which the experiment can be set. This can be easily adopted from the continuous case.

In order to reduce problem's dimension, we explore whole intervals of time and sequentially reduce the interval size for interesting time spaces. The input functions $f_i(t)$ are therefore designed as piecewise constant functions between fix knots $\{a_i^{(0)} = 0, a_i^{(1)}, \dots, a_i^{(n-1)}, a_i^{(n)} = 1\}$ similar to B-splines of order 1 (see [2])

$$f_i(t) = \sum_{k=1}^{n_i} z_i^{(k)} \mathbf{1}_{[a_i^{(k-1)}, a_i^{(k)}]}(t),$$

with $z_i^{(k)}$ the value of the function f_i at interval k .

In the sequential algorithm in the first step a low number of intervals is chosen and sensitivity analysis is performed on the variables $z_i^{(1)}, \dots, z_i^{(n_i)}$ for each functional input $i = 1, \dots, d_f$ and on the scalar inputs x_1, \dots, x_{d_s} . For all intervals that show a meaningful influence, a new knot is created inside the interval to split the interval into two parts. All other intervals are kept on a constant level and so are not further examined. Then sensitivity analysis is performed again as before. These steps can be repeated until the desired accuracy is reached. Several sensitivity analysis techniques can be considered in the steps, e.g. regression screening, Morris effects or Sobol indices (see e.g. [3]), depending on the complexity of the experiment and the possible number of evaluations. Since the intervals can have different sizes, their sensitivity has to be scaled according to the interval size in order to obtain comparable results.

The method has been successfully applied to analytical test functions and was also carried out in a springback simulation. Here the influence of two functional inputs has been examined, the friction between the sheet metal and the forming tool and the force of the blank holder which holds the sheet metal in the process. Three steps of sequential sensitivity analysis by regression screening have been performed using a total sum of 80 evaluations. The results (see Figure 2) show the regression coefficients of the time intervals, scaled by the interval size. A connected behaviour between both inputs is noticeable as well as a clear difference between the first and the second half of the punch time. In the first half the influence is positive where in the second half it is strongly negative with the most influence being in the very last time points. The results improved the engineers understanding of the time dependent influence and have been validated afterwards in real forming experiments.

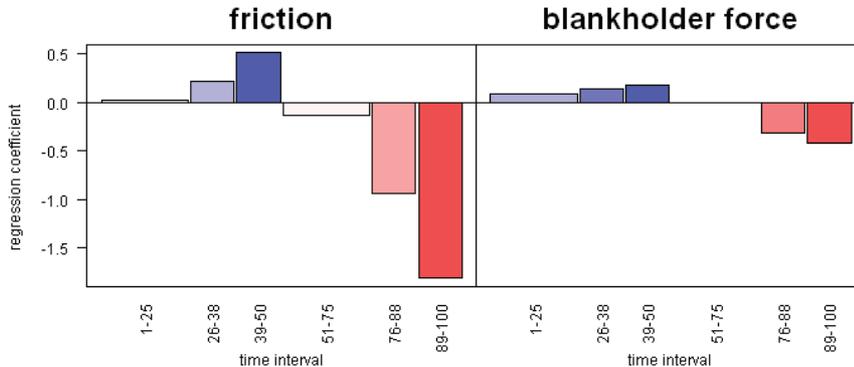


Figure 2: Coefficients of regression screening after three sequential steps of functional analysis

References:

- [1] M. Goesling, H. Kracker, A. Brosius, S. Kuhnt and A.E. Tekkaya, “Strategies for springback compensation regarding process robustness”, *Prod. Eng. Res. Devel.*, **5**(1):49–57, 2011.
- [2] J.O. Ramsay and B.W. Silverman, *Functional data analysis*, Springer, 2005.
- [3] A. Saltelli, K. Chan, E.M. Scott, *Sensitivity Analysis*, Wiley, 2000.

[Jana Fruth; Faculty of Statistics, TU Dortmund University, Vogelpothsweg 87, 44227 Dortmund, Germany]

[fruth@statistik.tu-dortmund.de – www.statistik.tu-dortmund.de/1553.html]

Master: TU Dortmund University

Ph.D. (04/2012 (*starting date*) — 04/2015 (*expected end*)): Sensitivity analysis and graph-based methods for black box functions with an application to sheet metal forming

Supervisor(s): S. Kuhnt and O. Roustant

PH.D. STUDENT PRESENTATION

Parallel optimization based on FANOVA graph decomposition

MOMCHIL IVANOV

TU Dortmund University, Dortmund, Germany

SONJA KUHN

TU Dortmund University, Dortmund, Germany

In the automobile industry sheet metal forming is used for the production of automobile body parts. Unfortunately, defects such as, tearing, wrinkling or spring back frequently occur in the formed parts. With increasing computing capabilities it has become customary to perform simulation studies beforehand, where the sheet-metal characteristics can be cheaply optimized. However, simulations generally take a long time to run, ranging from hours to days, making it impossible to perform direct optimization on the computer code. Instead, the simulator can be considered as a black-box function and an approximation model which is cheaper to evaluate is used to interpolate the simulation. Figure 1 shows a software which simulates a deep drawing process, which is optimized in this work.

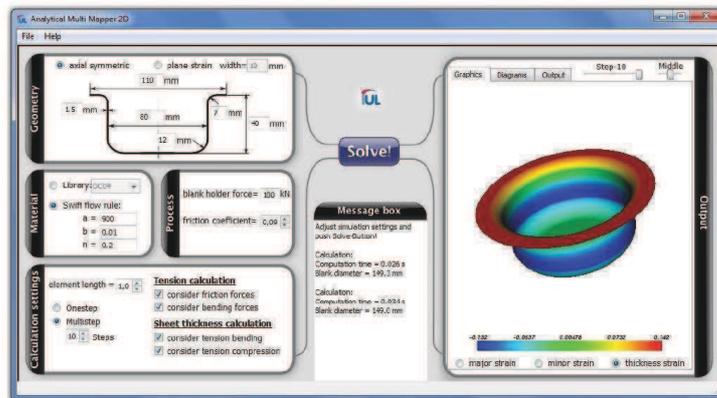


Figure 1: Simulation of a deep drawing process

The modeling and optimization of expensive to evaluate black-box functions is often performed with the help of metamodel-based sequential strategies. A popular choice is the efficient global optimization (EGO) algorithm which is based on the Kriging metamodel [1]. Although the EGO method has been reported to work efficiently with a reduced simulation budget [2], a big limitation of this procedure is the fact that it allows only one simulation at a time. Being able to take advantage of several simulators simultaneously, and thus parallelizing the optimization and potentially saving huge amounts of time, is a very appealing idea [3].

In this presentation a very elegant way to produce a parallel optimization procedure, based on a technique from the sensitivity analysis toolbox, called FANOVA graph [4] is presented. The FANOVA graph method studies the interactions between variables in a function and is able to recognize a block-additive interaction structure if it is present, i.e.:

$$f(x_1, \dots, x_d) = \sum_I f_I(x_I), \quad I \subset \{1, \dots, d\},$$

where f_I represent the additive blocks. In Figure 2 an example of a FANOVA graph decomposition for a function with 6 input variables is presented. The edges between the variables are based on

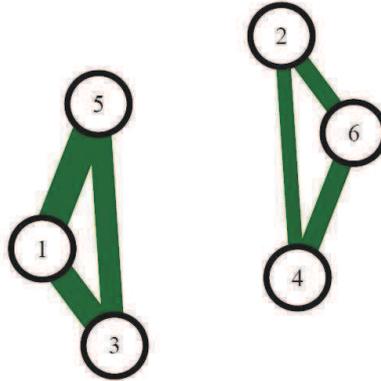


Figure 2: FANOVA graph example

the total interaction index [5]. It can be seen in the figure, that this example-function has a clear block additive structure. If a function has such an additive interactions structure, the blocks can be simultaneously optimized independent of each other, for example with the EGO algorithm, thus allowing the experimenter to use several simulators and to perform a parallel optimization:

$$\min_{x_1, \dots, x_d} f(x_1, \dots, x_d) = \sum_I \min_{x_1, \dots, x_d} f_I(x_I) = \sum_I \min_{\{x_i; i \in I\}} f(x_I, c_0, \dots, c_0), \quad I \subset \{1, \dots, d\},$$

where c_0 is a constant.

This talk presents the application of the described parallelization technique and discusses problems that may arise in the process as well as possible solutions. Different models, such as kernel interpolation [6] are also discussed as a basis for metamodel-driven optimization. In conclusion, some theoretical examples and a small simulation example of a sheet metal forming process is presented.

References:

- [1] Jones, D. R.; Schonlau, M.; Welch, W. J. (1998): Efficient global optimization of expensive black-box functions. In: Journal of Global optimization 13 (4), S. 455-492.
- [2] Knowles, J.; Corne, D.; Reynolds, A. (Hg.) (2009): Noisy multiobjective optimization on a budget of 250 evaluations. Evolutionary Multi-Criterion Optimization: Springer.
- [3] Ginsbourger, D.; Riche, R.; Carraro, L. (2010): Kriging is well-suited to parallelize optimization. In: Y. Tenne und C.-K. Goh (Hg.): Computational intelligence in expensive optimization problems. 1. Aufl. Berlin, New York: Springer, S. 131162.
- [4] Muehlenstaedt, T.; Roustant, O.; Carraro, L.; Kuhnt, S. (2012): Data-driven Kriging models based on FANOVA-decomposition. In: Statistics and Computing 22 (3), S. 723738.
- [5] Fruth, J.; Roustant, O.; Kuhnt, S. (2012): Total interaction index: A variance-based sensitivity index for interaction screening. Submitted.
- [6] Muehlenstaedt, T.; Kuhnt, S. (2011): Kernel interpolation. In: Computational Statistics & Data Analysis 55, S. 29622974.

[Momchil Ivanov; Faculty of Statistics, TU Dortmund University, Vogelpothsweg 87, 44227 Dortmund, Germany]

[ivanov@statistik.tu-dortmund.de – <http://www.statistik.tu-dortmund.de/1609.html>]

Master: TU Dortmund University

Ph.D. (05/2012 (*starting date*) — 05/2015 (*expected end*)): Multiobjective optimization based on metamodels for computer experiments

Supervisor(s): S. Kuhnt

PH.D. STUDENT PRESENTATION

Sequential Design of Experiments for Numerical Dosimetry.

MARJORIE JALA

Telecom ParisTech, Paris, and Orange Labs, Issy-les-Moulineaux, France

Over the past 30 years, wireless communication systems have been increasingly used. The number of mobile phones, WIFI boxes, antennas, etc., is growing together with a strong public concern over possible health problems related to the exposure to radio frequency (RF) electromagnetic fields (EMF). Among the questions related to exposure effects, we shall focus on the assessment of fetuses exposure to EMF from a numerical dosimetry point of view. This consists in virtually exposing pregnant woman and fetus 3D-models to one source of EMF; the Finite Difference in Time Domain (FDTD) method is used to compute the (whole body) Specific Absorption Rate (SAR) in the fetus, which is an evaluation of the rate at which energy is absorbed by the body of the fetus. We are interested in estimating a quantile of the distribution of the SAR in the fetus, given the fact that a numerical dosimetry simulation is expensive in terms of computational load, and that the preparation of this simulation could be complex and time-consuming. Thus, our goal is to propose a sequential sampling strategy where each computer trial is selected in order to estimate a quantile as accurately as possible by performing as few evaluations of the SAR in the fetus as possible.

We shall model the SAR in the fetus by $Y = f(\mathbf{X})$, where f is an unknown real-valued function and \mathbf{X} is a random vector of \mathbb{R}^d having a known distribution. For our sequential sampling approach, we shall adopt a Bayesian point of view, using a Gaussian Process path as a surrogate model for the unknown function f ; this surrogate model enables us to optimize a criterion which selects the next evaluation point of f among a fine grid \mathcal{X} of the space. Without loss of generality, we shall assume that \mathbf{X} is uniformly distributed on $[0, 1]^d$. Our prior on f is a zero-mean Gaussian process having k as a covariance function. For a sample of observations $\mathbf{y}_T = (Y_1, \dots, Y_T)^T$, the posterior over f is a GP again with mean $\mu_T(x)$ and covariance $k_T(x, x')$ are given by

$$\mu_T(x) = \mathbf{k}_T(x)' \mathbf{K}_T^{-1} \mathbf{y}_T, \quad (1)$$

$$k_T(x, x') = k(x, x') - \mathbf{k}_T(x)' \mathbf{K}_T^{-1} \mathbf{k}_T(x'), \quad (2)$$

where $\mathbf{k}_T(x) = [k(x_1, x) \dots k(x_T, x)]'$, x is in \mathcal{X} , $\mathbf{K}_T = [k(x_i, x_j)]_{1 \leq i, j \leq T}$ and $'$ denotes the matrix transposition. In our case, f models a quantity that is supposed to be very smooth, so for our applications we shall use the square exponential covariance function given by

$$k(x, x') = \exp\left(-\frac{(x - x')^2}{2\ell^2}\right), \quad x, x' \in [0, 1]^d, \ell > 0. \quad (3)$$

Indeed, with this covariance function, sampled GP paths are differentiable to any order almost surely. See [1] for further details.

Our strategy consists in adapting a classical kriging criterion to our quantile estimation problem. This criterion consists in selecting the point x_{T+1} of \mathcal{X} to add to the set of observed T points such that:

$$x_{T+1} = \arg \max_{x \in \mathcal{X}} \sigma_T(x), \quad (4)$$

with $\sigma_T(x)^2 = k_T(x, x)$ defined in (2). The novelty in our strategy is that it focuses on the areas of space that could be relevant for quantile estimation. Let us define the Upper Confidence Bound and its Lower Confidence Bound counterpart at step T by

$$\mu_T^U(x) = \mu_T(x) + \sqrt{\beta_T} \sigma_T(x),$$

$$\mu_T^L(x) = \mu_T(x) - \sqrt{\beta_T} \sigma_T(x),$$

with

$$\beta_T = 4 \ln(T) + 2 \ln\left(\frac{n}{\delta}\right), \quad n = |\mathcal{X}|, \delta \in (0, 1). \quad (5)$$

The set $\{[\mu_T^L(x); \mu_T^U(x)], x \in \mathcal{X}\}$ defines a confidence region for f at the confidence level $1 - \delta$, see [2]. Given those functions, we can estimate the upper α -quantile $q_{\alpha,T}^U$ and the lower α -quantile $q_{\alpha,T}^L$ by $\hat{q}_{\alpha,T}^U = (\mu_T^U(\mathcal{X}))_{([\lceil n\alpha \rceil])}$ and $\hat{q}_{\alpha,T}^L = (\mu_T^L(\mathcal{X}))_{([\lceil n\alpha \rceil])}$ where the i th smallest value taken by μ_T on the grid \mathcal{X} is denoted by $(\mu_T(\mathcal{X}))_{(i)}$. We denote $U_{\alpha,T}$ the set such that

$$U_{\alpha,T} = \{x : \mu_T^U(x) < q_{\alpha,T}^L\} .$$

A point x of \mathcal{X} shall be kept in the search space only if it is such that $\mu_T^U(x) \geq q_{\alpha,T}^L$ (i.e. if it belongs to $U_{\alpha,T}^C$). Indeed, if x is in $U_{\alpha,T}$, $f(x) < q_{\alpha}$ since $q_{\alpha} \geq q_{\alpha,T}^L$ and $f(x) \leq \mu_T^U$ with probability greater than $1 - \delta$.

We propose hereafter an application to real data. We estimate the 95% quantile of the SAR of a 26-week-old fetus inserted in an anatomically realistic woman model corresponding to the average dimensions of Japanese women. In our application, the pregnant woman model is exposed to 900 MHz vertically polarized electromagnetic plane waves with a 1 Volt per meter amplitude. The SAR (expressed in W/kg) of the fetus is considered as a function of the elevation and the azimuth of the incident wave. The surface plot of the SAR is displayed in Figure 1 (a). The average of the relative errors $|\hat{q}_{0.95} - \hat{q}_{\alpha,T}|/\hat{q}_{0.95}$ at each iteration T (50 Monte Carlo repetitions on the choice of the set of points to start with) is displayed in Figure 1 (b) where $\hat{q}_{0.95}$ is calculated with the 1000 points of the grid \mathcal{X} (selected with a maximin LHS obtained with the lhsdesign function of Matlab). A 5% relative error is reached on average after 7 iterations, which means 12 evaluations of the SAR, as we started with 5 points. Thus, with more than 80 times less points, the strategy provides an estimation of the quantile with 5% error.

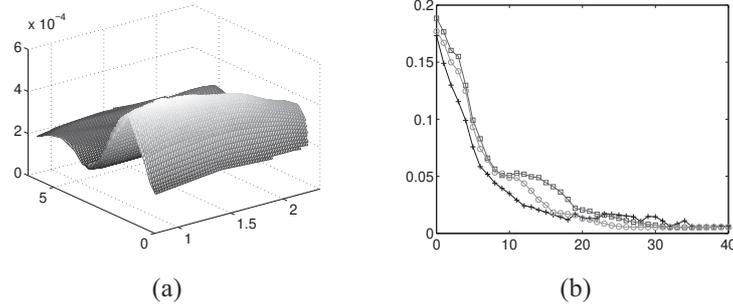


Figure 1: (a) Surface plot of the SAR in the fetus; (b) Average relative errors at each iteration T of the method.

References:

- [1] C. E. Rasmussen and C. K. I. Williams. Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning). The MIT Press, December 2005.
- [2] N. Srinivas, A. Krause, S.M. Kakade, and M. Seeger. Gaussian process optimization in the bandit setting: no regret and experimental design. Proceedings of the international conference on machine learning (ICML), 2010.

[Marjorie Jala; Telecom ParisTech and Orange Labs, Networks, and Careers]

[jala@telecom-paristech.fr -]

Master: UPMC

Ph.D. (11/2010 (*starting date*) — 11/2013 (*expected end*)): Our goal is to statistically characterize the fetus exposure to EMF throughout their development, for different types of use and source. This work is part of the French-Japanese ANR-JST FETUS project.

Supervisor(s): Céline Lévy-Leduc (AgroParisTech), Éric Moulines (Télécom ParisTech), Emmanuelle Conil and Joe Wiart (Orange Labs and Whist Lab)

PH.D. STUDENT PRESENTATION

Asymptotic analysis of the learning curve for Gaussian process regression with noisy observations

LOIC, LE GRATIET
Université Paris Diderot 75204 Paris Cedex 13, France
CEA, DAM, DIF, F-91297 Arpaçon, France

Gaussian process regression is a useful tool to surrogate an expensive computer code given some of its observations. For many realistic cases, computer codes provide observations tainted by noise. This happens for Monte-Carlo based simulators - also called stochastic simulators - which use Monte-Carlo or Monte-Carlo-Markov-Chain methods that solve a system of differential equations through its probabilistic interpretation. For such simulators, the level of the noise is inversely proportional to the number of Monte-Carlo particles used in the procedure. Thus, if we consider a constant budget defined as the total number of Monte-Carlo particles, a trade off between the number and the accuracy of the observations has to be made. The topic of this presentation is to minimize the generalization error - defined as the Integrated Mean Squared Error (IMSE) - of a Gaussian process regression with noisy observations and a given budget.

Let us suppose that we want to surrogate the output $y(x)$ of a stochastic simulator from noisy observations of it at points $(x_i)_{i=1,\dots,n}$ sampled from the design measure $\mu(dx)$. Furthermore, we consider that we have s_i Monte-Carlo particles at each point. We hence have $\sum_{i=1}^n s_i$ outputs of the form $z_j(x_i) = y(x_i) + \varepsilon_j(x_i)$ where $(\varepsilon_j(x_i))_{i=1,\dots,n;j=1,\dots,s}$ are independent random variables sampled from a Gaussian distribution with mean zero and variance $\sigma_\varepsilon^2(x_i)$. Then, the vector of observed values given by a stochastic simulator is $z^n = (\sum_{j=1}^s z_j(x_i)/s_i)_{i=1,\dots,n}$. The variance of an observation z_i^n is thus $\sigma_\varepsilon^2(x_i)/s_i$.

Now, let us consider a fixed budget $T = \sum_{i=1,\dots,n} s_i$. As a consequence, we can either decide to perform observations in lot of points (i.e. n large) but with an important noise variance (i.e. s_i small) or decide to perform few observations but with a small noise variance. Here, we consider the case n large. In our presentation, we are interested in two points:

1. For $n \gg 1$, how to determine the needed budget T to achieve a prescribed precision.
2. For T fixed, which sequence of integers $\{s_1, s_2, \dots, s_n\}$ minimizes the generalization error.

The main idea of the Gaussian process regression is to suppose that the objective function $y(x)$ is a realization of a Gaussian process $Y(x)$ with a known mean and a known covariance kernel $k(x, x')$. The mean can be equal to zero without loss of generality. In that framework, the Best Linear Unbiased Predictor (BLUP) given the observations z^n is:

$$\hat{y}(x) = k(x)^T (K + \Delta)^{-1} z^n, \quad \Delta = \text{diag} [(\sigma_\varepsilon^2(x_i)/s_i)_{i=1,\dots,n}] \quad (1)$$

where $k(x)$ is the vector containing the covariance between $Y(x)$ and $(Y(x_i))_{i=1,\dots,n}$ and K is the covariance matrix of $(Y(x_i))_{i=1,\dots,n}$. Furthermore, the Mean Squared Error (MSE) of the BLUP (1) is given by $\sigma^2(x) = k(x, x) - k(x)^T (K + \Delta)^{-1} k(x)$ and the generalization error equals $\text{IMSE} = \int \sigma^2(x) d\mu(x)$.

We handle the first point by considering an uniform allocation and a constant reduced noise variance, i.e. $s_i = s$ and $\sigma_\varepsilon^2(x) = \sigma_\varepsilon^2$. In this case, we have $T = ns$ and $\Delta = n\sigma_\varepsilon^2 I/T$ where I is the identity matrix. We have the following equality in probability:

$$\text{IMSE} \longrightarrow \sum_{p \geq 0} \sigma_\varepsilon^2 \lambda_p / (\sigma_\varepsilon^2 + T \lambda_p) \phi_p(x)^2, \quad n \rightarrow \infty \quad (2)$$

where $(\phi_p(x))_p$ is an orthonormal basis of $L_\mu^2(\mathbf{R}^d)$ consisting of eigenfunctions of $(T_{\mu,k} f)(x) = \int k(x, u) f(u) d\mu(u)$ and $(\lambda_p(x))_p$ is the sequence of corresponding non-negative eigenvalues sorted

in decreasing order. The convergence (2) has been addressed in the statistical and numerical analysis literature. In the statistical literature, [1], [2] and [3] present this result for degenerate kernels and in the numerical one, [4], [5] and [6] present it for a class of non-degenerate kernels and in dimension one and two. Our main result is the proof of the convergence (2) for non-degenerate and degenerate kernels and for any dimension. This result is of interest since classical kernels used in Gaussian process regression are non-degenerate and the dimension is usually larger than two. The equality (2) shows that the rate of convergence of the IMSE with respect to T depends on the one of the eigenvalues decay. Then, by knowing the eigenvalue decay, we can evaluate the budget T such that the IMSE reached a prescribed value.

For the second point, we consider as fixed the budget T and we consider a non-constant reduced noise $\sigma_\varepsilon^2(x)$ and a non-uniform allocation $\{s_1, s_2, \dots, s_n\}$ (the uniform allocation and a constant reduced noise are needed to determine T but does not provide an optimal allocation). Determining the optimal allocation whatever the Gaussian process for a heterogeneous noise is an open non-trivial problem. We handle it by considering the restricted condition K diagonal. The reader is referred to [7] for a proof of this proposition in a different framework (the proof uses the Karush-Kuhn-Tucker approach to solve the minimization problem with equality and inequality constraints).

The different results are illustrated on an industrial application to the safety assessment of a nuclear system containing fissile materials. The system is modeled by a neutron transport code called MORET [8].

References:

- [1] Opper, M. and Vivarelli, F., “General Bounds on Bayes Errors for Regression with Gaussian Processes”, *Advances in Neural Information Processing Systems, MIT Press*, **11**:302–308, 1999.
- [2] Sollich, P. and Halees, A., “Learning curves for Gaussian Process Regression: approximations and bounds”, *Neural computation*, **14**:1393–1428, 2002.
- [3] Picheny, V., “Improving Accuracy and Compensating for Uncertainty in Surrogate Modeling”, *Ecole Nationale Supérieure des Mines de Saint Etienne*, 2009.
- [4] Ritter, K., “Average-Case Analysis of Numerical Problems”, *Springer Verlag*, Berlin, 2000.
- [5] Ritter, K., “Almost Optimal Differentiation Using Noisy Data”, *Journal of approximation theory*, **86**:293–309, 2000.
- [6] Bozzini, M. and Rossini, M., “Numerical Differentiation of 2D Functions from Noisy Data”, *Computer and Mathematics with Applications*, **45**:309–327, 2003.
- [7] Munoz Zuniga, M. and Garnier, J. and Remy, E. and de Rocquigny, E., “Adaptative Directional Stratification for controlled estimation of the probability of a rare event”, *Reliability Engineering and System Safety*, **96**:1691–1712, 2011.
- [8] Fernex, F. and Heulers, L. and Jacquet, O. and Miss, J. and Richet, Y., “The MORET 4B Monte Carlo code New features to treat complex criticality systems”, *MandC International Conference on Mathematics and Computation Supercomputing, Reactor and Nuclear and Biological Application*, Avignon, France, 2005.

[Loic Le Gratiet; PARIS VII - CEA, DAM, DIF]

[loic.legratiet(at)cea.fr – www.proba.jussieu.fr/~legratiet]

Master: Ecole National Supérieure des Mines de Saint-Etienne & Université Jean-Monnet de Saint-Etienne.

Ph.D. (10/2010 — 10/2013): My applied and theoretical researches deal with uncertainty quantification and surrogate modeling for computer codes. In particular, they address the issue of building predictive models using data from codes with multiple levels of fidelity using a co-kriging based approach.

Supervisor(s): Josselin Garnier

PH.D. STUDENT PRESENTATION

A regression based method using sparse low rank approximations for uncertainty propagation

PRASHANT, RAI
MATHILDE, CHEVREUIL
ANTHONY, NOUY
*LUNAM Université, Ecole Centrale Nantes, Université de Nantes,
GeM, UMR CNRS 6183, Nantes, France*

RÉGIS, LEBRUN
*EADS Innovation Works- Applied Mathematics
EADS France Campus Engineering, Blagnac, France*

Uncertainty quantification has emerged as a crucial field of research in various branches of science and engineering. In the last ten years, considerable efforts have been made in the development of new methodologies based on a functional point of view in probability, where random output function $u(\xi)$ of d input random parameters $\xi = (\xi_1 \dots \xi_d)$ are approximated with suitable functional expansions

$$u(\xi) \approx u_P(\xi) = \sum_{\alpha=1}^P v_\alpha \phi_\alpha(\xi)$$

The so called non intrusive techniques allow the estimation of expansion coefficients by using evaluations $\{u(y^k)\}_{k=1}^Q$ of the numerical model at certain sample points $\{y^k\}_{k=1}^Q$, thus allowing the simple use of existing deterministic simulation codes. However the dimension P of classical approximation spaces has an exponential (or factorial) increase with d and hence the computational cost becomes prohibitively high as one needs to evaluate the model for a large number of samples $Q \approx P$.

In order to handle high-dimensional models, we propose a regression-based low rank tensor approximation method, which exploits the tensor structure of the stochastic function space. The underlying assumption is that the model output functional can be well represented in a low dimensional basis composed of elementary tensors (rank-one functions)[1][2]. Also, methods based on sparse regularization have been recently applied successfully for uncertainty quantification problems in a non intrusive setting[3]. The proposed method combines the ideas of regression with sparse regularization in suitable tensor subsets and demonstrates considerable advantages in resolving problems with high stochastic dimension especially when very few model evaluations can be performed.

We propose a sparse tensor approximation method that consists in approximating the model with a m -term representation $u_m = \sum_{i=1}^m \alpha_i w_i$, where the α_i are real coefficients and the w_i are selected in a suitable tensor subset \mathcal{S} of sparse rank-one elements.

A two step algorithm is proposed to perform greedy construction of canonical tensor approximation in \mathcal{S} for the quantity of interest. In the first step, the correction step, we compute w_m as the optimal correction of u_{m-1} in \mathcal{S} (which is a low dimensional manifold):

$$w_m \in \arg \min_{w \in \mathcal{S}} \|u - u_{m-1} - w\|_Q^2,$$

with $\|v\|_Q^2 = \frac{1}{Q} \sum_{k=1}^Q v(y^k)^2$. Sparsity of functions (on appropriate function bases) in this greedy approximation permits the use of only a few samples to recover better approximations by using suitable regularization techniques. It is followed by an update step in which an approximation u_m is computed in the linear space $V_m = \text{span}\{w_i\}_{i=1}^m$ using regularized regression:

$$u_m = \arg \min_{v \in V_m} \|u - v\|_Q^2 + \lambda \mathcal{R}(v)$$

where $\mathcal{R}(v)$ is a suitable regularization functional and λ is a regularization parameter.

We also show that expressions for Sobol indices can be readily derived from sparse low rank tensor approximations. Hence, once we have evaluated a tensor representation, the post processing required is almost costless. The proposed method is applied to several benchmark problems and its capabilities to treat both smooth and irregular functions using few samples with sufficient accuracy is demonstrated.

Keywords: Uncertainty quantification, high dimension, low rank approximation, sparse approximation, regression

References:

- [1] G. Beylkin, B. Garcke, and M.J. Mohlenkamp, “Multivariate regression and machine learning with sums of separable functions”, *Journal of Computational Physics*, 230:2345–2367, 2011.
- [2] A. Nouy, “Proper Generalized Decomposition and separated representations for the numerical solution of the high dimensional stochastic problems”, *Archives of Computational Methods in Engineering*, 17(4):403–434, 2010.
- [3] A. Doostan and H. Owhadi, “A non-adapted sparse approximation of pdes with stochastic inputs”, *Journal of Computational Physics*, 230(8):3015–3034, 2011.

[Prashant Rai; Ecole Centrale de Nantes, F-211, 1 rue de la Noë BP 92101 44115 Nantes Cedex 3 - France]

[prashant.raiec-nantes.fr –]

Master: Ecole Centrale de Nantes

Ph.D. (10/2011 — 09/2014): Non Intrusive Methods based on Separated Representations for Propagation of Uncertainty

Supervisor(s): A. Nouy and M. Chevreuil

Poster presentations

Alexandre A. Birolleau (CEA, France), Bayesian inference accelerated by iterative polynomial chaos for compressible fluid dynamics

Sarah Bouquet, Chantal de Fouquet, Dominique Bruel (Ecole des Mines de Paris, France), Optimization of CO2 storage assessment using selection of stochastic realizations

Guillaume Damblin, Merlin Keller, Alberto Pasanisi (EDF, France), Eric Parent, Pierre Barbillon, Jacques Bernier (AgroParisTech, France), A Bayes decision approach to code validation in an industrial context

Gregory A.J. Deman, Jaouher Kerrou, (University of Neuchatel, Switzerland), Hakim Benabderrahmane (ANDRA, France), Pierre Perrochet (University of Neuchatel, Switzerland), Comparison of Sensitivity Analysis methods applied on a groundwater flow and mass transport model

Jeanne Goffart (University of Savoie, France), Thierry A. Mara (Univ. La Réunion, France), E. Wurtz (CEA-INES, France), Impact of weather data on the performance evaluation of a passive house: an application of sensitivity analysis

Mathilde Grandjacques, Benoit Delinchant, Olivier Adrot (University of Grenoble, France), Clementine Prieur, Celine Helbert (LJK, France) Stochastic modelling of inputs

Markus Hainy, Werner G. Müller, Helga Wagner (University of Linz, Austria), Likelihood-free simulation-based optimal design

Sabra Hamza, Abderazik Birouche, Michel Basset (University of Mulhouse, France), Floriane Anstett-Collin (Lorraine University, France) Sensitivity analysis for the study of a tire model with correlated parameters and an arbitrary distribution

Benoit Jan, Julien Bect & Emmanuel Vazquez (SUPELEC, France), Fully Bayesian approach for the calculation of Sobol indices

Benjamin Lamoureux, Nazih Mechbal (Arts & Metiers ParisTech, France), Jean-Remi Masse (SAFRAN Snecma, France) Kriging-based Surrogate Modeling of Health Indicators for the Monitoring of a Turbofan Fuel System

Paul Lemaître (EDF/University of Bordeaux, France), A new sensitivity analysis method for failure probability

Vincent Moutoussamy (EDF/University of Toulouse, France), Nicolas Bousquet (EDF, France), Thierry Klein, Paul Rochet (IMT, France), Comparing conservative estimations of failure probabilities using sequential designs of experiments in monotone frameworks

Simon Nanty (CEA, France), Uncertainty quantification and visualization for functional data

Areski Cousin, Alexandre Janon, Ibrahima Niang, Veronique Maume-Deschamps (University of Lyon, France), Sensitivity analysis of long-term yield curves

Guillaume Piro, Julien Straubhaar, Philippe Renard (University of Neuchatel, Switzerland), Sensitivity Analysis of DeeSse, a recent Multiple-Point Statistic algorithm

Federico J. Zertuche (University of Grenoble, France), Celine Helbert (Ecole Centrale Lyon, France), Anestis Antoniadis (University of Grenoble, France) Learning with Gaussian Processes with high and Low Accuracy Observations

Juejing Zhao (University of Munich, Germany), Sensitivity Analysis with Application to Multi-Sensor Positioning

Philipp Ziegler, Sandro Wartzack (University of Nuremberg, Germany), A quality measure for comparing different feature deviations to perform sensitivity analysis in tolerancing

PH.D. STUDENT PRESENTATION

Bayesian inference accelerated by iterative polynomial chaos for compressible fluid dynamics

ALEXANDRE, A., BIROLLEAU
 DAM, DIF, F-91297 Arpajon, France, UMPC Paris VI, France

Context and first results

The thesis deals with uncertainty quantification and inverse problems for compressible fluid dynamics. Our work is motivated by the efficient Bayesian inference acceleration using the Polynomial Chaos made in [3] for diffusion processes. In [5] we emphasize some limitations to this strategy when applied to hydrodynamics (shocks and discontinuities). These limitations have motivated our new approach. This approach is an iterative algorithm based on the Polynomial Chaos (i-gPC) which will be presented.

At first, we have demonstrated the efficiency in terms of precision and convergence of this Bayesian framework coupled with the i-gPC algorithm on simpler one and two dimensional problems (cf. figures 1 and 2 and the submitted paper in CICP [5]).

Figure 1: Posterior. The discontinuity is better captured by i-gPC compared to gPC. A Gauss-Legendre quadrature level is 6 (33 points) and a polynomial order 7 are used.

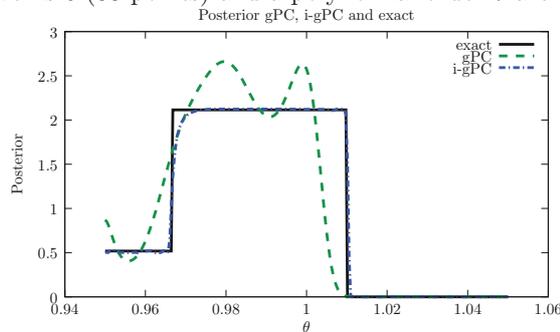
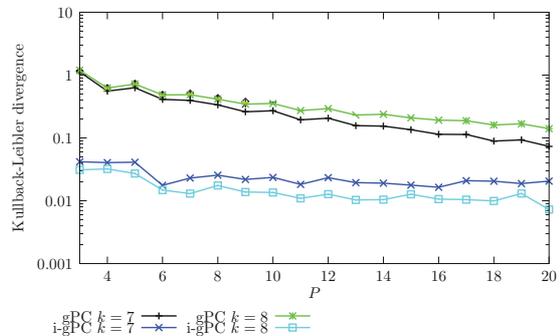


Figure 2: We quantify the acceleration on the Kullback-Leibler divergence convergence with the polynomial order P . We can see the i-gPC improvement compared to gPC inside the Bayesian framework



We will present our results on several toy problems and our first results on a complex flow introduced in the following.

Complex flow and Future Works

We focus now on the shock tube experiment (figure 3) in order to study Richtmyer-Meshkov instabilities [1]. During this experiment, a light fluid is shocked against a heavy fluid. In practice, the initial interface between the two fluids has defects. These defects have a very important influence on the dynamic process because of the sensitive character of the physics involved with respect to initial conditions. We study a shock tube with two fluids initially at rest. The physical

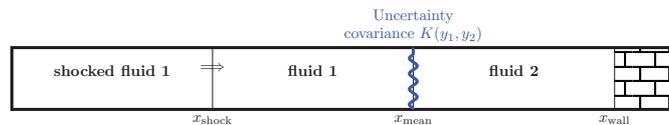


Figure 3: Shock tube experiment

process is modelled here by the 2D multi-fluid Euler equations [1]. This system is solved with a Lagrange and projection scheme [2].

The uncertainty comes with the initial condition. Indeed, due to, among other causes, industrial defects, the initial interface between the two fluids is not well known and is not a perfect straight line but rather a perturbed interface modeled by a stochastic process. The interface stochastic process is approximated by a truncated Karhunen-Loeve expansion with an exponential correlation Kernel

$$K(x, y) = \sigma^2 \exp\left(-\frac{|x - y|}{b}\right). \quad (1)$$

This kernel has two parameters: its standard deviation σ and its correlation length b . We use a Bayesian framework [3] in order to infer on these parameters based on experimental measurements [4] of the length of the mixing zone between the two fluids at different times.

The Bayesian framework is coupled with an uncertainty propagation step building an approximation of the physical model with i-gPC which capture the non linear effects. We will present our latest results about it.

References:

- [1] G. Poëtte, D. Lucor, H. Jourden, “A stochastic surrogate model approach applied to calibration of unstable fluid flow experiments”, *Comptes Rendus de l’Académie des sciences*.
- [2] H. Jourden, S. D. Pino, “Arbitrary High-Order Schemes for the Linear Advection and Wave Equations: Application to Hydrodynamics and Aeroacoustics”, *C.R. Acad. Sci. paris, Ser. I* 342 (2006) 441–446.
- [3] Y. Marzouk, H. Najm, L. Rahn, “Stochastic spectral methods for efficient bayesian solution of inverse problems, *Journal of Computational Physics*” 224 (2) (2007) 560–586.
- [4] F. Poggi, M. Thorembey, G. Rodriguez, “Velocity measurements in turbulent gaseous mixtures induced by richtmyer–meshkov instability”, *Physics of Fluids* 10 (1998) 2698.
- [5] D. Lucor, A. Birolleau, G. Poëtte, “Bayesian inference acceleration using iterative polynomial chaos”, *Communications In Computational Physics* (submitted).

[Alexandre Birolleau; CEA and UPMC Paris VI]

[alexandre.birolleau@cea.fr –]

Master: Centrale Paris

Ph.D. (03/2011 (*starting date*) — 03/2014 (*expected end*)):

Supervisor(s): Didier Lucor and Gaël Poëtte

PH.D. STUDENT PRESENTATION

Optimization of CO₂ storage assessment using selection of stochastic realizations

SARAH, BOUQUET

CHANTAL, DE FOUQUET

DOMINIQUE, BRUEL

MINES ParisTech, Geosciences centre, Fontainebleau, France.

Context & Objective:

Spatial variability of permeability is one of the main parameters that could significantly influence geological storage of CO₂. To evaluate uncertainties on injectivity and capacity, flow simulations are conducted on geostatistical simulations of permeability field. However, multiphase-flow simulations on thousands of grid cells is time-consuming and cannot be achieved on numerous sets of realizations. Is it possible to isolate a sub-set of realizations which reproduces minimum, maximum and mean behavior of an exhaustive set of realizations?

To test the feasibility of selection of stochastic realizations representative of mean, quantile, minimum and maximum behaviors, simplified flow simulation and selection procedures are combined. Efficiency of selections is compared between methods based on rankings or on distances of a specific criterion, calculated either from simplified or multiphase flow simulations.

Methods of selection:

Nicot et al. (2011) showed that pressure response from CO₂ injection in saline aquifer can be approximated by single phase flows which drastically reduces computational time. Since single-phase flow are approximated results, we propose to select realizations based on single-phase flow results on the exhaustive set and then simulate multiphase flow but only on the selected set.

Initiated for history-matching in oil & gas exploitation, the distance-based selection method relies on the following principle: if some realizations are close, either in term of geological model or in proxy-response, then they should behave similarly in terms of full flow simulations (Suzuky et al. (2008), Scheidt and Caer (2009), Alpak et al. (2010)). By using similarity distances between reservoir models, groups of similar realizations are created. Selecting one realization in each group should give similar statistics than the exhaustive set of reservoir models. Similarity distances between each couple of realizations can be calculated based on proxy-response of flow simulations. These distances are then projected in n-dimensional euclidean space by multidimensional scaling. Kernel K-means methods are used to identify clusters. Gaussian kernel function is used to calculate the distances between each realization and all centroids of clusters. Clusters are initialized by a spectral method and then iteratively modified. Once distances between realizations and centroids are minimized, the closest realization from the centroid of the corresponding cluster is selected for the sub-set of realization. As the number of realizations differed in each cluster, statistics from the sub-set will be weighted proportionally to their cardinals.

Application:

Spatial variability of permeability is simulated on 2D models by moving average. 200 realizations are generated. Permeability fields have a log-normal distribution with a mean of 100mD, median of 32.4mD and logarithmic standard deviation set at 1.5. Models are upscaled using geometric mean to obtain a finer mesh close to the injection well to limit calculation time. Variable of interest is the pressure response of the system because it is the limiting factor for injectivity and risks of interferences between CCS projects. Selection criterion is the pressure response at 6 months. Quality of this selection is evaluated by comparing results after one year of injection between exhaustive and selected sets. Two different methods of selection are tested:

1. Distance-based method, previously described, with a similarity distance based on pressure profile on the entire reservoir. So, it is a spatially variable response, contrary to previous studies,

for which selection criteria were scalar values such as produced volume (hydrocarbon or water-cut ratio), injected volume or well pressure.

2. Rankings of realizations according to its results proximity with respectively minimum, maximum or mean behavior of the 200 realizations. For minimum and maximum behaviors, criterion is the number of points similar to corresponding pressure envelop of the 200 realizations. For mean behavior, criterion is the area between pressure profile of the realization and mean pressure profile of all realizations. Then, number of selected realizations depends on deviation between statistic results of the selected set and exhaustive set. This error criterion is imposed by the user.

Selections are applied either directly on multiphase-flow response (reference scenario) or on a single-phase flow response (proxy-response), instead of streamlines simulations often used in studies of selection by distance-based methods. Selection results from each method are compared to results from 200 multiphase-flow simulations and results from random selections.

Results and comparison of methods:

51 realizations are selected from the minimum, maximum and mean ranking methods to honor the required maximum deviation based on multiphase-flow results. For a same number of realizations, maximum and minimum behaviors from rankings methods based on one-phase or multiphase flow simulations are closer to those from the exhaustive set of realizations than others methods. But errors from selected set based on similarity distances (51 realizations) are lower for quantiles (5, 25, 50, 75 and 95%), mean and standard deviation. Moreover, for these statistics, a set of 30 selected realizations via the distance based-method give equivalent or better estimate than ranking methods with 51 realizations. Distance-based selections indicate a clear improvement compared to random selections which is not the case for rankings methods for mean and median results. Overall, selection by statistic rankings methods would be more efficient to discriminate maximum/minimum cases. In another hand, if distance-based methods fail to approach minimum and maximum behavior, they could reproduce mean, median behavior and its dispersion.

For distance-based method, statistics results of selections based on single phase flow are close to those from multiphase flow (as well as for maximum and minimum results based on ranking). Nevertheless, except for the standard deviation, statistics from the 200 single-phase flow simulations poorly reproduce the 200 multiphase flow statistics compared to statistics from multiphase flow only on selected realizations (30 or 51). Therefore, using one-phase simulation as a proxy-response to select a sub-set of realizations for multiphase flow simulations would be the most efficient and accurate method to evaluate the uncertainty of geological parameters. Validity of this results are verified at 6 months and one year so selection could be predictive and period of simulation on the exhaustive set could be reduced.

References: [1] J.-P. Nicot, S. A. Hosseini and S. V. Solano, “Are single-phase flow numerical models sufficient to estimate pressure distribution in CO₂ sequestration projects?”, *Energy Procedia* (4):3919-3926, 2011.

[2] C. Scheidt and J. Caers, “Uncertainty quantification in reservoir performance using distances and kernel methods-application to a west Africa deepwater turbidite reservoir”, *SPE Journal* (14):680-692, 2009.

[3] S. Suzuki, G. Caumon, and J. Caers, “Dynamic data integration for structural modeling: Model screening approach using a distance-based model parameterization”, *Comput. Geosci.* (12): 105-119, 2008.

[4] F. O. Alpak, M. D. Barton, and J. Caers, “A flow-based pattern recognition algorithm for rapid quantification of geologic uncertainty”, *Comput. Geosci.* (14):603-621, 2010.

[Sarah Bouquet; MINES ParisTech, Geosciences centre, 45 rue Saint Honor, 77300 Fontainebleau, France.]

[sarah.bouquet@mines-paristech.fr –]

Ph.D. (11/2010 (*starting date*) — 11/2013 (*expected end*)): Interferences in between multiple injections of CO₂ in deep saline aquifer at industrial scale and at short to intermediate time scale

Supervisor(s): Dominique Bruel and Chantal de Fouquet

PH.D. STUDENT PRESENTATION

A Bayes decision approach to code validation in an industrial context.

GUILLAUME DAMBLIN, MERLIN KELLER, ALBERTO PASANISI
EDF R & D, 6 quai Watier, 78401 Chatou, France

ERIC PARENT, PIERRE BARBILLON, JACQUES BERNIER
AgroParisTech, 16 rue Claude Bernard, 75005 Paris, France

We deal with *validation* of costly computer models $y = y(x, \theta)$ seen as deterministic black box models depending on both a control variable $x \in \mathbb{R}^{d_x}$ and an unknown constant parameter $\theta \in \mathbb{R}^{d_p}$. The validation activities refer to the study of the predictive capability of the numerical model y with a special attention on how to take in account all sources of uncertainty. This is a major problem arousing a growing interest over years in industrial studies. When some experimental measures $z(x)$ from the true physical process simulated by y are unpractical, it may be convenient to perform numerical simulations $y(x, \theta)$ instead. The aim of validation is therefore to provide predictions of the real physical process coupled with tolerance bounds. It requires a set of experimental results $\{z(x_j)\}_{j=1, \dots, n}$ and is expressed as a statistical problem including both a calibration step (the fact of fitting θ) and a prediction step [2].

In many ways, pure-calibrated predictions $y(x, \hat{\theta})$ derived from a least squares equation are not relevant to be used. Indeed, a model bias $b(x)$ must be considered to take into account the inherent model uncertainty in addition to the measurement error ϵ . This idea comes from the reference framework introduced in 2001 by Kennedy and O’Hagan [1] that is based on a multi-fidelity regression equation : $z(x) = y(x, \theta) + b(x) + \epsilon$ with a Gaussian process as a prior assumption on both model y and bias b . It has been established that the inference based on this model can improve significantly the predictions [4]. However, in this general framework, several problems may occur, especially the lack of identifiability and the high computation cost as well.

After having illustrated these issues on toy functions, we think about how adjusting a validation framework so that it makes sense in some industrial cases. In particular, we focus on a simulator modelling the electrical consumption of a building for which we take advantage of the common Bayesian formulation [3] in order to derive a predictor based on the economic aspects of the problem.

References:

- [1] M. Kennedy and A. O’Hagan, “Bayesian Calibration of Computer Models”, *J. Roy. Statist. Soc.*, **B63**:425–464, 2001.
- [2] M.J Bayarri and al, “A Framework for Validation of Computer Models”, *Technometrics*, **B49**:138–154, 2007.
- [3] J.O Berger, “Statistical Decision Theory and Bayesian Analysis”, *Springer-Verlag*, 1985.
- [4] Loic Le Gratiet, “Bayesian analysis of hierarchical multi-fidelity codes”, 2012.

[Guillaume Damblin; EDF R & D and ABIES, AgroParisTech]
[guillaume.damblin@edf.fr – guillaume.damblin@gmail.com]

Master: University of Science and Technology, Lille 1

Ph.D. (09/2012 (*starting date*) — 08/2015 (*expected end*)): My PhD deals with Bayesian validation of “black box” computer models. A framework is proposed to achieve the predictive capability of the computer experiments. We also focus on bayesian decision theory to make estimators with regard to the social-economic consequences in an industrial case.

Supervisor(s): Eric Parent

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

[Alberto Pasanisi, Merlin Keller, Eric Parent; alberto.pasanisi@edf.fr, merlin.keller@edf.fr, eric.parent.agroparistech@gmail.com, pierre.barbillon@agroparistech.fr]

PH.D. STUDENT PRESENTATION

Comparison of Sensitivity Analysis methods applied on a groundwater flow and mass transport model

GREGORY, A.J., DEMAN

Centre of Hydrogeology and Geothermics (CHYN), University of Neuchâtel, Rue Emile Argand 11, 2000 Neuchâtel, Switzerland

JAOUHER, KERROU,

Centre of Hydrogeology and Geothermics (CHYN), University of Neuchâtel, Rue Emile Argand 11, 2000 Neuchâtel, Switzerland

HAKIM, BENABDERRAHMANE

ANDRA, 1-7 rue Jean Monnet, 92298 Châtenay-Malabry Cedex, France

PIERRE, PERROCHET

Centre of Hydrogeology and Geothermics (CHYN), University of Neuchâtel, Rue Emile Argand 11, 2000 Neuchâtel, Switzerland

The ANDRA (French National Agency for Nuclear Waste Management) considers a deep claystone formation of Callovo-Oxfordian age as a potential host for the disposal of high and intermediate-level and long lived radioactive waste. Prior to safety calculations, a three-dimensional groundwater flow and transport model was built and then used to conduct a sensitivity analysis study in order to assess the influence of the hydro-dispersive parameters uncertainties on the mass transport within the Meuse/Haute-Marne aquifer system. Considering the computational cost of running large scale models with high-dimensional inputs, the use of effective experimental designs is henceforth preferred over classical Monte Carlo schemes to explore the entire domain of uncertainty of the models' parameters.

Many researches were focused on selecting appropriate experimental designs to carry out sensitivity analysis on selected variables in order to rank their relative contribution to the total variance of the outputs of interest. Regression-based approaches, such as response surface method, coupled with statistical analysis (e.g. Student-test and Fisher-test) are useful techniques to assess the importance of interactions and high-order effects of inputs on model's outputs.

In this study, Latin Hypercube Designs have been selected to sample parameters over their respective statistical distributions. Linear and non-linear correlation analyses (e.g. Pearson and Kendall tests) were applied to assess the importance of input-output relationships. Then a multilinear regression model was derived to perform statistical tests. However, these quantitative measures can give misleading information in the event of having strong non-linear and non-monotonic relationships. As an alternative, the method of Morris [3] recently enhanced by Campolongo *et al.* [1] was applied. The method is an effective experimental design allowing the assessment of the influence of an individual input, as well as all-order interactions involving the latter, on the output of interest. This qualitative method has the advantage of giving a reliable overview of output's sensitivity to the model's parameters in a relatively small number of runs.

The case of Meuse/Haute-Marne site as potential repository for radioactive waste was considered to apply the above-mentioned methods to a three-dimensional groundwater flow and transport model. The finite elements hydrogeological model was developed by ANDRA in 2012 and consists of multi-layered aquifer system which integrates twenty-eight hydrogeological layers from the Trias to the Tertiary, with large contrasts of hydraulic parameters. A subset of 16 layers of interest was selected to

analyse the effects of uncertainty on their hydro-dispersive properties on the amount of mass reaching two target output zones. The uncertain parameters that have been tested are the dispersion parameters and the hydraulic conductivities (K) coupled with the effective porosities (Φ) in each of the sixteen layers of interest.

With respect to the dependency between both parameters, K and Φ are shifted from their nominal values according to the level given by the selected experimental design. The levels stand for a multiplicative factor that have been applied to the hydro-dispersive parameters : for the lowest level, hydraulic conductivities are multiplied by 0.01 and porosities are multiplied by 0.7, dispersion parameters are multiplied by 0.1 ; and for the highest level, hydraulic conductivities are multiplied by 100 and porosities are multiplied by 1.3, dispersion parameters are multiplied by 10.

A numerical realisation was selected for each combination of parameters. Hundreds of simulations of flow and mass transport were solved with the finite element simulator *GroundWater* [2]. Transient transport simulations were run for 1 Mio years with a source term located in the potential repository location within the claystone host formation. For every simulation the response of interest was the concentration in the target zones. Using ordinary least-squares method and classical statistic tools, sensitivity analysis was performed on the selected parameters to estimate their relative contribution to the total variance of the response. As a comparison tool the method of Morris-Campolongo was carried out to perform a screening exercise.

Both of the sensitivity analysis methods have identified the strong influence of two aquifers. In these aquifer formations groundwater flows mainly to North West while hydro-dispersive properties impacts the solute transit time and the associated solute mass at the outlet. However, the regression-based methods were less efficient in establishing a reliable relationship between inputs and outputs as we are in the presence of highly non-linear and non-monotonic behaviours.

References:

- [1] Campolongo, F., Cariboni, J., and Saltelli, A., “An effective screening design for sensitivity analysis of large models.”, *Environmental Modelling & Software*, in press., 2007.
- [2] Cornaton, F. J., “GroundWater: A 3-D Ground Water and Surface Water Flow, Mass Transport and Heat Transfer Finite Element Simulator. Reference Manual.”, *University of Neuchâtel*, 398 pp. 2007
- [3] Morris, M. D., “Factorial sampling plans for preliminary computational experiments.” *Technometrics* **33**: 161-174. 1991.

Grégory DEMAN

gregory.deman@unine.ch , <http://www2.unine.ch/chyn>

Master: Université Joseph Fourier, Grenoble, FRANCE

Ph.D. (01/2011 | 01/2015): Parametric uncertainty analysis by experimental design and surface response methods.

The purpose of this research is to test out the relevance of experimental design schemes when large numbers of hydrodynamic and hydrodispersive parameters are considered. The surface response methodology commonly used in petroleum engineering is used to analyze the uncertainties associated to large scale groundwater flow and transport models.

Supervisor(s): Jaouher Kerrou and Pierre Perrochet

PH.D. STUDENT PRESENTATION

Impact of weather data on the performance evaluation of a passive house: an application of sensitivity analysis

J., GOFFART

University of Savoy, Le Bourget du Lac, France

T., MARA

University of Reunion Island, St Denis, France

E., WURTZ

CEA-INES., Le Bourget du Lac, France

The context of the study is related to the need of decreasing building energy consumption. Indeed, building consumption represents more than 40% of total energy in Europe [1]. So far, much effort of the international community in Building Physics has concerned the development and validation of dynamic simulation software, (e.g. EnergyPlus [2]) used to design high-performance buildings.

Consequently, the models become more accurate and more detailed, increasing complexity and the number of inputs: more than hundreds. However during the design building process, most of the inputs are highly uncertain. This may involve significant uncertainties on the output and leads to poor energy performance of the final building project. In order to guarantee the consumption of the final building project, confidence on the results is crucial.

Another particularity of building energy models is that inputs can either be scalar or temporal. The scalar inputs generally characterize the building thermal behavior while the dynamic inputs are related to the weather data. All these inputs must be defined before running the model but are poorly known. We would like to perform the sensitivity analysis of the model by accounting for scalar and dynamic inputs simultaneously. The question we would like to answer is: Between the group of static inputs (building thermal properties) and the group of dynamic inputs (building solicitations) what preponderantly affects the energy model prediction?

The case study is a passive house (figure 1) in France, at the Bourget du Lac, near Lyon. This kind of construction tends to optimize the solar gain, that is why this kind of building is expected to be sensitive to weather data. We check this assumption by sensitivity analysis by accounting for dynamic inputs. The outputs are the heating demand for each thermal zone (the ground and the first floors) and the energy gains/losses by the windows for each facade and each zone.

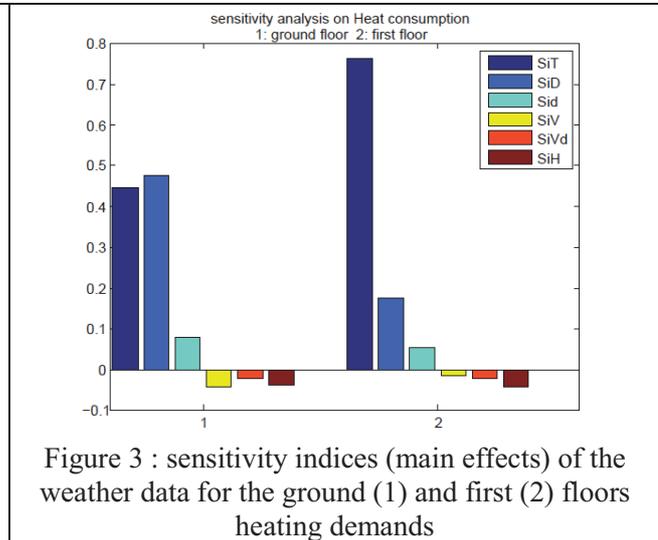
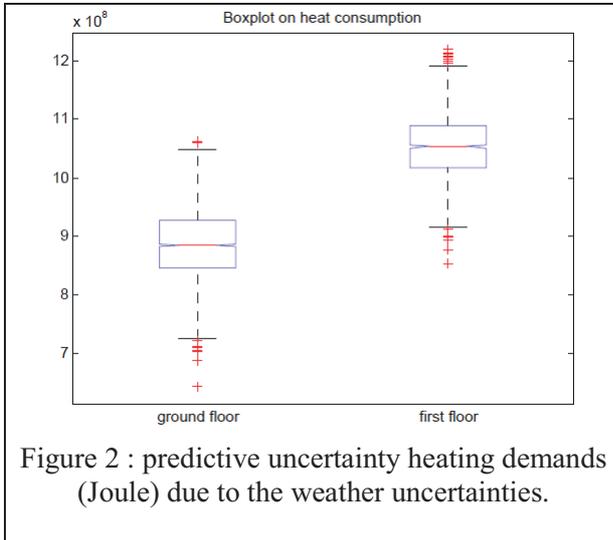


Figure 1: Passive house case study

Some results of a study with only variations on weather data are shown in the figures below. The uncertain inputs are dry bulb temperature (T), direct (D) and diffuse (d) radiation, relative humidity (H), wind direction (Vd) and wind velocity (V).

In figure 2, we show the non-negligible impact of the uncertain weather data at the ground and the first floors.

These results highlight the solar impact on the energy consumption and confirm that the passive design of the case study is efficient. In figure 3 we notice that at the ground floor the consumption is lower because the solar gain is more significant (bigger windows on south facade see figure 1).



References :

- [1] <http://epp.eurostat.ec.europa.eu> , Eurostat: European commission website for statistics.
- [2] Crawley et al., “Energy Plus: creating a new-generation building energy simulation program”, *Energy and Buildings*, 33 pp319-331, 2001.

[Jeanne Goffart; University of Savoy]
 [Jeanne.goffart@univ-savoie.fr]

Master: University of Joseph Fourier Grenoble

Ph.D. (11/2010 (*starting date*) | 11/2013 (*expected end*)) : The subject deals with the reliability on building energy simulation for low consumption building. More precisely it concerns the estimation of the impact of dynamic parameter as weather parameter and occupancy on output simulation as consumption and thermal comfort. Techniques of sensitivity analysis and uncertainty analysis are used to perform the aim of the thesis.

Supervisor(s): E. Wurtz and T. Mara

PH.D. STUDENT PRESENTATION

Stochastic modelling of inputs

MATHILDE GRANDJACQUES
 BENOIT DELINCHANT
 OLIVIER ADROT
G2ELab 38403 Saint Martin d'Hères, France

CLÉMENTINE PRIEUR
 CÉLINE HELBERT
LJK 38041 Grenoble, France

Energy consumption is becoming an important issue.

The building sector contributes up to 40% of total energy expenditure. An important part waste come from a poor design, the use of inappropriate technologies (e.g: choice of heating) or not reasoned behaviour.

Different models reflecting the design parameters and taking into account different inputs(temperature external, users' behaviour) are used today to improve performance energy of new buildings.

To analyse the quality of models or use them taking account inputs' uncertainty to forecast the consumption we therefore decided to perform a sensitivity analysis.

Inputs are dynamics, our approach was to calculate each time step Sobol's index by a Monte Carlo method.

Monte Carlo method require a lot of data, something that is difficult to obtain in quantity and quality. So an other solution is to find a modeling inputs quite close to reality.

Thanks to the study of a database of inputs we will explain in this study the mechanism generating the time series model of each inputs.

A building is composed by a thermal envelope, users, electrical equipments which can produce or store energy. Building model are state space representation:

$$\begin{cases} \dot{x} &= Ax + Bu \\ y &= Cx + Du \end{cases} \quad (1)$$

u : input vector (temperature, flux, ...)

x : state vector: wall temperature

y : output vector: internal temperature

A, B, C, D matrix containing the physical characteristics of the building.

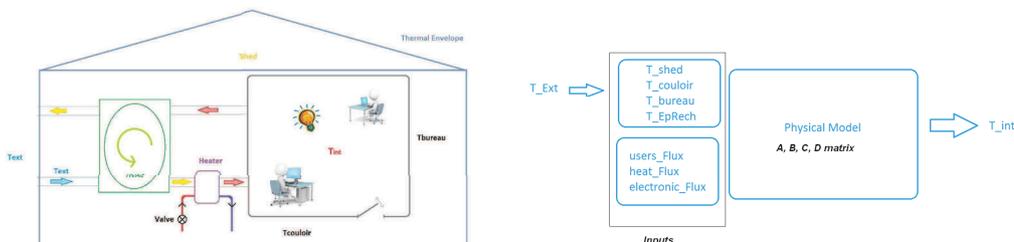


Figure 1: Buildings

We use a model of a monitoring platform. The internal temperature (T_{int}) is obtained from the temperature of different adjacent rooms and different heat flux (??).

Different sensors collect temperature and other information each hours creating a database.

Time series modelisation:

The most important informations for times series model is time dependence between inputs. A time series model can be decomposed as follows:

$$X(t) = Trend(t) + Seasonality(t) + Z(t) \quad \text{where } Z_t \text{ ARMA process} \quad (2)$$

Trend(t) and Seasonality(t) are determinism part.

Trend is estimated by least square method to a polynomial or spline method.

Seasonality can be a decomposition into Fourier series observing a strong period of 24 hours (e.g: Temperature) or a profile means that repeated every 24 hours. (e.g: Flux)

Once the trend and seasonality remove the time dependence is just inside Z_t . It's this dependence that we want to study. So, for that we survey the autocorrelation function, that is reflected the evolution of the time dependence.

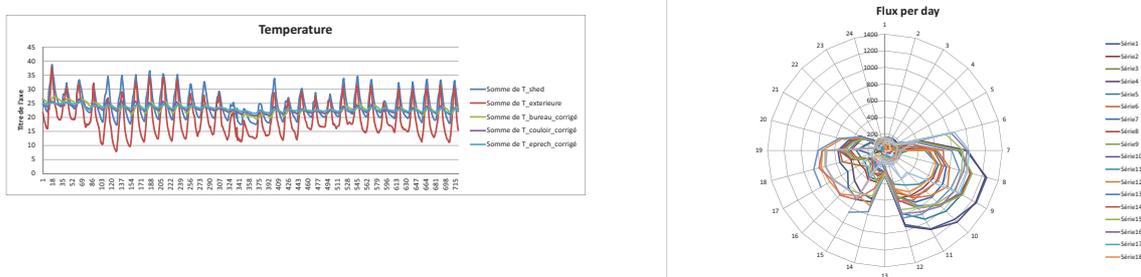


Figure 2: External temperature and heat flux:

In our model, all the temperature depend of T_{ext} so we have chose to create a second model linking all the other temperature to T_{ext} . We will model only T_{ext} by a time series

Regarding the heat flux, we have model also by a time series but we have to add some co-variable and take account the thermostat regulation (heat depend of the ventilation air flow (fixed by an algorithm) and occupancy).

Conclusion:

With this input's we want to to perform a sensitivity analysis.

First calculate Sobol's indices with T_{ext} , heat flux and some physical characteristics of the building's.

The second problem is that with this kind of model all the input's are correlated. How to perform a sensitivity analysis?

References:

- [1] Brockwell, Peter J. and Davis, Richard A., "Time Series: Theory and Methods Second Edition", Springer-Verlag New York Inc., 2009.
- [2] DRIGNEI, Dorin, Functional ANOVA in Computer Models With Time Series Output
- [3] A. Saltelli, K. Chan, and E. M. Scott. Sensitivity Analysis. Wiley, 1 edition, December 2008.

[Mathilde Grandjacques; G2ELab Domaine universitaire 961 rue de la Houille Blanche BP 46 38402 Saint Matrin d'Hères CEDEX]

[my e-mail address – mathilde.grandjacques@g2elab.grenoble-inp.fr]

Master: Université Joseph Fourier

Ph.D. (09/2011 (*starting date*) — 09/2014 (*expected end*)): The building sector represents a variable adjustment expenses energetic. So, we want to compute a sensitivity analysis on thermal models of the building to better manage expenses energetic.

Supervisor(s): Benoit Delinchant and Olivier Adrot

PH.D. STUDENT PRESENTATION

Likelihood-free simulation-based optimal design

MARKUS, HAINY
WERNER, G., MÜLLER
HELGA, WAGNER
Johannes Kepler University, Linz, Austria

Simulation-based optimal design techniques are a convenient tool for solving a particular class of optimal design problems. The goal is to find the optimal configuration of factor settings with respect to an expected utility criterion. This criterion depends on the specified probability model for the data and on the assumed prior distribution for the model parameters. We develop new simulation-based optimal design methods which incorporate likelihood-free approaches and utilize them in novel applications.

Most simulation-based design strategies solve the intractable expected utility integral at a specific design point by using Monte Carlo simulations from the probability model. Optimizing the criterion over the design points is carried out in a separate step. [3] introduce an MCMC algorithm which simultaneously addresses the simulation as well as the optimization problem. In principle, the optimal design can be found by detecting the mode of the sampled design points. Several improvements have been suggested to facilitate this task for multidimensional design problems (see e.g. [1]).

We aim to extend the simulation-based design methodology to design problems where the likelihood of the probability model is of an unknown analytical form but it is possible to simulate from the probability model. We further assume that prior observations are available. In such a setting it seems natural to employ approximate Bayesian computing (ABC) techniques in order to be able to simulate from the conditional probability model. In particular, we consider network planning for observations following a spatial extreme value distribution. It is not possible to obtain the analytical representation of the generalized extreme value distribution for dimensions greater than two (cf. [2]). We investigate the benefits and the limitations of our design methodology for this problem and analyze the sensitivity of the results on the ABC settings.

References:

- [1] B. Amzal, F.Y. Bois, E. Parent and C.P. Robert, “Bayesian-Optimal Design via Interacting Particle Systems”, *Journal of the American Statistical Association*, **101**(474):773–785, 2006.
- [2] R.J. Erhardt R.J. and R.L. Smith R.L., “Approximate Bayesian Computing for Spatial Extremes”, *Computational Statistics and Data Analysis*, **56**(6):1468–1481, 2012.
- [3] P. Müller, “Simulation Based Optimal Design”, In: J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith (eds.), *Bayesian Statistics 6*, Oxford University Press, New York:459–474, 1999.

[Markus Hainy; Johannes Kepler University Linz, Altenberger Straße 69, 4040 Linz, Austria]
[markus.hainy@jku.at – www.jku.at/ifas/]

Master: Johannes Kepler University Linz

Ph.D. (02/2012 (*starting date*) — 07/2015 (*expected end*)): Simulation-based optimal design algorithms with approximate Bayesian computing enhancements for intractable probability models.

Supervisor(s): Werner G. Müller and Helga Wagner

PH.D. STUDENT PRESENTATION

Sensitivity analysis for the study of a tire model with correlated parameters and an arbitrary distribution

SABRA HAMZA
 ABDERAZIK BIROUCHE
 MICHEL BASSET

Modélisation Intelligence Processus Systèmes (MIPS) laboratory, Haute-Alsace University, Mulhouse, France

FLORIANE ANSTETT-COLLIN
Centre de Recherche en Automatique de Nancy (CRAN), Lorraine University, France

Context

Tire is an essential element in vehicle dynamics, therefore passenger safety and comfort [4]. Indeed, it is the only contact between the vehicle and the ground. Depending on the application area, many models have been developed in the literature, to describe the behavior of the tire on the ground [5, 1]. These models contain numerous dependent input variables with an arbitrary distribution. The aim of this paper is to determine the parameters reflecting the variation of the tire forces in order to understand the mechanisms involved in dangerous situations.

Problem statement

Pacejka tire model is considered in this study [5]. It is one of the most employed model used to calculate forces and moments developed at the tire contact with the ground. An overview of inputs and outputs of tire Pacejka model is described below :

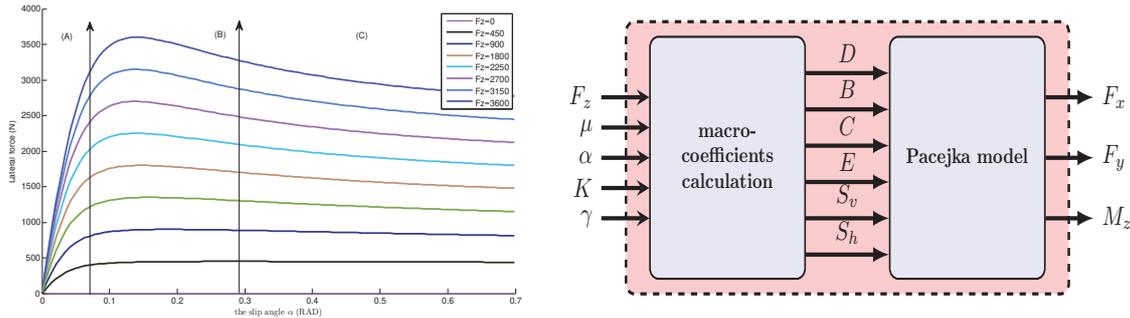


FIGURE 1 – a) Tire lateral force F_y function of slip α , for different values of F_z b) An overview of the structure of Pacejka tire model

Input variables considered are the vertical load F_z , the friction factor μ , the slip angle α , the lateral stiffness K_y and the camber angle γ . The coefficients B , C , D , E , S_h and S_v are the macro-coefficients. In this study, we will focus on the pure lateral force F_y which is given as follows :

$$F_y = \mu F_z \sin(C \arctan(B(\alpha + S_h) - E(B(\alpha + S_h) - \arctan(B(\alpha + S_h)))))) + S_v \tag{1}$$

The lateral stiffness K_y as a function of the vertical load F_z and the camber angle γ is defined as :

$$K_y = p_1 F_z \sin(2 \arctan(\frac{F_z}{(p_2 F_z)}))(1 - p_3 |\gamma|) \tag{2}$$

The parameters p_1 , p_2 and p_3 are determined experimentally to fit each type of tire. The contribution of this paper is an extension of the study and the quantification of the contribution of Pacejka tire input variables (F_z , α , μ , γ , K_y) on the model output F_y [2], by taking into account of real aspects of dependent input variables which are arbitrarily distributed.

Methodology

Two approaches are combined to study the sensitivity of Pacejka tire model dependant input variables which are arbitrarily distributed [3, 4]. The proposed approach consists of the decorrelation of the model input variables then in the determination of variance-based sensitivity indexes in terms of arbitrary polynomial chaos (aPC) coefficients. These latter are determined by constructing an orthonormal polynomial basis only in terms of data that characterize the model input variables.

The adopted methodology is summarized in 3 steps as follows :

- *Step 1* : Input decorrelation using Cholesky decomposition.
- *Step 2* : Orthonormal data basis construction in terms of non-central statistical moments of variables.
- *Step 3* : Calculation of sensitivity indexes explicitly from polynomial chaos coefficients.

Results

Based on the curve describing the behavior of the tire (the pure lateral force F_y as a function of the slip angle α (see FIGURE 1-a)), the methodology described above has been applied in the three zones (linear (A), nonlinear (B) and saturation zone (C)). The main effect of a parameter, its influence without its correlative contribution with other parameters and its uncorrelated contribution on the model output have been determined in each interval. In the linear zone (A), increasing the slip angle and the normal load on the tire also increases the force it generates. Consequently the normal load has the highest sensitivity index, explicitly the stiffness factor and followed by the slip angle. In the nonlinear zone (B), the lateral force can reach its maximum which depends on the normal load and the friction coefficient. Therefore normal load index remains high and friction coefficient index has increased. Finally, in the saturation zone (C), the normal load sensitivity index is slightly decreased and the contribution of the slip angle is almost negligible. The results obtained are physically consistent.

References

- [1] R. Kiebre, “Contribution to Modeling of Aircraft Tyre-road Interaction”, PhD thesis, Université de Haute-Alsace, 2010.
- [2] R. Kiebre, F. Anstett-Collin, and M. Basset, “Sensitivity analysis for tire/road interface model”. *International Conference on Sensitivity Analysis of Model Output (SAMO '10)*, Milan, Italy, 19–22 July 2010.
- [3] T. A. Mara and S. Tarantola, “Variance-based sensitivity indices for models with dependent inputs”. *Reliability Engineering & System Safety*, 107 :115–121, 2012.
- [4] W. F. Milliken and D. L Milliken, “Race Car Vehicle Dynamics”. *Society of Automotive Engineers Inc*, 1996.
- [5] S. Oladyshkin and W. Nowak, “Data-driven uncertainty quantification using the arbitrary polynomial chaos expansion”, *Reliability Engineering & System Safety*, 106 :179–190, 2012.
- [6] H. Pacejka, “Tire and vehicle dynamics”. Elsevier Ltd, 2002.

[Sabra HAMZA; Modelling, Intelligence, Process and Systems (MIPS) Laboratory, Mulhouse, France, sabra.hamza@uha.fr]

Master: University of Caen Basse-Normandie

Ph.D. (12/2011 — 11/2014): The objective of this thesis is to develop approaches for global sensitivity analysis for dynamic models. These approaches will be applied in the automobile domain.

Supervisor(s): Michel BASSET, Floriane ANSTETT-COLLIN and Abderazik BIROUCHE

PH.D. STUDENT PRESENTATION

**Fully Bayesian approach for the estimation
of first-order Sobol indices**

BENOIT JAN, JULIEN BECT AND EMMANUEL VAZQUEZ

SUPELEC, Gif-sur-Yvette, France

We consider the problem of estimating the first-order Sobol indices of a function that represents the output of a computer code. It is well known that Monte Carlo estimators of Sobol indices require many evaluations of the computer code (see, e.g., [1]). When running the computer code is time- or resource-consuming, it has become common practice [2-6] to use a metamodel of the expensive computer code. A natural question is, then, to quantify the error of approximation of the Sobol indices that is made when using the metamodel instead of the computer code. This question has been addressed in [4-6].

We focus in this work on the case of a Gaussian process-based (kriging) metamodel [3-5], where the posterior distribution of the Sobol indices provides an elegant answer to the above concern. Algorithms for drawing samples from this posterior distribution have been proposed, based on the simulation of sample paths of either the function itself [3], or its main effects [5]. In both papers, the maximum likelihood estimate of the parameters of the covariance function are simply plugged into the kriging equations.

We argue that the use of a plug-in approach for the parameters of the covariance function is a dangerous practice in this setting. We propose to adopt instead a fully Bayesian approach—i.e., to choose a prior distribution for the parameters of the covariance function—in order to take into account the uncertainty about the Sobol indices coming from the uncertainty about the parameters of the covariance. Samples from the posterior distribution of the parameters are drawn using a MCMC algorithm, and then used to generate samples from the posterior distribution of the (estimator of the) Sobol indices.

Our simulations show that the plug-in approach often produces overly confident credible intervals, while the Bayesian approach tends to produce larger (more conservative) intervals. Figure 1 shows an example of the distributions obtained for the g-function of Sobol in two dimensions. Examples in more than two dimensions will also be discussed.

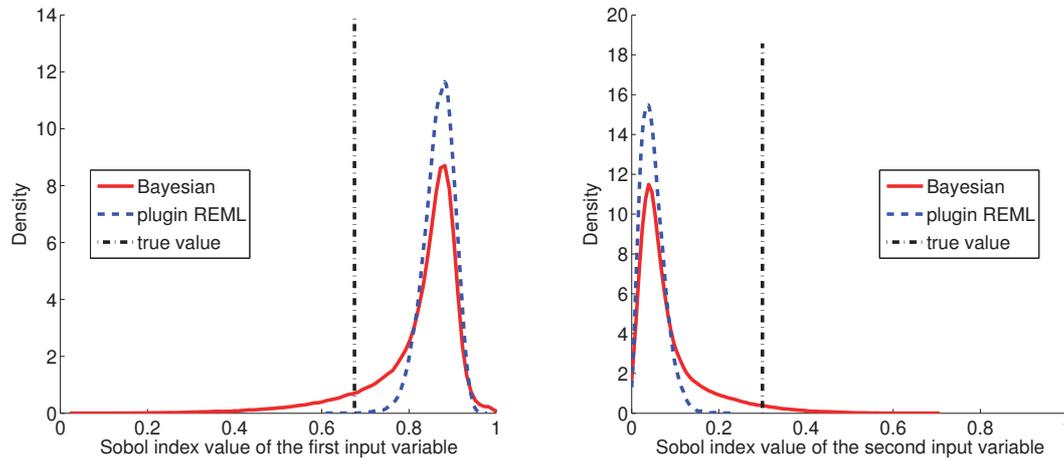
[Ph.D. Benoit Jan; SUPELEC, benoit.jan@supelec.fr]

Master: SUPELEC

Ph.D. (10/2012 (*starting date*)— 10/2015 (*expected end*)): The effectiveness of kriging strongly depends on the choice of an appropriate covariance structure and on the estimation of its parameters. When the number of input variables is large and the number of observations is small, this choice becomes difficult and kriging can perform badly. The aim of this thesis is to provide solutions to address some problems of kriging in high-dimensional settings, with a view towards applications in the sequential design of experiments.

Supervisor(s): Julien Bect and Emmanuel Vazquez

Figure 1: An example of distributions of first-order Sobol indices, for the g-function of Sobol with parameters $a_1 = 1$ and $a_2 = 2$, estimated using a Gaussian process conditioned on ten observations. Solid lines correspond to the Bayesian method; dashed lines correspond to the plug-in method. The vertical dash-dot lines are the true values of the Sobol indices.



References:

- [1] A. Janon, T. Klein, A. Lagnoux, M. Nodet and C. Prieur, “Asymptotic normality and efficiency of two Sobol index estimators”, Preprint, <http://hal.inria.fr/hal-00665048>, 2012.
- [2] W. Chen, R. Jin, and A. Sudjianto, “Analytical metamodel-based global sensitivity analysis and uncertainty propagation for robust design”, *Journal of Mechanical Design*, **127**:875–886, 2005.
- [3] C. B. Storlie, L. P. Swiler, J. C. Helton and C. J. Sallaberry, “Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models”, *Reliability Engineering & System Safety*, **94**(11):1735–1763, 2009.
- [4] J. E. Oakley and A. O’Hagan, “Probabilistic sensitivity analysis of complex models: a Bayesian approach”, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **66**(3):751–769, 2004.
- [5] A. Marrel, B. Iooss, B. Laurent and O. Roustant, “Calculations of Sobol indices for the Gaussian process metamodel”, *Reliability Engineering and System Safety*, **94**(3):742–751, 2009.
- [6] A. Janon, M. Nodet and C. Prieur, “Uncertainties assessment in global sensitivity indices estimation from metamodels”, Preprint, <http://hal.inria.fr/inria-00567977>, 2012.

PH.D. STUDENT PRESENTATION

Kriging-based Surrogate Modeling of Health Indicators for the Monitoring of a Turbofan Fuel System

BENJAMIN, LAMOUREUX

PIMM Laboratory, Arts & Métiers ParisTech, Paris, France

SAFRAN Snecma, Control Systems Division, Villaroche, France

NAZIH, MECHBAL

PIMM Laboratory, Arts & Métiers ParisTech, Paris, France

JEAN-REMI, MASSE

SAFRAN Snecma, Control Systems Division, Villaroche, France

Abstract:

For aircraft engine manufacturers, the scourge of modern time is the lack of availability because it induces expensive consequences such as unscheduled maintenance operations or delays & cancellations. Thus, research focuses on new approaches to increase the engine availability and one of the most promising one is Prognostics and Health Management (**PHM**). PHM is based on the monitoring of representative variables traducing the health condition of a given system and usually performs fault detection, fault classification and prognostic of the remaining useful life. It provides detailed asset health information which enables improved maintenance decision-making, starting point for novel maintenance strategies such as Condition-Based Maintenance (CBM) and Predictive Maintenance (PM).

In the past decade, PHM has gradually emerged as inevitable research topic, as evidenced by the numerous papers published in both the diagnostic [1] and the prognostic [2] domains. Many algorithmic methods have already been developed in order to perform the aforementioned tasks and nowadays, some efforts are even made to standardize the PHM concept [3]. Even if it originally focuses on structural applications [4], PHM is spreading to other areas of expertise, such as bearings monitoring [5], turbine blade creeping prediction [6] or in our context, complex aeronautic control system supervision.

PHM is mainly a problem of discrimination between two distributions, with false alarm and good detection performances [7]. Despite the abundance of mathematical tools, the development of PHM for complex systems is an intrinsic stochastic problem requiring a huge amount of data to validate items such as detection threshold selection or classification schemes. The best solution would be obviously to wait for the system to be worn enough to provide all the needed healthy and faulty data. However, this solution is not applicable in the case of highly reliable and costly complex systems for two reasons: first, due to the high reliability rate, it would take a very long time before recovering all the degraded data needed, and then the prohibitive controller retrofit costs make it impossible to wait for the data to program the embedded part of the PHM system.

In order to compensate this lack of data in the upstream development phases, it is necessary to create data from a numerical model of the system. Indeed, contrary to real data collection which can be done only during the in-service operations phase, modeling can be initiated earlier, during the design phase. For each degradation mode, a model is built in which all kinds of uncertainties, be it environmental ones or structural ones are taken into account in order to create a stochastic model of both the healthy behavior and the faulty behaviors of the system.

Thus, as explained above, the system modeling is the keystone of the PHM system validation. However, the high level of complexity of systems such as aircraft engines implies major difficulties in

modeling and even more in uncertainty management because of the huge amount of parameters. Indeed, when the model is expensive in computation time, the number of calls to the model should be limited and consequently, both advanced sensitivity analysis [8] and uncertainties propagation methods must be used, the former to optimize the selection of parameters and the latter to optimize the numerical design of experiment.

The first part of the document defines the basics of prognostic and health management and introduces in broad outlines the functions and architecture of the aircraft engine fuel system, on which the application is conducted. The model associated to this system is non linear, non monotonous and present some interactions between parameters. The challenge of the health monitoring for this kind of system is also addressed. The second part introduces the main sensitivity analysis methods, particularly the Morris method [9] and the Sobol Indices [10] based on surrogate modeling.

The third part is the applicative one and consists in applying the aforementioned methods to the application model. Firstly, a Morris method is used to perform a coarse sensitivity analysis of this model. Then, Due to the high computational cost of the model, surrogate modeling is undertaken in order to compute the Sobol Indices. Two methods are compared: one consisting in using a classical two levels based design of experiment associated to linear regression and the second consisting in using a Latin hypercube sampling associated to Kriging. It appears that the second method shows very good results. Thanks to the computed Sobol Indices, a hierarchization of the parameters is performed. Eventually, a Monte-Carlo algorithm is run on the recalibrated Kriging model to build the stochastic distribution of health indicators and validate the PHM system.

References:

- [1] R. Isermann, "Supervision, fault-detection and fault-diagnosis methods - An introduction," *Control Engineering Practice* vol.5, pp. 639-652, 1997.
- [2] P. W. Kalgren, C. S. Byington, M. J. Roemer, and M. J. Watson, "Defining PHM, a Lexical Evolution of Maintenance and Logistics," in *IEEE Autotestcon*, 2006, pp. 353-358.
- [3] J. W. Sheppard, M. A. Kaufman, and T. J. Wilmer, "IEEE Standards for Prognostics and Health Management," *Aerospace and Electronic Systems Magazine*, vol. 24, no. 9, p. 34-41, 2009.
- [4] N. Mechbal, M. Vergé, G. Coffignal, and M. Ganapathi, "Application of a combined active control and fault detection scheme to an active composite flexible structure," *Mechatronics* vol. 16, pp. 193-208, 2006.
- [5] M. J. Roemer and C. S. Byington, "Prognostics and health management software for gas turbine engine bearings," in *ASME Turbo Expo*, Montreal, 2007, pp. 795-802.
- [6] P. Baraldi, F. Mangili, and E. Zio, "Ensemble of Bootstrapped Models for the prediction of the Remaining Useful Life of a Creeping Turbine Blade," in *Proceedings of IEEE PHM*, Denver, 2012.
- [7] T. D. Wickens, *Elementary Signal Detection Theory*, N. Y. O. U. Press, Ed. 2002. (ISBN 0-19-509250-3).
- [8] B. Iooss, "Revue sur l'Analyse de Sensibilité Globale de Modèles Numériques," *Journal de la Société Française de Statistique*, vol. 152, no. 1, pp. 3-25, 2011.
- [9] M. D. Morris, "Factorial sampling plans for preliminary computational experiments," *Technometrics*, vol. 33, no. 2, pp. 161-174, 1991.
- [10] I. M. Sobol, "Sensitivity estimates for nonlinear mathematical models," *Mathematical Modelling and Computational Experiments*, vol. 1, p. 407-414, 1993.

Benjamin Lamoureux; PIMM Laboratory UMR 8006 □SAFRAN Snecma
benjamin.lamoureux@snecma.fr; bl9486@gmail.com

Master: Arts & Métiers ParisTech of Paris

Ph.D. (04/2011 | 03/2014): description of the subject (5 lines maximum)

Supervisor(s): P. Lorong (PIMM), N. Mechbal (PIMM) & J.-R. Massé (SAFRAN Snecma)

PH.D. STUDENT PRESENTATION

A new sensitivity analysis method for failure probability

PAUL, LEMAÎTRE

EDF R&D, Chatou, France; INRIA Bordeaux Sud-Ouest, Talence, France

Sensitivity analysis of a numerical model, for instance simulating physical phenomena, is a tool used to explore, understand and (partially) validate computer codes. It aims at explaining the outputs regarding the input uncertainties [1]. This communication proposes a sensitivity index, based upon the modification of the probability density function (pdf) of the random inputs, when the quantity of interest is a failure probability (probability that a model output exceeds a given threshold).

Let us consider a real-valued numerical model denoted by $G : \mathbb{R}^d \rightarrow \mathbb{R}$. In practice, each run of G can be CPU time consuming. We are interested in the event $G(\mathbf{X}) < 0$ (system failure) and in the complementary event $G(\mathbf{X}) \geq 0$ (system safe mode). $\mathbf{X} = (X_1, \dots, X_d)^T$ is a d -dimensional continuous random variable whose joint pdf is denoted f . For $i = 1, \dots, d$, let f_i denotes the distribution of X_i (the marginal pdf). We make the assumption that all components of \mathbf{X} are independent. The quantity of interest is the system failure probability:

$$P = \int \mathbf{1}_{\{G(\mathbf{x}) < 0\}} f(\mathbf{x}) d\mathbf{x}. \quad (1)$$

Given an unidimensional input variable X_i with pdf f_i and some perturbation parameter δ lying in a given subset of \mathbb{R} , let call $X_{i\delta} \sim f_{i\delta}$ the corresponding perturbed random input. Accordingly, the failure probability becomes:

$$P_{i\delta} = \int \mathbf{1}_{\{G(\mathbf{x}) < 0\}} \frac{f_{i\delta}(x_i)}{f_i(x_i)} f(\mathbf{x}) d\mathbf{x}. \quad (2)$$

Thus one can define the sensitivity index:

$$S_{i\delta} = \left[\frac{P_{i\delta}}{P} - 1 \right] \mathbf{1}_{\{P_{i\delta} \geq P\}} + \left[1 - \frac{P}{P_{i\delta}} \right] \mathbf{1}_{\{P_{i\delta} < P\}}. \quad (3)$$

These sensitivity indices can be computed using the sole set of simulations that has already been used to estimate the failure probability P with Monte-Carlo, thus limiting the number of calls to the numerical model. One can consistently estimate $P_{i\delta}$ by :

$$\hat{P}_{i\delta N} = \frac{1}{N} \sum_{n=1}^N \mathbf{1}_{\{G(\mathbf{x}^n) < 0\}} \frac{f_{i\delta}(x_i^n)}{f_i(x_i^n)}. \quad (4)$$

This property holds in the more general case when P is originally estimated by importance sampling rather than simple Monte Carlo, which is more appealing when G is time-consuming [2,3]. Asymptotical properties of the indices are derived [4], including a central limit theorem for $\hat{S}_{i\delta N}$, the plug-in estimator of $S_{i\delta}$.

We suggest to define a perturbed input density $f_{i\delta}$ as the closest distribution to the original f_i in the entropy sense and under some constraints of perturbation. Information-theoretical arguments [5] led us to choose the Kullback-Leibler (KL) divergence between $f_{i\delta}$ and f_i as a measure of the discrepancy to be minimized under some constraints. Several types of constraints are introduced. In the case of moment constraints, the perturbed density has an explicit form [6].

The relevance of this new sensitivity analysis method is analysed through three case studies.

References:

- [1] A. Saltelli, Sensitivity analysis for importance assessment, *Risk Analysis*, **22**(3):579–590, 2002.
- [2] T.C. Hesterberg. Estimates and confidence intervals for importance sampling sensitivity analysis, *Mathematical and Computer Modelling*, **23**(8):79–85, 1996.
- [3] R.J. Beckman and M.D McKay. Monte-Carlo estimation under different distributions using the same simulation. *Technometrics*, **29**(2):153–160, 1987.
- [4] P. Lemaître, E. Sergienko, A. Arnaud, N. Bousquet, F. Gamboa and B. Iooss . Density modification based reliability sensitivity analysis (Submitted) <http://hal.archives-ouvertes.fr/hal-00737978>
- [5] T.M. Cover and J.A. Thomas. Elements of information theory 2nd edition, *Wiley series in telecommunications and signal processing*. 2006.
- [6] I. Csiszár. I-divergence geometry of probability distributions and minimization problems. *The Annals of Probability*, **3**(1):146–158, 1975.

[Paul Lemaître; paul.lemaitre@edf.fr]

Master: INSA Toulouse, France

Ph.D. (10/2010 (*starting date*) — 10/2013 (*expected end*)): The subject of this PhD is *sensitivity analysis in the context of reliability studies*. In other words, the aim of this PhD is to quantify the impact of uncertainty on a binary output variable given some uncertainty on the input scalar variables. This PhD addresses the need to use phenomenologic computational codes, in order to assess the safety of EDF's nuclear installations.

Supervisor(s): P. Del Moral; A. Arnaud; B. Iooss

PH.D. STUDENT PRESENTATION

Comparing conservative estimations of failure probabilities using sequential designs of experiments in monotone frameworks

VINCENT, MOUTOUSSAMY

*EDF Research and Development, 6 quai Watier, 78401 Chatou, France
Institut de Mathématiques de Toulouse, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse, France*

NICOLAS, BOUSQUET

EDF Research and Development, 6 quai Watier, 78401 Chatou, France

THIERRY, KLEIN

PAUL, ROCHET

Institut de Mathématiques de Toulouse, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse, France

A key point of structural reliability studies is to estimate the probability of an undesirable event. Affecting a device, this estimation is made possible by using of a numerical code that represents the physical behavior of the studied phenomenon. Its uncertain inputs are modeled as random variables. This computer code is deterministic and time-consuming. The events considered are usually rare, accuring with a low probability. These constraints forbid in practice to use crude Monte Carlo methods. Variance reduction methods must be carried out to provide usable estimations in due time. This problem drives to the development and adaptation to few classical methods to reduce the number of calls of the code and keep a good precision at the same time.

In this context, we assume the computer code is monotone with binary output (see [4]). Then, one obtains two disjoint spaces: the *failure space* and the *safety space* separated by the *limit state*. Knowing the nature (safety or failure) of a set of points makes possible to construct two subspaces. The first is included in the failure space and the second one in the safety space. From these two subspaces, the monotone hypothesis enables to compute exact bounds to the probability (see [1]). The subspace defined as the complementary of the two failure and safety subspaces is called the non-dominated space.

A classical way to estimate a failure probability is to construct a sequential design of experiments such that the next point to be evaluated maximizes the gain in information. The objective is to select a criterion which chooses the best point within the *non dominated space* to accelerate the narrowing of the bounds. Two main approaches are explored. The first one is based on maximin criteria. Other approaches uses classification tools to have an approximation of the *limit surface*. Theses methods are tested and compared on the basis of toy examples and several dimensional settings.

References:

- [1] Bousquet, N., “Accelerated Monte Carlo estimation of exceedance probabilities under monotonicity constraints”, *Annales de la Faculté des Sciences de Toulouse*, **21**, 557:591, 2012.
- [2] Kroese, D.P., and Rubinstein, R.Y, “Simulation and the Monte Carlo Method” (2nd edition). *Wiley*, 2007
- [3] Madsen, H.O. and Ditlevsen, O., “Structural reliability methods”. *Wiley*, 1996
- [4] de Rocquigny, E, “Structural reliability under monotony: A review of properties of FORM and associated simulation methods and a new class of monotonous reliability methods (MRM)”. *Structural Safety*, **31**, 363-374, 2009

vincent.moutoussamy@edf.fr]

Master: University Paris Descartes

Ph.D. (12/2012 — 12/2015): To estimate the probability of an undesirable events, numerical code are made to simulate a structural physical behavior. The numerical code is usually considered as time-consuming and black box. In this thesis, we assume the numerical code is monotonic. This hypothesis, allow to compute exact bounds of the probability and to construct a better estimator than classical methods.

Supervisors: Nicolas Bousquet, Fabrice Gamboa, Bertrand Iooss, Thierry Klein

[Nicolas Bousquet; EDF Research and Development, 6 quai Watier, 78401 Chatou, France
nicolas.bousquet@edf.fr]

PH.D. STUDENT PRESENTATION

Uncertainty quantification and visualization for functional data

SIMON NANTY

Commissariat à l'Énergie Atomique, Cadarache (13), France

In the scope of nuclear safety and nuclear power plant lifetime program, physical modeling tools have been developed to assess the reliability of nuclear plants in numerous scenarios of use or of accident. These computer codes entail a lot of physical variables as inputs or outputs. Some of these variables can, in some cases, be functional ones. For example, parameters can be time or space dependent and represent the evolution of a physical variable during an experiment or be a map of a variable. The knowledge of the behavior of these functional parameters and of their uncertainties is of great interest for the study of these nuclear safety computer codes and the quantifying of the uncertainties. This enables to assess their impact on the results of the physical modeling. The modeling of a highly hypothetical type of accidental scenario is analyzed in this work. This modeling is computed thanks to two different computer codes: CATHARE2 and CAST3M. The first one models thermal-hydraulic behaviors inside the plant, and the later one computes the thermal-mechanical analysis of the system. CATHARE2 takes several scalar parameters as inputs, related to initial conditions in the power plant, and produces three functional variables depending on time only: temperature, pressure and flow rate. These variables along with some mechanical scalar parameters are used as parameters to CAST3M which gives a safety evaluation criterion as output.

Our objective in this work is to extract the features of the three studied functional variables and to quantify their uncertainties. This statistical characterization is meant to make us eventually able to generate new samples of these curves with the same features and obeying the same statistical laws. These new samples of data will be used then in future works of sensitivity analysis and uncertainty propagation on the CAST3M code to analyze the influence of these three curves on the safety criterion. Indeed, the first computer code, which is used to generate the three curves, is very costly to run. Therefore it allows producing only a few curves. A tool which enables to quickly produce this type of curves is needed to create a bigger sample of curves, and motivates so the characterization of the functional data. Methods of visualization of the curves and of their uncertainties are also proposed to help understanding them.

Among the techniques that could be applied for statistical characterization on functional data, one can cite the methods of functional decomposition. The functional principal components analysis introduced by Ramsay and Silverman [1] or non-linear principal component analysis used by Auder in [2] and the wavelet decomposition introduced by Donoho and Johnstone [3] and Mallat [4] are promising options of decomposition. More generally dimension reduction like multi-dimensional scaling method based on metric spaces and used by Scheidt and Caers in [5] can also be considered. All these methods could be combined to non-supervised classification, if clusters seem to appear among the data. A previous work by Auder [2] has shown the possibility to cluster the space of the temperature curve.

Some considered visualization tools rely on the results of dimension reduction methods like functional boxplot and bagplot proposed by Hyndman and Shang in [6]. The first one uses

density estimation techniques and the later one extends the multi-dimensional bagplot developed by Rousseeuw et al. [7]. Furthermore, the functional boxplot introduced by Sun and Genton [8] and based directly on functional data could be applied too on the data.

References :

- [1] J.O. Ramsay and B.W. Silverman, “Functional Data Analysis”, *Springer*, 2005.
- [2] B. Auder, “Classification et modélisation de sorties fonctionnelles de codes de calcul.”, *Phd Thesis*, 2011.
- [3] D.L. Donoho and I.M. Johnstone, “Ideal spatial adaptation via wavelet shrinkage”, *Biometrika* (81):425-455, 1994.
- [4] S. Mallat, “A theory for multiresolution signal decomposition: the wavelet representation”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (11):674-693., 1989.
- [5] C. Scheidt and J. Caers, “A new method for uncertainty quantification using distances and kernel methods: application to a deepwater turbidite reservoir”, *Society of Petroleum Engineers*, (14):680-692, 2008.
- [6] R. J. Hyndman and H. L. Shang, "Rainbow plots, bagplots, and boxplots for functional data", *Journal of Computational and Graphical Statistics*, **19**(1):29-45, 2010.
- [7] P. Rousseeuw, I. Ruts and J. W. Tukey, “The Bagplot: A Bivariate Boxplot,” *The American Statistician*, **53**(4):382–387, 1999.
- [8] Y. Sun and M.G. Genton, "Adjusted functional boxplots for spatio-temporal data visualization and outlier detection", *Environmetrics*, (23):54-64, 2012.

[Simon NANTY ; Commissariat à l'Energie Atomique, Cadarache (13), France ; simon.nanty@cea.fr]

Master: Université Joseph Fourier

Ph.D. (10/2012 (*starting date*) | 09/2015 (*expected end*)): The objective of the thesis is to quantify the uncertainties associated with a computer code with functional inputs and scalar output. The first step of the study is to statistically characterize the functional inputs of the code in order to be able to reproduce samples of these inputs. The two next steps consist in uncertainty propagation and sensitivity analysis in the context of functional inputs.

Supervisor(s): C. Helbert, A. Marrel, N. Perot and C. Prieur

PH.D. STUDENT PRESENTATION

Sensitivity analysis of long-term yield curves

ARESKE COUSIN

ALEXANDRE JANON

IBRAHIMA NIANG

VÉRONIQUE MAUME-DESCHAMPS

Université Claude Bernard (Lyon 1), ISFA, Laboratoire Sciences Actuarielle et Financière

We propose a sensitivity and uncertainty analysis of some long-term life insurance or risk management models. In risk management, present values of financial or insurance products are computed by discounting future cash-flows. In market practice, discounting is done by using the current yield curve, which gives the offered interest rate as a function of the maturity (time to expiration) for a given type of debt contract. This allows to consider different interest rates for different horizons.

Several operational methods are used in practice to reconstruct a yield curve from observed market data such as Euribor, FRA rates and Swap rates. It can be shown that the choice of the reconstruction method may have a significant impact on present values of cash-flows, especially for long-term horizons, for which few market prices are available to calibrate the yield curve. In order to analyze the uncertainty underlying the choice of the reconstruction method (in other words, the reconstruction model uncertainty), we consider a class of over-parametrized stochastic interest rate models which turns out to be auto-compatible with prices of standard debt products.

We introduce a reconstruction method based on the class of affine interest rate models with mean-reversion effect. More specifically, the current yield curve is given as an output of an instantaneous spot rate model with a dynamic given by the following stochastic differential equation:

$$dr_t = a(b(t) - r_t)dt + \sigma(r_t, t)dL_t,$$

where $a > 0$, L is a Lévy process, $b(\cdot)$ is a deterministic function of time, and $\sigma^2(\cdot, t)$ is an affine function, i.e., $\sigma^2(r_t, t) = \alpha(t) + \beta(t)r_t$. This specification allows us to consider several classical one-factor interest rate models such as Hull&White, extended CIR processes, or Ornstein-Uhlenbeck processes driven by a Lévy subordinator.

For given values of parameters a , σ , and the parameters that drive the Lévy process, we show that the b function can be chosen so that the model yield curve perfectly matches the market price of standard debt products. In other words, for each choice of (a, σ) , we can find a yield curve which is fully compatible with the market data.

We consider that b is a piecewise-constant function of time where discontinuities points correspond to the set of maturities for which market prices are known. We then use an iterative procedure to calibrate function b from observed market prices. Hence, we obtain a family of market-compatible yield curves which depend on the input parameters a and $\sigma(\cdot, \cdot)$.

The aim of this work is to study the sensitivity on the uncertain parameters a and σ of the present value of some insurance contracts such as mortgages and pensions, which are inherently long-term contracts. The indeterminacy of these parameters (the “free” parameters) causes an uncertainty on the yield-curve reconstruction, which in turn may have a huge impact on risk management decision.

In this talk, we propose to present the yield curve model we used, our calibration procedure, and to develop the methodology and the obtained results of the sensitivity and uncertainty analysis of operationally-pertinent functions of the reconstructed yield curve, with respect to the “free” parameters.

References:

- [1] D. Brigo, F. Mercurio, “Interest Rate Models – Theory and Practice”, *Springer*, 2007.

[2] M. Rainer, “Calibration of stochastic models for interest rate derivatives”, Optimization **58** (2009).

[Ibrahima Niang; Université Lyon 1, ISFA, 50 avenue Tony Garnier, 69007 Lyon, France]
[niang.brahima@gmail.com –]

Master: Université du Maine (Le Mans)

Ph.D. (10/2012 (*starting date*) — 10/2015 (*expected end*)): The thesis aims at the modelisation of long-term dependence in heavy-tailed processes, and the impact of this dependence in insurance, especially regarding capital allocation and credit risk, in the context of the application of the European directive *Solvency II*.

Supervisor(s): A. Cousin, V. Maume-Deschamps

[Alexandre Janon; Université Lyon 1, ISFA, 50 avenue Tony Garnier, 69007 Lyon, France]
[alexandre.janon@univ-lyon1.fr – <http://isfaserveur.univ-lyon1.fr/~janona/>]

PH.D. STUDENT PRESENTATION

Sensitivity Analysis of DeeSse, a recent Multiple-Point Statistics algorithm

GUILLAUME, PIROT
JULIEN, STRAUBHAAR
PHILIPPE, RENARD

Centre for Hydrogeology and Geothermics, University of Neuchâtel, Switzerland

The Direct Sampling algorithm (DeeSse) [1] is a multiple-point statistical (MPS) tool allowing to simulate a very broad range of multi-variate stochastic fields. Regarding classical MPS techniques, the particularity of DeeSse consists in scanning directly patterns in the training image (TI) rather than building a pattern histogram. In addition to bring a high degree of realism in simulations as for many MPS algorithm, it offers the advantages of dealing with both continuous and discrete variables, and handling multiple variables as well as a wide range of non stationarity constraints.

As DeeSse is a recent algorithm, it seems important to assess its capacities and understand the impact of its main parameters. While previous MPS publications were essentially illustrated by simple structures such as channels, our aims here are to extend the study to a greater variety of patterns and to proceed a systematic performance analyze of the method when changing its main parameters : the acceptance threshold, the number of neighbours, and the scanned fraction of the TI. We analyse the performance in terms of speed of calculation and quality of reproduction of the patterns. These results provide guidelines for the users of DeeSse [2].

For this investigations we use ice wedge polygons, a microscope view of a marble slice and a 3D model of concrete as TI for categorical variables and continuous variables. For each TI, we generate non-conditional simulations exploring the three dimensions of the main input parameters space. By comparing statistical data such as variogram, probability distribution function, connectivity [3] and CPU time between the TI and the realizations, we assess the quality of the realizations in the results space. Then for each evaluation criteria and each TI, we provide a quality map in the 3D space.

The results show that substantial computational gains can be obtained without degrading the quality by reducing the scanned fraction of the TI allows as long as the TI contains enough reproducible patterns. On the other hand increasing the acceptance threshold or reducing the number of neighbours offer better CPU times at the expense of the pattern reproduction quality.

References:

- [1] G. Mariethoz, P. Renard and J. Straubhaar, “The Direct Sampling method to perform multiple-point geostatistical simulations”, *Water Resources Research*, **Vol. 46**, W11536, 2010.
- [2] E. Meerschman, G. Pirot, G. Mariethoz, J. Straubhaar, M. Van Meirvenne, P. Renard, “A practical guide to performing multiple-point statistical simulations with the Direct Sampling algorithm”, *Computers & Geosciences*, **Vol. 52**:307–324, 2013.
- [3] P. Renard and D.Allard, “Connectivity metrics for subsurface flow and transport”, *Advances in Water Resources*, in press.

[Ph.D. Guillaume Pirot; Centre d’Hydrogéologie et de Géothermie, rue Emile Argand 11, CH-2000 Neuchâtel]

[guillaume.pirot@unine.ch – <http://www2.unine.ch/philippe.renard/page-28516.html>]

Master: University of Neuchâtel

Ph.D. (03/2011 (*starting date*) — 02/2014 (*expected end*)): description of the subject (5 lines maximum)

Supervisor(s): J. Straubhaar and P. Renard

PH.D. STUDENT PRESENTATION

Learning with Gaussian Processes with high and Low Accuracy Observations.

FEDERICO, J., ZERTUCHE
Université Joseph Fourier, Grenoble, France

CELINE, HELBERT
École Centrale, Lyon, France

ANESTIS, ANTONIADIS
Université Joseph Fourier, Grenoble, France

The problem studied is the prediction of an output given an input by learning from different types of observations produced by an experiment with adjustable precision.

Each level of precision and the difference between two consecutive levels are modeled by Gaussian Processes with free parameters. The predictor is built by making an assumption about the relationship between two consecutive levels and by integrating the estimations of the parameters by using a plug-in method. This allows for a flexible and easy to use method that integrates different types of observations.

The relationship between two consecutive levels of precision is studied. A linear and a non-parametric method for building the relationship are proposed. In the linear model case when the observations of two consecutive levels are nested, all the parameters estimators and the form of the predictor and error of prediction are fully described.

An E.M. algorithm is proposed to extend the linear model to the case when the observations of two consecutive levels are not nested.

References:

- [1] Peter Z. G. Qian and C. F. Jeff Wu., “Bayesian Hierarchical Modeling for Integrating Low-accuracy and High-accuracy Experiments’.”, *J. Technometrics* , **50**: 192-204, 2008.
- [2] Loic Le Gratiet, “Bayesian analysis of hierarchical multi-fidelity codes. ”, *J. arXiv:1112.5389v2*, 2011.
- [3] Kennedy, M. and O’Hagan, A., “Predicting the output from a complex computer code when fast approximations are available. ”, *J. Biometrika* , **87**: 1-13., 2001.
- [4] Forrester, Alexander I.J., Sbester, Andrs and Keane, Andy J., A., “Multi-fidelity optimization via surrogate modelling. ”, *J. Proceedings of the Royal Society A*, **463**, (2088): 3251-3269., 2007.

[Federico Zertuche; Laboratoire Jean Kuntzmann 51 rue des Mathématiques, BP 53 38041 Grenoble cedex 9 France]

[zertuche.federico@imag.fr – <https://team.inria.fr/moise/federico-zertuche/>]

Master: Université Jean Monnet.

Ph.D. (10 /2011 (*starting date*) — 10/2014 (*expected end*)): The objective of the thesis is to study and implement multi-fidelity methods in the context of extending the life of aging nuclear plants by studying a model of the cooling system of the plant.):

Supervisor(s): Celine Helbert and Anestis Antoniadis

[Celine Helbert, Anestis Antoniadis;]

[Celine.Helbert@imag.fr, Anestis.Antoniadis@imag.fr – <http://www-ljk.imag.fr/membres/Anestis.Antoniadis/>]

PH.D. STUDENT PRESENTATION

Sensitivity Analysis with Application to Multi-Sensor Positioning System

JUEJING ZHAO

University of Applied Sciences Munich, Munich, Germany

Nowadays, Global Positioning System (GPS), which can be widely used for the commercial and military markets, is becoming the most important part of positioning. Although GPS is a very helpful and accurate navigation system, it is difficult to rely much on this technology because of the unavailability of the GPS signal in different areas. The integration of GPS with other positioning systems has been shown that it is part of a solution to overcome this problem. Thus, a multi-sensor system consisting of four sensors (GPS, odometer, accelerometer and gyroscope) is used. Some examples related to the multi-sensor system are given in [4] and [6].

Further to this, an integrated multi-sensor system can be combined with different sensors or different accuracy level of the sensors according to the intended use, for example, navigation systems, advanced driver assistance systems and mobile mapping and so on.

In order to evaluate the performance of the integrated multi-sensor positioning system and for designing a best sensor combination with optimal accuracy level, the variance-based sensitivity analysis is utilized.

This PH.D. thesis deals with processing the acquired data of the measurement vehicle and setting up approaches of Kalman filter approach for fusing the different sensor data of a positioning system in order to compute the 2D-coordinates of the measurement vehicle.

The research focuses on an application of the applied geodesy field: kinematic positioning of vehicles. For this reason a Kalman filter was developed that is capable to integrate different sensors in a common evaluation model. The Kalman filter which is a common sensor fusion algorithm for the determination of the trajectories is used as a model in this thesis. The state quantities of the Kalman filter are the model outputs which are used in sensitivity analysis afterwards. The algorithm was developed in [1] and [5].

To evaluate the performance of the positioning system two variance-based sensitivity analysis are applied: Sobol' and Extended Fourier Amplitude Sensitivity Test (E-FAST), see [2] and [3]. The variance-based sensitivity analysis is a very practical tool for the analysis of the models and optimization of the known models. The relationship between the measurement quantities (input) and the estimated coordinates computed by the Kalman filter (output) is investigated by sensitivity analysis. Another emphasis in this paper is on the comparison between Sobol' method and E-FAST. The differences and the similarity of the two methods are presented. Results obtained using Sobol' is adopted as benchmark for testing the latter method.

The optimization of model for the trajectory estimation is realized by improving the sensor accuracy level and changing the combination of the sensors.

For the application the sensitivity analysis either Sobol' method or E-FAST shows, that how the inputs influence the outputs. Furthermore it is shown that the variance-based sensitivity analysis is well suited to detect the share of the influence of the input quantities on the output quantities, here the estimated positions. Even more important, the changes in the models lead to obvious effects in the respective sensitivity measures. This emphasizes the possibility to optimize the filter models by use of the variance-based sensitivity analysis tools.

References:

[1] A., Eichhorn, " Ein Beitrag zur Identifikation von dynamischen Strukturmodellen mit Methoden der adaptiven KALMAN-Filterung", *Deutsche Geodtische Kommission*, Reihe C, Heft 585, 2005.

[2] A., Saltelli, K., Chan and M., Scott(Eds.), "Sensitivity Analysis", John Wiley and Sons Ltd, 2000.

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

[4] K., Ramm and V., Schwieger, "Multisensorortung fr Kraftfahrzeuge", *In V., Schwieger and K., Foppe (Ed., 2004), "Kinematische Messmethoden Vermessung in Bewegung", DVW Schriftenreihe, Band45*, Winer Verlag, Augsburg, 2004.

[5]T., Aussems, "Positionsschtzung von Landfahrzeugen mittels KALMAN-Filterung aus Satelliten- und Koppelnavigationsbeobachtungen", *Verffentlichungen des Geodtischen Instituts der Rheinisch-Westflischen Technischen Hochschule Aachen*, Nr. 55, 1999.

[6] V., Schwieger, R., Schollmeyer, and K., Ramm, "Integrated Positioning by MOPSY", *Proceedings on 5th European Congress and Exhibition on Intelligent Transport Systems and Services*, Hannover, 01.-03.06. 2005.

[Juejing Zhao; University of Applied Sciences Munich, Karlstrasse 6, 80333, Munich]

[jzhao@hm.edu – www.hm.edu]

Master: University Stuttgart

Ph.D. (07/2010 — 07/2013): The research focuses on variance-based sensitivity analysis with applications of the applied geodesy field: kinematic positioning of vehicles. Kalman filter approach is used for fusing the different sensor data of a positioning system and producing the output quantities for the sensitivity analysis. Finally to design a best sensor combination with optimal accuracy level by using variance-based sensitivity analysis tools.

Supervisor(s): C. Tiede and V. Schwieger

PH.D. STUDENT PRESENTATION

A quality measure for comparing different feature deviations to perform sensitivity analysis in tolerancing

PHILIPP, ZIEGLER

Chair of Engineering Design, Friedrich-Alexander-University Erlangen-Nuremberg, Germany

SANDRO, WARTZACK

Chair of Engineering Design, Friedrich-Alexander-University Erlangen-Nuremberg, Germany

During manufacturing in mechanical industry, the real product geometry always varies from its nominal geometry. Additional varying product geometry of assemblies arises from variations in the interface positions between its components (e.g. fixture points). The products geometry is closely related to the functionality of the product. To ensure the functionality, functional product requirements have to be translated into functional geometric requirements of the product, which are restrictions for the appearing geometrical variations. Tolerance management is the process of restricting and analyzing the products variations to ensure the product meeting the functional geometric requirements [1]. Henceforth, basic geometry elements (planes, axis, etc.) are called features. The functional geometric requirements are restrictions to particular features (e.g. the axis of a cylinder) to a reference (e.g. the surface of a flange). The restricted feature is called functional key characteristic (FKC). Geometric product variations are composed from a lot of single variations, due to the separate manufacturing and assembly steps. The considered varying features of the product form a tolerance chain. An important task of tolerance management today is to identify the main contributing deviations (resulting in variations of the FKC) and to control them.

Clearance deviations of the component interface points will be not considered in the following. The mathematical representation of tolerances is a challenging task because of the (historically) grown complexity of the tolerance formulation. Component deviations can be represented by small orientation and position variations of the nominal component features coordinate systems. These variations can be represented by the small displacement torsor (SDT) [2]. The SDT has up to six entries (e.g. three for a plane: one translation, two rotations), which can be seen as the degrees of freedom (DOF) of the feature, as seen in figure 1. Most math models for tolerances base on restrictions of feature SDT's, although there are also different approaches [3] (which just differ in detail). The deviation domain approach by GIORDANO [3] is used here as spatial math model for representing the geometry deviation.

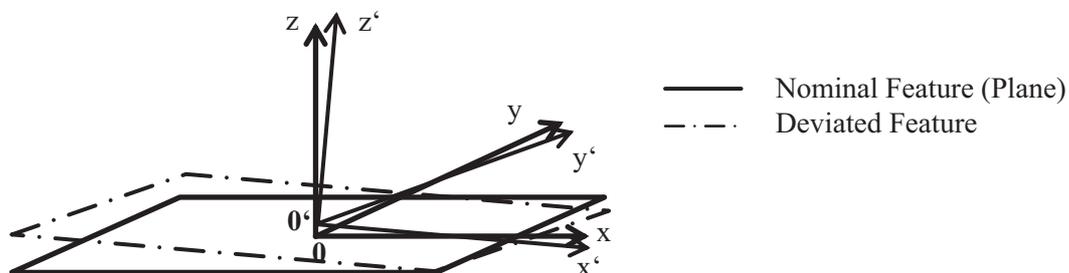


Figure 1: Geometric variation of a plane, represented with the small displacement torsor

If the deviation of a feature is restricted by geometric tolerances (e.g. parallelism, position), this has to be translated into restrictions for its SDT. The restrictions for individual DOFs are in general dependent on each other. The functional geometric requirements in this context are also restrictions for

the SDTs of the FKCs. Adopting statistical tolerance analysis means to sample the SDTs of single geometry elements and to check if the FKC holds the restrictions.

In this context, for performing variance-based sensitivity analysis [4], a major problem arises: Every FKC has up to six dependent parameters (its degrees of freedom (DOF)) and to have sensitivities for all these parameters would be bulky. Therefore, a quality measure for the FKC-SDT is presented in the following to have a measure which evaluates the total displacement quality of the FKC-SDT.

The restrictions of the FKC (the functional geometric requirements) can be represented by the volume Ω_{FKC} in the SDT parameter space. This volume is convex and includes the 0. The FKC-SDT of an assembly is a point in the parameter space. If an assembly fulfills the geometric functional requirements, the FKC-SDT is included in the volume, otherwise not. The idea is, to shrink (or inflate in the second case) the volume until the FKC-SDT is just in the volume, where the scaling parameter of the volume is the quality measure

$$\lambda(x) = \min \{ \mu \mid x \in \mu \cdot \Omega_{FKC}, \mu > 0 \}.$$

The quality measure is nonnegative, zero if the FKC has the nominal form (with respect to the datum), between 0 and 1 if the functional geometric requirements are fulfilled and larger than 1 otherwise. For tight restricted SDT parameters it is sensitive to small variations (fig. 2 $\delta\theta$), for broadly restricted ones it is less sensitive (fig. 2 δz). The two deviated surfaces in figure 2 fulfill the tolerance requirement, so they are inside the deviation domain. They both have $\lambda=0.7$, the skewed because of the parallelism tolerance with datum A.

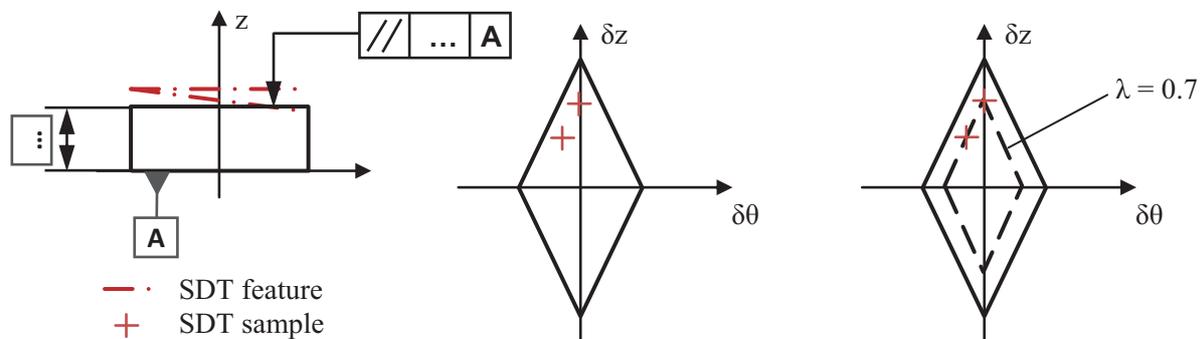


Figure 2: Quality measure for a two dimensional line with parallelism tolerance to datum A

The quality measure can also be used for all deviating features of the assembly. The main aim is to develop a feature based sensitivity analysis method. Actual sensitivity analysis in tolerancing is based on distributions of single parameters (or even just their range), so one feature can have up to six (!) sensitivity values. With this parameter reduction, the number of sensitivity indices is manageable. Following, a group sampling method for multiple SDT parameters of the features will be developed, which handles parameter areas (figure 2). Also interface deviations are considered in a next step.

References:

- [1] J.-Y. Dantan, N. Anwer and L. Mathieu, "Integrated Tolerancing Process for conceptual design", *CIRP Annals*, **52**: 135-138, 2003.
- [2] P. Bourdet, L. Mathieu, C. Lartigue, A. Ballu, "The concept of the small displacement torsor in metrology", *Advanced Mathematical Tools in Metrology II*, Edited by Work Scientific Publishing Company, Series Advances in Mathematics for Applied Sciences, **40**, 110-120, 1996
- [3] G. Ameta, S. Serge and M. Giordano, "Comparison of Spatial Math Models for Tolerance Analysis: Tolerance-Maps, Deviation Domain, and TTRS", *Jour. of Comp. and Inf. Sc. in Eng.*, **11**(2): 021004, pp. 8, 2011.
- [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 and Sons, England, 2008.

[Philipp Ziegler; Chair of Engineering Design, University of Erlangen-Nuremberg]

[ziegler@mfk.fau.de - www.mfk.fau.de]

Diplom: University of Erlangen-Nuremberg

Ph.D. (09/2011 | 09/2016): Statistical methods in tolerancing

Supervisor: Prof. Sandro Wartzack