Real-Time Optimization of Speed Control To Limit Train Energy Consumption

R. Jorge Do Marco[†,1,2], G. Perrin[§,1], C. Funfschilling[2], C. Soize[§,3]

[†] PhD student (presenting author). [§] PhD supervisor

PhD expected duration: Dec. 2022 – Dec. 2025

[1] Université Gustave Eiffel, COSYS, 5 Bd Descartes, 77454 Marne-La-Vallée, France
guillaume.perrin@univ-eiffel.fr

[2] SNCF, Direction Technologies, Innovation et Projets Groupe, 1-3 Av. François Mitterrand, 93574 Saint-Denis, France
{romain.jorge-domarco, christine.funfschilling}@sncf.fr

[3] Université Gustave Eiffel, MSME UMR 8208, 5 Bd Descartes, 77454 Marne-La-Vallée, France
christian.soize@univ-eiffel.fr

Abstract

Reducing the ecological footprint has become a priority for many industries, including the railway industry which is one of the most energy-intensive industries. Indeed, this industry faces two paradoxical needs ; on the first hand, it must decrease its energy consumption, meeting both an environmental goal and a financial objective, and on the other hand, it must not only maintain but increase the circulation of train, thus allowing a larger part of the population to use the most ecological means of land transport. In order to respond to this double problem, multiple actions can be undertaken, and the present work focuses on the driver control. Historically, this problem has been handled using speed profile along the track [2]. Although this concept may be relevant as it allows one to avoid modelling control law, it is not possible to use such profile directly in industry as the speed is not an input to the train.

In addition to the previous goal, which can be described as the *minimization* of the energy consumption, the driver must respect multiple *constraints* like *punctuality*, *safety* and *comfort*. As the dynamic and energy behaviors rely on non-linear differential equations, the problem can be considered as a black-box one. Finally, *uncertainties* exist both in the input parameters of the train and in the equations which are simplifications of the physic. This problem, included in the family of the black-box constrained optimization under uncertainties, has been treated previously and it has been shown that for each trip about 25% of energy could be saved [3]. The main weakness is the time needed to optimize the driver command, as it could require multiple days to converge. Accelerating this resolution in order to be able to adapt the control commands in *real time* is a major challenge of the work presented here.

Beyond a "simple" immediacy and in a first time, real-time will allow the control to use the *measures* continuously acquired by the multiple sensors placed on the train notably in the framework of an *automatic train operation*. Using these new data, it would be possible to refine the model parameters as the train moves along the railway track and therefore better adjust the commands. Some environmental parameters, such as wind speed, which could only be predicted in previous work, could also be better considered in the method. Moreover, real-time is expected to allow us to handle unexpected one-off events such as slowdowns or late departures to ensure optimal energy consumption in any context.

As the system under study is the train, the notion of real time is linked to characteristic calculation time, the magnitude of which is approximately 1 µs compared to the initial state of art, the magnitude of which is approximately 1 day.

To this end, this work will focus on building a new solver which will take well-represented input allowing fast calculation. Starting from a refined expensive optimization, it will therefore be possible to make real time adjustments. The whole method will use *experimental data* to update parameters.

From this use case, this work develops an original method for the real-time optimal control under uncertainties that is able to learn from experimental measures to adapt to random phenomena. Finally, this work could also be used to propose a new calculation of the regulatory margin using a *multi-objective optimization* between *punctuality* and *energy consumption* ; i.e. a journey could be optimized considering that the travel time is not a constraint but a new objective [1].

Short biography (PhD student)

Romain Jorge Do Marco is currently an engineering PhD student at Gustave Eiffel University. He is supervised by Prof. Guillaume Perrin, Prof. Christian Soize and Dr. Christine Funfschilling. His thesis focuses on the real-time optimization under uncertainties of control for train to limit energy consumption. The work is supported by the French railway company SNCF.

References

[1] Adrián Fernández-Rodríguez, Antonio Fernández-Cardador, Asunción P. Cucala, María Domínguez, and Tad Gonsalves. Design of robust and energy-efficient ato speed profiles of metropolitan lines considering train load variations and delays. *IEEE Transactions on Intelligent Transportation Systems*, 16(4):2061–2071, 2015.

[2] Daniel Lancien and Michèle Fontaine. Calcul de marches de trains économisant l'énergie de traction - le programme MAREO. *Revue générale des chemins de fer*, pages 1–14, 1981.

[3] Julien Nespoulous, Christian Soize, Christine Funfschilling, and Guillaume Perrin. Optimisation of train speed to limit energy consumption. *Vehicle System Dynamics*, 60(10):3540–3557, 2022.

On optimal predictive maintenance planning informed by data-driven prognostics

A. Kamariotis[†,1], K. Tatsis[2], K. Goebel[3], E. Chatzi[§,2], D. Straub[§,1]

[†] PhD student (presenting author). [§] PhD supervisor

PhD expected duration: May 2019 – Apr. 2023

[1] Engineering Risk Analysis Group, Technical University of Munich
{antonis.kamariotis,straub}@tum.de

[2] Department of Civil, Environmental and Geomatic Engineering, ETH Zurich
{tatsis,chatzi}@ibk.baug.ethz.ch

[3] Palo Alto Research Center (PARC)
kgoebel@parc.com

Abstract

Engineering components/systems are subject to degradation of various forms that can adversely affect their intended performance. In an effort to optimize associated maintenance, monitoring information can be harnessed via the deployment of sensors on the degrading components/systems. An optimal maintenance plan should optimally balance the risk of failure and the benefit of extending the life-cycle of the component/system. The maintenance plan needs to determine the type of repair action as well as the time for performing a preventive repair action.

Prognostics can support such tasks by delivering an estimate of the remaining useful life (RUL) of a degrading component/system utilizing monitoring data [3]. In the modern data-rich engineering landscape, prognostics solutions are commonly based on the use of data-driven methods, e.g., artificial intelligence (AI) approaches or statistical process-based approaches. Data-driven prognostic schemes have paved the way for establishment of the data-driven predictive maintenance (PdM) paradigm [3]. Multiple sources of uncertainty enter the process of prognostics. Therefore, PdM can be formulated as a problem for decision-making under uncertainty.

In this work, PdM decision policies are discussed for planning preventive replacement actions. These rely on input from data-driven prognostic algorithms. We employ two distinct PdM policies of varying complexity, and we rigorously investigate them in terms of optimality of the resulting decisions. The first is a simple heuristic PdM policy, which operates on the basis of prognostics outcomes that are expressed in the form of probability of RUL exceedance in certain decision-relevant time intervals. The second investigated PdM policy seeks to account for the full probability distribution function (PDF) of the RUL predictions, and is somewhat more complex. Both PdM policies are evaluated via the quantification of the long-run expected maintenance cost per unit time [2]. Based on this quantification, we propose a metric that assesses data-driven prognostic algorithms according to the PdM decisions that are triggered by their outcome. This decision-oriented metric can provide the basis for a paradigm shift when it comes to optimizing the training process of prognostic algorithms (e.g., hyperparameter tuning) from a prediction-based training approach to a decision-based one.

We numerically investigate the above on an actual case study by employing a widely used prognostic data set related to degrading turbofan engines [1]. Following the flowchart presented

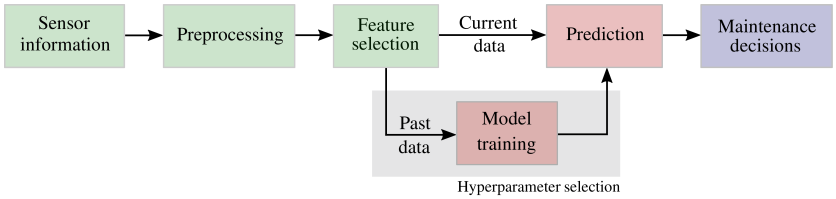


Figure 1: Flowchart of the adopted data-driven PdM decision process

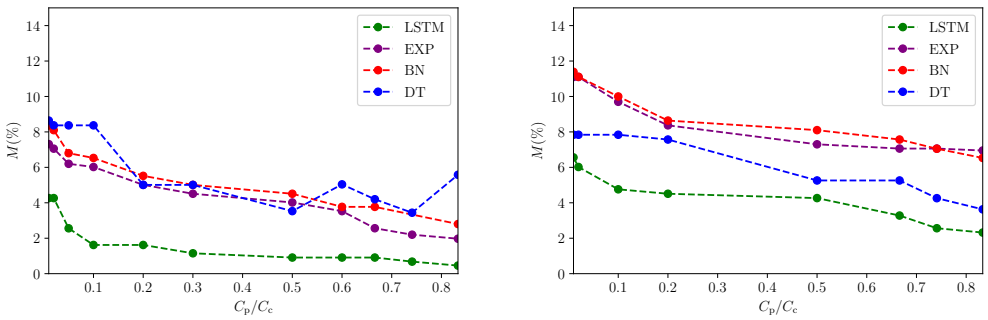


Figure 2: Evaluation of the decision-oriented metric M for different cost ratios. Left panel: simple heuristic PdM policy. Right panel: PdM policy which accounts for the PDF of the RUL predictions.

in Fig. 1, we train data-driven prognostic algorithms of different nature on this data set for performing RUL predictions, which we subsequently provide as input to the two PdM policies mentioned above. The four investigated models are: i) Long Short-Term Memory (LSTM) networks for classification, ii) Bayesian Networks (BN) classifiers, iii) Decision Trees (DT) classifiers and iv) Bayesian filtering of an exponential degradation model (EXP) for regression. Fig. 2 plots the values of the decision-oriented metric M that we obtain for different assumed C_p/C_c cost ratios with each of the two investigated PdM policies. C_p denotes the cost of a preventive replacement action, while C_c is the cost of a corrective replacement action, with $C_c > C_p$. A smaller value of the metric M is preferable. For this data set, and for the specific prognostic models that we implement, the heuristic PdM policy is found to lead to better decisions than the alternative more involved PdM policy. The LSTM prognostic classifier seems to deliver the best performance with respect to PdM decision-making on preventive replacement planning.

Short biography (PhD student)

Antonios received his Dipl. in Civil Engineering (2016) from the Aristotle University of Thessaloniki, and his MSc in Computational Mechanics (2019) from TU Munich. He is currently doing his PhD on "Decision support with structural health monitoring", in a collaborative project between TU Munich and ETH Zurich, funded by the TUM Institute for Advanced Study.

References

[1] NASA Ames Prognostics Data Repository, NASA Ames Research Center. https://www.nasa.gov/content/prognostics-center-of-excellence-data-set-repository.

[2] P. Do, A. Voising, E. Levrat, and B. Iung. A proactive condition-based maintenance strategy with both perfect and imperfect maintenance actions. *Reliability Engineering & System Safety*, 133:22–32, 2015.

[3] O. Fink, Q. Wang, M. Svensén, P. Dersin, W.-J. Lee, and M. Ducoffe. Potential, challenges and future directions for deep learning in prognostics and health management applications. *Engineering Applications of Artificial Intelligence*, 92:103678, 2020.

Robustness Analysis for Uncertainty Quantification by Optimization on Riemannian Manifolds

B. Ketema[†,1,2], F. Costantino[§,1], F. Gamboa[§,1],
R. Sueur [§,2], N. Bousquet[§,2], B. Iooss [§,1,2]

[†] PhD student (presenting author). [§] PhD supervisors

PhD expected duration: Oct. 2022 – Sept. 2025

[1] Institut de mathématiques de Toulouse, Paul Sabatier University,
{baalu-belay.ketema,francesco.costantino,fabrice.gamboa}@univ-toulouse.fr

[2] EDF R&D,
{roman.sueur,nicolas.bousquet,bertrand.iooss}@edf.fr

Abstract

EDF wishes to study the sensitivity of a real-valued numerical code G, coming from the domain of nuclear safety, where the inputs $X = (X_1, ..., X_d)$ are uncertain physical quantities that are seen as random variables. This procedure aims to justify that a certain quantile of the output $Y = G(X)$ does not go beyond a certain regulatory threshold [2]. The specificity of this problem is that the input distribution of X is itself uncertain. Our goal will be to perform "perturbations" on the input X of G and compute robustness indices on a quantile of Y to understand how robust the quantile is to "perturbations" of X. The notion of "perturbation" will be defined using Information Geometry i.e. using the distance induced by the Fisher matrices on the parameter set of possible distributions of X. More precisely, our goal is to identify parameters corresponding to the distributions that maximize (or minimize) the robustness indices for a given level of perturbation.

This is an optimization problem on a set of probability distributions that, when endowed with the Fisher information matrices, becomes a Riemannian manifold. In addition, due to the complexity of the numerical code G, the distribution of Y cannot be determined analytically therefore we can only estimate it with some statistical sampling method. To sum up, we are dealing with a stochastic optimization problem on a Riemannian manifold.

The one dimensional case of the Gaussian distribution has been studied in [1] as well as during my master's internship. We built an algorithm based on the usual Riemannian gradient descent method that we adapted to the specific case of our problem. We applied this algorithm to optimize toy functions on a given closed ball, see Figure 1 and 2 for illustration. For what follows, our goal is to study more general families of distributions like the Gaussian distribution in higher dimensions or more generally, exponential families, and try to identify the optimization algorithms that are more suited to our problem.

In this communication, I would first like to explain briefly where our mathematical problem comes from and why we decided to formulate it in this particular way. Then, I will briefly present what I have previously been working on during my master's degree internship where we have made moderate simplifications to the initial mathematical problem to give ourselves an achievable goal. Then, I will explain what we were able to do in the first few months of PhD and lastly, I will mention the short and medium-term objectives that we plan to achieve.

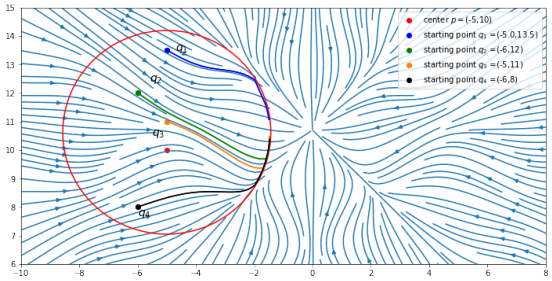


Figure 1: Descent trajectories for $f(z) = 5\cos(x)\sin(y) + x^2 + (y-10)^2$ on a closed ball.

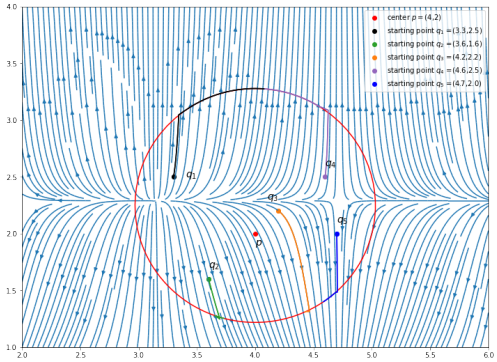


Figure 2: Descent trajectories for $h(z) := \sin(x)^2 + y^2\sin(y)$ on a closed ball.

Short biography (PhD student)

Formerly a student of Paul Sabatier University in Toulouse, I did my bachelor's and master's degree in mathematics in research specializing in probability, statistics and optimization. I did my master's internship in the R&D division of the company EDF, while being supervised by researchers from both EDF and the mathematics lab of Toulouse IMT. Now, I am a first year PhD student and this PhD thesis is a continuation of my master's internship and our goal is to perform robustness analysis for uncertainty quantification using information geometry. I am funded by the Labex CIMI (Toulouse) for the first few months of my PhD and by EDF for the remaining duration.

References

[1] Clement Gauchy, Jerome Stenger, Roman Sueur, and Bertrand Iooss. An information geometry approach to robustness analysis for the uncertainty quantification of computer codes. *Technometrics*, 64(1):80–91, 2022.

[2] Bertrand Iooss and Amandine Marrel. Advanced methodology for uncertainty propagation in computer experiments with large number of inputs. *Nuclear Technology*, 2019.

Multi-output surrogate modeling for inverse uncertainty quantification in random neutronics

P. Lartaud[1,2], P. Humbert[2], J. Garnier[1],

PhD expected duration: Oct. 2021 – Sep. 2024

[1] CMAP, CNRS, École polytechnique, Institut Polytechnique de Paris, 91120 Palaiseau, France
{paul.lartaud,josselin.garnier}@polytechnique.edu

[2] CEA DAM-DIF, Bruyères-le-Châtel 91297 Arpajon, France
philippe.humbert@cea.fr

Abstract

Neutron noise analysis [4] describes a set of techniques used in nuclear safeguards and nuclear security to identify an unknown fissile material based on the temporal correlations of the detected neutrons. The observations are often noisy due to the stochastic nature of the underlying physical processes, which makes the resolution of this inverse problem complex. Moreover, the uncertainty quantification of the estimation of the fissile material parameters has not been studied thoroughly, though it is crucial for such an ill-posed inverse problem. On top of this, the analytical direct model widely used to describe neutron correlations is based on strong physical assumptions, which are never met in a real-world scenario. The impact of these simplifications should be discussed and quantified. Overall, the knowledge of the quality of the predictions is crucial for the decision makers.

This work addresses dual objectives. First of all, surrogate models specifically designed for multiple correlated outputs are built to extend and improve the biased analytical model currently used, while providing a reliable quantification of the errors associated with the predictions. The surrogate models are based on multi-fidelity [6], multi-output Gaussian process regression. Since the observations are strongly correlated, the Gaussian processes for the individual outputs must not be taken independent. Instead, a Linear Model of Coregionalization [2] and Convolutional Gaussian processes [1] are investigated to build the surrogate model. The performance of the surrogate models are evaluated with different metrics, and the coverage probabilities are estimated to guarantee the reliability of the predictions and their uncertainty quantification.

In order to cope with the inherent limitations of Gaussian process surrogate models, this work also investigates other approaches to build efficient surrogate models such as Bayesian Neural Networks [7].

Finally, the inverse problem is solved with a Bayesian approach which takes into account the covariances of the measurements and the predictive means and covariances of the Gaussian process surrogate model. Thus, the general methodology presented in this work takes into account both sources of error, measurement noise and model bias [5]. The resulting posterior distribution more accurately reflects the inferred knowledge about the material properties. The Bayesian inference also allows to effortlessly include additional information (gamma correlations, expert knowledge) into the posterior distribution. To highlight the improvements brought up for neutron noise analysis applications, this method is applied to a specific test case coming from a neutron multiplicity benchmark [3].

Short biography

I am a second-year PhD student working at CEA DAM Île-de-France and École polytechnique under the supervision of Josselin Garnier. This work is funded by the CEA and aims at providing robust uncertainty quantification methods for the identification of nuclear materials.

References

[1] Mauricio A Alvarez and Neil D Lawrence. Computationally efficient convolved multiple output gaussian processes. *The Journal of Machine Learning Research*, 12:1459–1500, 2011.

[2] Edwin V Bonilla, Kian Chai, and Christopher Williams. Multi-task gaussian process prediction. *Advances in neural information processing systems*, 20, 2007.

[3] J Blair Briggs, Lori Scott, and Ali Nouri. The international criticality safety benchmark evaluation project. *Nuclear science and engineering*, 145(1):1–10, 2003.

[4] Richard Phillips Feynman, F De Hoffmann, and Robert Serber. Dispersion of the neutron emission in U-235 fission. *Journal of Nuclear Energy (1954)*, 3(1-2):64–IN10, 1956.

[5] Paul Lartaud, Philippe Humbert, and Josselin Garnier. Multi-output gaussian processes for inverse uncertainty quantification in neutron noise analysis. *arXiv preprint arXiv:2211.02465*, 2022.

[6] Loic Le Gratiet and Josselin Garnier. Recursive co-kriging model for design of computer experiments with multiple levels of fidelity. *International Journal for Uncertainty Quantification*, 4(5), 2014.

[7] Radford M Neal. *Bayesian learning for neural networks*, volume 118. Springer Science & Business Media, 2012.

Chained Gaussian processes to estimate battery health degradation with uncertainties

B. Larvaron[†,1,2], M. Clausel[§,1], A. Bertoncello[§,2], S. Benjamin[3], G. Oppenheim[4]

[†] PhD student (presenting author). [§] PhD supervisor

PhD expected duration: May. 2021 – May. 2024

[1] Université de Lorraine, CNRS, IECL, Nancy, France
{benjamin.larvaron,marianne.clausel}@univ-lorraine.fr

[2] TotalEnergies OneTech, 91120 Palaiseau, France
antoine.bertoncello@totalenergies.com

[3] SAFT, 33074 Bordeaux, France
sebastien.benjamin@.saftbatteries.com

[4] Université Paris-Est-Marne la Vallée, Département de Mathématiques, Marne-la-Vallée, France
georges.oppenhein@gmail.com

Abstract

Thanks to falling costs and high physical performances, Lithium-ion batteries are currently the solution of choice for many applications, including the thriving markets of electric vehicles and smart grids. However, electric batteries age with time. Depending on their composition and conditions of use, their performance degrades until they are considered unfit for their designed use. For manufacturers, knowledge of this aging process is a essential issue [1], since it is a pivotal element in determining its price.

To quantify the aging behavior of a new kind of battery, manufacturers perform a series of aging tests in controlled experimental conditions on a batch of batteries. These tests are expensive and time-consuming, so only few batteries are generally tested. The degradation is quantified thanks to state of health (SoH) indicators, typically capacity or internal resistance. Fig.1 displays an example of such data, from a dataset provided by the Aachen university [2].

Our objective is to predict the average degradation of the state of health. As in the example of the Aachen dataset, we consider a setting of identical batteries, tested in similar experimental conditions, such that we characterize the typical degradation of a battery design at this experimental condition. Moreover, as an essential point, predictions have to include an uncertainties quantification. This is particularly important for manufacturers in order to assess the financial risk related to performance guarantees.

Related to uncertainties, a significant phenomenon in battery health degradation is the inter-battery variability. Considering a batch of batteries with a similar design and cycled at identical conditions, there is often an important range of SoH at a fixed time. Besides, we can see in Fig. 1 that this uncertainty increases with time: at cycle 100, the capacity range is approximately 1% of the nominal capacity, whereas at cycle 1300, it is close to 20%. This effect is often observed by manufacturers and is not specific to this dataset. The main challenge in this work is to model the inter-battery variability accurately.

Gaussian processes have already been used in this context, see for example [3]. This approach is well suited to the problematic; as a non-parametric Bayesian method, Gaussian processes can

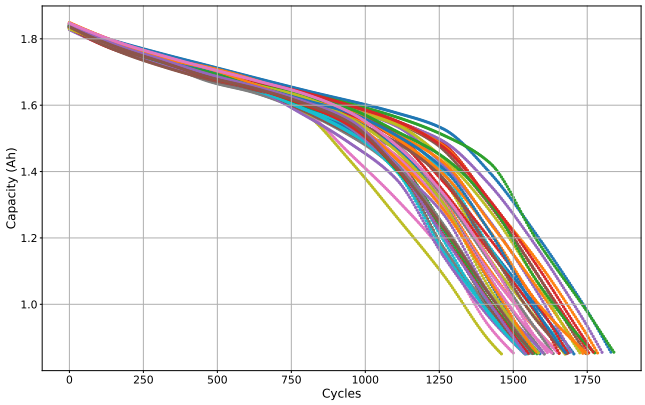


Figure 1: Capacity degradation curves of the Aachen dataset

fit complex functions while naturally providing confidence intervals. Moreover, they are known to perform well even when little data are available which is generally the case.

However, Gaussian processes regression often imposes constraints on uncertainties with a stationary prior, which supposes a constant variance. In our context, this is too restrictive since, as we have seen, the variance of the state of health indicator commonly increases with time. That is why we proposed an extended model relying on the Chained Gaussian processes framework [4], an approximated model based on variationnal inference. Coupling several Gaussian processes, this model allows a simultaneous estimate of the time evolution of the mean and the variance.

Short biography (PhD student)

Before starting my PhD, I studied statistics at ENSAI near Rennes, with the "Genie statistique" specialization, turned toward industrial applications. After an end-of-study internship at Framatome. I started a CIFRE thesis with TotalEnergies and Lorraine university, related to the needs of SAFT, a battery manufacturer, and a subsidiary of TotalEnergies.

References

[1] Xiaosong Hu, Le Xu, Xianke Lin, and Michael Pecht. Battery lifetime prognostics. *Joule*, 4(2):310–346, 2020.

[2] Weihan Li, Neil Sengupta, Philipp Dechent, David Howey, Anuradha Annaswamy, and Dirk Uwe Sauer. One-shot battery degradation trajectory prediction with deep learning. *Journal of Power Sources*, page 230024, 2021.

[3] Robert R Richardson, Michael A Osborne, and David A Howey. Gaussian process regression for forecasting battery state of health. *Journal of Power Sources*, 357:209–219, 2017.

[4] Alan D Saul, James Hensman, Aki Vehtari, and Neil D Lawrence. Chained gaussian processes. In *Artificial Intelligence and Statistics*, pages 1431–1440. PMLR, 2016.

3D Principal Component Analysis to characterize geological fields and their influence on seismic ground motion

F. Lehmann[†,1,2], F. Gatti[§,2], M. Bertin[§,1], D. Clouteau[§,2]

[†] PhD student (presenting author). [§] PhD supervisor

PhD expected duration: Oct. 2021 – Sep. 2024

[1] CEA/DAM/DIF, Arpajon, France
michael.bertin@cea.fr

[2] LMPS - Laboratoire de Mécanique Paris-Saclay, Université Paris-Saclay, CentraleSupélec, ENS Paris-Saclay, CNRS, Gif-sur-Yvette, France
{fanny.lehmann,filippo.gatti,didier.clouteau}@centralesupelec.fr

Abstract

Thanks to the continuous increase of computational resources, high-fidelity numerical simulations are becoming more affordable, in seismology as in other fields. Despite the amount of work devoted to optimizing numerical schemes, their computational cost in complex settings prohibits uncertainty quantification analyses. Therefore, alternative methods are needed to explore the uncertainties in the simulation parameters and, more importantly, their influence on the simulation results.

This study uses a High-Performance Computing (HPC) code based on the Spectral Element Method, SEM3D, to solve the 3-dimensional (3D) wave propagation equation [4]. We focus on the properties of the propagation domain (*i.e.* the wave velocities) as the primary source of uncertainty. Indeed, geological parameters have a crucial influence on the generated ground motion. Furthermore, they are described by a large number of variables (10^4 to 10^8).

The geological uncertainties were modeled with random fields. A von Karman correlation model was chosen and several correlation lengths and coefficients of variation were explored to build a database of 100,000 3D heterogeneous geological fields. Our database is publicly available at https://doi.org/10.5281/zenodo.6983053. Due to the curse of dimensionality inherent to the three dimensions, more than 32,000 variables are needed to describe each sample of the geological database. Therefore, reducing this number of variables is necessary to characterize the relationship between the geological parameters and the surface ground motion.

We compared the performances of the Principal Component Analysis (PCA) and an auto-encoder neural network named 3D UNet [1] to perform the dimensionality reduction. Although the PCA treats 3D fields as 1D vectors, the PCA basis shows robust spatial coherency. To reflect that, we called the basis elements *eigengeologies*. We also found that the *eigengeology* basis was expressive enough to represent geological fields very different from those in the original database, for example, 3D domains containing sedimentary basins. Despite the greater complexity of the 3D UNet neural network, we could not reach a larger dimensionality reduction than the PCA. In addition, the latent space elements (analogs of the PCA basis) were less interpretable.

Concerning the reduction percentage, 1000 PCA components were sufficient to reconstruct 3D geological fields with very good accuracy. The Root Mean Square Error was 90m/s, corresponding to 8.6% (resp. 2%) of the minimum (resp. maximum) velocity value [3]. To study the

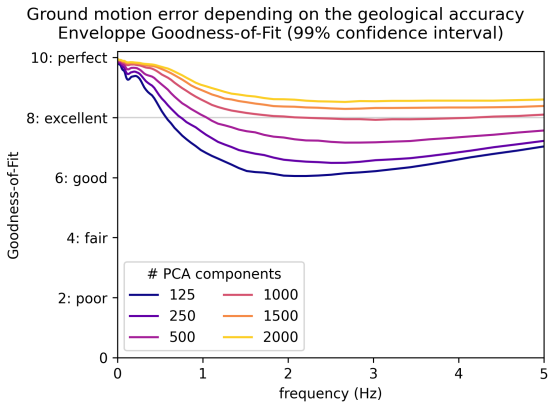


Figure 1: 99th percentile of the frequency-dependent GOF for the East-West velocity recorded by 256 virtual sensors located at the surface of 250 geological fields. For each field, the GOF is assessed between the reference ground motion and the one obtained after the PCA reconstruction with the specified number of PCA components. Computational cost of the HPC simulations: 96,000h CPUs.

impact of the PCA on ground motion, we randomly chose 250 geological fields and simulated the propagation of seismic waves through each of them. Then, each geological field was decomposed on the *eigengeology* basis (with 125 to 2000 PCA components). By propagating seismic waves through each reconstructed field, we showed that the ground motion was very close to the reference one when using more than 1000 PCA components. Figure 1 indeed depicts GOF larger than 8 for all frequencies.

In conclusion, the PCA led to a dimensionality reduction factor larger than 32 for all 3D geological fields while ensuring a good reconstruction accuracy for i) the geological field and ii) the resulting ground motion. This shows that the PCA components can be used instead of the entire 3D fields to characterize ground motion in future uncertainty analyses.

Short biography (Fanny Lehmann)

I graduated from the mathematics department of the ENS Paris-Saclay in 2021. I am doing my PhD at CEA DAM, which partners with academic institutions like Centrale Supélec to perform seismic hazard analyses. This PhD builds on several years of development of the HPC code SEM3D to design a meta-model enabling fast and accurate predictions of seismic ground motion. The meta-model will then be used to investigate site effects for seismic hazard analyses.

References

[1] Özgün Çiçek et al. "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*. Ed. by Sebastien Ourselin et al. Vol. 9901. Springer International Publishing, 2016.

[2] Miriam Kristeková, Jozef Kristek, and Peter Moczo. "Time-Frequency Misfit and Goodness-of-Fit Criteria for Quantitative Comparison of Time Signals". In: *Geophysical Journal International* 178.2 (Aug. 2009). DOI: 10.1111/j.1365-246X.2009.04177.x.

[3] Fanny Lehmann et al. "Machine Learning Opportunities to Conduct High-Fidelity Earthquake Simulations in Multi-Scale Heterogeneous Geology". In: *Frontiers in Earth Science* 10 (2022). DOI: 10.3389/feart.2022.1029160.

[4] Sara Touhami et al. "SEM3D: A 3D High-Fidelity Numerical Earthquake Simulator for Broadband (0–10 Hz) Seismic Response Prediction at a Regional Scale". In: *Geosciences* 12.3 (2022), p. 112. DOI: 10.3390/geosciences12030112.

Multifidelity approaches for solving inverse problems relying on computer codes with functional outputs - Application to thermal evaluation of buildings.

H. Nasser

Gustave Eiffel University

Supervisor(s): G. Perrin (Gustave Eiffel University, IMSE, 5 Bd Descartes - 77447 Marne-la-Vallée, France), Julien Waytens (Gustave Eiffel University, IMSE, 5 Bd Descartes - 77447 Marne-la-Vallée, France), R. Chakir (Gustave Eiffel University, IMSE, 5 Bd Descartes - 77447 Marne-la-Vallée, France), S. Demeyer (LNE, 1 r Gaston Boissier, 75015 Paris, France).

PhD expected duration: October 2022 – October 2025

Address: 9 rue Paul Belmondo, 93160 Noisy-Le-Grand, France

E-mail: hadi.nasser@univ-eiffel.fr

Abstract

In a context of energy and environmental renovation, major advances are expected and necessary in the building sector. For existing buildings, the reduction of energy consumption requires a better assessment of the energy performance of buildings and its improvement through rehabilitation actions. Particular attention must be paid to the evaluation and in-situ control of the thermal performance of buildings before and after a rehabilitation action in order to prevent any defects and thus obtain a building with the expected performance. The thermal insulation of buildings is a key factor to guarantee the thermal comfort of the occupants and to achieve high-energy savings. Thus, the control of the thermal resistance of walls (indicator of the insulation level) by in-situ measurement on existing and new buildings or during renovation is a growing demand. Although many of the techniques in the literature are at the building scale [1-2], we have chosen to focus on the wall scale in order to propose a solution that is easily deployable, minimally intrusive and low sensitive to external weather conditions.

We propose an interdisciplinary technical solution combining physical modeling, statistical modeling, numerical simulation and measurements for a better in-situ characterization of the energy performance of conventional and bio-based walls.

To identify the thermal resistance of highly insulated and bio-based walls, both experimental and numerical works will be investigated. Concerning the experimental part, an active solicitation on the inside face of the wall is retained to perform an accelerated identification (within 72h) and to limit the influence of weather conditions. Physical modelling and numerical simulations are of particular interest for the “Optimal Experimental Design” stage [3] and for the identification process using inverse modelling techniques. As we observed in the previous ANR project named “RESBATI”, simplified physical model like RC or 1D heat equation are not suitable for an accurate identification of the thermal resistance for highly insulated walls and for some wall typologies [4, 5]. It is due to transverse heat fluxes in the wall which are not taken into account in these models.

Consequently, to extend previous works to highly insulated and bio-based walls, more sophisticated physical models are to be considered. We will study 2D/2D axi-symmetrical thermal simulations and hygro-thermal physical models. Moreover, the computational time being more expensive for these direct simulations, we propose to use Reduced Basis Techniques to ensure real-time simulations. We focus on non-intrusive reduced basis methods which are particularly adapted to interface with existing and commercial softwares. Then, we will take advantage of the real-time simulations in a Bayesian inversion framework to identify the thermal resistance. A Bayesian sequential multi-fidelity statistical approach will be developed to automatically select the most appropriate version of the physical or surrogate model at each iteration of the inversion algorithm. This model operates the best compromise between the computational cost and the expected uncertainty reduction to properly identify the thermal resistance of a given wall. For that, both measurement and modelling errors will be estimated and their associated uncertainties will be propagated into the inverse process. We can underline that the influence of model error in the inversion process has been studied in previous works [5, 6]. Several directions can be investigated to minimize the impact of this model error on the inversion results at a given total computational cost. For instance, Bayesian sequential strategies inspired by those presented in [7] can be implemented to automatically select the level of fidelity but also the position where to evaluate the code to optimize an uncertainty reduction criterion on the quantities of interest.

By considering these uncertainties, the objective is to avoid biased identification due to over-fitting. Furthermore, the use of Bayesian methods will lead to a robust confidence interval on the identified thermal resistance.

References

[1] S. Roels, IEA Annex 58 Seminar – 847, Real building energy perf. assessment, 2014.

[2] F. Alzetto et al, Energy and Buildings, 174, 124-133, 2018.

[3] V. V. Fedorov, Theory of optimal experiments. Academic Press, 1972.

[4] T-T. Ha, L. Ibos, V. Feuillet, Y. Garcia, V. Le Sant, A. Koenen, L. Peiffer, R. Bouchie, K. Zibouche, J. Waeytens, SFT Conference, 2020a.

[5] T-T. Ha, V. Feuillet, J. Waeytens, K. Zibouche, S. Thebault, R. Bouchie, V. Le Sant, L. Ibos, 224, 110-130, Energy and Buildings, 2020b.

[6] J. Waeytens et al, J. Fluids Eng., 137, 9, 2015.

[7] G. Perrin, Reliability Engineering & System Safety, 196, 2020

Scientific background:

In 2017, I started a license in mathematics and computer science at the University of Montpellier. Then, I realized a master of applied mathematics, statistics at the university of Clermont Auvergne in Clermont-Ferrand. At the end of the master 1, I did a 5 months research internship at the Magmas et Volcans laboratory. Then, I did a one-year apprenticeship during the master 2, at the Conseil Départemental du Puy de Dôme, in Clermont-Ferrand.

Fixed-budget online adaptive mesh learning for physics-informed neural networks. Towards parameterized problem inference.

Thi Nguyen Khoa Nguyen[†,1,2,3], Thibault Dairay[§,2], Raphaël Meunier[§,2], Christophe Millet[§,1,3] Mathilde Mougeot[§,1,4]

[†] PhD student (presenting author). [§] PhD director/supervisor

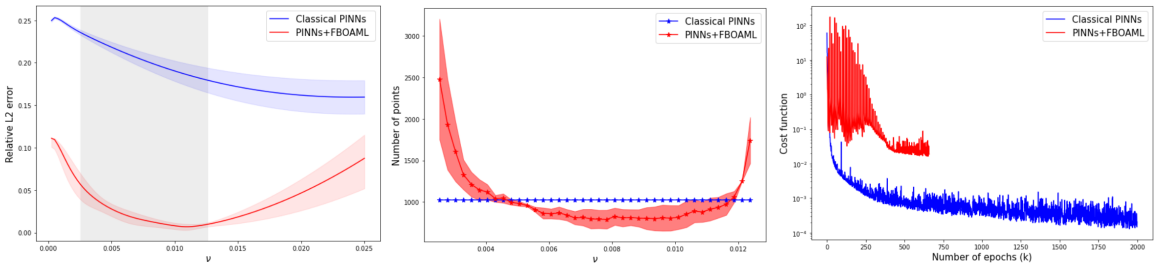
PhD expected duration: Sept. 2021 – Aug. 2024

[1] Universite Paris-Saclay, ENS Paris-Saclay, CNRS, Centre Borelli, Gif-sur-Yvette, 91190 France
[2] Michelin, Centre de Recherche de Ladoux, Cébazat, 63118 France
[3] CEA, DAM, DIF, F-91297 Arpajon, France
[4] ENSIIE, Évry-Courcouronnes, 91000 France
Corresponding author email: `thi.nguyen2@ens-paris-saclay.fr`

Abstract

Physics-Informed Neural Networks (PINNs) [5] have gained much attention in various fields of engineering thanks to their capability of incorporating physical laws into the models. PINNs integrate the physical constraints by minimizing the partial differential equations (PDEs) residuals on a set of collocation points. The distribution of these collocation points appears to have a huge impact on the performance of PINNs [3] and the assessment of the sampling methods for these points is still an active topic. To the best of the authors knowledge, the first work that showed the improvement of PINNs performance by modifying the set of collocation points is introduced in [1]. This work proposed the Residual-based Adaptive Refinement (RAR) that adds new training collocation points to the location where the PDE residual errors are large. RAR has been proved to be very efficient to increase the accuracy of the prediction but however leads to an uncontrollable amount of collocation points and computational cost at the end of the training process.

Motivated by this work, we propose a Fixed-Budget Online Adaptive Mesh Learning (FBOAML) [2, 4] method, which decomposes the domain into sub-domains, for training collocation points based on local maxima and local minima of the PDEs residuals so that the number of total collocation points remains the same during the training. The stopping criteria is based on a data set of reference, which leads to an adaptive number of iterations for each specific problem. The effectiveness of FBOAML is demonstrated in the context of non-parameterized and in particular parameterized problems. The impact of the hyper-parameters in FBOAML is investigated in this work. The comparison with other adaptive sampling methods is also illustrated. The numerical results demonstrate important gains in terms of accuracy of PINNs with FBOAML over the classical PINNs with non-adaptive collocation points. We also apply FBOAML in a complex industrial application involving coupling between mechanical and thermal fields. We show that FBOAML is able to identify the high-gradient location and even give better prediction for some physical fields than the classical PINNs with collocation points taken on a pre-adapted finite element mesh.



(a) Relative $\mathcal{L}^2$ error (b) Number of collocation points (c) Cost function during the training

Figure 1: *Parameterized Burgers equation (i.e. the viscosity ν is varied and is considered as an input of PINNs): comparison of classical PINNs and PINNs with FBOAML.* The curves and shaded regions represent the geometric mean and one standard deviation of five runs. In (a) the zone in gray is the learning interval for ν (interpolation zone). In (b) the number of collocation points for each ν in FBOAML can be varied, however the total number of collocation points remains the same. In (c) FBOAML needs less iterations to meet the stopping criteria. The classical PINNs minimize the cost function better since the collocation points are fixed during the training, which leads to over-fitting on the training points. While with FBOAML, after every certain number of iterations, the collocations points are adaptive, which leads to the jumps in the cost function.

Short biography (PhD student)

T.N.K. Nguyen obtained Master degree in applied mathematics at INSA Toulouse in 2020. Her PhD project, which is funded by Michelin and CEA through the Industrial Data Analytics and Machine Learning chair of ENS Paris-Saclay, focuses on physics-informed learning methods and their application to multi-physical simulation in industrial contexts.

References

[1] Lu Lu, Xuhui Meng, Zhiping Mao, and George Em Karniadakis. Deepxde: A deep learning library for solving differential equations. *SIAM Review*, 63(1):208–228, 2021.

[2] Thi Nguyen Khoa Nguyen, Thibault Dairay, Raphaël Meunier, Christophe Millet, and Mathilde Mougeot. Fixed budget online adaptive mesh learning for physics-informed neural networks. towards parameterized problem inference. *arXiv preprint arXiv:2212.11776*, 2022.

[3] Thi Nguyen Khoa Nguyen, Thibault Dairay, Raphaël Meunier, and Mathilde Mougeot. Physics-informed neural networks for non-newtonian fluid thermo-mechanical problems: an application to rubber calendering process. *Engineering Applications of Artificial Intelligence*, 114:105176, 2022.

[4] Thi Nguyen Khoa Nguyen, Christophe Millet, Thibault Dairay, Raphaël Meunier, and Mathilde Mougeot. Adaptive physics-informed deep learning for the parameterization of orographic gravity waves. In *AGU Fall Meeting*, 2022.

[5] Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

Sensitivity Analysis for Incompresible Navier-Stokes Equations

N. Nouaime [†,1,2], B. Després [§,2], M. A. Puscas [§,1], C. Fiorini [§,3]

[†] Ph.D. student (presenting author). [§] Ph.D. supervisor

Ph.D. expected duration: Oct. 2021 – Oct. 2024

[1] CEA, Centre de Saclay, DEN/SAC/DANS/DM2S/STMF/LMSF, Saclay, France.
nathalie.nouaime@cea.fr

maria-adela.puscas@cea.fr

[2] LJLL, Sorbonne Université, Paris, France
bruno.despres@sorbonne-universite.fr

[3] M2N, Conservatoire National des Arts et Métiers,Paris, France.
camilla.fiorini@lecnam.net

Abstract

Sensitivity analysis (SA) studies how changes in the input of a model affect the output. It answers what-if questions, i.e., what happens to the solution of the model if the input parameters change [6]. This task can be performed in many different ways, depending on the nature of the model considered. The present study focuses on sensitivity analysis of incompressible Navier–Stokes equations using the polynomial chaos method (PCM).
Let Ω denote an open bounded subset of $\mathbb{R}^2$, the incompressible Navier-Stokes system with homogeneous Dirichlet boundary conditions for this domain are

$$\begin{cases} \partial_t \mathbf{u}_0(\mathbf{x},t) - \nu\Delta\mathbf{u}_0(\mathbf{x},t) + (\mathbf{u}_0(\mathbf{x},t).\nabla)\mathbf{u}_0(\mathbf{x},t) + \nabla p_0(\mathbf{x},t) = \mathbf{f}_0(\mathbf{x},t) & \Omega,\ t>0, \\ \nabla.\mathbf{u}_0(\mathbf{x},t) = \mathbf{0} & \Omega,\ t>0, \\ \mathbf{u}_0(\mathbf{x},0) = \mathbf{0} & \Omega,\ t=0, \\ \mathbf{u}_0(\mathbf{x},t) = \mathbf{0} & \text{on} \quad \Gamma=\partial\Omega,\ t>0. \end{cases} \tag{1}$$

where $\mathbf{u}_0 = (u^x, u^y)$ is the velocity, p_0 the pressure, $\mathbf{f}_0$ the external force.
The PCM provides the treatment of a large variety of stochastic variables that can be described by probability density functions (PDF). Each stochastic variable of the model is described as a linear combination of stochastic modes. The modes are generally orthogonal polynomial functions of standardized random variables with known statistical properties. Assume that $a \sim \mathcal{N}(\mu, \sigma^2)$ is the uncertain parameter, and let $Y(\mathbf{x}, t; a)$ be a physical variable, i.e., the horizontal or vertical velocity, or the pressure.

The PCM, the uncertain variables can be decomposed on a basis of complete orthogonal polynomials, the so-called polynomial chaos expansion (PCE) [3], [4]. E.g. $Y(\mathbf{x}, t; a)$ can then be expressed by its PCE

$$Y(\mathbf{x}, t; a) = \sum_{i=0}^{n} Y_i(\mathbf{x}, t)\psi_i(a). \tag{2}$$

The unknown Y_i are deterministic coefficients and represent the random mode i of the physical variable component Y, and ψ_i are the orthogonal polynomials of degree i. The first order sensitivity equations, referred to as the sensitivity equations in short, are:

$$\begin{cases} \partial_t \tilde{\mathbf{u}}_1(\mathbf{x},t) - \nu\Delta\tilde{\mathbf{u}}_1(\mathbf{x},t) + \tilde{\mathbf{u}}_0(\mathbf{x},t).\nabla\tilde{\mathbf{u}}_1(\mathbf{x},t) + \tilde{\mathbf{u}}_1(\mathbf{x},t).\nabla\tilde{\mathbf{u}}_0(\mathbf{x},t) & \\ \quad + 2\sigma\tilde{\mathbf{u}}_1(\mathbf{x},t).\nabla\tilde{\mathbf{u}}_1(\mathbf{x},t) + \nabla\tilde{p}_1(\mathbf{x},t) = \tilde{\mathbf{f}}_1(\mathbf{x},t) & \Omega, t>0, \\ \nabla.\tilde{\mathbf{u}}_1(\mathbf{x},t) = \mathbf{0} & \Omega, t>0, \\ \tilde{\mathbf{u}}_1(\mathbf{x},0) = \mathbf{0} & \Omega, t=0, \\ \tilde{\mathbf{u}}_1(\mathbf{x},t) = \mathbf{0} & \text{on} \quad \Gamma, t>0. \end{cases} \tag{3}$$

where $\mathbf{u}_0$ is the solution of Eq. (1) and $\tilde{\mathbf{u}}_0 = \mathbf{u}_0 - \sigma\tilde{\mathbf{u}}_1$. The first order senstivity velocity and pressure are $\mathbf{u}_1 = \sigma\tilde{\mathbf{u}}_1$ and $p_1 = \sigma\tilde{p}_1$ respectively, and $\tilde{\mathbf{f}}_1$ is the sensitivity external force.
First, a stability estimate is provided for the Navier Stokes system and for its sensitivity system (3). The state and the sensitivity systems are discretized based on a Finite Elements Volumes Method (FEVM) [5, 7]. We provide a stability estimation for the discrete sensitivity of the Navier Stokes system. This numerical scheme is integrated into the open-source numerical fluid mechanics simulation software called "TrioCFD" [2] promoted by CEA.

Short biography (Ph.D. student)

I'm a second-year Ph.D. student; my thesis is entitled "Sensitivity analysis for thermohydrodynamics: uncertainty analysis and estimation of parameters", and the CEA [1] finances it. I'm working on the SA for the incompressible Navier-Stokes equations, discretizing these equations using FEVM, and developing this methods in TrioCFD. Previously, I did a bachelor and a M1 in mathematics at the Lebanese University, and a M2 in numerical modeling and analysis at the University of Montepellier.

References

[1] INSTN. Webpage. https://instn.cea.fr/these/.

[2] P. E. Angeli, M. A. Puscas, G. Fauchet, and A. Cartalade. FVCA8 benchmark for the Stokes and Navier-Stokes equations with the TrioCFD code – benchmark session. In *Finite Volumes for Complex Applications VIII - Methods and Theoretical Aspects*, pages 181–202, 2017.

[3] H. Bijl, D. Lucor, S. Mishra, and C. Schwab. *Uncertainty Quantification in Computational Fluid Dynamics*. Lecture Notes in Computational Science and Engineering. Springer International Publishing, 2013.

[4] P. Bonnaire, P. Pettersson, and C. F. Silva. Intrusive generalized polynomial chaos with asynchronous time integration for the solution of the unsteady navier–stokes equations. *Computers and Fluids*, 223:104952, 2021.

[5] P. Emonot. *Méthode de volumes éléments finis : applications aux équations de Navier-Stokes et résultats de convergence*. PhD thesis, Université Claude Bernard – Lyon I, 1992.

[6] C. Fiorini, B. Després, and M. A. Puscas. Sensitivity equation method for the navier-stokes equations applied to uncertainty propagation. *International Journal for Numerical Methods in Fluids*, pages 1–23, 2020.

[7] S. Heib. *Nouvelles discrétisations non structurées pour des écoulements de fluides à incompressibilité renforcée*. PhD thesis, Université Paris VI, 2003.

Dictionary-based model reduction for state estimation

A. Pasco[†,1], A. Nouy[§,1]

[†] PhD student (presenting author). [§] PhD supervisor

PhD expected duration: Dec. 2022 – Dec. 2025

[1] Centrale Nantes, Nantes Université, Laboratoire de Mathématiques Jean Leray, UMR CNRS 6629
{Alexandre.Pasco,Anthony.Nouy}@ec-nantes.fr

Abstract

We consider the problem of state estimation from a finite number of linear measurements, where the state $u(\xi)$ to recover is the solution of a parameterized partial differential equation for an unknown parameter ξ. A variational approach called PBDW and based on linear model order reduction (MOR) was proposed in [3]. This approach relies on the assumption that the solution manifold $\mathcal{M} := \{u(\xi) \ : \ \xi \in \mathcal{P}\}$ can be well approximated by a low-dimensional linear subspace V, which can be constructed by, e.g., proper orthogonal decomposition or greedy algorithms. The state is approximated by a linear combination of an element in V and a correction term in a space related to the linear measurements. Given m linear measurements, the dimension of V is bounded by m and the approximation error related to the contribution in V is therefore lower bounded by the Kolmogorov m-width

$$d_m(\mathcal{M}) := \inf_{\dim(V)=m} \sup_{v\in\mathcal{M}} \|v - P_V v\|. \tag{1}$$

Thus, when only a small number of measurements are available, the PBDW is not adapted to problems where $d_m(\mathcal{M})$ has a slow decay with m. This can be circumvented by using non linear reduced order models, such as the multi-space approach from [2], where the authors proposed an online model selection of a linear space using a residual-based surrogate distance to the manifold as selection criterion.

In this paper, we propose two contributions inspired from [1]. Firstly, we propose a randomized surrogate distance to manifold for model selection, allowing to estimate the residual-based distance with a low online cost while preserving robustness to round-off errors, which can be an issue for classical residual norm estimation. Secondly, we propose a dictionary-based approach for state estimation, which is a nonlinear approach that has shown several advantages over other linear and nonlinear methods in model order reduction. For a (unknown) parameter value ξ, we use as background space a low dimensional space $V(\xi)$ spanned by vectors from a dictionary $\mathcal{D}_K$ of rather large size K. Those vectors may be for example snapshots associated with random parameter values, or parameter values selected by a greedily algorithm. The selection of $V(\xi)$ is based on a least-squares formulation of the original PBDW formulation with ℓ_1 (sparsity-inducing) regularization. Finally, we propose two ways to select the regularization amplitude: a classical offline selection on some training set of snapshots, and an online selection using a randomized surrogate distance to the manifold as in [2]. Figure 1 shows good performances of our approach on a simple numerical example.

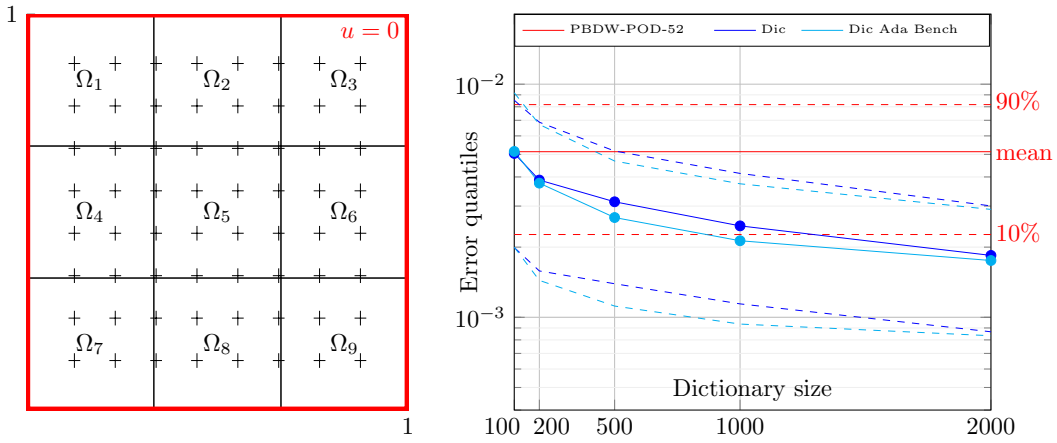


Figure 1: Dictionary-based state estimation applied to the thermal block problem parameterized by the thermal conductivity of the different subdomains (left). Sensors are located on a grid of $m = 64$ points. On the right, we compare our approach to the classical PBDW (red) using the best truncated POD modes as background space, on a test set of 100 snapshots. We distinguish the offline (blue) and the online (cyan) model selection of the regularization parameter.

Short biography (PhD student)

I graduated from Ecole Centrale de Nantes in 2022 as generalist engineer. It included 2 years focused on applied mathematics, from machine learning to numerical analysis. I continued at ECN by starting my PhD thesis on December 2022, funded by ANR COFNET, focusing on compositional function networks for nonlinear model reduction, for forward and inverse problems.

References

[1] Oleg Balabanov and Anthony Nouy. Randomized linear algebra for model reduction—part ii: minimal residual methods and dictionary-based approximation. *Advances in Computational Mathematics*, 47(2):1–54, 2021.

[2] Albert Cohen, Wolfgang Dahmen, Olga Mula, and James Nichols. Nonlinear reduced models for state and parameter estimation. *SIAM/ASA Journal on Uncertainty Quantification*, 10(1):227–267, 2022.

[3] Yvon Maday, Anthony T Patera, James D Penn, and Masayuki Yano. A parameterized-background data-weak approach to variational data assimilation: formulation, analysis, and application to acoustics. *International Journal for Numerical Methods in Engineering*, 102(5):933–965, 2015.

Global sensitivity analysis for meteorological and traffic-related uncertain variables in microscale pollutant dispersion CFD simulations

M. Pasquier[†,1], S. Jay[§,1], D. Sinoquet[§,1], J. Jacob[§,2], P. Sagaut[§,2]

[†] PhD student (presenting author). [§] PhD supervisor

PhD expected duration: Oct. 2020 – Oct. 2023

[1] IFP Energies nouvelles, Institut Carnot IFPEN Transports Energie, 1-4 avenue de Bois-Préau, 92852 Rueil-Malmaison, France
{mathis.pasquier,stephane.jay,delphine.sinoquet}@ifpen.fr

[2] Aix Marseille Univ, CNRS, Centrale Marseille, M2P2 UMR 7340, 13451 Marseille, France
{jerome.jacob,pierre.sagaut}@univ-amu.fr

Abstract

Road traffic accounts for a significant part of air quality damage in urban areas and thus represents a major threat to public health, since the population living in cities is expected to increase in the future. The recent field of urban physics uses numerical simulations and experimental outdoor/wind-tunnel measurements to better understand the physics of pollutant dispersion. Numerical simulations traditionally rely on fast-response low-fidelity models such as Lagrangian approaches, Gaussian plume models and street-network models. These models are fast and enable efficient decision-making for large scale urban domains (such as entire cities) but the increasing computational capabilities of modern hardware now allow the use of high-fidelity methods such as computational fluid dynamics (CFD). Different fidelity levels exist within CFD methods : solving only for the time-averaged flow field (Reynolds-averaged Navier-Stokes, or RANS) or capturing the transient turbulence dynamics (Large-eddy simulation, or LES) come at different computational costs and rely on specific numerical methods. RANS methods were extensively used for pollutant dispersion simulation and also to construct metamodels for uncertainty propagation and bayesian inversion [7, 8]. On the other hand, traditional LES methods remain very costly to be used for uncertainty quantification (UQ) studies. However, the recently developed Lattice-Boltzmann method (LBM) [2] now stands as a credible alternative to traditional LES in terms of computational cost and accuracy.

Growing attention is devoted to the study of epistemic and aleatory uncertainties in expensive numerical simulations. However, to our knowledge no one in the literature has yet studied microscale traffic-related urban pollutant dispersion while accounting for aleatory uncertainties associated to meteorological inputs *and* traffic emissions. Existing UQ studies focus on global scale simulations with fast-response models applied to traffic emissions [4] or on local point-source pollutant release with RANS [3]. Yet, high-fidelity simulations could be used to compare meteorological and traffic-related uncertainties with each other at the local scale and help decide whether or not to include them in future studies.

In this work, we use an enhanced LBM open-source code coupled with a traffic simulator and a physical engine emissions model to simulate traffic-related pollutant dispersion at the urban neighbourhood scale [6]. The simulation of 10 minutes of physical time is of the order of 6

to 8 hours on 960 cores, which remains costly but is a considerable cost reduction compared to traditional techniques. Thus, we use proper orthogonal decomposition (which can be seen as a deterministic version of the Karhunen-Loève decomposition) along with gaussian processes (POD-kriging or POD-GP) [1] to construct response surfaces for spatial time-averaged pollutant concentration fields in a realistic urban setting, simulated with the LBM solver (see fig. 1). We also compare this technique with a metamodeling method based on an anchored ANOVA (c-ANOVA) decomposition of the quantity of interest, which was initially designed to reduce the number of expensive simulation calls [5]. Our final objective is to carry out a global sensitivity analysis with 5 uncertain variables: 2 for the meteorological inputs and 3 for the traffic emissions.

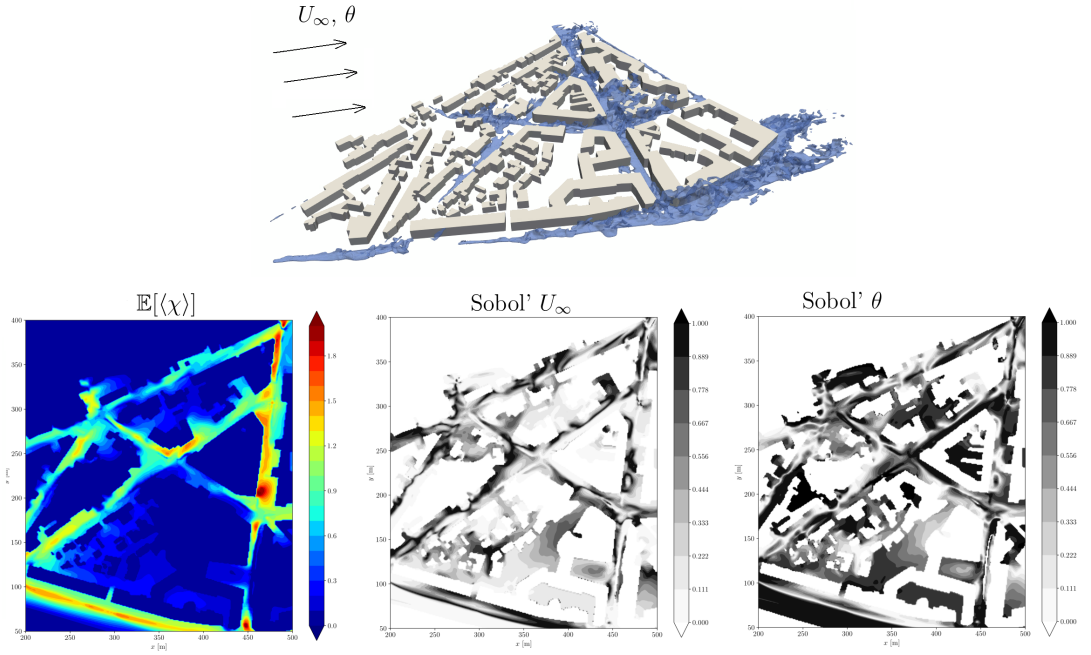


Figure 1: Study on an urban neighbourhood in the greater Paris area. **Top:** computational domain with instantaneous road traffic-related pollutant concentration iso-contours with the wind coming from the left ; the quantity of interest (QOI) is the time-averaged concentration $\langle\chi\rangle$ level [g/m^3] in a horizontal plane 1 meter high above the ground. **Bottom, from left to right:** QOI sample mean and first order Sobol' indices for incoming time-averaged wind speed and wind direction ; U_∞ and θ vary uniformly in $[6; 10]$ m/s and $[-0.2; +0.2]$ rad respectively.

Short biography (PhD student)

Mathis Pasquier graduated in 2020 from the Ecole des Mines de Saint Etienne with a generalist engineering diploma and from Cranfield university with a MSc. in computational fluid dynamics. He started a PhD at IFPEN in 2020 on microscale traffic-related pollutant dispersion and uncertainty quantification. The project focuses on the applicability of state-of-the art CFD methods for local air quality simulation along with uncertainty quantification for expensive codes.

References

[1] Thierry Braconnier, Marc Ferrier, J.-C Jouhaud, Marc Montagnac, and Pierre Sagaut. Towards an adaptive pod/svd surrogate model for aeronautic design. *Computers & Fluids*, 40:195–209, 2011.

[2] Shiyi Chen and Gary Doolen. Lattice boltzmann method for fluid flows. *Annual Review of Fluid Mechanics*, 30:329–364, 2003.

[3] Clara Garcia-Sanchez, G. Tendeloo, and Catherine Gorlé. Quantifying inflow uncertainties in rans simulations of urban pollutant dispersion. *Atmospheric Environment*, 161, 2017.

[4] J. Hammond, R. Chen, and V. Mallet. Meta-modeling of a simulation chain for urban air quality. *Advanced Modeling and Simulation in Engineering Sciences*, 7:37, 2020.

[5] Luca Margheri and Pierre Sagaut. A hybrid anchored-anova - pod/kriging method for uncertainty quantification in unsteady high-fidelity cfd simulations. *Journal of Computational Physics*, 324, 2016.

[6] Mathis Pasquier, Stéphane Jay, and Pierre Sagaut. Lattice-boltzmann simulations of traffic-related atmospheric pollutant dispersion in urban areas. *8th ECCOMAS Congress*, 2022.

[7] Jorge Sousa, Clara Garcia-Sanchez, and Catherine Gorlé. Improving urban flow predictions through data assimilation. *Building and Environment*, 132, 2018.

[8] Jorge Sousa and Catherine Gorlé. Computational urban flow predictions with bayesian inference: Validation with field data. *Building and Environment*, 154, 2019.

Kriging approaches: dealing with uncertainty in modeling aircraft trajectories

R. Perrichon[†,1], T. Klein [§,1,2], X. Gendre [§,2,3]

[†] PhD student (presenting author). [§] PhD supervisor

PhD expected duration: Oct. 2021 – Sep. 2024

[1] ENAC—Ecole Nationale de L'Aviation Civile, Université de Toulouse, 31030 Toulouse, France
remi.perrichon@enac.fr
thierry01.klein@enac.fr

[2] Institut de Mathématiques de Toulouse (UMR5219), Université de Toulouse, 31030 Toulouse, France
xavier.gendre@math.univ-toulouse.fr

[3] ISAE SUPAERO—Institut Supérieur de L'aéronautique et de L'espace, Université de Toulouse, 31030 Toulouse, France

Abstract

In transportation research, few studies are focusing on aircraft trajectories. One reason is the irrelevance of the usual statistical frameworks to achieve so.

An adapted framework to study trajectories is the one of Functional Data Analysis (FDA), popularised by [10] and [2]. FDA is an active field of statistics as highlighted by the development of inference procedures by [3], or more recently, by the blooming of the curve registration geometric framework of [11].

The promotion of FDA to study aircraft trajectories was early made by [8]. The statistical literature focusing on aircraft trajectories revolves around Functional Principal Component Analysis (FPCA) carried out by [7] and is applied to the detection of atypical energy behaviours by [4].

The first step of most FDA studies is to reconstruct the continuous realisations from the raw discrete data (it is the so-called dense framework as opposed to the sparse one).

A smoothing approach has prevailed in the literature following the work of [9] and [10] and was viewed as a pre-processing step. A review of four popular non-parametric techniques is given by [13].

Given operational concerns, it is often valuable to consider so-called augmented trajectories, that is to say trajectories for which weather dimensions are added, for example the wind speed and direction.

1 Kriging for augmented trajectories

Most weather data come as a set of regular grids indexed by time. Matching the usual dimensions of a trajectory (longitude, latitude, altitude) with weather data is never perfect but uncertainties can be taken into account. Kriging is especially popular in geostatistics and have been extended to the spatio-temporal framework [6]. Using kriging equations allow to predict the weather values at the non-observed points at which the aircraft has flown. It is the first objective of the poster.

2 Co-kriging models

When trajectories are viewed as open parametrized curves of $\mathbb{R}^n$, co-kriging models can be used to reconstruct trajectories taking into account uncertainties. Co-kriging refers to extension of the kriging model for multivariate functions, as pinpointed by [5]. It was first developed in geostatistics [1] [12]. The second objective of the poster is to assess if co-kriging can be used to recontruct aircraft trajectories which has not been done in the past.

Short biography (PhD student)

My background revolves around Statistics and Econometrics (Master's degree at Toulouse School of Economics, Magistère diploma at Université Paul Sabatier (Toulouse), apprenticeship at Airbus Helicopters). My PhD is funded by the ENAC-Isae-Supaéro-ONERA foundation and is entitled *Statistical Modeling of Plane Trajectories for Classification and Prediction*.

References

[1] Jean-Paul Chilès and Pierre Delfiner. *Geostatistics: Modeling Spatial Uncertainty*. Wiley Series in Probability and Statistics. 2012.

[2] Frederic Ferraty and Philippe Vieu. Nonparametric Functional Data Analysis: Theory and Practice. 2006.

[3] Lajos Horváth and Piotr Kokoszka. *Inference for Functional Data with Applications*, volume 200 of *Springer Series in Statistics*. Springer-Verlag New York Inc., 2012.

[4] Gabriel Jarry, Daniel Delahaye, Florence Nicol, and Eric Feron. Aircraft atypical approach detection using functional principal component analysis. *Journal of Air Transport Management*, 84:101787, May 2020.

[5] Loic Le Gratiet. *Multi-fidelity Gaussian process regression for computer experiments*. PhD thesis, Université de Paris-Diderot, 2013.

[6] José Montero, Gema Fernández-Avilés, and Jorge Mateu. *Spatial and Spatio-Temporal Geostatistical Modeling and Kriging*. Wiley Series in Probability and Statistics. 2015.

[7] Florence Nicol. Statistical Analysis of Aircraft Trajectories: a Functional Data Analysis Approach. page pp.51. Alldata 2017, The Third International Conference on Big Data, Small Data, Linked Data and Open Data, April 2017.

[8] Stéphane Puechmorel and Daniel Delahaye. 4D trajectories : a functional data perspective. page pp 1.C.6. IEEE, October 2007.

[9] J. O. Ramsay and B. W. Silverman. *Applied Functional Data Analysis: Methods and Case Studies*. Springer-Verlag New York Inc., June 2002.

[10] J. O. Ramsay and B. W. Silverman. *Functional data analysis*. Springer series in statistics. Springer-Verlag New York Inc., 2nd edition, 2005.

[11] Anuj Srivastava and Eric P. Klassen. *Functional and Shape Data Analysis*. Springer-Verlag New York Inc., 1st edition, October 2016.

[12] Hans Wackernagel. *Multivariate Geostatistics*. Springer Berlin, Heidelberg, 2003.

[13] Jin-Ting Zhang. *Analysis of Variance for Functional Data*. Chapman and Hall/CRC, 1st edition, July 2013.

Bayesian Inference for Inverse Problems with Hyperparameters Selection of the Field Covariance Function

N. Polette[†,1,2], P. Sochala[§,1], A. Gesret[§,2], O. Le Maître[§,3]

[†] PhD student (presenting author) [§] PhD supervisor

PhD expected duration: Nov. 2022 – Oct. 2025

[1] CEA, DAM, DIF, F-91297 Arpajon, France
{nadege.polette,pierre.sochala}@cea.fr

[2] MINES Paris, PSL University, Centre de Géosciences, France
alexandrine.gesret@minesparis.psl.eu

[3] Centre de Mathématiques appliquées, CNRS and INRIA, Ecole Polytechnique
olivier.le-maitre@polytechnique.edu

Abstract

Inverse problems are encoutered in many applications whenever one search for information about a physical system based on measurements [7]. The Bayesian inference is an attractive approach for solving such type of problem since it provides a full estimation of the unknown parameters distributions. However, the convergence of the posterior distribution sampled with Markov Chain Monte–Carlo (MCMC) [1] can be difficult to reach especially when dealing with a high-dimensional search space.

In this work, we are interested in estimating a physical field by using a set of undirect observations. In order to accelerate the MCMC sampling, the forward model predictions are replaced by surrogate models based on polynomial chaos (PC) expansions [8, 2]. In order to reduce the input dimension of the surrogate model, a parsimonious representation of the field is introduced by means of the Karhunen-Loève (KL) decomposition, considering that the field of interest is a particular realization of a Gaussian random field. The interest of this decomposition is to entirely characterize a random field by a low-dimensional set of parameters, namely its mean, its autocovariance function, and its coordinates [3]. In particular, the decomposition basis is obtained by solving the eigenvalue problem related to the given autocovariance function. In practice, the autocovariance function depends on hyperparameters that are often fixed a priori. We propose here another approach that consists in inferring the hyperparameters during the MCMC sampling. The interest of this approach is twofold: i) to allow a weak parametrization of the field and ii) to avoid overconfident results.

Since the hyperparameters are inferred, their values change from one MCMC step to another, inducing a change of basis for the KL decomposition, as well as a strong nonlinear dependency of the PC expansions to the hyperparameters. In [6], a method has been proposed to avoid this dependency and therefore to mitigate the construction cost of the surrogate model. The idea is to rely on a reference basis that does not depend on hyperparameters. This basis is obtained by solving the eigenvalue problem averaged with respect to the hyperparameters prior and is used to perform a change of coordinates. Despite the efficiency of this method, several drawbacks are raised in [6, 4]. In particular, the change of coordinates requires to solve a new eigenvalue problem at each MCMC step, which turns out to be expensive. An attempt to use

a surrogate model for the eigenelement computation was not convincing due to non-smooth hyperparameters dependency [4].

Our contribution to alleviate this difficulty is to redesign the change of coordinates to a change of measure. In this new method, the coordinates of the field in the reference basis become Gaussian centered distributed, with a covariance matrix that depends smoothly on the hyperparameters and does not require to solve the eigenvalue problem at each step. This change of measure method is applied to a seismic tomography problem, where we infer a seismic wave velocity field with the first-arrival traveltimes at given locations. This application to a continuous velocity model generalizes the work on a layered velocity model realized in traveltime tomography [5]. More realistic predictions than when fixing the hyperparameters at constant values are obtained. Moreover, this method allows for various field shapes, while keeping the implementation computationally tractable.

Short biography (PhD student)

Nadège Polette studied applied mathematics at École des Ponts ParisTech as well as at Sorbonne Université (Master of Science in mathematical modeling). Her PhD is funded by CEA. The supervision is ensured by Alexandrine Gesret (Geosciences center, École des Mines de Paris), Pierre Sochala (CEA), and Olivier Le Maître (CNRS/INRIA/CMAP). Her PhD falls within the objective of the CTBTO to detect and analyse seismic events in order to oversee the respect of the international treaties. The goal is to develop and implement numerical methods for solving inverse problems applied to geophysical events analysis.

References

[1] A. Doucet, A. Smith, N. de Freitas, and N. Gordon. *Sequential Monte Carlo Methods in Practice*. Information Science and Statistics. Springer New York, 2013.

[2] R. G. Ghanem and P. D. Spanos. *Stochastic Finite Element Method: Response Statistics*, pages 101–119. Springer New York, New York, NY, 1991.

[3] M. Meyer and H. G. Matthies. Efficient model reduction in non-linear dynamics using the karhunen-loève expansion and dual-weighted-residual methods. *Computational Mechanics*, 31:179–191, 2003.

[4] A. Siripatana, O. Le Maître, O. Knio, C. Dawson, and I. Hoteit. Bayesian inference of spatially varying manning's n coefficients in an idealized coastal ocean model using a generalized karhunen-loève expansion and polynomial chaos. *Ocean Dynamics*, 70, 2020.

[5] P. Sochala, A. Gesret, and O. Le Maitre. Polynomial surrogates for bayesian traveltime tomography. *GEM - International Journal on Geomathematics*, 12(1), December 2021.

[6] I. Sraj, O. P. Le Maître, O. M. Knio, and I. Hoteit. Coordinate transformation and polynomial chaos for the bayesian inference of a gaussian process with parametrized prior covariance function. *Computer Methods in Applied Mechanics and Engineering*, 298:205–228, 2016.

[7] A. Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation*, volume xii. 01 2005.

[8] N. Wiener. The homogeneous chaos. *American Journal of Mathematics*, 60(4):897–936, 1938.

Robust calibration of a hydrological model with stochastic surrogates

Katarina Radišić[†,1,2], Claire Lauvernet[§,1], Arthur Vidard[§,2]

[†] PhD student (presenting author). [§] PhD supervisor

PhD expected duration: Dec. 2021 – Dec. 2024

[1] INRAE, RiverLy, 69625, Villeurbanne, France
{katarina.radisic,claire.lauvernet}@inrae.fr

[2] Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France
arthur.vidard@inria.fr

Abstract

Misspecifying external forcings (such as rain) on a hydrological model can directly affect subsequent parameter calibrations. Indeed, by using classical calibration and problem inversion methods, the error in the external forcings is propagated to the model output, and then, if not treated correctly, this error is compensated by overcalibrating the model parameters. As a consequence, parameter values that were found optimal for one value of the external forcings, are not granted to be optimal for another one. Ideally however, estimated parameter values (that describe time-invariant soil properties) should be the same no matter the value of the external forcing.

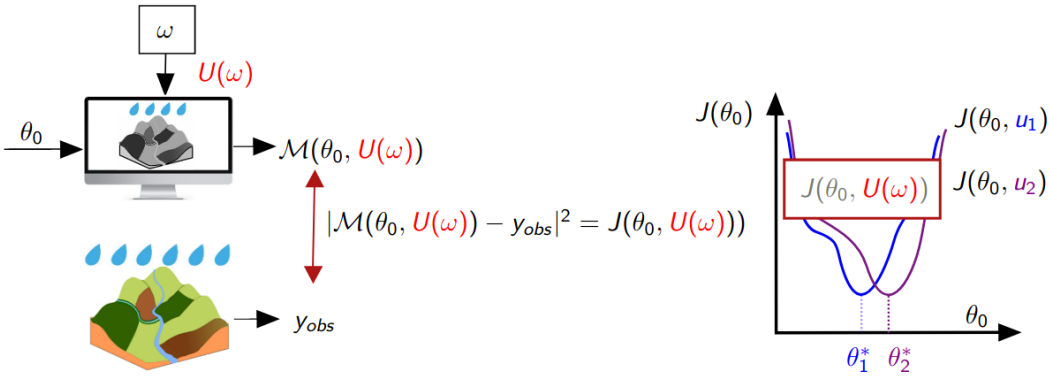


Figure 1: The influence of stochastic external forcings $U(\omega)$ on the calibration. The aim is to estimate the parameter θ_0 that minimizes the distance between model simulations $\mathcal{M}$ and terrain observations y_{obs}. This distance is quantified through a cost function $J(\theta_0, U(\omega))$. However, the cost function depends on the realization of the random variable $U(\omega)$, hence its minimizer θ^* does too. The aim of robust calibration is to annul this dependence.

Robust calibration was proposed to reduce the dependence of the estimated parameter values on the external forcings [3]. The aim of robust calibration is to propose parameter estimators

which are satisfactory over a large set of values of the external forcings. We model the external forcing (here, the rain) as a random variable, and want the estimated parameter to be "optimal enough" over a set of probable values of the random variable. We will be interested in the following robust estimators: the minimizer of the mean of the cost function, the minimizer of its variance, and the Pareto front of these two objectives.

As robust calibration is difficult to apply on distributed hydrological models, mainly due to their high dimension and computational cost, we use a surrogate model for the cost function. The stochastic surrogate for the cost function is constructed as proposed in [1]. Indeed, in a context where rain forcings are considered stochastic, the cost function to be minimised in the calibration setting is itself stochastic, Figure 1. Furthermore, this approach is non-intrusive as the structure of the stochasticity in the rain is never used, only an external stochastic rain simulator is needed.

We present the robust calibration of the PESHMELBA [2] distributed, process based, hydrological pesticide transfer model used for the simulation of pesticide fate on small agricultural catchments. The case study is a small virtual catchment in the Beaujolais region. Here, we focus on the calibration of two parameters, the soil content at saturation in deep soil layers, and the Van Genuchten infiltration parameter. Our results confirm the interest of the robust approach. Indeed, we show that the estimated parameters satisfy a larger set of rain realizations than when compared with the classic calibration of PESHMELBA for the same rain errors.

Short biography (PhD student)

I have completed my master studies in applied mathematics - statistical learning in Politecnico di Milano. My PhD focuses on the calibration of a hydrological model for pesticide transfers, developed to be used as a decision making tool in landscape management, as the latter conditions long-term water quality in rivers surrounded by pesticide-treated agricultural areas. The PhD is co-funded by the research centers INRAE and Inria.

References

[1] Nora Lüthen, Stefano Marelli, and Bruno Sudret. A spectral surrogate model for stochastic simulators computed from trajectory samples, July 2022. arXiv:2207.05653 [cs, math, stat].

[2] Emilie Rouzies, Claire Lauvernet, Christine Barachet, Thierry Morel, Flora Branger, Isabelle Braud, and Nadia Carluer. From agricultural catchment to management scenarios: A modular tool to assess effects of landscape features on water and pesticide behavior. *Science of The Total Environment*, 671:1144–1160, June 2019. doi: 10.1016/j.scitotenv.2019.03.060.

[3] Victor Trappler, Élise Arnaud, Arthur Vidard, and Laurent Debreu. Robust calibration of numerical models based on relative regret. *Journal of Computational Physics*, 426:109952, February 2021. doi: 10.1016/j.jcp.2020.109952.

Diffusion approximations for rank-based estimators of first order Sobol' Indices

D. Remot[1,2,†], R. Chhaibi[1,2,§], F. Gamboa[1,2,§], C. Pellegrini[1,2,§],
C. Oger[3,††], V. Oliveira[3,††]

[†] PhD student (presenting author). [§] Academic PhD supervisor [††] Industrial PhD supervisor

PhD expected duration: Oct. 2022 – Sep. 2025

[1] Institut de Mathématiques de Toulouse (Toulouse)
www.math.univ-toulouse.fr

[2] ANITI
www.aniti.univ-toulouse.fr

[3] NXP
www.nxp.com

Abstract

We establish a diffusion approximation theorem for the Chatterjee estimator of the Sobol' indices of first order (see [3] for a definition). This kind of estimator has been proposed by Chatterjee in [1]. It allows the estimation of conditional quantities using only a single sample. Indeed, classical procedures, such as the so-called Pick-Freeze method [4], needs more than a single sample.

A first central limit theorem is proposed in [1], under restrictive assumptions of independence. Several generalizations are discussed in [2]. Nevertheless, the vast majority of these works assume independence between the input and the output. More recently, Lin & Han [5] get rid of the independence hypothesis, using the tool of Hájeck projection. It is the first work of that kind.

In our work, we get a diffusion approximation of the estimator. Our result improves the previous central limit theorems, providing a general approximation along the sampling time. Furthermore, we deal with the multivariate input case, without independence assumptions between inputs and outputs.

Short biography (PhD student)

My PhD deals with sensitivity analysis, machine learning and their applications to accelerate the evaluation of analog semiconductor circuits' performances. It is in joint supervision between the Mathematical Institute of Toulouse (IMT), the AI institute (ANITI) and NXP, a leading company in the design and manufacturing of semi-conductors.

References

[1] Sourav Chatterjee. A new coefficient of correlation. *Journal of the American Statistical Association*, 116(536):2009–2022, 2021.

[2] Sourav Chatterjee. A survey of some recent developments in measures of association. *arXiv preprint arXiv:2211.04702*, 2022.

[3] Fabrice Gamboa, Pierre Gremaud, Thierry Klein, and Agnès Lagnoux. Global sensitivity analysis: A novel generation of mighty estimators based on rank statistics. *Bernoulli*, 28(4):2345–2374, 2022.

[4] Fabrice Gamboa, Alexandre Janon, Thierry Klein, A Lagnoux, and Clémentine Prieur. Statistical inference for sobol pick-freeze monte carlo method. *Statistics*, 50(4):881–902, 2016.

[5] Zhexiao Lin and Fang Han. Limit theorems of chatterjee's rank correlation. *arXiv preprint arXiv:2204.08031*, 2022.

Optimizing the diffusion of overdamped Langevin dynamics

R. Santet[†,1], T. Lelièvre[§,1], G. Pavliotis[2], G. Robin[3], G. Stoltz[§,1],

[†] PhD student (presenting author). [§] PhD supervisors

PhD expected duration: Oct. 2021 – Sep. 2024

[1] CERMICS, École des Ponts ParisTech, France
MATHERIALS, Inria Paris, France
{regis.santet,tony.lelievre,gabriel.stoltz}@enpc.fr

[2] Department of Mathematics, Imperial College, London, United-Kingdom
g.pavliotis@imperial.ac.uk

[3] CNRS & Laboratoire de Mathématiques et Modélisation d'Évry, Évry, France
genevieve.robin@cnrs.fr

Abstract

Predicting properties of materials and macroscopic physical systems in the framework of statistical physics [1], or obtaining the distribution of parameter values for Bayesian inference problems [2], both rely on sampling high dimensional probability measures. Popular methods rely on stochastic dynamics, in particular Markov Chain Monte Carlo (MCMC) methods. We focus on the overdamped Langevin dynamics, defined on a d-dimensional configuration space $\mathcal{Q}$, which are ergodic with respect to the Boltzmann–Gibbs distribution

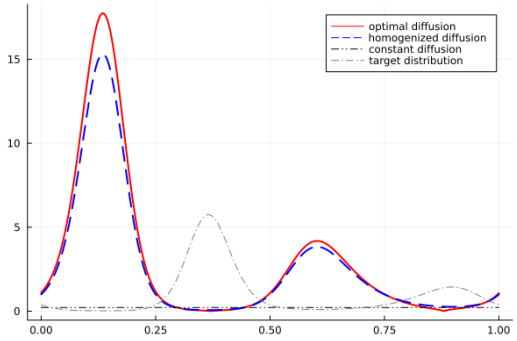
$$\mathrm{d}\mu = Z^{-1}\mathrm{e}^{-\beta V(q)}\mathrm{d}q, \qquad Z = \int_{\mathcal{Q}} \mathrm{e}^{-\beta V} < +\infty, \tag{1}$$

where $V \in C^\infty(\mathcal{Q})$ and $\beta > 0$. The overdamped Langevin dynamics are the following stochastic differential equations

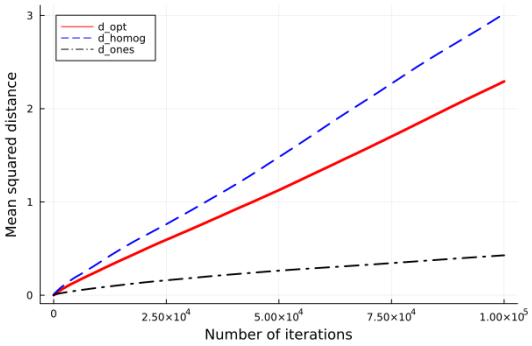
$$\mathrm{d}q_t = \left(-\mathcal{D}(q_t)\nabla V(q_t) + \beta^{-1}\mathrm{div}\,\mathcal{D}(q_t)\right)\mathrm{d}t + \sqrt{2\beta^{-1}}\mathcal{D}(q_t)^{1/2}\mathrm{d}W_t, \tag{2}$$

where $\mathcal{D}(q_t) \in \mathbb{R}^{d\times d}$ is a symmetric positive definite matrix, $\mathcal{D}^{1/2}$ is defined by spectral calculus, and $\mathrm{div}\,\mathcal{D}$ is the vector whose i-th component is the divergence of the i-th column of the matrix $\mathcal{D} = [\mathcal{D}_1, \ldots, \mathcal{D}_d]$, *i.e.* $\mathrm{div}\,\mathcal{D} = (\mathrm{div}\,\mathcal{D}_1, \ldots, \mathrm{div}\,\mathcal{D}_d)^\mathsf{T}$. Dynamics such as (2) are then discretized to create a Markov chain that samples (1) (up to a bias related to the time discretization error, and which can be cancelled using a Metropolis–Hastings correction). The convergence of MCMC methods may however be quite slow because the target measure is typically concentrated on a few high probability modes separated by low probability regions. A convenient choice usually made in (2) is to set $\mathcal{D}(q_t) = \mathrm{I}_d$ for all $q_t \in \mathcal{Q}$. However, this may not be optimal, as the rate of convergence towards (1) is related to the spectral gap of the operator of the dynamics (2)(see for instance [3]), which is parameterized by the diffusion coefficient $\mathcal{D}$.

Our objective is to compute, explicitly or numerically, the optimal diffusion function leading to the largest spectral gap, and thus to the fastest convergence rate. In particular, we discuss the normalization of the diffusion coefficient, since a too large diffusion coefficient has to be compensated by very small time steps. We formalize the maximization of the convergence rate of the overdamped Langevin dynamics (2) with respect to the diffusion coefficient $\mathcal{D}$ as a convex optimization program, and we show its well-posedness. We also propose in low dimensional scenarios a numerical procedure combining a finite element parameterization and an optimization



(a) Optimal diffusion obtained numerically, optimal diffusion in the homogenized limit, constant diffusion and target probability measure.



(b) Mean squared displacement averaged over 10^3 simulations for the three diffusion coefficients.

Figure 1: Numerical results for the optimization procedures for $\mathcal{Q} = \mathbb{T}$ (one-dimensional torus), $V(q) = \sin(4\pi q)(2 + \sin(2\pi q))$ under L^2 constraints on the diffusion coefficients.

algorithm, to compute the optimal diffusion in practice. The numerical procedure is illustrated on simple one-dimensional examples which show the benefits of having a position-dependent diffusion coefficient, see Figure 1. All the methods are provided in an open source Python and Julia code available at `https://github.com/rsantet/Optimal_Overdamped_Langevin_Diffusion_Python` and `https://github.com/rsantet/Optimal_Overdamped_Langevin_Diffusion_Julia`. We also study the behaviour of the optimal diffusion in the homogenized limit and show it has an analytical expression, proportional to the inverse of the target density, which is in accordance with various previous heuristics (see for example [4]). This simple limiting behaviour can be used as an initial guess in an optimization algorithm, or as a proxy for the optimal diffusion which does not require costly convex optimization procedures.

Short biography (Régis Santet)

R. Santet obtained an engineering degree at École des Ponts ParisTech as well as a Master's degree at Sorbonne Université (UPMC), specializing in functional and numerical analysis, PDEs & SDEs and computational statistical physics. His PhD is funded by École des Ponts and supervised by Tony Lelièvre and Gabriel Stoltz, researchers at CERMICS, École des Ponts & MATHERIALS, Inria Paris. The objective of his PhD is to improve sampling methods in the context of reversible and nonreversible dynamics, for instance in the framework of linear responses for transport coefficients computations in molecular dynamics.

References

[1] D. Frenkel and B. Smit *Understanding Molecular Simulation, From Algorithms to Applications* Academic Press, 2nd edition, 2002.

[2] C. P. Robert *The Bayesian Choice* Springer Texts in Statistics, 2nd edition, 2006.

[3] T. Lelièvre, F. Nier and G. A. Pavliotis *Optimal non-reversible linear drift for the convergence to equilibrium of a diffusion* Journal of Statistical Physics, 152(2):237–274, 2013.

[4] G. O. Roberts and O. Stramer *Langevin Diffusions and Metropolis–Hastings* Methodology And Computing In Applied Probability, 4:337–357, 2002.

Advances in surrogate modelling of dynamical systems with high-dimensional inputs

S. Schär[†,1], B. Sudret[§,1], S. Marelli[§,1]

[†] PhD student. [§] PhD supervisor

PhD expected duration: Oct. 2021 – Sep. 2025

[1] Chair of Risk, Safety & Uncertainty Quantification, ETH Zürich
{styfen.schaer,marelli,sudret}@ibk.baug.ethz.ch

Abstract

In recent decades, there has been a strong trend toward sustainable and environmentally friendly energy production. Wind energy, especially in offshore wind farms, has emerged as a promising candidate to meet this need [3], attracting attention and investments from both the academic and industrial research communities. However, the design of wind turbines remains challenging due to the need for costly simulations and the high uncertainties in environmental conditions.

Simulation-based cost-effective design of complex systems under uncertain conditions has been extensively studied in the engineering literature. Many state-of-the-art approaches use a small number of simulator outputs to train surrogate models (or emulators) that can accurately approximate one or more quantities of interest from the full simulator at a relatively low computational cost. As of today, however, most of these surrogates can only deal with time-independent problems. To model dynamical systems, a special class of models has been developed, namely that of non-linear autoregressive models with exogenous inputs (NARX). NARXs take time series inputs and return the corresponding dynamic system response as output. These models can capture temporal dependencies in the data by combining their auto-regressive predictions with the exogenous input time series to predict one or more future time steps. However, they tend to be highly susceptible to the curse of dimensionality.

Wind turbine simulators require a detailed parametrization of the wind input, also known as a turbulence box, hence falling into the class of time-dependent systems with high-dimensional inputs. The turbulence box $\boldsymbol{v} : \mathcal{T} \to \mathbb{R}^{3\times\nu_y\times\nu_z}$ itself is often represented as a multi-dimensional random field indexed by a discrete time axis $\mathcal{T}$. At every time step it consists of a longitudinal, transverse and upward wind velocity at each point on the spatial grid of size $\nu_y \times \nu_z$. An accurate surrogate of this dynamical behaviour $\widetilde{\mathcal{M}}$ must therefore be able to approximate any simulator output $f_i : \mathcal{T} \to \mathbb{R}$ of the simulator $\mathcal{M}$ as a function of $\boldsymbol{v}$, up to and including time t: $f_i(t) = \mathcal{M}(\boldsymbol{v}(\leq t)) \approx \widetilde{\mathcal{M}}(\boldsymbol{v}(\leq t))$.

To tackle this challenge, we have developed a novel surrogate modelling approach that combines spectral compression techniques and NARX. This new algorithm, which we call manifold NARX (mNARX) [1], first reduces the spatial components of the turbulence box to a small number of time-dependent spectral coefficients. These coefficients are then used to construct NARX models of simple auxiliary quantities related to the physics of the system. We then combine the original coefficients and the auxiliary quantities to form an exogenous input manifold, which allows us to use NARX and predict the final quantities of interest.

This incremental approach achieves a high prediction accuracy over long time horizons even when trained on a small experimental design. However, its ability to predict extreme values, e.g. peak

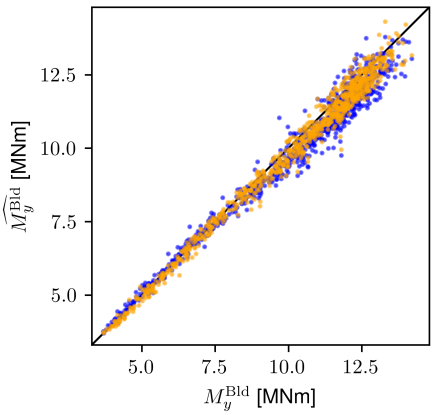


Figure 1: Maximum value of the bending moment at the blade root extracted from the simulator output M_y^{Bld} and that from the surrogate prediction $\widehat{M}_y^{\mathrm{Bld}}$. The results of the original mNARX are shown in blue, while the results of the modified mNARX are depicted in orange.

loads, within a simulation or across multiple simulations is highly dependent on the training data and strategy. To mitigate this problem, we extended the original mNARX algorithm in two different ways. First, we divide the original simulations into segments of shorter duration, which we classify based on the average velocity of the corresponding turbulence box section. Second, we subsample from the training data to ensure uniform coverage of the entire output quantile space. For each class, we then train a dedicated mNARX model.

In Fig. 1, we compare the emulated peak blade root bending moment M_y^{Bld} from the original mNARX and the modified mNARX. Both surrogates were trained using 100×600 s simulations and validated using an out-of-sample validation set of 900 simulations. The updated mNARX clearly provides a more accurate prediction of the peak moments, in particular for the higher quantiles of the peak moment.

To further improve mNARX, we will use the recently developed DRSM algorithm (dimensionality reduction for surrogate modelling) [2] to automate the construction of the NARX models, by optimizing the dimensionality reduction, subsampling and classification steps, making it more accessible to a less specialized audience.

Short biography

Styfen Schär is a second-year PhD student at the Chair of Risk, Safety and Uncertainty Quantification at ETH Zürich. He is developing new surrogate modeling techniques for predicting the long-term response of dynamical systems with high-dimensional exogenous inputs. Within the European HIPERWIND project, he is using them to facilitate the wind turbine design process.

References

[1] N. Dimitrov, S. Marelli, and S. Schär. Novel surrogate modelling approaches for wind turbine reliability assessment. H2020 Project HIPERWIND. Deliverable D4.1, 2022.

[2] C. Lataniotis, S. Marelli, and B. Sudret. Extending classical surrogate modelling to high-dimensions through supervised dimensionality reduction: a data-driven approach. *International Journal for Uncertainty Quantification*, 10(1):55–82, 2020.

[3] T. Tröndle, J. Lilliestam, S. Marelli, and S. Pfenninger. Trade-offs between geographic scale, cost, and infrastructure requirements for fully renewable electricity in europe. *Joule*, 4(9):1929–1948, 2020.

Sparse Bayesian learning for rational polynomial chaos expansions

F. Schneider[†,1], I. Papaioannou[§,2], G. Müller[§,1],

[†] PhD student (presenting author). [§] PhD supervisor

PhD expected duration: Nov. 2019 – Dec. 2023

[1] Chair of Structural Mechanics, Technical University of Munich
{felix.w.schneider,gerhard.mueller}@tum.de

[2] Engineering Risk Analysis Group, Technical University of Munich
iason.papaioannou@tum.de

Abstract

Surrogate models enable efficient propagation of uncertainties in computationally demanding models of physical systems. In linear structural dynamic models, the system response can be described by the frequency response function. We employ surrogate models that draw upon polynomial bases to model the stochastic response of structural dynamics systems. Specifically, in [2] we proposed a rational approximation that expresses the system frequency response as a rational of two polynomials with complex coefficients.

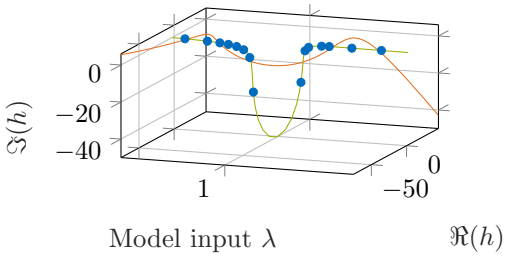


Figure 1: Graphical comparison between the approximation accuracy of standard PCE (—) and rational PCE (—) based on 15 samples (•) of the model $h(\lambda) = (\lambda - 1 + \mathrm{i}0.02)^{-1}$, where λ is log-normally distributed. The PCE is built with polynomial order $m = 5$, rational PCE with $m_p = m_q = 1$.

The inclusion of a denominator polynomial is motivated by the slow convergence of standard linear polynomial chaos expansion surrogate models for frequency response models. The rational surrogate structure is especially suitable for the representations of frequency response functions as they are mappings of rational form. To estimate the coefficients of the approximation, a non-intrusive regression approach that can be coupled easily with existing deterministic solvers was introduced. A comparison between standard and rational PCE in the approximation of a rational function on a one-dimensional input λ is given in Fig. 1.

However, to accurately estimate all coefficients of the polynomials in the numerator and the denominator, a large number of system evaluations is required in order to avoid overfitting. This is especially critical in high dimensional problems where the number of coefficients becomes excessively high. In order to extend the applicability of the proposed surrogate model to higher dimensional problems, we introduced a sparse learning approach that retains only the polynomial terms that contribute significantly to the predictability of the surrogate in [1]. In particular, we employ a sparse Bayesian learning approach with a hierarchical prior construction that follows the formalism of the relevance vector machine approach in [3]. Due to the rational form of the surrogate model, the problem becomes non-linear in the denominator coefficients and a

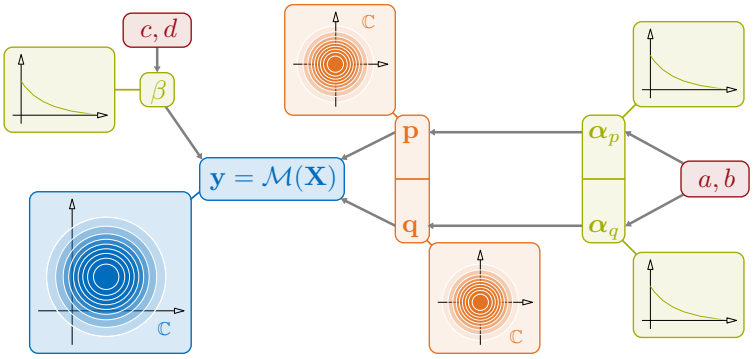


Figure 2: Illustration of the hierarchical Bayesian model for the rational PCE from [1]. Coefficients $\mathbf{p}$ and $\mathbf{q}$ are modeled through complex Gaussian distributions with precision parameters $\boldsymbol{\alpha}_\mathbf{p}$ and $\boldsymbol{\alpha}_\mathbf{q}$, which are again assigned Gamma distributions in order to induce sparsity into the model.

closed form solution for the coefficients is no longer available.. To circumvent this problem, we first find the posterior distribution of the numerator coefficients conditional on the denominator coefficients and all hyperparameters. Subsequently, we approximate the posterior distribution of the denominator coefficients through a Dirac at the maximum-a-posteriori (MAP) estimate of the denominator coefficients. We employ a gradient-descent algorithm to find the MAP estimate of the denominator coefficients and derive the necessary gradients analytically using $\mathbb{CR}$-calculus. On the basis of this MAP approximation, the optimal hyperparameters are found through maximizing the marginal likelihood. To accelerate convergence, we apply a pruning of the coefficients following the algorithm of [3]. The method is able to approximate the original model response accurately while using a low number of data samples. Thereby, the performance is improved in comparison to the previously proposed least squares approach. In an extension of the method we investigate further pruning and basis selection strategies, e.g., based on the fast marginal likelihood maximization algorithm in [4]. We compare the performance and accuracy of the methods on algebraic and finite element models of structural dynamic problems.

Short biography (PhD student)

I am a Ph.D. student at the Chair of Structural Mechanics at TU Munich (TUM), Germany in collaboration with the Engineering Risk Analysis group at TUM since 2019. I studied Civil Engineering at Technical University of Munich from 2011 to 2018. My Ph.D. thesis focuses on frequency domain based Uncertainty Quantification for structural dynamic models. My research interests include surrogate modeling, especially for rational models and Bayesian updating.

References

[1] Felix Schneider, Iason Papaioannou, and Gerhard Müller. “Sparse Bayesian Learning for Complex-Valued Rational Approximations”. In: *International Journal for Numerical Methods in Engineering* n/a.n/a (). DOI: https://doi.org/10.1002/nme.7182.

[2] Felix Schneider et al. “Polynomial chaos based rational approximation in linear structural dynamics with parameter uncertainties”. In: *Computers & Structures* 233 (2020), p. 106223. ISSN: 0045-7949. DOI: https://doi.org/10.1016/j.compstruc.2020.106223.

[3] Michael E Tipping. “Sparse Bayesian learning and the relevance vector machine”. In: *Journal of machine learning research* 1.Jun (2001), pp. 211–244.

[4] Michael E Tipping and Anita C Faul. “Fast marginal likelihood maximisation for sparse Bayesian models”. In: *International workshop on artificial intelligence and statistics*. PMLR. 2003, pp. 276–283.

Learning functions defined over sets of vectors with kernel methods

B. SOW[†,1], R. Le RICHE[§,2], S. Zannane[§,3], J. PELAMATTI[§,3], M. KELLER[§,3]

[†] PhD student (presenting author), babacar.sow@emse.fr
[§] PhD supervisors

PhD expected duration: Nov. 2021 – Oct. 2024

[1] LIMOS, Ecole Nationale Superieure des Mines de SAINT-ETIENNE
[2] CNRS/LIMOS [3] EDF R&D

Abstract

We consider the task of estimating functions from a restricted number of observations where the inputs are in the form of varying-size sets of vectors. A classical method in this expensive functions context is to approximate the true expensive function with a Gaussian process that relies on semi-definite positive kernels.

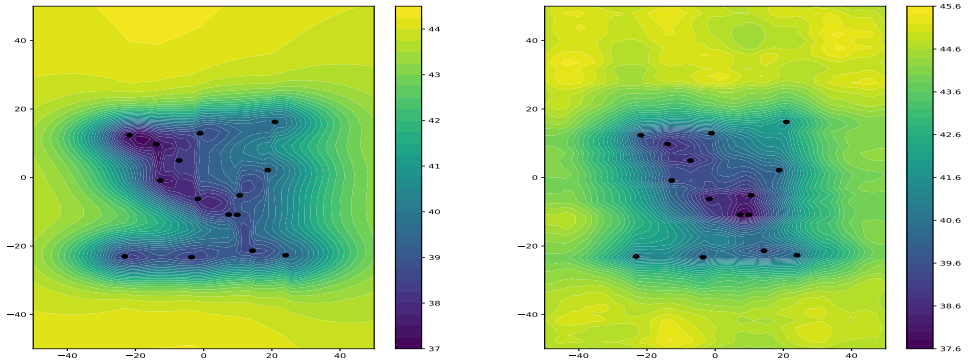
Varying-size sets of vectors have some peculiar characteristics : the number of vectors is not fixed and the vectors are not ordered. These properties prevent the use of classical, fixed dimension, kernels.

Sets of vectors can be represented as probability distributions or as vectors composed of well-chosen characteristic features. The distribution associated with each set of vectors can be either a (non-informative) discrete uniform distribution supported by the points of the set, or an empirical Gaussian distribution characterized by a mean and a covariance matrix estimated from the vectors of the set as in [5].

With this representation, we can define semi-definite positive kernels between distributions or vectors. The kernels can be classified into explicit ones, with known feature map and scalar product, and implicit kernels where the feature maps and the scalar product come mixed together. For implicit kernels, we restrict ourselves to distance substitution kernels with an exponential form. It is known that this approach for constructing kernels requires the substitute to be conditionally semi-definite negative, see [1]. In this context, it is necessary and sufficient to have an Hilbertian metric (isometric to a L^2 norm) [4]. The distances between distributions verifying the latter condition which we consider are the sliced Wasserstein distance [3], the maximum mean discrepancy, the approximated Gaussian Wasserstein (see [2]), and the Euclidean distance for vectors of features. For explicit kernels, we investigate the probability product kernel between Gaussian distributions [5] and the scalar product of embeddings of discrete uniform laws.

The prediction performance of theses kernels are compared on three types of analytical functions. The first ones mimic wind-farms productions for different winds, the second function is the inertia of a set of vectors and the third is the maximin function, well-known in the design of experiments. We first analyze the predictions over sets of vectors drawn uniformly over the same rectangular search space. Afterwards, we investigate the predictions over sets of vectors modified through geometrical transformations such as dilatation, rotation and translation. We show that some geometrical properties of the wind-farm functions can be better learned by MMD-based kernels than others. As an example, the MMD-based kernel prediction together with the true function is represented in Figure 1.

Figure 1: Left : Output of a function over a set. Right : MMD-based kernel prediction. 15 points were set at fixed positions and one had varying coordinates.



Short biography (PhD student)

Babacar SOW obtained his engineering degree from Telecom Paris and a Master's degree in Data Science and Applied Statistics at Institut Polytechnique de Paris. He is currently a PhD student in Applied Mathematics at Ecole des Mines de Saint-Etienne. The PhD is funded by the French National Research Agency (ANR), within the SAMOURAI project.

References

[1] Christian Berg, Jens Peter Reus Christensen, and Paul Ressel. *Harmonic analysis on semigroups: theory of positive definite and related functions*, volume 100. Springer, 1984.

[2] Thi Thien Trang Bui, Jean-Michel Loubes, Laurent Risser, and Patricia Balaresque. Distribution regression model with a reproducing kernel hilbert space approach. *arXiv preprint arXiv:1806.10493*, 2018.

[3] Mathieu Carriere, Marco Cuturi, and Steve Oudot. Sliced Wasserstein kernel for persistence diagrams. In *International conference on machine learning*, pages 664–673. PMLR, 2017.

[4] Bernard Haasdonk and Claus Bahlmann. Learning with distance substitution kernels. In *Joint pattern recognition symposium*, pages 220–227. Springer, 2004.

[5] Tony Jebara and Risi Kondor. Bhattacharyya and expected likelihood kernels. In *Learning theory and kernel machines*, pages 57–71. Springer, 2003.

Cost-free and optimized trade-off between simulation and physical experiment with variance based sensitivity analysis

C. Surget[†,1,2], S. Dubreuil[§,1], J. Morio[§,1], C. Mattrand[§,2], J-M. Bourinet[§,2], N. Gayton[§,2]

[†] PhD student (presenting author). [§] PhD supervisor

PhD expected duration: May 2021 – May 2024

[1] ONERA/DTIS, F-31055 Toulouse, France
{charles.surget,sylvain.dubreuil,jerome.morio}@onera.fr

[2] SIGMA Clermont, F-63000 Clermont–Ferrand, France
{charles.surget,cecile.mattrand,jean-marc.bourinet,nicolas.gayton}@sigma-clermont.fr

Abstract

Uncertainty quantification is an objective that is pursued to control the performances or the reliability of an industrial system. A complex simulation model can be viewed as a costly black box function $\phi : \mathcal{X} \to \mathbb{R}$. Its input $\boldsymbol{X}$ is a continuous random vector defined on the domain $\mathcal{X} \subseteq \mathbb{R}^d$ and by its probability density function (PDF) $f_{\boldsymbol{X}}$. Its output Y is a random variable such that $Y = \phi(\boldsymbol{X})$. One could be interested in assessing a given expectation of a particular function τ of Y such as a mean or a probability of failure. The actual expectation $\mathbb{E}_{f_{\boldsymbol{X}}}$ can be estimated by Monte Carlo (MC) simulation:

$$\mathbb{E}_{f_{\boldsymbol{X}}}\left[\tau\left(\phi\left(\boldsymbol{X}\right)\right)\right] = \int_{\mathcal{X}} \tau\left(\phi\left(\boldsymbol{x}\right)\right) f_{\boldsymbol{X}}(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \approx \frac{1}{N_{\boldsymbol{X}}} \sum_{j=1}^{N_{\boldsymbol{X}}} \tau\left(\phi\left(\boldsymbol{X}^{(j)}\right)\right), \tag{1}$$

with $\boldsymbol{X}^{(j)} \overset{i.i.d.}{\sim} f_{\boldsymbol{X}}$ and $N_{\boldsymbol{X}}$ the size of the MC sample. The input probabilistic model $f_{\boldsymbol{X}}$ is seldom known and may be deduced as a result of physical experiments [4]. In such a data driven context [3], the knowledge of the input probabilistic model is restricted to a $N_{\boldsymbol{D}}$-sample $\widetilde{\boldsymbol{D}} = \{\boldsymbol{D}^{(i)}, i = 1, \ldots, N_{\boldsymbol{D}}\} \sim f_{\widetilde{\boldsymbol{D}}}$, defined on the domain $\mathcal{D} \subseteq \mathbb{R}^{d \times N_{\boldsymbol{D}}}$, with $\boldsymbol{D}^{(i)} \overset{i.i.d.}{\sim} f_{\boldsymbol{X}}$. Realizations of $\boldsymbol{D}^{(i)}$ are assumed to be obtained from costly experimental tests, thus inducing a small-data context. The expectation that considers the fluctuations in the database can be expressed as follows [1]:

$$\mathbb{E}_{f_{(\boldsymbol{X},\widetilde{\boldsymbol{D}})}}\left[\tau\left(\phi(\boldsymbol{X})\right)\right] = \int_{\mathcal{D}} \int_{\mathcal{X}} \tau\left(\phi(\boldsymbol{x})\right) f_{\boldsymbol{X}|\widetilde{\boldsymbol{D}}}(\boldsymbol{x}|\widetilde{\boldsymbol{d}}) f_{\widetilde{\boldsymbol{D}}}(\widetilde{\boldsymbol{d}})\mathrm{d}\boldsymbol{x}\ \mathrm{d}\widetilde{\boldsymbol{d}} \tag{2}$$

$$\approx \frac{1}{N} \sum_{k=1}^{N} \frac{1}{N_{\boldsymbol{X}}} \sum_{j=1}^{N_{\boldsymbol{X}}} \tau\left(\phi\left(\boldsymbol{X}_k^{(j)}\right)\right) \tag{3}$$

where $\boldsymbol{X}_k^{(j)} \overset{i.i.d.}{\sim} \hat{f}_{\boldsymbol{X}|\widetilde{\boldsymbol{D}}_k}$ and $\hat{f}_{\boldsymbol{X}|\widetilde{\boldsymbol{D}}_k}$ is the joint PDF estimate of the true PDF $f_{\boldsymbol{X}}$, built from the database $\widetilde{\boldsymbol{D}}_k$ with $k = 1, ..., N$. The estimator is then subject to a bi-level uncertainty source. The first level comes from the MC estimate whereas the second one comes from the identification of the joint PDF.

The main objective here is to assess the variance of the given estimator with respect to those two sources of uncertainty in a small-data context. This variance can be reduced by increasing the database and MC sample sizes, respectively $N_{\boldsymbol{D}}$ and $N_{\boldsymbol{X}}$, i.e. by performing new physical experiments and running new simulations. Experimental tests and black-box evaluations however have a non-negligible cost for complex industrial applications. Hence, it is desirable to determine how the budget should be allocated: is it better to carry out new experiments or perform new simulations to optimize the variance decrease ?

The proposed work thus focuses on the simulation and physical experiment trade-off. The novelty of this contribution is to tackle this issue by means of a sensitivity analysis in order to determine the contribution of each uncertainty source in the variance of the estimator. A bootstrap resampling method is applied here to generate new databases $\widetilde{\boldsymbol{D}}_k$ and then to estimate several PDFs $\hat{f}_{\boldsymbol{X}|\widetilde{\boldsymbol{D}}_k}$. Sobol' indices [5] are subsequently computed to guide the investment between the experiment database or the MC sample. A multiple importance sampling method [2] is implemented to allow a cost-free approach. The relevance of an enrichment in data guided by sensitivity analysis is then illustrated on academic examples.

Short biography (PhD student)

Charles Surget obtained a master's degree in mechanical engineering from SIGMA Clermont in 2021 and then started his PhD thesis in May 2021. The research work is funded by Université Clermont Auvergne and ONERA. The framework of the thesis is to consider a bi-level of uncertainties obtained from database identifications and Monte Carlo Simulations. This issue is of a major concern in an industrial context. The goal is to suggest an approach able to propose a cost-free and optimized trade-off between simulation and physical experiment.

References

[1] Vincent Chabridon. *Analyse de sensibilité fiabiliste avec prise en compte d'incertitudes sur le modèle probabiliste-Application aux systèmes aérospatiaux*. PhD thesis, Université Clermont Auvergne (2017-2020), 2018.

[2] Art Owen and Yi Zhou. Safe and effective importance sampling. *Journal of the American Statistical Association*, 95(449):135–143, 2000.

[3] Gabriel Sarazin. *Analyse de sensibilité fiabiliste en présence d'incertitudes épistémiques introduites par les données d'apprentissage*. PhD thesis, Toulouse, ISAE, 2021.

[4] Bernard W Silverman. *Density estimation for statistics and data analysis*. Routledge, 2018.

[5] Ilya M Sobol'. Sensitivity analysis for non-linear mathematical models. *Mathematical modelling and computational experiment*, 1:407–414, 1993.

Goal-Oriented Reduction of Porous-Media Flows

Jana Tarhini[†,1,2], Sébastien Boyaval[§,2,3], Guillaume Enchéry[§,1], Quang Huy Tran[§,1]

[†] PhD student (presenting author). [§] PhD supervisor

PhD expected duration: December 2021 – November 2024

[1] IFP Energies nouvelles, 1 et 4 avenue de Bois-Préau, 92852 Rueil-Malmaison, France
{jana.tarhini,guillaume.enchery,quang-huy.tran}@ifpen.fr

[2] Laboratoire d'hydraulique Saint-Venant. Ecole des Ponts ParisTech

[3] Matherials Inria Paris ; EDF Lab Chatou,6 quai Wattier, 78401 Chatou Cedex, France
sebastien.boyaval@enpc.fr

Abstract

In hydrogeology, a flow simulator is considered as a practical tool to optimize injection-production scenarios and evaluate the impact of the uncertainties of some model parameters on particular quantities of interests (fluxes along internal boundaries, well flow rates...). This simulator implements a numerical discretization of parameterized partial differential equations. For specific case studies, it should be run for several values of the parameters, which induces costly computational efforts especially for large domains. Techniques like the reduced basis method [1, 4, 5, 6] can significantly lower the overall simulation cost.

In this work, we evaluate the reduced basis method for a single phase flow model. Classically, simulators implement the following linearized mass balance, where the fluid velocity is given by Darcy's law:

$$\begin{cases} \varphi \partial_t p - \boldsymbol{\nabla}.(\Lambda(\boldsymbol{\nabla} p + \rho g \boldsymbol{\nabla} z)) &= q & \text{in } (0,T)\times\Omega, \\ -\Lambda(\boldsymbol{\nabla} p + \rho g \boldsymbol{\nabla} z)\cdot\nu &= 0 & \text{on } (0,T)\times\Gamma_{\mathrm{N}}, \\ p &= p_{\mathrm{D}} & \text{on } (0,T)\times\Gamma_{\mathrm{D}}, \\ p(\boldsymbol{x}, t=0) &= p^0(\boldsymbol{x}) & \text{in } \Omega, \end{cases} \tag{1}$$

where φ is the product of the porosity with the total compressibility, p the pressure, Λ is the mobility tensor, ρ the mass density, g the gravity constant, z the depth and q the well flow rate term. ν is the outer normal vector to the domain boundary $\partial\Omega = \Gamma_{\mathrm{D}} \cup \Gamma_{\mathrm{N}}$ and T the total duration of simulation. Let us note that φ and Λ vary across Ω.

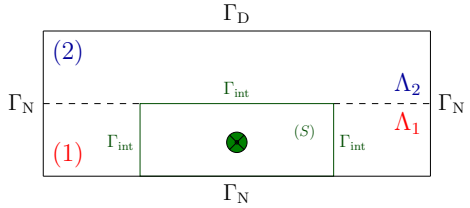


Figure 1: Typical domain configurations.

In this methodological study, we consider a rectangular domain Ω (see figure 1) composed of two areas (a reservoir one and a cap-rock) with two different mobilities Λ_1 and Λ_2. As quantity of interest, we take the flux s over the interior boundaries Γ_{int} given by

$$s = -\int_{\Gamma_{\text{int}}} \Lambda(\nabla p + \rho g \nabla z) \cdot \nu \, ds. \tag{2}$$

We have to solve (1) and (2) at all time steps and for different values of Λ_1 and Λ_2. We construct a reduced basis by means of a POD-Greedy algorithm as introduced in [3] and a goal oriented error estimator as in [2]. We discuss numerical results that illustrate the methodology and the efficiency of the a posteriori error estimate.

Short biography (PhD student)

I am second year PhD student at IFP Energies Nouvelles. I received a Master degree in PDEs and numerical analysis. The goal of my thesis is to accelerate simulations of CO2 underground storage. The single phase flow problem is a preliminary step. This enables the prediction of boundary conditions along the storage zone surrounded by the boundary Γ_{int} and reduces the computational effort by removing a part of the mesh.

References

[1] Nguyen Ngoc Cuong, Karen Veroy, and Anthony T. Patera. Certified real-time solution of parametrized partial differential equations. In *Handbook of Materials Modeling: Methods*, pages 1529–1564, 2005.

[2] Martin A. Grepl and Anthony T. Patera. A posteriori error bounds for reduced-basis approximations of parametrized parabolic partial differential equations. *ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique*, 39:157–181, 2005.

[3] Bernard Haasdonk and Mario Ohlberger. Reduced basis method for finite volume approximations of parametrized linear evolution equations. *ESAIM: Mathematical Modelling and Numerical Analysis*, 42(2):277–302, 2008.

[4] Jan S. Hesthaven, Gianluigi Rozza, and Benjamin Stamm. *Certified Reduced Basis Methods for Parametrized Partial Differential Equations.* Springer International Publishing, 2015.

[5] Christophe Prud'Homme, Dimitrios V. Rovas, Karen Veroy, Luc Machiels, Yvon Maday, Anthony T. Patera, and Gabriel Turinici. Reliable Real-Time Solution of Parametrized Partial Differential Equations: Reduced-Basis Output Bound Methods. *Journal of Fluids Engineering*, 124:70–80, 2001.

[6] A. Quarteroni, A. Manzoni, and F. Negri. *Reduced Basis Methods for Partial Differential Equations: An Introduction.* Springer International Publishing, 2015.

Bayesian estimation of seismic fragility for industrial structures and equipment

A. Van Biesbroeck[†,1,2], J. Garnier[§,1], C. Feau[§,2], C. Gauchy[§,3]

[†] PhD student (presenting author). [§] PhD supervisor

PhD expected duration: Oct. 2022 – Sep. 2025

[1] CMAP, CNRS, École polytechnique, Institut Polytechnique de Paris, 91120 Palaiseau, France
{antoine.van-biesbroeck, josselin.garnier}@polytechnique.edu

[2] Université Paris-Saclay, CEA, SEMT, 91191 Gif-sur-Yvette, France
cyril.feau@cea.fr

[3] Université Paris-Saclay, CEA, STMF, 91191 Gif-sur-Yvette, France
clement.gauchy@cea.fr

Abstract

As part of the assessment of the seismic safety of industrial installations, seismic fragility curves of mechanical structures are key quantities of interest for probabilistic seismic risk assessment studies [1]. They express the probability of failure of the mechanical demand conditional to a scalar value derived from the seismic ground motions, coined intensity measure.

Evaluation of these curves using Monte Carlo methods and mechanical numerical simulations with artificial seismic signals is becoming common [2]. Nevertheless, in this context, when resorting to a complex and detailed modeling, implying in addition a nonlinear behavior, the number of data available is limited due to the calculation burden.

The Bayesian point of view allows more efficient learning conditioned on the prior distribution of the parameters that can be assumed to constitute fragility curves. Indeed, in practice, fragility curves are assumed to follow the c.d.f. of a log-normal distribution [1, 3, 4, 2]. Under a small data-set, the choice of the prior has a non-negligible influence on the posterior distribution [5]. The proper choice of the prior probability distribution is thus of primary importance.

Non informative and objective priors are answers to this problematic and can be chosen as the maximal or minimal arguments of some metrics [6], such as the expected mutual information for which the optimum is asymptotically the Jeffreys prior w.r.t. the data-set size [7]. A study of those and their application to the seismic fragility curves framework are the main purposes of this PhD. We present an implementation of the Jeffreys prior under the usual log-linear probit model for *a posteriori* simulations and estimations of seismic fragility curves (Figure 1). This method is compared with other common methods in the literature.

Short biography (PhD student)

During my studies in applied mathematics at the Ecole Normale Supérieure Paris-Saclay I chose to be specialized into statistics and machine learning being passionate by the probability theory and their abilities to describe phenomena from data. At the end of my Master's degree I chose to apply to a end-of-study internship at the CEA on the subject "Objective Bayesian estimation

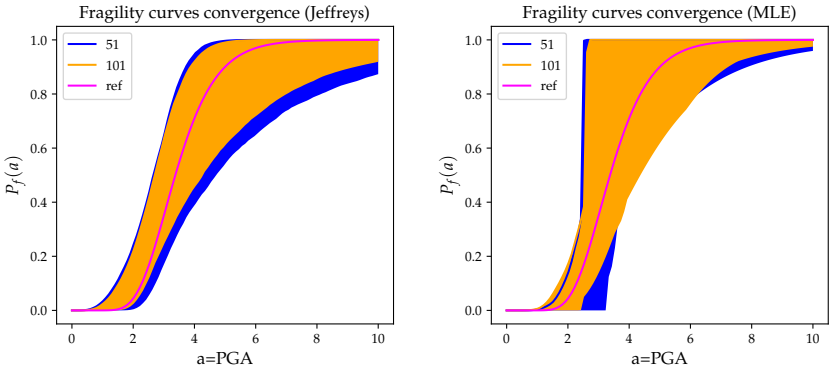


Figure 1: Fragility curves estimation using Jeffreys prior (on the left), and the classical maximum likelihood estimator method (on the right). The resulting 95% confidence intervals for the fragility curves of two data-set samples of different sizes (51 in blue and 101 in orange) are plotted. The system studied is a single degree of freedom oscillator with a nonlinear restoring force (see [8]). The considered model for the fragility curve is the log-linear probit with respect to the parameters α, β as $P_f(a) = \Phi(\beta^{-1} \log \frac{a}{\alpha})$ where Φ is the c.d.f. of the standard Gaussian distribution.

of seismic fragility curves". As I appreciated very well the subject and as its results led to interesting and open questions we decided to pursue it further as a PhD thesis 1 year later, financed by the CEA under the SEISM Institute (https://www.institut-seism.fr/en/).

References

[1] Robert P. Kennedy, C. Allin Cornell, Robert D. Campbell, Stan J. Kaplan, and Harold F. Perla. Probabilistic seismic safety study of an existing nuclear power plant. *Nuclear Engineering and Design*, 59(2):315–338, 08 1980.

[2] Irmela Zentner, Max Gündel, and Nicolas Bonfils. Fragility analysis methods: Review of existing approaches and application. *Nuclear Engineering and Design*, 323:245–258, 11 2017.

[3] Daniel Straub and Armen Der Kiureghian. Improved seismic fragility modeling from empirical data. *Structural Safety*, 30(4):320–336, 07 2008.

[4] Guillaume Damblin, Merlin Keller, Alberto Pasanisi, Pierre Barbillon, and Eric Parent. Approche décisionnelle bayésienne pour estimer une courbe de fragilité. *Journal de la Société Française de Statistique*, 155(3):78–103, 07 2014. hal-01545648.

[5] Andrew Gelman, John B. Carlin, Hal S. Stern, and David B. Dunson. *Bayesian data analysis*. Texts in Statistical Science. Chapman and Hall/CRC, 3 edition, 2013.

[6] Robert E. Kass and Larry Wasserman. The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, 91(435):1343–1370, 09 1996.

[7] James O. Berger, José M. Bernardo, and Dongchu Sun. The formal definition of reference priors. *The Annals of statistics*, 37(2):905–938, 05 2009.

[8] Rémi Sainct, Cyril Feau, Jean-Marc Martinez, and Josselin Garnier. Efficient methodology for seismic fragility curves estimation by active learning on support vector machines. *Structural Safety*, 86:101972, 09 2020.