# Bayesian calibration using Gaussian surrogate model of the likelihood function: application to train suspensions monitoring

D. Lebel
*Université Paris-Est*

**Supervisor(s):** Prof. C. Soize (Université Paris-Est), Dr. C. Fünfschilling (SNCF) and Dr. G. Perrin (CEA)

**Ph.D. expected duration:** Jan. 2016 - Dec. 2018

**Adress**: SNCF I&R, 40 avenue des Terroirs de France, 75012 Paris

**Email**: david.lebel@sncf.fr

**Abstract:**

The objective of the work presented here is the development of a Bayesian calibration method for a simulation-based model with stochastic functional input and output, involving the representation of the likelihood function by a Gaussian surrogate model.

Classical Bayesian calibration implies the computation of the likelihood function using a stochastic model of the system output and the available experimental data. This likelihood function is then used to estimate the posterior distribution of the model parameters. Typically, this estimation step is performed by Markov Chain Monte Carlo (MCMC) algorithms. Such algorithms require numerous calls to the likelihood function, and consequently numerous simulation runs. We here consider the case when the latter are expensive, which results in unaffordable computational costs.

Such issues may be adressed by relying on surrogate models of the system output. However, in the present case, because the system output is functional, its representation by a surrogate model is a complex task. Instead, we chose to directly build a gaussian surrogate model of the scalar likelihood function. This surrogate model is built in two steps. First, a global training step using a space-filling design of experiment in the multidimensional parameter space determines the global dependance of the likelihood function to the model parameters. Second, an enrichment step refines the surrogate model in the areas of interest. In the present case, we are interested in the most probable values of the parameters. The refinement is thus performed around the location of the likelihood function maximum. For this purpose, optimization algorithms can be used, such as the EGO algorithm proposed in [2] or the KGCP policy proposed in [4] if the likelihood observations are noisy.

With the deterministic likelihood function represented by a random surrogate model, the most straighforward solution to perform the subsequent MCMC step is to use the gaussian surrogate model mean function as the best predictor of the likelihood function. Although it may provide interesting results, such an approach tends to overestimate the accuracy of the calibration because it does not take into account the new type of uncertainty introduced by the use of a surrogate model. To include the surrogate model uncertainty in the posterior distribution estimation, we propose to perform a Monte Carlo sampling of trajectories of the surrogate model. The MCMC can then be performed on each one of these trajectories, and the posterior distribution estimated from the resulting samples. However, drawing a trajectory of gaussian process indexed on a multidimensional space can be expensive. Moreover, we do not know *a priori* at which points of the parameter space the value of the trajectory is needed. Instead of computing each trajectory exactly, we propose a way to approximate them. This approximation consists in the expectation of the surrogate model conditioned by the value of the trajectory at the points of a conditoning set of limited size. The choice of this conditoning set is crucial. Its goal is to reduce as much as

possible the variance of the surrogate model in the areas of interest of the parameter space, so that its expectation is close to any trajectory in these areas.

The method is applied to a railway case, for the state health monitoring of high-speed train suspensions using in-service measurements by embeded accelerometers. The system consists of a rolling train, excited by the track geometric irregularities. They consist of small displacements of the rails relatively to the theoretical track design. The observed quantities are accelerations at various points in the train, gathered in what we call the train dynamic repsonse. The model parameters to identify by calibration are the ones describing the mechanical characteristics of the suspensions.

Deterministic train dymanics simulation is used to build the stochastic train response model. In [3], it has been shown that the input of the system, the track geometric irregularities, can be modeled as a non-stationary random field. An output predictive error consisting of a gaussian stochastic process is added to the simulation output. Its purpose is to globally take into account the measurement noise and the various model errors, such as the modeling simplification, the discretization or the uncertainty about certain model parameters (other than the suspension parameters to identify). This output predictive error must be carefully identified in a preliminary step. For this application, the MCMC algorithm that was used is the Transitional MCMC proposed in [1] because the probability density functions to estimate were very peaked.

The method was validated on a numerical experiment, with simulated data for which the value of the parameters is known. Its application on actual experimental data gave very promising results.

## References

[1] Wolfgang Betz, Iason Papaioannou, and Daniel Straub. Transitional Markov Chain Monte Carlo: Observations and Improvements. *Journal of Engineering Mechanics*, 142(5):04016016, 2016.

[2] Donald R. Jones, Matthias Schonlau, and William J. Welch. Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global Optimization*, 13:455–492, 1998.

[3] Guillaume Perrin, Christian Soize, Denis Duhamel, and Christine Funfschilling. Track irregularities stochastic modeling. *Probabilistic Engineering Mechanics*, 34:123–130, 2013.

[4] Warren Scott, Peter I. Frazier, and Warren B. Powel. The Correlated Knowledge Gradient for Simulation Optimization of Continuous Parameters using Gaussian Process Regression. *SIAM Journal on Optimization*, 21(3):996–1026, 2011.

**Short biography** – Engineering degree from ENSTA ParisTech.
This Ph.D. is funded by SNCF Innovation & Research as part of a CIFRE contract.