Integrating hyper-parameter uncertainties in a multi-fidelity Bayesian model for the estimation of a probability of failure

Rémi Stroh, Julien Bect, Séverine Demeyer, Nicolas Fischer, Emmanuel Vazquez

Integrating hyper-parameter uncertainties in a multi-fidelity Bayesian model for the estimation of a probability of failure

R. Stroh[†,◇,*], J. Bect[◇], S. Demeyer[†], N. Fischer[†] and E. Vazquez[◇]

[†] Mathematics and Statistics Department,
Laboratoire National de métrologie et d'Essais (LNE), Trappes, France

[◇]Laboratoire des Signaux et Systèmes (L2S), CentraleSupélec, Univ. Paris-Sud,
CNRS, Université Paris-Saclay, Gif-sur-Yvette, France

[*]E-mail: remi.stroh@lne.fr

Abstract

A multi-fidelity simulator is a numerical model, in which one of the inputs controls a trade-off between the realism and the computational cost of the simulation. Our goal is to estimate the probability of exceeding a given threshold on a multi-fidelity stochastic simulator. We propose a fully Bayesian approach based on Gaussian processes to compute the posterior probability distribution of this probability. We pay special attention to the hyper-parameters of the model. Our methodology is illustrated on an academic example.

1 Introduction

In this article, we aim to estimate the Probability of Failure (PoF) of a system described by a multi-fidelity numerical model. Multi-fidelity simulators are characterized by the fact that the user has to make a trade-off between the realism of the simulation and its computational cost, for instance by tuning the mesh size when the simulator is a finite difference simulator. An expensive simulation gives a high-fidelity result, while a cheap simulation returns a low-fidelity approximation. A multi-fidelity approach combines different levels of fidelity to estimate a quantity of interest. A method for estimating probabilities of exceeding a threshold of a stochastic multi-fidelity numerical model is proposed in [6]. In this paper, we extend the methodology to a fully Bayesian approach.

A stochastic multi-fidelity simulator can be seen as a black-box, which returns an output modeled by a random variable Z from a vector of inputs $(x, t) \in \mathbb{X} \times \mathbb{R}^+$, $\mathbb{X} \subset \mathbb{R}^d$. The vector x is a set of input parameters of

the simulation, and the scalar t controls the fidelity of the simulation. The fidelity increases when t decreases. We denote by $\mathbb{P}_{x,t}$ the probability distribution of the output Z at (x,t). We assume that an input distribution $f_{\mathbb{X}}$ on the input space $\mathbb{X}$ and a critical threshold z^{crit} are also provided. The PoF is the probability that the output exceeds the critical threshold

$$P = \int_{\mathbb{X}} \mathbb{P}_{x,t^{\text{ref}}}(Z > z^{\text{crit}}) f_{\mathbb{X}}(x) \mathrm{d}x, \tag{1}$$

where t^{ref} is a reference level where we would like to compute the probability. We use a Bayesian approach based on a multi-fidelity Gaussian process model of Z in order to compute a posterior distribution of the PoF. Prior distributions are added on the hyper-parameters of the Gaussian process, so we expect that the posterior distribution of the PoF has better predictive properties. This approach is compared to a classical plug-in approach.

The paper is organized as follows. Section 2 explains the Bayesian multi-fidelity model. Section 3 describes how to take into account the hyper-parameter uncertainties to compute the posterior density of the PoF. Section 4 illustrates the methodology on an academic example.

2 Multi-fidelity Gaussian process

In this section, we present the model proposed in [6]. The output Z at x,t is assumed conditionally Gaussian

$$Z|\xi,\lambda \sim \mathsf{N}(\xi(x,t),\lambda(t)), \tag{2}$$

with $\xi(x,t)$ and $\lambda(t)$ the mean and variance functions, the latter being assumed independent of x for simplicity. Knowing ξ and λ, two different runs of the simulator produce independent outputs. Bayesian prior models are independently added on ξ and λ.

For the mean function ξ, we use the multi-fidelity model proposed by [3, 7]. This model decomposes the Gaussian process $\xi(x,t)$ in two independent Gaussian processes:

$$\xi(x,t) = \xi_0(x) + \epsilon(x,t), \tag{3}$$

where the process ξ_0 describes an ideal simulator, which would be the result at $t=0$, and ϵ represents the numerical error of the simulator. The model imposes $\mathbb{E}\left[\epsilon(x,0)^2\right] = 0$. Moreover, as the fidelity increases when t decreases, the variance of ϵ according to t is decreasing when t decreases.

The ideal process ξ_0 is a stationary Gaussian process with constant mean m and stationary covariance c_0. The error process ϵ is a centered Gaussian process with a separable covariance between x and t, independent of ξ_0. Thus, the distribution of ξ is

$$\xi \sim \mathsf{GP}(m, c_0(x-x') + r(t,t') \cdot c_\epsilon(x-x')). \tag{4}$$

The prior distribution of m is a uniform improper distribution on $\mathbb{R}$, which is a classical assumption in ordinary kriging (see [5]). Following the recommendations of [7], a Matérn 5/2 covariance function is selected for c_0 and c_ϵ:

$$c_0(h) = \sigma_0^2 \mathcal{M}_{5/2}\left(\sqrt{\sum_{k=1}^{d}\left(\frac{h_k}{\rho_k^0}\right)^2}\right), c_\epsilon(h) = \sigma_0^2 G\, \mathcal{M}_{5/2}\left(\sqrt{\sum_{k=1}^{d}\left(\frac{h_k}{\rho_k^\epsilon}\right)^2}\right), \tag{5}$$

and a distorted Brownian covariance function for the fidelity covariance:

$$r(t,t') = \left(\frac{\min\{t,t'\}}{t^{\mathrm{LF}}}\right)^L, \tag{6}$$

with $\sigma_0^2, G, L, (\rho_k^0, \rho_k^\epsilon)_{1\le k\le d}$ $2d+3$ positive hyper-parameters, t^{LF} the lowest level of fidelity (to ensure $r(t,t') \le 1$), and $\mathcal{M}_{5/2}$ the covariance function $\mathcal{M}_{5/2}(h) = \left(1+\sqrt{5}h+\frac{5}{3}h^2\right)e^{-\sqrt{5}h}$.

In this article, even if the simulator could be observed at any level t, we assume that only S levels $t_1 > t_2 > \cdots > t_S > 0$ are actually observed. Thus, instead of inferring on the whole function $\lambda(t)$, we consider only the parameters $(\lambda(t_s))_{1\le s\le S}$. The vector of hyper-parameters $\theta = \left\{\sigma_0^2, (\rho_k^0)_{1\le k\le d}, G, L, (\rho_k^\epsilon)_{1\le k\le d}, (\lambda(t_s))_{1\le s\le S}\right\}$ therefore has length $2d+3+S$.

3 Dealing with hyper-parameters

In order to carry out a fully-Bayesian approach, prior distributions are added on these hyper-parameters. To simplify the estimations and the inference, the hyper-parameters are expressed in log-scale $l_\theta = \log(\theta)$, and the joint prior distribution of l_θ is chosen to be a multivariate normal distribution. The hyper-parameters of the mean function ξ are assumed mutually independent, and independent of the noise variance λ. An approximate value r^{out} of the range of the output is assumed known, and the input domain $\mathbb{X}$ is assumed to be an hyper-rectangle $\mathbb{X} = \prod_{k=1}^{d}[a_k; b_k]$. We propose, for the model described in Section 2, the following prior distributions:

$$l_{\sigma_0^2} \sim \mathsf{N}\left(\log\left(\frac{r^{\mathrm{out}\,2}}{100^2}\right), \log(100)^2\right), \tag{7a}$$

$$l_G \sim \mathsf{N}\left(\log(1), \log(100)^2\right), \tag{7b}$$

$$l_{\rho_k^0}, l_{\rho_k^\epsilon} \sim \mathsf{N}\left(\log\left(\frac{b_k-a_k}{2}\right), \log(10)^2\right),\ 1\le k\le d, \tag{7c}$$

$$l_L \sim \mathsf{N}(\log(4), \log(3)^2), \tag{7d}$$

$$\left(l_{\lambda(t_s)}\right)_{1\le s\le S} \sim \mathsf{N}\left(\log\left(\frac{r^{\mathrm{out}\,2}}{100^2}\right)\mathbb{1}_S, \log(100)^2\cdot((1-c)I_S + cU_S)\right). \tag{7e}$$

with c the correlation between two noise variances, $\mathbb{1}_S$ the vector of ones of length S, I_S the identity matrix of size S, and U_S the square matrix of ones with size S. To select the prior distributions, we propose a reference value for each hyper-parameter, and add a large prior uncertainty to get weakly-informative prior distributions. The parameters σ_0^2 and $G\sigma_0^2$ are assumed to be approximatively equal to $\left(\frac{r^{\text{out}}}{100}\right)^2$. The range parameters $(\rho_k^0, \rho_k^\epsilon)_{1\leq k\leq d}$ are assumed to be about the half of the domain $\rho_k \approx \frac{b_k - a_k}{2}$. For the degree parameter L, the mean is a value recommended by [7].

The noise variances are assumed to be about $\left(\frac{r^{\text{out}}}{100}\right)^2$, with a large standard deviation. However, we also assume that the prior uncertainty on the difference between two log-noise variance are really small with respect to the uncertainty of the noise variance, $\mathbb{V}\text{ar}\left[\log(\lambda(t_1)) - \log(\lambda(t_2))\right] \ll \mathbb{V}\text{ar}\left[\log(\lambda(t_1))\right]$. Consequently, we assume [6] a strong correlation between log-noise variances, which is set to $c = 99\%$. This assumption helps to estimate noise variance on the levels with few observations.

Once the prior distribution are defined, we can compute the posterior distribution conditionally to observations using Bayes theorem. Let $\chi_n = (x_i, t_i; z_i)_{1\leq i\leq n}$ denote n observations of the simulator. Because of the assumption of normal output distribution and Gaussian process with unknown mean (Equations (2) and (4)), the prior and posterior processes conditioned by θ are Gaussian. Thus, for any vector of outputs Z at given input vectors, $\pi(Z|\chi_n, \theta)$ and $\pi(\chi_n|\theta)$ are Gaussian multivariate distributions, whose mean and covariance are given by the kriging equations [5].

The posterior distribution of θ can be expressed with Bayes formula up to a normalizing constant: $\pi(\theta|\chi_n) \propto \pi(\chi_n|\theta) \cdot \pi(\theta)$. As there is no close expression of this posterior distribution, we sample it using a Monte-Carlo method. More precisely, we use the adaptive Metropolis-Hastings algorithm proposed by [2] to get samples $(\theta_j)_{1\leq j\leq p}$, distributed according to $\pi(\theta|\chi_n)$.

The sampled hyper-parameters are used to compute the probability distribution of the PoF P (1). Since the density of P is intractable, we use a Monte-Carlo method to draw samples from the posterior distribution $\pi(P|\chi_n)$. At each fixed θ_j, first, m inputs are drawn according to the input distribution $X^{(j)} = \left(x_i^{(j)}\right)_{1\leq i\leq m}$, $x_i^{(j)} \sim f_{\mathbb{X}}$. Then, we draw q Gaussian sample paths at the inputs $X^{(j)}$ at the reference level $\left(\xi^{(l)}_{\chi_n,\theta_j}\left(x_i^{(j)}, t^{\text{ref}}\right)\right)_{1\leq i\leq m,\, 1\leq l\leq q}$ and compute the probability function $p_j^{(l)}\left(x_i^{(j)}, t^{\text{ref}}\right) = \Phi\left(\frac{\xi^{(l)}_{\chi_n,\theta_j}\left(x_i^{(j)}, t^{\text{ref}}\right) - z^{\text{crit}}}{\sqrt{\lambda^{(l)}_{\theta_j}(t^{\text{ref}})}}\right)$. Finally, the samples $(P_{j,l})_{1\leq j\leq p,\, 1\leq l\leq q}$ are computed by averaging on the input space, $P_{j,l} = \frac{1}{m}\sum_{i=1}^m p_j^{(l)}\left(x_i^{(j)}, t^{\text{ref}}\right)$. With this sample, we can estimate the PoF with a measure of uncertainty, for instance, by computing the empirical median and a 95% confidence interval.

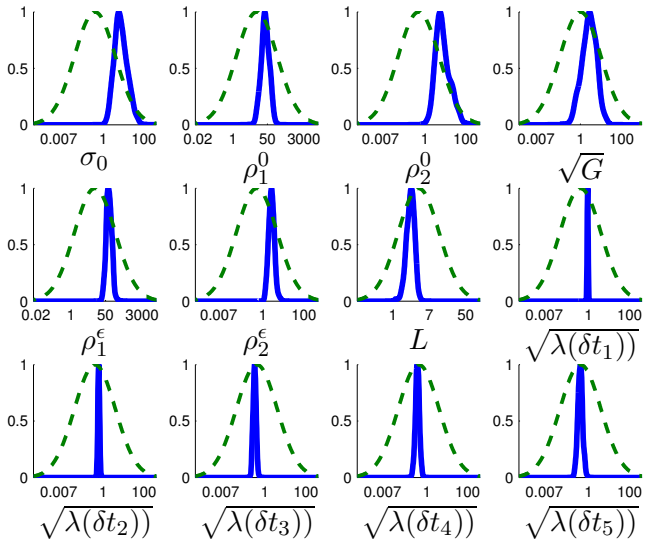


Figure 1: Normalized densities of the hyper-parameters of the multi-fidelity Gaussian process. The solid blue lines are the posterior densities, and the green dashed lines the prior densities. The abscissa axes are in logarithmic scale.

4 Application

The algorithm is illustrated on a random damped harmonic oscillator from [1]. Consider $X(t)$ the solution of the second-order differential stochastic equation, driven by a Brownian motion with spectral density equal to one, and with $X(t=0)=0$ and $\dot{X}(t=0)=0$ as initial conditions. The parameters of the differential equation, the natural pulse ω_0 and the damping ratio ζ, are the $d=2$ inputs of the simulator. The stochastic equation is solved on a period $t \in [0; t^{\text{end}}]$, with $t^{\text{end}} = 30$s, by an explicit exponential Euler scheme, which approximates X by a sequence $\widetilde{X}_n \approx X(n \cdot \delta t)$. The time step δt is the fidelity parameter. The multi-fidelity simulator is

$$f : (\omega_0, \zeta, \delta t) \mapsto \max_{0 \leq n \leq \left\lceil \frac{t^{\text{end}}}{\delta t} \right\rceil} \left\{ \log\left(|\widetilde{X}_n| \right) \right\}, \tag{8}$$

with $\omega_0 \in [0; 30]$ rads^{-1}, $\zeta \in [0; 1]$ and $\delta t \in [0; 1]$ s. The cost of this simulator is linear in $1/\delta t$: observing the level δt (in seconds) costs $C(\delta t) = \frac{2.61}{\delta t} + 5.45$ (in milliseconds). The approximate output range is $r^{\text{out}} = 40$.

For this article, we consider $S = 5$ levels of fidelity: $\delta t = 1$, 0.5, 0.1, 0.05, and 0.01 s. The multi-fidelity design is a Nested Latin Hypercube Sampling (NLHS) with respectively 168, 56, 28, 14 and 7 points at each level

of fidelity, generated with the algorithm of [4] and a maximin optimization. The adaptive Metropolis algorithm is applied to draw $p = 10^3$ vectors θ. The figure 1 represents the normalized marginal prior and posterior distributions of θ, the latter being estimated with a kernel density method.

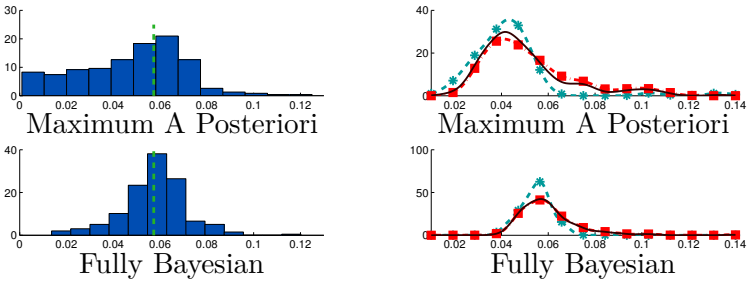
The marginal posterior distributions are more concentrated than their prior counterparts, indicating that the observations χ_n brings information about the hyper-parameters. Particularly, for noise variances ($\lambda(\delta t_s)$), the strong correlation between levels allows to reduce the uncertainties of all noise variances, including those from levels with few observations. The value of L is rather well-estimated, an observation which is opposite to the one in [7], which recommends to fix the value.

With the sampling of hyper-parameters, we can estimate the posterior distribution of the PoF. The input distribution $f_{\mathbb{X}}$ is an uniform distribution on the input space $[0; 30]\text{rads}^{-1} \times [0; 1]$, the critical threshold is $z^{\text{crit}} = 1$. In order to make a comparison, we estimate the PoF at an observable fidelity-level, fixed to $\delta t^{\text{ref}} = 0.01\text{s}$. We compute a reference value $P^\star = 5.73\%$.

We compare two different methods of estimation: a Fully Bayesian (FB) approach, and a plug-in approach, where hyper-parameters are replaced by their Maximum A Posteriori (MAP). Our methodology is applied on 240 independent experiments. On these experiments, the input and outputs observations change, but the models and their priors are fixed. For each experiment, the posterior density of the PoF of the four models is sampled, which gives 240×2 posterior densities. From these posterior densities of the PoF, we compute the median, and the 95% confidence intervals.

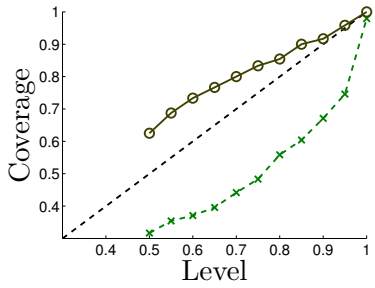
Figure 2(a) displays the empirical histograms of the 240 medians of the posterior distributions of the PoF. We can see that the medians returned by FB approach vary less from an experiment to another than those returned by MAP. Figure 2(b) plots the empirical densities of the 240 lengths of the 95% confidence intervals, estimated by kernel density regression. We can see that, for all approaches, the failing intervals have a smaller length than the successful intervals. We can also see that the FB approach always provides non-zero confidence intervals, opposite to MAP approach. Figure 2(c) presents the capacity of the models to catch the reference value. Each curve corresponds to one approach. Each point of the curve at abscissa p is the coverage, the proportion of the confidence intervals of level p which contain the reference, according to the associated approach. We can see that the FB approach provides much more conservative intervals than MAP approach.

The three Figures 2(a), 2(b) and 2(c) suggest that, on this example, the FB approach returns a better posterior distribution than the MAP approach.



(a) Posterior medians

(b) Lengths of 95% intervals



(c) Coverage rate versus level of the confidence intervals

Figure 2: (a) Histograms of the medians. The vertical dotted line is the reference. (b) Estimated densities of the lengths of the 95% confidence interval. The dashed lines with squares correspond to intervals which contain the reference value, the dashed lines with stars to those which miss it, and the solid lines to all intervals. (c) Coverage of confidence intervals at level p. The coverage is the proportion of cases where the reference value is inside the confidence interval. The solid and dashed-crossed lines corresponds respectively to the FB and MAP approaches.

5 Conclusion

In this article, we propose a Bayesian model for stochastic multi-fidelity numerical model. The model is based on a Gaussian process, completed with prior distributions on the hyper-parameters of the covariance function and on noise variances. By comparing prior and posterior hyper-parameter distributions, we see that observations bring informations about the hyper-parameters. Using sampling algorithms, we can sample the posterior distribution of the quantity of interest, here a Probability of Failure (PoF). By comparing the Fully Bayesian approach with Maximum A Posteriori plug-in approach, we can see that, on an academic example, the Fully Bayesian approach provides more robust confidence intervals of the PoF. However, the priors require care when using the models. Future work will focus on assessing the impact of the different prior modeling choices on the posterior distributions of hyper-parameters and of quantities of interest.

References

[1] Siu-Kui Au and James L. Beck. Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics*, 16(4):263–277, 2001.

[2] Heikki Haario, Eero Saksman, and Johanna Tamminen. An adaptive metropolis algorithm. *Bernoulli*, 7(2):223–242, 2001.

[3] Victor Picheny and David Ginsbourger. A nonstationary space-time gaussian process model for partially converged simulations. *SIAM/ASA Journal on Uncertainty Quantification*, 1(1):57–78, 2013.

[4] Peter Z. G. Qian and C. F. Jeff Wu. Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments. *Technometrics*, 50(2):192–204, 2008.

[5] Thomas J. Santner, Brian J. Williams, and William I. Notz. *The Design and Analysis of Computer Experiments*. Springer Series in Statistics. Springer, New York, 2003.

[6] Rémi Stroh, Julien Bect, Séverine Demeyer, Nicolas Fischer, Marquis Damien, and Emmanuel Vazquez. Assessing fire safety using complex numerical models with a bayesian multi-fidelity approach. *Fire Safety Journal*, 91:1016–1025, 2017.

[7] Rui Tuo, C. F. Jeff Wu, and Dan Yu. Surrogate modeling of computer experiments with different mesh densities. *Technometrics*, 56(3):372–380, 2014.