13th International Conference on Applications of Statistics and Probability in Civil Engineering, ICASP13
Seoul, South Korea, May 26-30, 2019

Reliability assessment by adaptive kernel-based surrogate models – Approximation of non-smooth limit-state functions

Jean-Marc Bourinet
Associate Professor, Institut Pascal, Sigma Clermont, Clermont Ferrand, France

ABSTRACT: Adaptive surrogate models are of practical use for reliability analysis based on costly-to-evaluate limit-state functions. The quality of the approximation made depends on both the selected type of surrogate model (and its related assumptions) and the adaptive scheme applied for the construction of the approximate model. Most of surrogate models assume some degree of smoothness, which allows them to be learned with a not so large set of input-output data pairs. This paper investigates the use of Matérn kernels in the context of support vector regression, with tuned regularity parameters. This kernel is used in an adaptive scheme based on MCMC sampling, whose objective is to progressively sample the failure domain. The proposed approach is applied to both a smooth and a non smooth limit-state functions, showing the benefits of using such a highly flexible kernel.

1. INTRODUCTION

The focus of this paper is on reliability estimation in problems involving costly-to-evaluate models (referred to as true models) such as those encountered e.g. in the numerical solving of PDEs. In this context we may try to learn the true model from a set of selected input-output data pairs and use the trained approximate model as a surrogate to the costly-to-evaluate model. Let us assume that the true model is defined by a scalar function $y : \mathscr{X} \to \mathscr{Y}, \mathrm{x} \mapsto y = y(\mathrm{x})$ where $\mathscr{X} \subseteq \mathbb{R}^n$ denotes the input space and $\mathscr{Y} \subseteq \mathbb{R}$ the output space. The surrogate of y will be denoted $\widetilde{y} : \mathrm{x} \mapsto \widetilde{y} = \widetilde{y}(\mathrm{x})$. $\widetilde{y}$ should be viewed as conditional on the N data pairs $\mathscr{T} = \{(\mathrm{x}_i, y_i) \in \mathscr{X} \times \mathscr{Y}, 1 \leq i \leq N\}$ it is trained on.

Learning the true function y from data is conditional on a set of predefined hypotheses on $\widetilde{y}$, e.g. polynomial basis for a polynomial response surface or a polynomial chaos expansion, type of kernel for kriging or support vector regression, etc. If this set of predefined hypotheses is not wide enough w.r.t. the complexity of the true function y, the trained approximate model $\widetilde{y}$ is expected to have a poor predictive capacity (e.g. approximation of true function y of a degree strictly greater than one by a linear approximate model $\widetilde{y}$). If this set is too wide (high complexity of $\widetilde{y}$), the learning process is prone to overfitting, which is also an unwanted situation. The contruction of surrogate models is therefore to find a good tradeoff between the choice of hypotheses selected for $\widetilde{y}$ and the unknown complexity of y. This choice of hypotheses for $\widetilde{y}$ also has consequences on the training process. A complex approximate model has numerous parameters, which are harder to tune in the context of training on small datasets. In this paper, we investigate the use of a kernel of complexity greater than that used in most common kernel-based approaches.

Training on an initial set of N datapairs $\mathscr{T}$ is known to not be the most efficient technique for the purpose of optimization or reliability assessment. Sequentially updated surrogate models are often preferred and the reader may refer to (Bourinet, 2018, section II–3.1) for a short review of such works in the context of reliability assessment. Sequential techniques were applied with several types of surrogate models, e.g. polynomial response surfaces, kriging (Bichon et al., 2008; Echard et al., 2011; Bect et al., 2012; Hu and Mahadevan, 2016),

support vector classification (Most, 2007; Basudhar and Missoum, 2010; Bourinet et al., 2011), support vector regression (Dai et al., 2012; Bourinet, 2016, 2017), polynomial chaos expansion (Marelli and Sudret, 2018), generalized ANOVA (Chakraborty and Chowdhury, 2016). In the following, s will denote the iteration number of such a sequential scheme, $\widetilde{y}_s$ the approximate model trained on a set of data pairs $\mathscr{T}_s$ at iteration s. This sequential construction is carried out until some accuracy criteria are met.

In this work, the reliability problem to solve is supposed expressed in the so-called standard normal space and the unknown failure probability p_f is given by the followwing n-fold integral:

$$p_\mathrm{f} = \mathbb{P}(G(\mathrm{U}) \leq 0) = \int_{G(\mathrm{u})\leq 0} \varphi_n(\mathrm{u})\,\mathrm{d}\mathrm{u} \qquad (1)$$

where $G : \mathbb{R}^n \to \mathbb{R}$ is the costly-to-evaluate limit-state function (LSF), φ_n is the n-dimensional standard normal PDF and $\mathrm{d}\mathrm{u} = \mathrm{d}u_1 \ldots \mathrm{d}u_n$.

Section 2 introduces the type of surrogate model used in the present work with its main assumptions and details about its tuning. The proposed sequential training close to that proposed in previous papers of the author (Bourinet, 2016, 2017) is presented in Section 3. Two application examples are addressed in Section 4, which points out the benefits of using the selected kernel. Concluding remarks and perspectives are given in Section 5.

2. SURROGATE MODEL

The proposed surrogate-based reliability assessment method makes use of support vector regression (Vapnik, 1995; Drucker et al., 1997). This technique shares some similarities with Gaussian process emulators a.k.a. kriging (see, e.g., Bourinet, 2018, section II–4). From a given set of training data pairs $\mathscr{T} = \{(\mathrm{x}_i, y_i), 1 \leq i \leq N\}$ where $y_i = y(\mathrm{x}_i)$ for $i = 1, \ldots, N$, the objective is to predict a scalar output $y \in \mathbb{R}$ at any new point $\mathrm{x} \in \mathbb{R}^n$. The univariate regression problem is solved by means of an approximation $\widetilde{y}$ of y in the form:

$$\widetilde{y}(\mathrm{x}) = h(\mathrm{x}) + b \;\text{ where }\; h(\mathrm{x}) = \sum_{i=1}^{N} c_i k(\mathrm{x}_i, \mathrm{x}) \qquad (2)$$

and where $\mathrm{c} = (c_1, \ldots, c_N)^\mathrm{T} \in \mathbb{R}^N$ is a vector of unknown expansion coefficients, $b \in \mathbb{R}$ is an unknown unregularized bias term and $k : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is a selected positive definite kernel. Such a representation is the solution of the minimization problem:

$$\min_{\mathrm{c},b} C \sum_{i=1}^{N} \ell(y_i, (\mathrm{Kc})_i + b) + \frac{1}{2}\mathrm{c}^\mathrm{T}\mathrm{Kc} \qquad (3)$$

where $\mathrm{K} = [k(\mathrm{x}_i, \mathrm{x}_j)]_{1\leq i,j\leq N}$ is the so-called Gram matrix, where $(\mathrm{Kc})_i = \sum_{j=1}^{N} c_j k(\mathrm{x}_j, \mathrm{x}_i)$, where $\ell(y,u)$ is taken here as the ε-insensitive loss function such that $\ell(y,u) = 0$ if $|y-u| < \varepsilon$, $|y-u| - \varepsilon$ otherwise (specific case of L1-SVR), and where C is a regularization parameter. This parameter C controls the tradeoff between finding a function h (and therefore $\widetilde{y}$) of low complexity and fitting well the training data, which thus avoids overfitting.

Several types of kernels k can be selected for the analysis. It is important to point out that the choice made for k is part of the hypotheses which control the accuracy of the approximation $\widetilde{y}$ to be constructed. The practice may differ according to the problems of interest. In the SVM community, the objective is to learn a model from a large number of data pairs and kernels with a very few parameters to tune are often preferred. In the context of small datasets such as encountered in surrogate-based approximation of costly-to-evaluate models, kernels with more parameters to tune may be considered. A common practice with kriging model is to consider e.g. anisotropic Gaussian or Matérn kernels, which are more complex therefore more flexible than their isotropic counterparts (see, e.g., Moustapha et al., 2018, for a comparison of kriging and SVR with isotropic/anisotropic kernels).

In the following, we investigate the use an anisotropic form of the Matérn kernel (Stein, 1999) as a tentative to approximate non smooth functions. The common practice with such a kernel is to consider a single scale parameter γ (isotropic form) or one scale parameter γ_i per input dimension of x (anisotropic form), along with an apriori selected regularity parameter ν which controls the smoothness of the approximation, most often $\nu = 5/2$. It is reminded that the Matérn kernel includes the Gaussian RBF kernel as the limiting case $\nu \to \infty$.

The idea here is to use a Matérn kernel which is anisotropic both in terms of the scale parameter γ and the smoothness parameter ν. The following kernel k is considered, for $\mathrm{x}, \mathrm{x}' \in \mathbb{R}^n$:

$$k(\mathrm{x}, \mathrm{x}'; \theta) = \prod_{i=1}^{n} k_{\text{Matérn-1D}}(x_i, x_i'; \nu_i, \gamma_i) \qquad (4)$$

where $\theta = \{(\nu_i, \gamma_i), i = 1, \ldots, n\}$ is a set of parameters to tune and $k_{\text{Matérn-1D}}$ is the 1D Matérn kernel:

$$k_{\text{Matérn-1D}}(x, x'; \nu, \gamma) = \ldots \\ \frac{\left(\gamma\sqrt{2\nu}\,|x-x'|\right)^{\nu}}{2^{\nu-1}\Gamma(\nu)} K_\nu\left(\gamma\sqrt{2\nu}\,|x-x'|\right) \qquad (5)$$

where $x, x' \in \mathbb{R}$, $\nu \geq \frac{1}{2}$ is a smoothness parameter, $\gamma > 0$ is a scale parameter, K_ν is the modified Bessel function of 2nd kind of order ν and Γ is the Gamma function.

Training an approximate SVR model $\widetilde{y}$ requires a proper tuning of its hyperparameters, here the $2n$ kernel parameters gathered in θ, additionally to the regularization parameter C and the width ε of the so-called ε tube specific to the ε-insensitive loss function. The hyperparameter selection problem is solved by minimizing an approximation of the leave-one-out (LOO) error proposed by Chang and Lin (2005), as a good proxy to the generalization error. The optimization problem to solve is non smooth and has $2n+2$ parameters to find. For these reasons, we favor a stochastic search with the covariance matrix adaptation evolution strategy (CMA-ES) algorithm (Hansen, 2016), such as described by Moustapha et al. (2018). The search for the smoothness parameters $(\nu_i, 1 \leq i \leq n)$ is bounded to the range $[1,4]$. We select each ν_i as the closest values to $3/2, 5/2$ or $7/2$ for which the kernel has a simple expression (Abramowitz and Stegun, 1965, Eq. 10.2.15).

3. ADAPTATIVE METHOD

The proposed adapative method consists in constructing a sequence of SVR models in the standard normal space $\{\widetilde{G}_s : \mathrm{u} \mapsto \widetilde{G}_s(\mathrm{u}), 1 \leq s \leq s_{\max}\}$, whose last elements are used to define an estimate of the failure probability p_f, see Eq. (1). The main steps of the proposed algorithm are given in Table 1. This algorithm is close to that proposed by Bourinet (2016), with some of the modifications brought in (Bourinet, 2017). The whole algorithm has three main phases: phase 1 until the intermediate LSF level y_s gets lower than zero, phases 2 and 3 where y_s is set to zero. Phase 2 ends when the convergence criteria described later in this section become lower than some prescribed limit values. Phases 3 is run over a given maximal number of iterations s.

1	Define initial dataset $\mathscr{D} = \{ (\mathrm{u}_i, G(\mathrm{u}_i)) , 1 \leq i \leq N \}$ and initial LSF level y_1
2	Iterate on s until convergence
3	Define training set $\mathscr{T}_s = \{ (\mathrm{u}_i, G(\mathrm{u}_i)) , i \in \mathscr{I}_{\text{train},s} \}$
4	Train LSF surrogate $\widetilde{G}_s$ on $\mathscr{T}_s$
5	Compute $\widehat{p}_s$ estimate of $p_s = \mathbb{P}(\overline{\widetilde{G}}_s(\mathrm{U}) \leq y_s)$ where $\overline{\widetilde{G}}_s$ is a weighted averaged surrogate, see Eq. (6)
6	Select N_a new training points, add N_a associated data pairs to $\mathscr{D}$ and update y_s
7	Compute convergence criteria
8	Compute failure probability estimate $\widehat{p}_\mathrm{f}$

Table 1: Flowchart of the proposed algorithm.

(a) *Training set*. The initial training set $\mathscr{T}_1$ is composed of $N = 50$ standard Gaussian samples. $N_\mathrm{a} = 5$ new points are defined at each iteration, see (e), and the associated data pairs become part of the set $\mathscr{D}$ of available data pairs. A key idea of the proposed method is to set the size of $\mathscr{T}_s$ to $N_{\min} \leq \#\mathscr{D}$, where # denotes the cardinal of a set, i.e. to construct a local approximation by regression. $N_{\min}$ is initially set to 100 and increased if $r_{\overline{\mathscr{D}}_1 \mathscr{D}_2; 0.5}(s)$ defined in (f) exceeds a threshold value of 0.3. The training set $\mathscr{T}_s$ is composed of the data pairs of $\mathscr{D}$ with the lowest LSF evaluations.

(b) *Intermediate LSF level y_s*. The method consists in sampling the standard normal space until the failure domain is reached. This is achieved by decreasing an intermediate LSF level y_s with iteration s. We select $y_s = y^{(25)}$ where $\{y^{(i)}, 1 \leq i \leq \#\mathscr{D}\}$ are the LSF values of the data pairs in $\mathscr{D}$ sorted in ascending order.

(c) *Weighted averaged SVR models*. At each iteration s, a SVR model $\widetilde{G}_s$ is trained on $\mathscr{T}_s$. We also consider the weighted averaged SVR

model over the last 5 iterations:

$$\overline{\widetilde{G}}_s(\mathrm{u}) = \sum_{k=s-4}^{s} \omega_k \widetilde{G}_k(\mathrm{u}) \,, \tag{6}$$

for any $\mathrm{u} \in \mathbb{R}^n$, where $\omega_k = (5+k-s)/15$ for $k = (s-4), \ldots, s$. The deviation between the two hypersurfaces $\{\mathrm{u} \in \mathbb{R}^n : \overline{\widetilde{G}}_s(\mathrm{u}) = y_s\}$ and $\{\mathrm{u} \in \mathbb{R}^n : \widetilde{G}_s(\mathrm{u}) = y_s\}$ is used as a convergence criterion, see ratio $r_{\overline{\mathscr{D}}_1\mathscr{D}_2;0.5}(s)$ defined in (f). The weighted averaged SVR model is moreover used for the approximation of the probability $\mathbb{P}(G_s(\mathrm{U}) \leq y_s)$ at each iteration s, as it is found of greater accuracy than $\widetilde{G}_s$.

(d) *Subset simulation, sample sets $\mathscr{U}_{\mathrm{SS};s}$ and $\overline{\mathscr{U}}_{\mathrm{SS};s}$.* Subset simulation (SS) (Au and Beck, 2001) is applied at each iteration s with the following LSFs:

$$\widetilde{G}_s(\mathrm{u}) - y_s \quad \text{and} \quad \overline{\widetilde{G}}_s(\mathrm{u}) - y_s \tag{7}$$

The settings are the same as those of Bourinet (2017): sample set per level of size 100,000 and target probability level of 0.5. The probability estimated by SS based on $\overline{\widetilde{G}}_s(\mathrm{u}) - y_s$ is denoted $\widehat{p}_s$. The samples of the last SS level satisfying $\widetilde{G}_s(\mathrm{u}) < y_s$, resp. $\overline{\widetilde{G}}_s(\mathrm{u}) < y_s$, are denoted $\mathscr{U}_s$, resp. $\overline{\mathscr{U}}_s$. These samples are used to define the convergence criteria of the algorithm, see (f). New training points at iteration $s+1$ are samples of the set $\mathscr{U}_{\mathrm{SS};s}$, see (e).

(e) *Enrichment strategy.* The number N_a of new points added to the set of available data pairs $\mathscr{D}$ is set to 5 at each iteration s. In phase 1, these N_a points are uniformly drawn from the subsets $\mathscr{B}_0$ and $\mathscr{B}_1$ represented in Figure 1. The number of samples drawn from each subset is respectively proportional to $\#\mathscr{B}_0/\#\mathscr{U}_s$ and $\#\mathscr{B}_1/\#\mathscr{U}_s$. In phase 2 and 3, we restrict the selection to samples in $\mathscr{B}_1$ and $\mathscr{B}_2$.

(f) *Convergence criteria.* The algorithm iterates over s first until y_s decreases down to zero (phase1). y_s is set to zero for subsequent iterations (phase 2 and 3). The objective of phases 2 and 3 is to decrease the following ratios (see Bourinet, 2017, Figure 1):

$$\begin{aligned}
r_{\mathscr{B}_1\mathscr{B}_2}(s) &= \left(\#\mathscr{B}_1/\#\mathscr{U}_s + \#\mathscr{B}_2/\#\mathscr{U}_{s-1}\right)/2 \,, \\
r_{\overline{\mathscr{B}}_1\overline{\mathscr{B}}_2}(s) &= \left(\#\overline{\mathscr{B}}_1/\#\overline{\mathscr{U}}_s + \#\overline{\mathscr{B}}_2/\#\overline{\mathscr{U}}_{s-1}\right)/2 \,, \\
r_{\overline{\mathscr{D}}_1\mathscr{D}_2}(s) &= \left(\#\overline{\mathscr{D}}_1/\#\overline{\mathscr{U}}_s + \#\mathscr{D}_2/\#\mathscr{U}_{s-1}\right)/2 \,,
\end{aligned} \tag{8}$$

which quantify the closeness between:

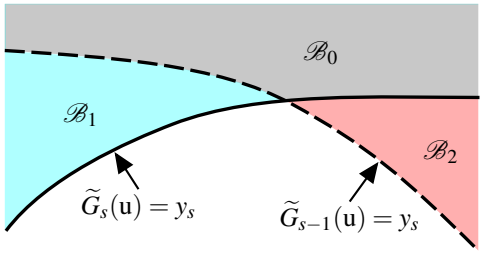
- $\widetilde{G}_s(\mathrm{u}) = y_s$ and $\widetilde{G}_{s-1}(\mathrm{u}) = y_s$ for $r_{\mathscr{B}_1\mathscr{B}_2}(s)$,
- $\overline{\widetilde{G}}_s(\mathrm{u}) = y_s$ and $\overline{\widetilde{G}}_{s-1}(\mathrm{u}) = y_s$ for $r_{\overline{\mathscr{B}}_1\overline{\mathscr{B}}_2}(s)$,
- $\overline{\widetilde{G}}_s(\mathrm{u}) = y_s$ and $\widetilde{G}_s(\mathrm{u}) = y_s$ for $r_{\overline{\mathscr{D}}_1\mathscr{D}_2}(s)$.

Phase 2 iterates until:

$$\begin{aligned}
& r_{\mathscr{B}_1\mathscr{B}_2;0.5} < 0.2 \\
\text{and} \quad & r_{\overline{\mathscr{B}}_1\overline{\mathscr{B}}_2;0.5} < 0.03 \\
\text{and} \quad & r_{\overline{\mathscr{D}}_1\mathscr{D}_2;0.5} < 0.2
\end{aligned} \tag{9}$$

where the subscript 0.5 indicates a median value of these ratios over 5 iterations. A maximal number of 20 iterations is carried out in phase 3, unless $r_{\overline{\mathscr{B}}_1\overline{\mathscr{B}}_2;0.5} < 0.01$ which ends the algorithm.

(g) *Failure probability estimate $\widehat{p}_\mathrm{f}$.* The estimation of p_f is obtained as an average of the last 5 estimates $\widehat{p}_s$, which are themselves based on the associated weighted averaged surrogates $\overline{\widetilde{G}}_s$.



$$\mathscr{B}_0 = \left\{\mathrm{u} \in \mathscr{U}_s : \widetilde{G}_s(\mathrm{u}) < y_s \;\&\; \widetilde{G}_{s-1}(\mathrm{u}) < y_s\right\}$$
$$\mathscr{B}_1 = \left\{\mathrm{u} \in \mathscr{U}_s : \widetilde{G}_s(\mathrm{u}) < y_s \;\&\; \widetilde{G}_{s-1}(\mathrm{u}) > y_s\right\}$$
$$\mathscr{B}_2 = \left\{\mathrm{u} \in \mathscr{U}_{s-1} : \widetilde{G}_{s-1}(\mathrm{u}) < y_s \;\&\; \widetilde{G}_s(\mathrm{u}) > y_s\right\}$$

Figure 1: Sample sets $\mathscr{B}_0$, $\mathscr{B}_1$ and $\mathscr{B}_2$.

An efficient computer implementation is needed to keep acceptable the training time needed by the proposed algorithm. In the work we use both parallel and GPU-accelerated computations. The whole training time can be reduced at the following two levels:

- The λ quadratic programming problems to solve at each iteration of the CMA-ES algorithm for minimizing the LOO error estimate are sent in parallel on a multicore CPU.
- The several SVR evaluations on new u-samples are run on a GPU. This drastically reduce the testing time resulting from the choice of model averaging over 5 iterations and a large sample size at each SS level (here 100,000).

4. APPLICATION EXAMPLES

4.1. Example 1: a smooth limit-state surface

The first example is the two-degree-of-freedom primary-secondary system under white noise base acceleration initially proposed by De Stefano and Der Kiureghian (1990). We focus here on the reliability problem initially studied by Bourinet et al. (2011) with a mean force capacity F_s of the secondary spring set to 27.5, which results in a failure probability equal to 3.78×10^{-7}. This problem has 8 independent lognormally-distributed random variables and has an analytically-defined LSF (see Bourinet, 2018, Section I–3.4.2). It is characterized by a strongly curved limit-state surface (LSS) at a single most probable failure point (MPFP). The LSS albeit highly nonlinear is smooth, which makes it easy to approximate with a Gaussian kernel, either in its isotropic or anisotropic form (Bourinet, 2017). This example is selected to investigate how the selected Matérn kernel with anisotropic regularity parameters fits this problem in comparison with the Gaussian kernel.

The result obtained with the proposed approach is given in Table 2, along with those obtained in previous works. The evolution of $\widehat{p}_s$ with the number of LSF evaluations N_s is plotted in Figure 2. The failure probability approximation obtained as the average of the last 5 $\widehat{p}_s$-estimates is represented in Figure 3. The proposed method has an efficiency comparable with that of previous instances of the method based on a Gaussian kernel. The cost in terms of LSF evaluations is found lower that based on the anisotropic Gaussian kernel (575 vs. 690), but still greater than that based on the isotropic Gaussian kernel (540 LSF evaluations).

Table 2: Example 1. Results

Method	p_f estimate (# LSF calls)
SS (reference)	3.78×10^{-7}
^{2}SMART [(a)]	3.66×10^{-7} (4011)
Meta-IS [(b)]	3.76×10^{-7} (680)
ASVR(-iso) [(c)]	3.81×10^{-7} (648)
iASVR-iso [(d)]	3.72×10^{-7} (540)
iASVR-ani [(d)]	3.73×10^{-7} (690)
Proposed method	3.78×10^{-7} (575)

[(a)]: (Bourinet et al., 2011)
[(b)]: (Dubourg et al., 2013)
[(c)]: (Bourinet, 2016)
[(d)]: (Bourinet, 2017)

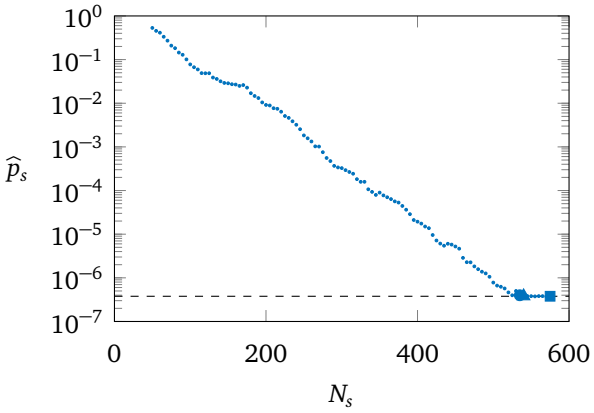


Figure 2: $\widehat{p}_s$ vs. number of LSF evaluations N_s (solid circle: end of phase 1, solid triangle: end of phase 2, solid square: end of phase 3).

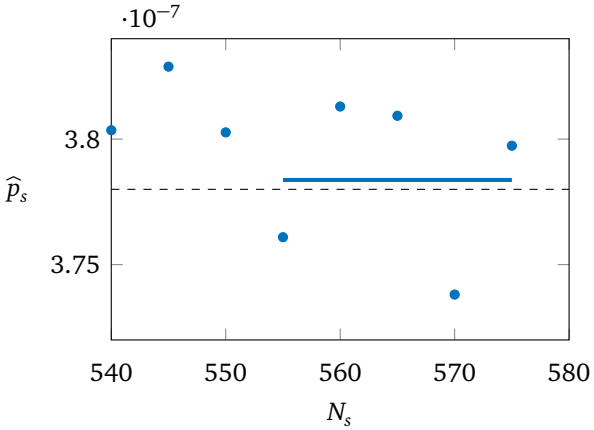


Figure 3: Last iterations of the algorithm and failure probability approximation $\widehat{p}_f$ (solid line).

4.2. Example 2: a non smooth limit-state surface

This second example is an electromagnetic compatibility reliability problem studied by Kouassi et al. (2016). It investigates a lossy transmission line of length L, diameter d and attenuation coefficient α. The line is placed at a uniform height h (considered as random in the present study) above a perfectly conducting ground plane and loaded at both ends by two impedances Z_0 and Z_L. This line is illuminated at a frequency f by a linearly polarized plane wave with incidence angles ϕ_p (azimut angle) and θ_p (elevation angle). The polarization angle and the magnitude of the electric field $\mathbf{E}$ are denoted θ_e and a_e, respectively.

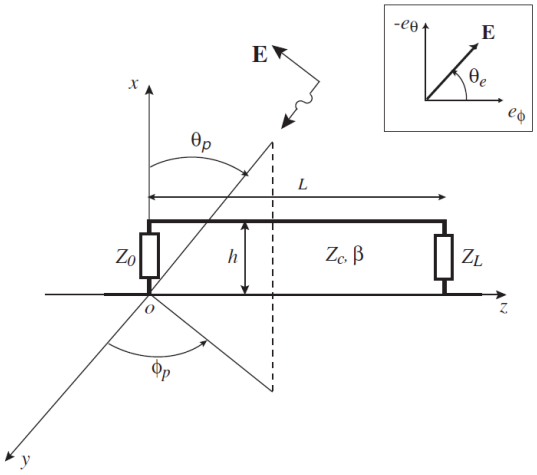


Figure 4: Transmission line under study.

The LSF is given by:

$$g(\mathrm{x}) = I_{\mathrm{cr}} - I(\mathrm{x}) \ , \tag{10}$$

where $I_{\mathrm{cr}} = 1.5 \times 10^{-4}$ A is a given current magnitude level to be not exceeded, where $I(\mathrm{x})$ is the magnitude of the electric current in the output load impedance Z_L (see Bourinet, 2018, appendix A, for its analytically-defined expression), and where $\mathrm{x} = (L, h, d, Z_L, Z_0, a_e, \theta_e, \theta_p, \phi_p, f, \alpha)$ is the vector of input parameters. The 11 input parameters are modeled as mutually independent random variables (see Bourinet, 2018, appendix A, for the distribution parameters selected in the analysis).

This example is characterized by a non smooth LSS, which can be explained by an electric current highly sensitive to the azimuth angle ϕ_p. Failures are lileky to occur for values of ϕ_p close to zero and π, which results in a failure domain with very sharp peaks pointing towards the origin, see bottom plots of Figure 5. As a consequence such a LSS is extremely hard to approximate with usual types of surrogate models, which in general are based on the assumption of smooth solutions.

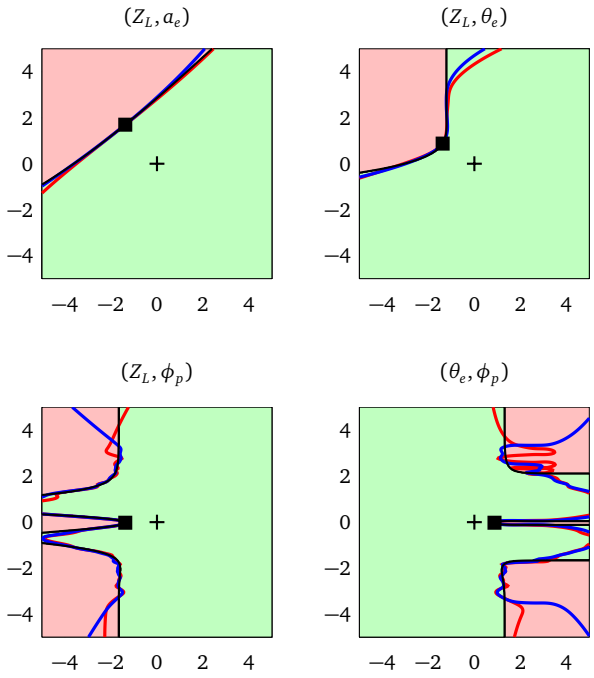


Figure 5: (u_i, u_j)-pairwise cross-cuts passing through the unique MPFP in standard normal space. Black square: MPFP, black cross: projection of standard normal space origin onto (u_i, u_j) cross-cut plane, black line: true LSS, green area: safe domain, pink area: failure domain, blue line: approximate LSS $\{\mathrm{u} \in \mathbb{R}^n : \widetilde{G}_s(\mathrm{u}) = 0\}$ at end of phase 3, red line: approximate LSS $\{\mathrm{u} \in \mathbb{R}^n : \widetilde{\widetilde{G}}_s(\mathrm{u}) = 0\}$ at end of phase 3.

The failure probability approximation obtained by the proposed method is $\widehat{p}_{\mathrm{f}} = 2.16 \times 10^{-4}$ with a total number of 730 LSF evaluations. The relative error on p_{f} is less than 4%, compared with the reference solution of 2.24×10^{-4} obtained by crude Monte Carlo with 10^9 samples. As represented in Figure 5, the Matérn kernel with tuned regularity parameters is able to fit the peaks of the LSS at MPFP in the standard normal space. It is noticed that the smoothness parameter ν_i of the input ϕ_p is almost always tuned to $3/2$, compared to those of other inputs which are most often found to be equal

to $5/2$ or $7/2$. This is consistent with a lack of regularity of the LSS w.r.t. the input ϕ_p.

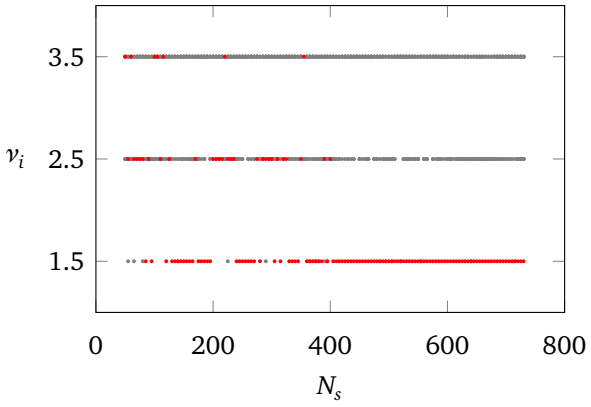


Figure 6: Smoothness parameter ν_i vs. number of LSF evaluations N_s. ν_i of ϕ_p plotted in red, ν_i of all other inputs plotted in grey.

5. CONCLUSION

This paper investigates the use of an anisotropic Matérn kernel with both a scale and a smoothness parameter per input dimension. This kernel has been used in the adaptive enrichment strategy already proposed in previous works of the author. The proposed kernel is able to handle smooth LSS, without loosing efficiency w.r.t. a Gaussian kernel. We also show the benefits of using such a kernel with controlled regularity, making it possible to approximate a non smooth LSS which cannot be handled by other types of surrogate models so far from the author's knowledge. Using such an approximation with high flexibility has a price to pay: the approximate model has many parameters to tune, which requires efficient computer implementations to keep training times to acceptable limits. As about perspectives, we will investigate the performances of the proposed method on other reliability problems hard to solve by means of surrogate models (e.g. high dimensional problem and multiple MPFPs).

6. REFERENCES

M. Abramowitz and I. A. Stegun, eds. (1965). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. Dover.

Au, S.-K. and Beck, J. L. (2001). "Estimation of small failure probabilities in high dimensions by subset simulation." *Probabilistic Engineering Mechanics*, 16(4), 263–277.

Basudhar, A. and Missoum, S. (2010). "An improved adaptive sampling scheme for the construction of explicit boundaries." *Structural and Multidisciplinary Optimization*, 42(4), 517–529.

Bect, J., Ginsbourger, D., Li, L., Picheny, V., and Vazquez, E. (2012). "Sequential design of computer experiments for the estimation of a probability of failure." *Statistics and Computing*, 22(3), 773–793.

Bichon, B. J., Eldred, M. S., Swiler, L. P., Mahadevan, S., and McFarland, J. M. (2008). "Efficient global reliability analysis for nonlinear implicit performance functions." *AIAA Journal*, 46(10), 2459–2468.

Bourinet, J.-M. (2016). "Rare-event probability estimation with adaptive support vector regression surrogates." *Reliability Engineering & System Safety*, 150, 210–221.

Bourinet, J.-M. (2017). "Anisotropic-kernel-based support vector regression for reliability assessment." *Proc. 12th International Conference on Structural Safety and Reliability (ICOSSAR 2017), Vienna, Austria, August 6–10, 2017*, TU Verlag.

Bourinet, J.-M. (2018). "Reliability analysis and optimal design under uncertainty – focus on adaptive surrogate-based approaches." HDR report, Université Clermont Auvergne, France, Université Clermont Auvergne, France, <https://tel.archives-ouvertes.fr/tel-01737299/document>.

Bourinet, J.-M., Deheeger, F., and Lemaire, M. (2011). "Assessing small failure probabilities by combined subset simulation and support vector machines." *Structural Safety*, 33(6), 343–353.

Chakraborty, S. and Chowdhury, R. (2016). "Sequential experimental design based generalised ANOVA." *Journal of Computational Physics*, 317, 15–32.

Chang, M.-W. and Lin, C.-J. (2005). "Leave-one-out bounds for support vector regression model selection." *Neural Computation*, 17(5), 1188–1222.

Dai, H., Zhang, H., Wang, W., and Xue, G. (2012). "Structural reliability assessment by local approximation of limit state functions using adaptive Markov chain simulation and support vector regression." *Computer-Aided Civil and Infrastructure Engineering*, 27(9), 676–686.

De Stefano, M. and Der Kiureghian, A. (1990). "An efficient algorithm for second-order reliability analysis." *Report no. UCB/SEMM-90/20*, Dept of Civil and Environmental Engineering, University of California, Berkeley.

Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A., and Vapnik, V. (1997). "Support vector regression machines." *Advances in Neural Information Processing Systems 9*, M. Mozer, M. Jordan, and T. Petsche, eds., MIT Press, 155–161.

Dubourg, V., Sudret, B., and Deheeger, F. (2013). "Metamodel-based importance sampling for structural reliability analysis." *Probabilistic Engineering Mechanics*, 33, 47–57.

Echard, B., Gayton, N., and Lemaire, M. (2011). "AK-MCS: An active learning reliability method combining kriging and Monte Carlo simulation." *Structural Safety*, 33(2), 145–154.

Hansen, N. (2016). "The CMA evolution strategy: a tutorial." *ArXiv e-prints*.

Hu, Z. and Mahadevan, S. (2016). "Global sensitivity analysis-enhanced surrogate (gsas) modeling for reliability analysis." *Structural and Multidisciplinary Optimization*, 53(3), 501–521.

Kouassi, A., Bourinet, J.-M., Lalléchère, S., Bonnet, P., and Fogli, M. (2016). "Reliability and sensitivity analysis of transmission lines in a probabilistic EMC context." *Electromagnetic Compatibility, IEEE Transactions on*, 58(2), 561–572.

Marelli, S. and Sudret, B. (2018). "An active-learning algorithm that combines sparse polynomial chaos expansions and bootstrap for structural reliability analysis." *Structural Safety*, 75, 67–74.

Most, T. (2007). "An adaptive response surface approach for structural reliability analyses based on support vector machines." *Proc. 11th International Conference on Civil, Structural and Environmental Engineering Computing, St. Julians, Malta, September 18–21, 2007*, B. H. V. Topping, ed., Civil-Comp Press, <http://dx.doi.org/10.4203/ccp.86.94>.

Moustapha, M., Bourinet, J.-M., Guillaume, B., and Sudret, B. (2018). "Comparative study of kriging and support vector regression for structural engineering applications." *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 4(2), 04018005.

Stein, M. L. (1999). *Interpolation of spatial data. Some theory for kriging*. Springer Series in Statistics. Springer, New York, NY, USA, <http://dx.doi.org/10.1007/978-1-4612-1494-6>.

Vapnik, V. (1995). *The nature of statistical learning theory*. Springer-Verlag, New York, NY, USA.