

Surrogate model based sequential sampling estimation of conformance probability for computationally expensive systems: application to fire safety science *

Titre: Estimation de probabilité de conformité par échantillonnage séquentiel sur métamodèle

Séverine Demeyer¹, Nicolas Fischer¹ and Damien Marquis¹

Abstract: The use of complex simulation systems has become common practice when physical experiments are not feasible or when too few are feasible. The statistical modelling of numerical experiments with kriging models yields a probabilistic decision framework to assess the probability of failure of the system. Combining fast low-fidelity simulations with costly high-fidelity simulations has proved an efficient method to decrease the burden of costly simulations when predicting the output of a system. In addition, sequential design is commonly used to estimate the probability of failure of a system modelled by kriging. In this work, a methodology is derived to benefit from sequential design in a multi-fidelity framework to predict the probability of failure of a computationally expensive system and its uncertainty. The methodology is applied to a fire safety engineering case study to assess the probability of non-conformity of a smoke control system from complex numerical fire tools.

Résumé : Le recours à la simulation numérique est devenue courant lorsque les expériences réelles sont impossibles ou réalisables qu'en très petit nombre. La modélisation statistique d'expériences numériques à partir de modèles de krigeage offre un cadre de décision probabiliste pour évaluer la probabilité de défaillance d'un système. La combinaison de simulations rapides de basse fidélité avec des simulations coûteuses de haute fidélité s'est avérée une méthode efficace pour diminuer le coût en simulations lors de la prévision de sorties d'un système. Par ailleurs, l'échantillonnage séquentiel est couramment utilisé pour estimer une probabilité de défaillance d'un système modélisé par krigeage. Dans cette étude, une méthodologie est exposée d'utilisation d'un plan séquentiel dans un cadre multi-fidélité pour prédire la probabilité de défaillance d'un système numérique coûteux et son incertitude. La méthodologie est appliquée à un cas d'étude en ingénierie de la sécurité incendie pour évaluer la probabilité de non-conformité d'un système d'évacuation de fumée à partir d'outils numériques complexes de simulation incendie.

Keywords: expériences numériques, méthode Monte Carlo, co-krigeage, échantillonnage séquentiel, probabilité de dépassement de seuil, évaluation de conformité

Mots-clés : numerical experiments, Monte Carlo Method, co-kriging, sequential sampling, probability of exceeding a threshold, conformity assessment

AMS 2000 subject classifications: 62L12, 65C05, 60G15, 62P30

* This work was part of a Joint Research Project within the European Metrology Research Programme EMRP under Grant Agreement No. 912/2009/EC.

¹ Laboratoire national de métrologie et d'essais, 29 avenue Roger Hennequin, 78197 Trappes
E-mail: severine.demeyer@lne.fr

1. Introduction

When dealing with a computational system, one may be interested in propagating the uncertainties associated with the input quantities to evaluate the uncertainty associated with the output quantity.

In the decision theoretical framework pertaining to conformity assessment, one is interested also in the position of the output variable with respect to a given threshold (regulatory threshold...). The problem of knowing whether the output of a computationally expensive model exceeds a given threshold is very common for reliability analysis and safety-critical applications such as aerospace, nuclear power stations, civil engineering (models of bridges and buildings), fire safety engineering etc. The probability that the system output exceeds a threshold (probability of failure) gives a reliable measure of non conformity.

The probabilistic model associated to a simulator output is defined as $Y = F(X)$ where $F(\cdot)$ denotes a generic simulator, viewed as a deterministic black box describing a physical model, and $X = (X_1, \dots, X_d)^T$ denotes the column vector of d input variables, Y is the model output. Let us denote \mathcal{D} the domain of variation of the d input variables ($\mathcal{D} \subset \mathbb{R}^d$).

The probability of failure p_f of a system F over the input domain \mathcal{D} is defined as

$$p_f = P(\{x \in \mathcal{D} : F(x) > s\}) = \int \mathbf{1}_{F(x) > s} f(x) dx \quad (1)$$

where f is the joint density of the input variables defined on the domain \mathcal{D} , $\mathbf{1}_{F_2(x) > s} = 1$ if $F_2(x) > s$ and $\mathbf{1}_{F_2(x) > s} = 0$ otherwise.

Monte Carlo (MC) methods are usually used to perform uncertainty quantification (UQ) of physical systems. Due to the simulation time, MSE (mean square error) reduction techniques like importance sampling in [Rubinstein and Kroese \(2008\)](#) can decrease the number of simulations needed to achieve the same accuracy but the simulation time may remain prohibitive. A more efficient MSE reduction method consist in taking advantage of the physical model by using multilevel Monte Carlo (MLMC) methods ([Giles, 2013](#)) and distributing the simulation budget between different levels where each level corresponds to a given resolution and numerical cost and thus limiting the recourse to simulations at finest resolution. MLMC can be applied quite generally to physical models, for instance when a finite spatial or temporal resolution is used to solve numerically a stochastics differential equation. This setting is a special way of creating multifidelity levels.

More generally, multifidelity engages an expensive high fidelity physical model and at least one cheap version (low fidelity model) which can be obtained e.g. by simplifying the physical model or changing the discretization model. [Geraci et al. \(2015\)](#) applied MLMC method to both high fidelity and low fidelity physical models and showed that the multifidelity MLMC method is more efficient than MLMC for the same number of simulations. A review of multifidelity approaches building on Monte Carlo methods for UQ is given in [Peherstorfer et al. \(2016\)](#). A general review of the evolving use of multifidelity methods is given in [Fernández-Godino et al. \(2016\)](#).

It is often the case that the discretized versions of a complex system (the high fidelity model) are still too time consuming to allow a proper UQ, e.g. when computational fluid dynamics (CFD) equations are tuned by a mesh size to create fidelity levels to estimate a probability of failure. To tackle this particular issue, [Stroh et al. \(2017\)](#) showed that using surrogate models of each fidelity level allows to perform UQ with a given accuracy.

Indeed, when direct sampling approaches from a system become intractable, a common practice is to capture the main features of interest of the system in surrogate models built from a number of observations. The most popular ones are the response surface models (see [Myers and Montgomery, 2015](#)) and Gaussian process (kriging) models (see [Sacks et al., 1989](#); [Santner and Notz, 2003](#); [Rasmussen and Williams, 2006](#)) that produce a global approximation of the system.

However, the problem of computing a probability of exceeding a threshold is closely related to local approximation of a function in the failure region and to the estimation of the associated contour and can be seen as the volume of the failure domain [Bect et al. \(2012\)](#). This problem is even more critical as the simulation budget is very limited (case of CFD simulators) and the probability is low. The kriging based methods that are usually used to compute low order moments (mean, variance) of the outputs based on initial observation points need to be adapted to compute the probability of exceeding a threshold. The key idea is to design adaptive kriging algorithms based on the evaluation of smart new points involving sequential sampling strategies.

The aim of this paper is to propose a methodology to apply sequential sampling in targeted failure regions to multifidelity co-kriging metamodels in order to estimate the probability of conformity and its associated uncertainty. This work is mostly derived for two levels of fidelity and provides references to extend it to more levels.

The paper is organized as follows. Section 2 presents the motivating fire engineering application. Section 3 presents the Gaussian process modelling of numerical experiments. Section 4 presents the co-kriging multifidelity surrogate model used to combine two levels of simulations. Section 5 addresses the computation of the probability of non conformity based on the predictions from the surrogate model. Section 6 derived the implementation of the sequential co-kriging procedure to improve the estimation of the probability of non conformity. Section 7 provides an illustration of the methodology in a fire safety engineering case study.

2. Motivating fire engineering application

The French fire safety regulation allows simulation based conformity assessment for establishment receiving the public.

High-fidelity fire propagation simulations are based on CFD equations and are very expensive, which drastically limits the number of critical configurations on which the conformity assessment is based. Less accurate (low-fidelity) numerous cheap simulations are obtained with a zone model solving ordinary differential equations (simplified physical model).

Only a few statistical works based on low-fidelity simulations can be found in the literature. An uncertainty propagation with Monte Carlo simulations through a zone model code was performed in [Kong et al. \(2012\)](#). An inverse problem approach is tackled in [Overholt and Ezekoye \(2012\)](#) to characterize an input quantity after fires have been simulated. The computation of a probability of failure with subset sampling is derived in [Au et al. \(2007\)](#).

The fusion of two fidelity levels allows to construct a probabilistic framework for conformity assessment based on high-fidelity simulations of fire.

3. Gaussian process modelling of numerical experiments

Gaussian process modelling of deterministic numerical experiments makes the assumption that the simulator F is a realization of a random function Y such that $F(x) = Y(x, \omega)$, $x \in \mathcal{D}$, $\omega \in \Omega$ where Ω is the probability space and Y is of the form

$$Y(x) = Z(x) \quad (2)$$

where $Z(\cdot)$ is a stationary Gaussian process having mean μ , variance σ_Z^2 and covariance function $k(\cdot)$.

A Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution, so that a Gaussian process is uniquely determined by a mean function $\mu = E(Z)$ and a positive definite covariance function $\Sigma = \text{Cov}(Z) = (k(x, x'))_{x, x' \in \mathcal{D}}$. In this paper, we use the shorthand notation $Z \sim GP(\mu, \Sigma)$.

Various assumptions on μ yield to *simple kriging* (μ is a known constant), *ordinary kriging* (μ is an unknown constant independent of x), *universal kriging* (μ is modelled as a linear combination of basis functions). For details about Gaussian process modelling of numerical experiments, the reader is referred to Picheny et al. (2010), Rasmussen and Williams (2006) and Santner and Notz (2003). Interpretations of the covariance function and the mean function are provided in appendices A and B respectively.

Measurement noise or model error can be modelled as a Gaussian process having zero mean added to the right hand part of eq. (2) as in Perdikaris et al. (2015).

Kriging model (2) can be extended in a multifidelity framework by considering Z_1, \dots, Z_L various surrogate fidelity levels of Z sorted by increasing order of fidelity where Z_L is the highest-fidelity level. The setting of this paper requires that at least two fidelity versions of F exist and that surrogates are obtained from Gaussian process modelling, yielding the so called co-kriging model presented next section. The lower fidelity levels may be obtained by simplification of the highest level physical model or a coarse discretization of the problem (e.g. by tuning a mesh size).

4. Modelling numerical experiments from two fidelity levels with co-kriging models

4.1. Initial design of experiments

Denote $\mathcal{D}_1 = (x_{11}, \dots, x_{1n_1})$ and $\mathcal{D}_2 = (x_{21}, \dots, x_{2n_2})$ the designs of experiments at low-fidelity and high-fidelity levels respectively, with $\mathcal{D}_2 \subset \mathcal{D}_1 \subset \mathcal{D}$, $n_1 = \#\mathcal{D}_1$ and $n_2 = \#\mathcal{D}_2$, where $\#\mathcal{D}$ denotes the cardinality of the set \mathcal{D} .

Let $y^T = (y_1^T, y_2^T)$ be the output data, where $y_1^T = F_1(\mathcal{D}_1) = (F_1(x_{11}), \dots, F_1(x_{1n_1}))$ and $y_2^T = F_2(\mathcal{D}_2) = (F_2(x_{21}), \dots, F_2(x_{2n_2}))$ denote the outputs of the low-fidelity system $F_1(\cdot)$ (fast approximations) and high-fidelity system $F_2(\cdot)$ (expensive and accurate) respectively.

If information or an educated guess are available about the failure domain, some of the points or some of the coordinates of the points should be fixed accordingly. The remaining points of the initial designs of experiments should explore the input space and can be obtained with nested latin hypercube designs.

The number of initial points and their distribution on either the high-fidelity level or the low-fidelity level depends on a compromise between the time of a run at each level and its relevance

to evaluate the failure probability (which depends on the number of input variables and the sought probability). If higher-fidelity experiments are too expensive, lower-fidelity experiments can be used to predict initial failure points to be evaluated at the high-fidelity level.

4.2. Co-kriging model

Model outputs y_1 (low fidelity) and y_2 (high fidelity) are respectively modelled as observations from a Gaussian process Z_1 and Z_2 . The co-kriging multifidelity model presented by Kennedy and O’Hagan (2000) is based on the following first-order auto-regressive relation

$$Z_2(x) = \rho Z_1(x) + \delta(x) \tag{3}$$

where ρ measures the correlation between model outputs, $\delta(\cdot)$ is a Gaussian process independent from $Z_1(\cdot)$ with $\delta \sim GP(\mu_\delta, \Sigma_\delta)$ and $Z_1 \sim GP(\mu_1, \Sigma_1)$.

As a consequence of model (3), $Z_2(\cdot)$ is a Gaussian process $Z_2 \sim GP(\mu_2, \Sigma_2)$ where $\mu_2 = \rho\mu_1 + \mu_\delta$ and $\Sigma_2 = \rho^2\Sigma_1 + \Sigma_\delta$.

For a vector of n_0 locations x_0 , the joint distribution of $Z_2(x_0)$ and y is the following multivariate Normal distribution

$$\begin{pmatrix} Z_2(x_0) \\ y \end{pmatrix} \sim N_{n_0+n_1+n_2} \left(\begin{pmatrix} \mu_2(x_0) \\ m \end{pmatrix}, \begin{pmatrix} \Sigma_2(x_0) & r(x_0)^T \\ r(x_0) & V \end{pmatrix} \right)$$

where $m = (\mu_1(\mathcal{D}_1)^T, \mu_2(\mathcal{D}_2)^T)^T$ is the mean vector of length $n_1 + n_2$ of point at the two levels, the $n_0 - by - n_1 + n_2$ matrix $r(x_0)$ contains the covariances between points x_0 and the points at the two levels, the $n_1 + n_2 - by - n_1 + n_2$ covariance matrix V represents the cross-covariances between the points at the two levels. These matrices are described in appendix C in the case where Z_1 and δ are modelled as universal kriging models.

It follows that the conditional distribution of $Z_2(x)$ w.r.t. the vector of observations y and the parameters in the mean and covariance function is the following Gaussian process

$$Z_2(x)|y \sim GP(\mu_{2,co}(x), \Sigma_{2,co}(x)) \tag{4}$$

where

$$\mu_{2,co}(x) = \mu_2(x) + r(x)^T V^{-1}(y - m) \tag{5}$$

$$\Sigma_{2,co}(x) = \Sigma_2(x) - r(x)^T V^{-1} r(x) \tag{6}$$

The predictor (posterior mean) $\mu_{2,co}(x)$ can be used as a co-kriging surrogate of F_2 and the posterior variance $\sigma_{2,co}^2 = \rho^2\sigma_1^2 + \sigma_\delta^2 - r(x)^T V^{-1} r(x)$ is the mean square error (MSE) of the predictor.

These formulas require the inversion of a possibly large data covariance matrix V which may become ill-conditioned as the number of points increases. This is particularly challenging for higher number of levels as V contains all cross covariances of points between all levels.

Le Gratiet and Cannamela (2015) take advantage of nested designs to propose a recursive formulation of the posterior mean and variance at each level which allows to invert L data

covariance matrices of points at each level instead of a matrix of dimension $\sum_{l=1}^L n_l - by - \sum_{l=1}^L n_l$. This formulation decomposes the huge co-kriging problem into L smaller kriging problems.

We illustrate this process on two levels in the universal kriging framework of appendix C.

Let us denote \tilde{Z}_1 the conditional Gaussian process at low-fidelity level with mean μ_{Z_1} and covariance Σ_{Z_1} and replace Z_1 by \tilde{Z}_1 in eq.(2) with the same hypothesis on δ

$$Z_2(x) = \rho \tilde{Z}_1(x) + \delta(x) \quad (7)$$

For a vector of n_0 locations x_0 , the joint distribution of $Z_2(x_0)$ and y_2 is the following multivariate Normal distribution

$$\begin{pmatrix} Z_2(x_0) \\ y_2 \end{pmatrix} \sim N_{n_0+n_2} \left(\begin{pmatrix} \rho \mu_{Z_1}(x_0) + H_2(x_0) \\ \rho y_1(\mathcal{D}_2) + H_2(\mathcal{D}_2) \end{pmatrix} \beta_2, \begin{pmatrix} \rho^2 \Sigma_{Z_1}(x_0) + \Sigma_\delta(x_0) & r_2(x_0)^T \\ r_2(x_0) & V_{22} \end{pmatrix} \right) \quad (8)$$

where $r_2(x_0)$ is the vector of covariance between x_0 and points in \mathcal{D}_2 , $V_{22} = \rho^2 \Sigma_{Z_1}^2 + \Sigma_\delta$ is the data covariance matrix of points in \mathcal{D}_2 .

The recursive conditional distribution of $Z_2(x)$ w.r.t. the vector of observations y is the following Gaussian process

$$\tilde{Z}_2(x)|y \sim GP(\mu_{Z_2}(x), \Sigma_{Z_2}(x)) \quad (9)$$

with

$$\mu_{Z_2}(x) = \rho \mu_{Z_1}(x) + h_2(x)^T \beta_2 + r_2(x)^T V_{22}^{-1} (y_2 - \rho y_1(\mathcal{D}_2) - H_2 \beta_2) \quad (10)$$

$$\Sigma_{Z_2}(x) = \rho^2 \Sigma_{Z_1}(x) + \Sigma_\delta - r_2(x)^T V_{22}^{-1} r_2(x) \quad (11)$$

where $y_1(\mathcal{D}_2)$ is the vector containing the low fidelity observations at points in \mathcal{D}_2 .

It can be noticed that the recursive conditional co-kriging formulas involve conditional kriging formulas at low-fidelity and a correction term depending only on high-fidelity points (e.g. the matrix of regression functions H_2 , the covariance matrix between high-fidelity points V_{22} , covariance with high-fidelity points r_2). Extension of these formulas for $L > 2$ fidelity levels can be found in [Le Gratiet and Cannamela \(2015\)](#).

In the remaining of this paper, credible values for $F_2(x)$ given the observed data y and the estimated parameters and their uncertainties are embedded in the posterior predictive distribution (metamodel) based on the recursive formulation eq.(9) now denoted $M(x) = [\tilde{Z}_2(x)|y] = (\tilde{Z}_2(x)|obs)_{x \in \mathcal{D}}$.

For the estimation of recursive co-kriging model parameters the reader is referred to [Le Gratiet and Cannamela \(2015\)](#) and to the R package [MuFiCokriging \(Le Gratiet, 2012\)](#).

4.3. Toy example

In this example, we consider the two fidelity level toy functions used by [Le Gratiet \(2013\)](#) where $F_2(x) = (6x - 2)^2 \sin(12x - 4)$ and $F_1(x) = 0.5F_2(x) + 10(x - 0.5) - 5$.

Figure 1a shows the kriging estimate and the 95% credible bounds of F_2 based on 5 observation points. It can be observed that no observation allows the metamodel to capture the fall of F_2 between $x = 0.6$ and $x = 1$, so that the kriging metamodel is wrong in that range.

Figure 1b shows the effect of a low-fidelity function (in grey) whose behaviour shows an inflection in this range of values. As a result, the co-kriging estimate (in red) now reaches lower values, but still remains above the minimum. Moreover, the large number of low-fidelity simulations reduces the uncertainty over all the domain. It remains that the information contained in the cheap version F_1 is unable to render the magnitude of the fall in F_2 , since it is biased. New observations from F_2 are needed in this range.

This simple example illustrates a challenge of multifidelity co-kriging. As co-kriging allows to perform more cheap simulations, an issue is to determine when costly but unbiased observations with higher fidelity are required. This issue is more and more challenging as the number of levels increases.

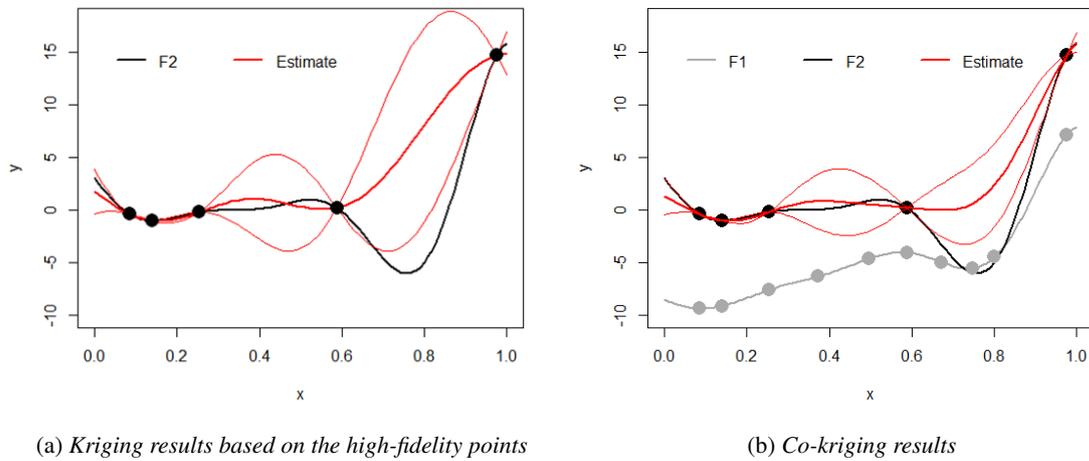


FIGURE 1: Comparison between kriging at high-fidelity and co-kriging.

5. Computation of the probability of non conformity of an expensive system

This section derives a generic methodology to compute the probability of non conformity p_f of an output from a computationally expensive system F modelled either by kriging or by co-kriging with a cheap approximation. Given a regulatory threshold s , p_f is defined as

$$p_f = \int 1_{F(x) > s} f(x) dx \tag{12}$$

where $f(\cdot)$ is the joint distribution of the input variables X_1, \dots, X_d , $1_{F(x) > s} = 1$ if $F(x) > s$ and $1_{F(x) > s} = 0$ otherwise.

The objective of this section is to provide an estimate \hat{p}_f of p_f with an associated uncertainty $u(\hat{p}_f)$, based on a metamodel M of F obtained either by kriging or by co-kriging.

Given a metamodel M approximating F , the best estimator p_f^{MSE} of p_f that minimizes the mean squared error $E_M \left((p_f - p_f^{MSE})^2 \right)$ is (see Bect et al., 2012)

$$p_f^{MSE} = \int P(M(x) > s) f(x) dx = \int \pi(x) f(x) dx \quad (13)$$

where $\pi(x)$ is the probability of excursion defined for location $x \in \mathcal{D} \setminus \mathcal{D}_n$ as

$$\pi(x) = P(M(x) > s) = P\left(\frac{M(x) - \hat{m}(x)}{\hat{\sigma}(x)} > \frac{s - \hat{m}(x)}{\hat{\sigma}(x)}\right) = \Phi\left(\frac{\hat{m}(x) - s}{\hat{\sigma}(x)}\right) \quad (14)$$

where $\hat{m}(x)$ and $\hat{\sigma}^2(x)$ are the predicted mean and variance at point x respectively, Φ is the cumulative distribution function of the standard Gaussian distribution.

Since M is a conditional Gaussian process, p_f^{MSE} is a random variable so that p_f and $u^2(p_f)$ are respectively estimated by the mean $E_M(p_f^{MSE})$ and variance $\text{Var}_M(p_f^{MSE})$. The subscript M indicates that integration is performed with respect to the metamodel uncertainty.

Let us denote $M_{(k)}$ a simulated trajectory from M , Monte Carlo estimates of $E_M(p_f^{MSE})$ and $\text{Var}_M(p_f^{MSE})$ respectively read

$$\hat{p}_f = \frac{1}{KL} \sum_{k=1}^K \sum_{l=1}^L 1_{M_{(k)}(x_l) > s} \quad (15)$$

$$u^2(\hat{p}_f) = \frac{1}{K} \sum_{k=1}^K \left(\frac{1}{L} \sum_{l=1}^L 1_{M_{(k)}(x_l) > s} - \hat{p}_f \right)^2 \quad (16)$$

Define $cv(\hat{p}_f) = \frac{u(\hat{p}_f)}{\hat{p}_f}$ the relative uncertainty of \hat{p}_f (also called coefficient of variation), as an indicator of the accuracy of the estimated probability of non conformity that can be used as a stopping criterion for the sequential computation of the probability of failure.

For a large number of draws L , say $L > 10,000$, it is preferable to sample trajectories with the adaptation proposed by [Oakley and A.O'Hagan \(2002\)](#).

6. Sequential planning for co-kriging probability of failure

Over the last decade, sequential approaches taking advantage of multifidelity have been developed for optimization ([Kandasamy et al., 2016](#); [Huang et al., 2006](#); [Courrier et al., 2014](#); [Gano et al., 2006](#)), approximation ([Xiong et al., 2013](#); [Le Gratiet and Cannamela, 2015](#)) and contour line estimation ([Chen et al., 2013](#)).

These approaches propose various strategies to select the best new point and the fidelity level of the simulation. Some of them introduce cost functions to better allocate resources, see [Le Gratiet and Cannamela \(2015\)](#) for global approximation problems and [Huang et al. \(2006\)](#) for optimization problems.

This approach is new for sequential multifidelity co-kriging based probability of failure. This section proposes a review of literature on sequential planning for kriging based evaluation of probabilities of failure ([Bect et al., 2012](#); [Picheny et al., 2010](#)) and contour line estimation ([Ranjan et al., 2008](#)) followed by a review of multifidelity methods tackling the choice of the level. The objective of this section is to give enough material to carry out a complete sequential procedure for multifidelity estimation of probability of failure. An implementation of a such a methodology is provided in the application.

6.1. Kriging-based sequential designs

This review and the organization of the section are mainly based on [Bect et al. \(2012\)](#) and [Chevalier et al. \(2014b\)](#) and distinguish criteria based on the marginal posterior distribution (metamodel) or on partially updated metamodels.

6.1.1. Criteria based on the marginal posterior distribution

In order to simplify the notations, at step n of the sequential algorithm, denote M_n the current metamodel, X_n and Y_n generic notations respectively for the collection of data points and their evaluation at all levels. Although unnecessary, an explicit conditioning of the metamodel is used to distinguish from partially updated metamodels, for $x \in \mathcal{D}$

$$M_n(x)|X_n, Y_n \sim N(\hat{m}_n(x)|X_n, Y_n, \hat{\sigma}_n^2(x)|X_n) \tag{17}$$

where $\hat{m}_n(x)$ and $\hat{\sigma}_n^2(x)$ are the predicted mean and kriging variance at point x respectively.

A common feature of the selected criteria is to compare the metamodel for each candidate point x to the threshold s and to sample points in a neighbourhood of the contour where the kriging variance is high.

The targeted mean square error (tMSE) criterion introduced by [Picheny et al. \(2010\)](#) reads

$$tMSE(x) = \hat{\sigma}_n^2(x)W_n(x) \tag{18}$$

where

$$W_n(x) = \frac{1}{\sqrt{2\pi(\hat{\sigma}_n^2(x) + \sigma_\varepsilon^2)}} \exp\left(-\frac{1}{2}\left(\frac{m_n(x) - s}{\sqrt{\hat{\sigma}_n^2(x) + \sigma_\varepsilon^2}}\right)^2\right) \tag{19}$$

This criterion gives more weight to points closed to the estimated contour through the weight function W_n which is obtained as

$$W_n(x) = E_{M_n}(f_\varepsilon(M_n(x) - s)) \tag{20}$$

with f_ε the pdf of $N(0, \sigma_\varepsilon^2)$.

The resulting convolution yields expression (19).

$$W_n(x) = \int f_\varepsilon(u - s)f_{M_n(x)(u)}du \tag{21}$$

The tuning parameter σ_ε controls the exploration area around the predicted contour line $\tilde{\Gamma} = \{\tilde{Y} = s\}$. For instance, $\sigma_\varepsilon = 0$ forces simulation in a close neighbourhood of the predicted frontier whereas $\sigma_\varepsilon > 0$ allows a larger exploration.

The criterion proposed in [Echard et al. \(2011\)](#) for contour estimation relies on the risk that $\hat{m}_n(x) - s$ changes sign as a function of the kriging variance $\hat{\sigma}_n(x)^2$. A learning function U is introduced such as

$$\hat{m}_n(x) - U(x)\hat{\sigma}_n(x) = s \tag{22}$$

The resulting criterion (to minimize) reads

$$U(x) = \frac{|\hat{m}_n(x) - s|}{\hat{\sigma}_n(x)} \quad (23)$$

The corresponding probability criterion (probability that $\hat{m}_n(x) - s$ changes sign) is given in [Bect et al. \(2012\)](#) as

$$J_n^{EGL}(x) = 1 - \Phi\left(\frac{|s - \hat{m}_n(x)|}{\hat{\sigma}_n(x)}\right) \quad (24)$$

This criterion is the core of the AK-MCS (Active learning reliability method combining Kriging and Monte Carlo Simulation) methodology developed by the authors.

According to [Echard et al. \(2011\)](#), this criterion gives more weight to points close to the threshold than to further ones with high kriging variance and favours regions with high input density. This is attained by choosing the new point among a large population sampled with a Monte Carlo sampling in the input distributions. This criterion is suitable for high non linearity and rather complex limit states (like non-connex or non-convex), and was tested for dimensions up to 100 input variables.

[Bect et al. \(2012\)](#) demonstrated that the *bichon* and *ranjan* criteria are special cases of the following criterion, obtained with $\delta = 1$ and $\delta = 2$ respectively

$$J_n^{RB}(x) = E_{M_n}(\max(0, \varepsilon(x)^\delta - |s - M_n(x)|^\delta)) \quad (25)$$

where $\varepsilon(x) = \alpha \hat{\sigma}_n(x)$ defines a neighbourhood around the contour which is a function of the standard deviation of the prediction at x with $\alpha, \delta > 0$. Note that, unlike [Echard et al. \(2011\)](#) criterion, α does not depend on the candidate point. The future point is chosen as to maximize $J_n^{RB}(x)$.

[Bichon et al. \(2008\)](#) expected feasibility criterion favours points with the highest probability that the standardized deviation to the threshold of the marginal posterior distribution $\frac{|s - M_n(x)|}{\hat{\sigma}_n(x)}$ lies in the interval $[-\alpha, +\alpha]$.

[Ranjan et al. \(2008\)](#) expected improvement criterion favours points with the highest probability that the square of standardized deviation to the threshold of the marginal posterior distribution $\left(\frac{|s - M_n(x)|}{\hat{\sigma}_n(x)}\right)^2$ lies in the interval $[-\alpha, +\alpha]$. Compared with *bichon* criterion, *ranjan* should favour sparse regions. In practice, *bichon* and *ranjan* criteria have similar behaviours.

A main difference with [Echard et al. \(2011\)](#) criterion is that both *bichon* and *ranjan* criteria are averaged over the metamodel uncertainty.

Close form expressions of these criteria can be obtained ([Ranjan et al., 2008](#); [Bect et al., 2012](#)) The optimization of the criterion is discussed in [Ranjan et al. \(2008\)](#) and is computationally fast. For instance, [Ranjan et al. \(2008\)](#) use a genetic algorithm to obtain starting values for package optimizers.

An implementation of the tmse, Bichon, and Ranjan criteria is available in the R package KrigInv in functions *tmse_optim*, *bichon_optim* and *ranjan_optim* respectively.

6.1.2. Criteria based on partially updated marginal posterior distributions

At step n of the sequential algorithm, the marginal posterior distribution $M_n(x)$, $x \in \mathcal{D}$ can be partially updated since co-kriging variance eq.(11) does not depend on y , which allows to update the kriging variance.

In this paper, the terminology "partially updated marginal posterior distribution" refers to the marginal posterior distribution where only the kriging variance is updated.

For any candidate point x_{new} , the updated kriging variance is denoted $\hat{\sigma}_{n+1}^2(x)|X_n, x_{new}$ and does not depend on $F(x_{new})$

$$M_n(x)|X_n, Y_n, x_{new} \sim N(\hat{m}_n(x)|X_n, Y_n, \hat{\sigma}_{n+1}^2(x)|X_n, x_{new}) \tag{26}$$

Picheny et al. (2010) extend the IMSE (Integrated Mean Square Error) criterion to targeted IMSE criterion ($IMSE_T$)

$$IMSE_T(X_n, Y_n, x_{new}) = \int_D \hat{\sigma}_{n+1}^2(x|X_n, x_{new}) W_n(x|X_n, Y_n) f(x) dx \tag{27}$$

where W_n is the weight function given eq (19).

The optimization of the criterion is discussed in Picheny et al. (2010) and is computationally demanding. Integration requires numerical methods for high dimensions and the optimization can be performed with global optimization methods (population-based,...) to avoid local minima. According to Picheny et al. (2010) the tIMSE is not suited for dimensions higher than 10 due to numerical integration

Stepwise uncertainty reduction (SUR) strategies require a measure of uncertainty on the quantity of interest, which is a probability of failure in our case.

The conditional variance $Var_n(\alpha)$ is a natural choice and yields the following SUR sampling criterion (Chevalier et al., 2014a)

$$J_n^{(\alpha)}(x_{n+1}) = E_{M_n}(Var_{n+1}(\alpha)|X_{n+1} = x_{n+1}) \tag{28}$$

Chevalier et al. (2014a) derived a numerically tractable form of the SUR criterion $J_n^{(\alpha)}$.

The random variable $1_{\tilde{Y}(x) > s}$ has conditional mean $p_n(x) = P(\tilde{Y}(x) > s|X_n, Y_n) = \Phi(\frac{m_n(x)-s}{\hat{\sigma}_n^2})$ and conditional variance $p_n(x)(1-p_n(x))$ so that $\int p_{n+1}(1-p_{n+1})dP_X$ is a measure of uncertainty targeting the contour.

The corresponding SUR sampling criterion is

$$J_n^{(\Gamma)} = E_{M_n} \left(\int p_{n+1}(1-p_{n+1})dP_X|X_{n+1} = x_{n+1} \right) \tag{29}$$

Criterion $J_n^{(\Gamma)}$ can be estimated with Gauss-Hermite quadrature (Bect et al., 2012; Chevalier et al., 2014a).

Bect et al. (2012) compares the performance of $IMSE_t$ and SUR criteria in terms of relative MSE of the probability of failure. They show, on a simulated example in low dimension, that the performances of several SUR strategies including $J_n^{(\Gamma)}$ (but not $J_n^{(\alpha)}$ which was intractable in Bect et al., 2012) are similar to $IMSE_t$ when the tuning parameter σ_ϵ^2 is close to zero.

6.2. Choice of the fidelity level

This section proposes a short review of the multifidelity sequential approaches developed by [Huang et al. \(2006\)](#), [Le Gratiet and Cannamela \(2015\)](#) and [Chen et al. \(2013\)](#). The first two references apply to more than two levels (denoted $l = 1, 2, \dots, L$) and introduce cost functions combining the individual costs of each simulator to better allocate resources. In this section, the relative cost between levels $l + 1$ and l is denoted $B_{l+1/l} = C_{l+1}/C_l$ where C_l denotes the cost of level l .

For global approximation problems, [Le Gratiet and Cannamela \(2015\)](#) operate a trade-off between expected global uncertainty reduction when adding points at level l versus level $l + 1$ and the cost of each level. The adaptation of [Le Gratiet and Cannamela \(2015\)](#) criterion to expected reduction of uncertainty of the probability of failure would read

$$(1 + B_{l+1/l})MSE_{red}^l(\hat{p}_f|X_n, Y_n, x_{new}) > MSE_{red}^{l+1}(\hat{p}_f|X_n, Y_n, x_{new}) \quad (30)$$

where $MSE_{red}^l(\hat{p}_f|X_n, Y_n, x_{new})$ denotes the expected reduction of uncertainty obtained with the partially updated posterior distribution if the point is observed at level l . Starting from the low-fidelity level, the procedure returns the lowest-fidelity level satisfying the inequality. This integral criterion is computationally expensive.

For optimization problems, [Huang et al. \(2006\)](#) propose an augmented expected improvement function used to optimize simultaneously the location and the level for any number of levels based on the current metamodel. In short, an expected improvement function at the highest-fidelity level is weighted by the relative cost $B_{l+1/l}$ and the correlation between posterior means at both levels. An additional multiplicative factor should be used for noisy simulators. An analytical form can be derived for the expected improvement function and optimization tools are required to maximize the full criterion in both the point and the level. For the same expected gains, this criterion favours lower-fidelity levels.

For contour estimation problems, when only two fidelity simulators are available, [Chen et al. \(2013\)](#) propose a simulator selection criterion which selects the high-fidelity simulator if the new point is close enough to the contour. For $\varepsilon > 0$, $F_2(x_{new})$ should be evaluated if $|\hat{m}_n^{(2)}(x_{new}) - s| < \varepsilon$. According to the authors, the tuning parameter ε can be adjusted by taking into account the estimated value of the auto-regressive coefficient between the two simulators at each step of the sequential algorithm. If this coefficient is close to 1, the criterion favours the lower-fidelity level. Since it tackles the related issue of contour estimation, this criterion can be considered as a baseline criterion for co-kriging based probability of failure problems.

6.3. Extension to sample a batch of points

The criteria section 6.1 can be extended to sample batches of new points based of the constant liar heuristic ([Chevalier et al., 2014b](#)) to allow parallel computations. The procedure to add R points $x_{new} = \{x_{n+1}^*, \dots, x_{n+R}^*\}$ at step n of the sequential procedure is displayed in algorithm 1 using the tMSE criterion.

The SUR strategy in [Bect et al. \(2012\)](#) has been extended in [Chevalier et al. \(2014a\)](#) to run a batch of simulations in parallel at a reduced numerical cost.

Algorithm 1 Sampling batches of new points**Require:** current database of size n **Ensure:** batch of R additional points $x_{n+1}^*, \dots, x_{n+R}^*$

- 1: initialize : $r=0$;
- 2: **repeat**
- 3: build or update the current metamodel with point $(x_{n+r}^*, \hat{m}_{n+r-1}(x_{n+r}^*))$;
- 4: sample a large number of candidate points \mathcal{C} ;
- 5: compute $x_{n+1+r}^* = \arg \max_{x \in \mathcal{C}} \hat{\sigma}_{n+r}^2(x) W_{n+r}(x)$;
- 6: add x_{n+1+r}^* to the current database;
- 7: $r \leftarrow r + 1$
- 8: **until** ($r < R$)

6.4. Iterative computation of the probability of non conformity

An immediate extension of the aforementioned methods to multifidelity probability of failure is to use sampling criteria to select the best future point at the highest-fidelity level and then to select the level on which to evaluate this point. The co-kriging metamodel is then updated with this new point and the evaluation(s) on F_1 or F_2 . The probability of failure, its associated uncertainty and its relative uncertainty are computed according to section 5. If the stopping criterion based on the relative uncertainty is not met then a new point is investigated based on the updated metamodel, and so on until the stopping criterion is met. A more efficient approach would consist to simultaneously determine the point and the level.

The iterative estimation of the probability of non conformity is given at algorithm 2.

Algorithm 2 Iterative computation of the probability of non conformity**Require:** initial nested databases $\mathcal{D}_2 \subset \mathcal{D}_1$ and the output vector $y^T = (y_1^T, y_2^T)$, where $y_1 = F_1(\mathcal{D}_1)$ and $y_2 = F_2(\mathcal{D}_2)$ (see 4.1);**Ensure:** estimates of the probability of non conformity \hat{p}_f and its associated uncertainty $u(\hat{p}_f)$

- 1: initialize: set prior distributions Z_1 and Z_2 for code outputs F_1 and F_2 (see section 4.2);
- 2: **repeat**
- 3: sample additional point(s) $x.new$ (either one point or a batch of points) in \mathcal{D} with a sequential procedure from section 6.1 and algorithm 1;
- 4: select the fidelity level(s) of the new point(s);
- 5: set $\mathcal{D}_1 \leftarrow \mathcal{D}_1 \cup \{x.new\}$ and $\mathcal{D}_2 \leftarrow \mathcal{D}_2$ for low-fidelity observation(s) only;
- 6: set $\mathcal{D}_1 \leftarrow \mathcal{D}_1 \cup \{x.new\}$ and $\mathcal{D}_2 \leftarrow \mathcal{D}_2 \cup \{x.new\}$ for low and high-fidelity observations;
- 7: update the vector y with new observations $y.new = (y_1.new^T, y_2.new^T)^T$
- 8: build or update the co-kriging metamodel M ;
- 9: compute Monte Carlo estimates \hat{p}_f and $u(\hat{p}_f)$ based on simulated trajectories from M ;
- 10: **until** ($cv(\hat{p}_f) > stop.crit$)

A flowchart of the complete procedure of sequential co-kriging is given at figure 2.

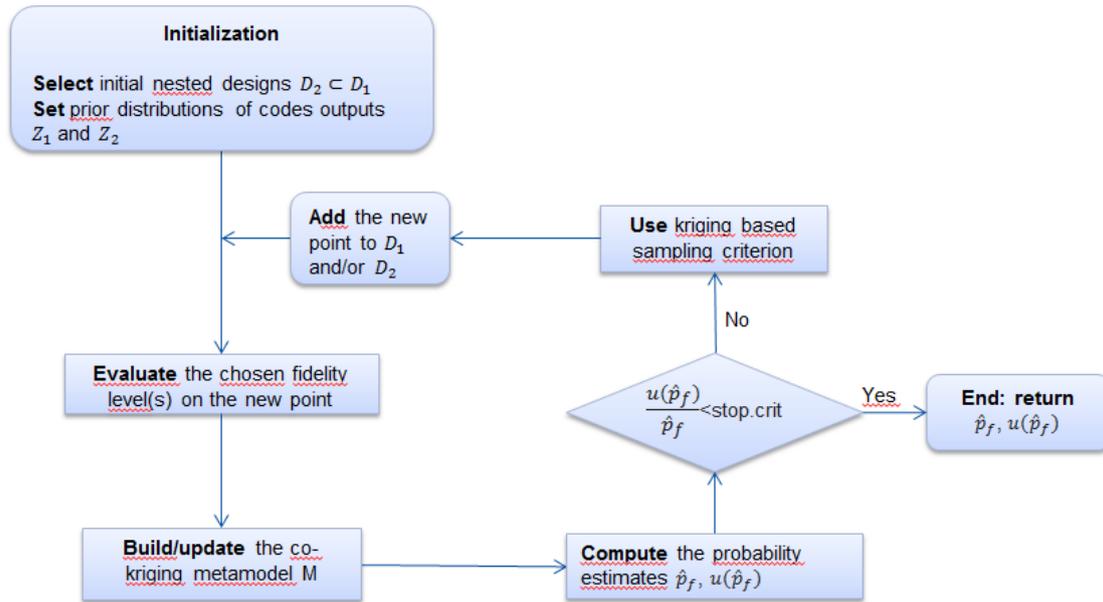


FIGURE 2: Flowchart of the sequential co-kriging procedure.

7. Application to the fire engineering case study

7.1. Conformity in fire safety engineering

For fire safety engineers, one of the major issues is to limit the fire risk in large space such as building, shopping malls, car park, theatres, etc. The strategy adopted consists to reduce its occurrence and the design fire load or to limit its extent through the use of fire suppression systems such as sprinklers (Carlotti and Lamalle, 2013). This approach is currently achieved by a second line of defense which consists to ensure that the thermal stratification is high enough so that occupant are not trapped in the smoke and may escape easily to a place of refuge. In other words, the objective is therefore to guarantee that the environment at heights useful for evacuation is clear enough to allow occupant to evacuate before reaching compromised tenability conditions in term of heat and toxicity of smoke. In order to achieve this objective, adapted smoke removal systems are used in order to keep the thermal stratification at a sufficient height for a time long enough.

The prediction of the evolution of tenability conditions in an environment in fire can be estimated using simplified fire models like zone model (like CFAST see section 7.2) or computational fluid dynamics tool (like FDS see section 7.2). Tenability criteria ensure that the occupants will not be exposed to untenable conditions and are based on the best available scientific judgement of the consequences of human exposure to fire effluents (see the international standard ISO 13571, 2012). According to this standard, during the people evacuation, the thermal risks generated by fire and smoke can lead to incapacitation or lethality among populations and are linked to the heat flux associated to the radiation of the smoke and the temperature of the atmosphere. The tenability limit for exposure of skin to radiant heat flux is approximately 2.5 kW.m^{-2} . Below this incident heat flux level, exposure can be tolerated for 30 min or longer without significantly affecting

tenability. The radiant heat flux limit of 2.5 kW.m^{-2} may be reached when the temperature T rises above 200°C .

In fire safety studies, the typical methods for ensuring whether the environment on the evacuation path is practicable enough to allow occupant to locate emergency exit, consist, in particular, to check that the radiant heat from the fire and the hot smoke layer does not exceed the skin pain threshold of 2.5 kW.m^{-2} and therefore the temperature T remains under 200°C .

7.2. Codes

The Consolidated Model of Fire and Smoke Transport (Jones et al., 2009), CFAST, is a two zone model solving a system of ordinary differential equations based on very strong simplifications. CFAST is a fire model which relies on the assumption that a volume is subdivided in two zones, perfectly mixed and with homogeneous properties in terms of temperature and composition: a hot layer with combustion products, located near the ceiling, and a cold layer with fresh clean air at the bottom, separated by a moving interface.

The Fire Dynamics Simulator version 6:1:2 (McGrattan et al., 2014), FDS, is a computational fluid dynamics CFD model of fire-driven fluid flow. It solves an approximation of the Navier-Stokes equations appropriate for low-Mach number, thermally driven flows. Software places particular emphasis on the description of smoke and heat transport from fires, as illustrated in figure 3.

7.3. Case study

For the purpose of the study, a real building is considered. As a consequence, its dimensions are known (dimensions: $19.75 \text{ m}(\text{length}) \times 12 \text{ m}(\text{width}) \times 16.50 \text{ m}(\text{height})$) along with those of the openings with negligible uncertainty. These quantities are then assigned a fixed value. The thickness and the thermal properties of the walls are defined in accordance with the present structures. The test hall is equipped with two doors assumed as open and two natural smoke removal systems. The room mesh is defined from the actual dimensions of the test hall. In the case of FDS, the grid is uniform, and the cell dimensions are 25 cm on each side. This cell size is a compromise between flow resolution and computational time.

As the boundary condition, the gas in the computation domain was set still with ambient temperature. At the free side, static pressure boundary condition was employed. The fire source is placed in the centre of the test hall.

7.4. Input variables

Apart from changes in environmental conditions (such as external temperature T_{ext} and ambient temperatures T_{amb} and atmospheric pressure P_{atm}), the properties of the fire (fire source area A_f , fire growth rate α , heat release rate (HRR) per unit area \dot{Q}'' and the fire load density q_f'' in the hall test is governed by the physical and chemical process evolved.

Multiple interactions between these input variables at different times during the fire may affect the pattern of the fire growth and lead to uncertainties. For this reason, there is a need to determine the uncertainty (probability) with which input variables may affect a real fire in a known building.

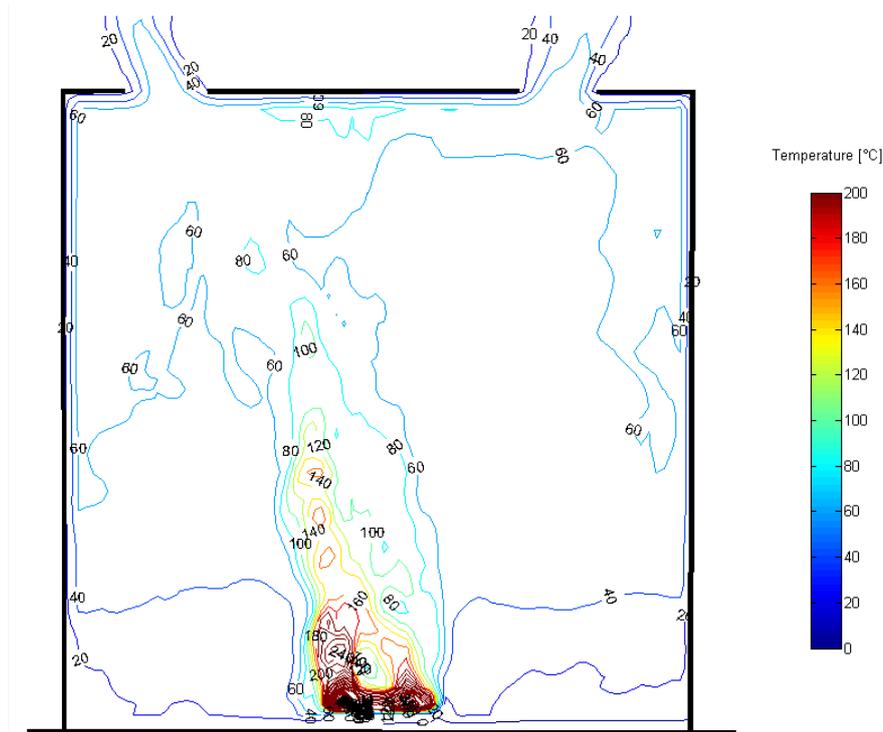


FIGURE 3: Iso-temperature located at the term source ($y = 0$) at $t = 200s$ Results obtained from a fire area of $13.75m^2$ and a heat release rate of $487kW.m^{-2}$.

In the present framework, the environmental conditions and the properties of the fire are therefore randomly determined. Their distributions and their range of variation pertaining to this study are displayed in table 3 when available.

The fire scenarios are representative of the scenarios commonly validated in the context of the studies in fire safety engineering, according to the [French regulation \(2004\)](#). In the smoke control chapter of this regulation, the fire to be taken into account to assess smoke removal systems is conventionally defined as a fire area called A_f ([French regulation, 2004](#)) whose surface may be $9m^2$, $18m^2$, $36m^2$ depending on the main purpose of the building. In the present study, a fire area from $2m^2$ to $20m^2$ has been considered using a uniform distribution.

The fire power is classically associated with the heat release rate per unit area \dot{Q}'' and its evolution is usually described by empirical design fire curves describing a growth step, a steady state and a decay period. For smoke control engineering studies, usage is to consider \dot{Q}'' that lies between $300kW.m^{-2}$ and $500kW.m^{-2}$, which correspond to a fire of $300kW$ to $18MW$. The upper bound has been increased to $600kW.m^{-2}$ for this study. The chosen associated law is a uniform one.

According to [Alpert \(2008\)](#), the rate of fire development can be approximated by a parabolic growth (t^2 , time) after the ignition reference time t_i as follows:

$$\dot{Q}_f = \alpha(t - t_i)^2 \quad (31)$$

where α is a fire growth coefficient ($kW.s^{-2}$), and t the time (s). The coefficient α varies between $0.011338 kW.s^{-2}$ for a very slowly developing fire and $0.2 kW.s^{-2}$ for very fast fire growth [Drysdale \(2011\)](#). The chosen associated law is a uniform one.

The fire load densities were typically presented in MJ per unit area of the surfaces bounding the fire building, as follow:

$$q_f'' = \int_{t_i}^{t_e} \dot{Q}'' dt \quad (32)$$

In the present case, the value of q_f'' have an impact of the combustion duration. Usage is to consider q_f'' that lies between $300 MJ.m^{-2}$ and $500 MJ.m^{-2}$. The chosen associated law is a uniform one.

The probability distributions for the environmental conditions are also representative of the climatic conditions encountered in France all over the year. External mean temperature (denoted as T_{ext}) could vary between $-10^\circ C$ (263.15 K) and $+30^\circ C$ (303.15 K), which are mean extreme values found in France according to meteorological survey. The chosen distribution law is normal with a standard deviation equal to the third of the semi length of the interval of variation. Ambient temperature (denoted as T_{amb}) could vary between $+16^\circ C$ (290 K) and $+30^\circ C$ (303.15 K) according to extrema found in [ISO 7730 \(2006\)](#) standard . The ambient temperature T_{amb} is correlated with the external temperature T_{ext} according to the following formula

$$T_{amb} \sim N \left(295.65 + \rho \times 2.5 \times \frac{T_{ext} - 283.15}{6.66}, 2.5 \times \sqrt{1 - \rho^2} \right) K \quad (33)$$

with $\rho = 0.8$.

For simplicity purposes, the effects of the wind are not considered in this study. The distribution of the atmospheric pressure is obtained from weather reports at Trappes (France) and is uniformly taken.

Finally, apart from the external and ambient temperature which are correlated, the input quantities are supposed to be independent.

TABLE 1. Description of the input variables of the fire code (N: normal, U: uniform). HRR: heat release rate (kW).

Variable	Name	Unit	Range	Distribution
P_{atm}	Atmospheric pressure	Pa	[98000, 102000]	N
T_{ext}	External temperature	K	[263.15, 303.15]	N
T_{amb}	Ambient temperature	K	[290, 303.15]	N
α	Fire growth rate	$kW.s^{-2}$	[0.011338, 0.20]	U
A_f	Fire area	m^2	[2, 20]	U
\dot{Q}''	Characteristic HRR per unit area	$kW.m^{-2}$	[300, 600]	U
q_f''	Design fire load density per unit area	$MJ.m^{-2}$	[300, 600]	U

A graphical sensitivity study performed in a report of the EMRP joint research project NEW04 ([Demeyer et al., 2015](#)) showed that among the 7 input variables described in table 1, only the fire area A_f and the characteristic heat release rate (HRR) per unit area \dot{Q}'' have an impact on exceeding the threshold.

From now on, the fire safety engineering case study will thus be treated in a 2D framework in the domain $\mathcal{D} = A_f \times \dot{Q}'' = [1, 20](m^2) \times [300, 500](kW.m^{-2})$ with the other variables kept constant.

7.5. Output variable and quantity of interest

Denote $\bar{T}_U(A_f, \dot{Q}'', t)$, the mean temperature of the sensors in the hot layer at time t as a function of the input variables A_f (fire area in m^2) and \dot{Q}'' (characteristic heat release rate in $kW.m^{-2}$).

The performance function F is here taken as the maximum mean temperature of the hot layer over the simulation time

$$F(A_f, \dot{Q}'') = \max_t \bar{T}_U(A_f, \dot{Q}'', t) \quad (34)$$

and yields the non conformity criterion $F(A_f, \dot{Q}'') > 200^\circ C$ and the associated probability of non conformity $p_f = P(F(A_f, \dot{Q}'') > 200^\circ C)$.

7.6. Monte Carlo simulations

Due to the simulation time of FDS (1 run a day on 1 core), Monte Carlo estimators of the probability of conformity, its uncertainty and its coefficient of variation are intractable. Indeed, nearly 50,000 computer experiments are required to reach $cv = 1\%$, which is impracticable : if 100 cores/day are available, then the total simulation time amounts to 500 days. Besides, as the sought probability gets smaller, the number of Monte Carlo simulations needs to be increased to maintain a given cv. For instance, for a probability of 5%, reaching $cv = 1\%$ requires more than 200,000 computer experiments.

However, as described in the body of the paper, kriging surrogate based Monte Carlo simulations can be driven instead.

7.7. Kriging and co-kriging models

Initial database The initial database evaluated by CFAST (F_1) comprises 14 points (black circles and grey triangles) displayed in figure 4, among them 9 points (black circles) are also evaluated by FDS (F_2). The grey triangles are only evaluated by CFAST. The points are displayed so as to cover the input space, while allowing a better estimation of parameters. For instance, grey triangles near black points ensure a better estimation of the relationship between the two codes and clusters of points should ensure a better estimation of the covariance parameters.

The level plot of the predictions of the code output $F(A_f, \dot{Q}'')$ (in $^\circ C$) obtained with the co-kriging model is displayed in figure 4. This representation displays a predicted zone where $F(A_f, \dot{Q}'') > 200^\circ C$ (dark area) with its uncertainty (gradient area).

Parameters of the Gaussian processes Each Gaussian process is modelled with the mean function $h(\cdot) = (1A_f)$ (see section B) to account for the linear effect of A_f on the output, and

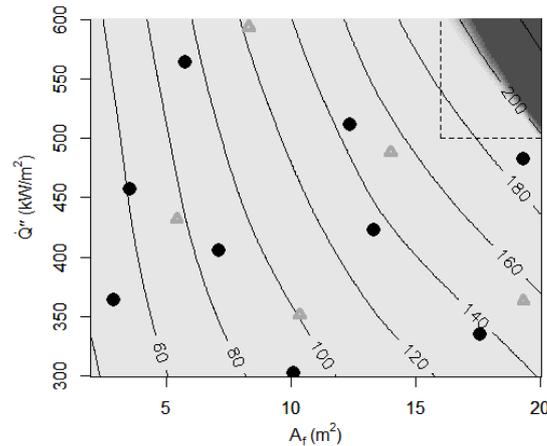


FIGURE 4: Initial database for co-kriging: grey triangles are the points only evaluated by CFAST and black circles are the points evaluated by both CFAST and FDS. The level plot of the output quantity of interest $F(A_f, \dot{Q}'')$ ($^{\circ}\text{C}$) modelled with the co-kriging model is also displayed.

with the Matérn-5/2 covariance function (defined at Eq.35 section A) to account for smoothness of the output.

The parameters of the co-kriging models have been estimated with the R package MuFiCokriging (Le Gratiot, 2012). The parameters of the kriging models have been estimated with the R package DiceKriging (Roustant et al., 2012).

Nugget effect The co-kriging metamodel of FDS as well as the kriging model of FDS (and CFAST) require the estimation of a nugget effect (Peng and Wu, 2014) to capture numerical instabilities, although the code can be seen as deterministic. Indeed, for the same values of the inputs FDS produces the same value of the output, but a slight variation in the input can produce a jump in the output value. This phenomenon and its treatment in kriging metamodels are reported in Roustant et al. (2012) and result in the estimation of an additional variance parameter that adds to the predicted variance of each point. In the following fire safety engineering application, the nugget effect (denoted τ^2) is given for each model as its value is not negligible and its effect can be seen in the output graphs.

Number of kriging based Monte Carlo simulations to compute probabilities of non conformity With the notations of section 5, $K = 1000$ trajectories and $L = 1600$ points on a 40×40 grid.

7.8. Interpretation of the co-kriging results versus kriging results

A comparison of various kriging based methods has been carried out on the fire engineering case study to show the influence of the number of points in the database and the influence of their location on the estimates of the probability of conformity and its accuracy (relative uncertainty). Results are displayed in table 2.

The analysis conducted on the initial database shows that, for the same number of FDS experiments (9 experiments) co-kriging dramatically reduces the relative uncertainty of the estimated probability \hat{p}_f by taking into account the information brought by additional CFAST points. Table 2 shows that a relative uncertainty divided by 8 ($cv(\hat{p}_f) = 0.1099$ under the co-kriging model versus $cv(\hat{p}_f) = 0.8928$ under the kriging model) was obtained at the cost of only 5 additional cheap CFAST simulations. This shows the positive impact of combining expensive experiments with cheap experiments.

The evolution of the contour plot of the posterior probability of excursion function on the full domain provides a graphical tool to assess the efficiency of the co-kriging method with respect to the kriging method. Indeed figure 5a obtained under kriging shows a large dispersion of the level lines whereas figure 5b obtained under co-kriging shows reduced uncertainty with the appearance of a non conformity area (dark area). Note that the nugget effect of the co-kriging model is $\tau_{co}^2 = 0.267$ whereas the nugget effect of the kriging model is $\tau^2 = 1.095e^{-5}$ (negligible). The appearance of the nugget effect with the co-kriging model corresponds to the addition of CFAST points to the initial FDS points. Indeed, CFAST points are transformed into mean point estimates of FDS values during the estimation of parameters and so induce instability between true FDS points and predicted FDS values. This artificial instability is meant to decrease as more points are added to the co-kriging model.

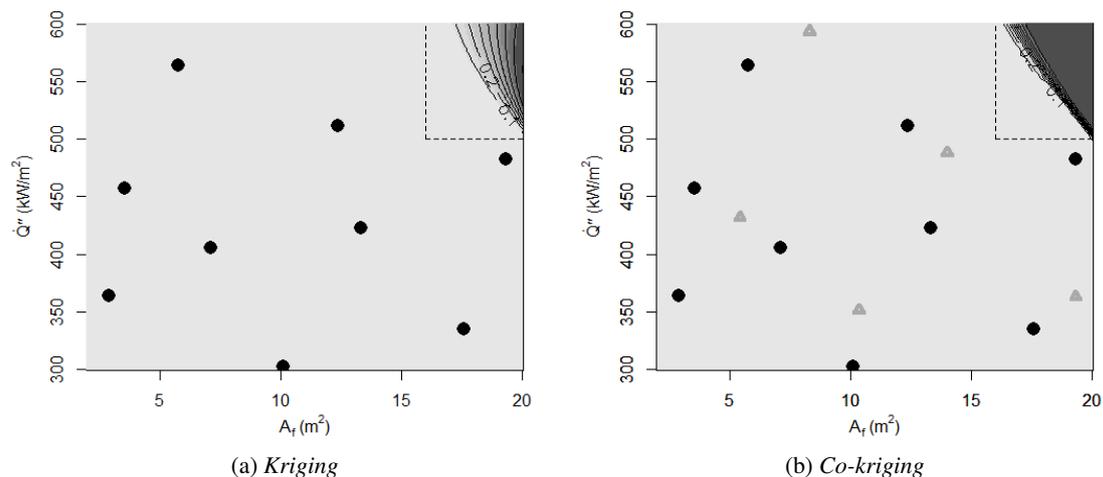


FIGURE 5: Contour plot of the probability of excursion based on kriging the FDS values of the points (black circles) (a) and the updated contour plot obtained with co-kriging after 5 CFAST points (grey triangles) have been added (b). Note that black circles are evaluated by both CFAST and FDS. The dashed rectangle defines the working area on which future results will be displayed.

7.8.1. Analysis of the sequential procedure

Table 3 gives the coordinates of the points iteratively produced by the sequential co-kriging procedure to improve the estimation of the probability of non conformity p_f . Due to the targeted MSE criterion, all the points have an expected mean \hat{m}_2 close to the threshold 200°C .

TABLE 2. Summary results of the probability of non conformity \hat{p}_f , its associated uncertainty $u(\hat{p}_f)$ and its coefficient of variation (relative uncertainty) $cv(\hat{p}_f)$ obtained with various kriging based methods.

number of FDS experiments	type of analysis	\hat{p}_f	$u(\hat{p}_f)$	$cv(\hat{p}_f)$
9	kriging	0.01354	0.01209	0.89283
9	co-kriging	0.03508	0.00385	0.1099
9 + 3	sequential co-kriging	0.03331	0.00054	0.01618

Between each iteration, the co-kriging metamodel is updated with real observations of CFAST (F_1) and FDS (F_2) and is used to provide a new point. For instance, for the second point, the actual evaluation of the codes matches the predicted mean, which means that the updated co-kriging model is suited for predicting failure events.

Besides, the predicted standard deviations decrease as observations are brought into the co-kriging model, with a significant decrease after the first iteration.

TABLE 3. Table giving the coordinates (A_f and \dot{Q}'') of the points obtained at each iteration of the sequential co-kriging procedure. For each point, the predicted means \hat{m}_1 and \hat{m}_2 of the codes F_1 and F_2 respectively, the predicted standard deviation $\hat{\sigma}_2$ and the evaluations of the codes F_1 and F_2 are displayed.

iter.	A_f (m^2)	\dot{Q}'' ($kW.m^{-2}$)	\hat{m}_1 ($^{\circ}C$)	\hat{m}_2 ($^{\circ}C$)	$\hat{\sigma}_2$ ($^{\circ}C$)	F_2 ($^{\circ}C$)	F_1 ($^{\circ}C$)
1	16.6875	599.3357	291.2835	199.1431	2.829	197.94	290.84
2	18.3125	551.8552	294.6508	199.7153	0.928	199.54	294.297
3	19.875	508.9258	295.7965	200.2372	0.729	200.40	295.574

The estimates \hat{p}_f , $u(\hat{p}_f)$ and $cv(\hat{p}_f)$ obtained at each iteration of the sequential co-kriging procedure are displayed in table 4.

TABLE 4. Summary results of the probability of non conformity \hat{p}_f , its associated uncertainty $u(\hat{p}_f)$ and its coefficient of variation (relative uncertainty) $cv(\hat{p}_f)$ obtained at each iteration of the sequential procedure.

iteration	\hat{p}_f	$u(\hat{p}_f)$	$cv(\hat{p}_f)$
1	0.0335	0.00144	0.04309
2	0.03306	0.00074	0.02239
3	0.03331	0.00054	0.01618

7.8.2. Graphical interpretation of the sequential procedure

As previously, the evolution of the contour plot of the posterior probability of excursion function provides a graphical tool to assess the efficiency of the iterative algorithm. In the following, for clarity, contour plots are displayed for a fire area A_f ranging between $16m^2$ and $20m^2$ and a characteristic heat release rate \dot{Q}'' between $500kW.m^{-2}$ and $600kW.m^{-2}$. Figure 6a reproduces the initial contour plot of the probability of excursion function displayed on the full domain on figure 5b restricted to $[16m^2 - 20m^2] \times [500kW.m^{-2} - 600kW.m^{-2}]$. Note that no point from the initial database has been sampled in this area.

The initial contour plot displayed on figure 6a shows a large dispersion of the level lines, representing the uncertainty on the predicted frontier between non conformity (black area) and

conformity (black area). For instance, the predicted 0.5 line indicates that there is 50% chance that the non conformity domain lies above this line and the predicted 0.9 line indicates that there is 90% chance that the non conformity domain lies above this line.

The first iteration of the sequential procedure samples a point in the uncertainty region where the lines are the most spread (highest predicted variances) on the 0.5-probability line. Once evaluated by FDS and CFAST the point is added to the initial database and an updated contour plot of the probability of excursion is obtained at figure 6b. Since the FDS value for this point is below the threshold ($F_2(16.6875, 599.3357) = 197.94$), this point belongs to the conformity area and the level lines are shifted towards upper values of A_f and \dot{Q}'' . A 1-probability of non conformity zone appears at the top right hand corner. Note that the nugget effect is $\tau_{(1)}^2 = 0.168$.

The second iteration provides a point which has the effect of reducing the uncertainty in the middle of the domain. Note that the nugget effect is $\tau_{(2)}^2 = 0.1096$.

A third iteration is needed to reach a coefficient of variation less than 2%. Note that the nugget effect is $\tau_{(3)}^2 = 0.067$. It can be noticed that, as points are added to the co-kriging model, the nugget effect decreases. Since the nugget effect is added to the predicted variance when computing the probability of excursion, the nugget is responsible of the uncertainty at observation points. This is particularly visible for the third point added such that $F_2(19.875, 508.9258) = 200.40$ which is non-conform but still lies under the 1-probability line. If there were no nugget effect, all the level lines should be pinched at this point.

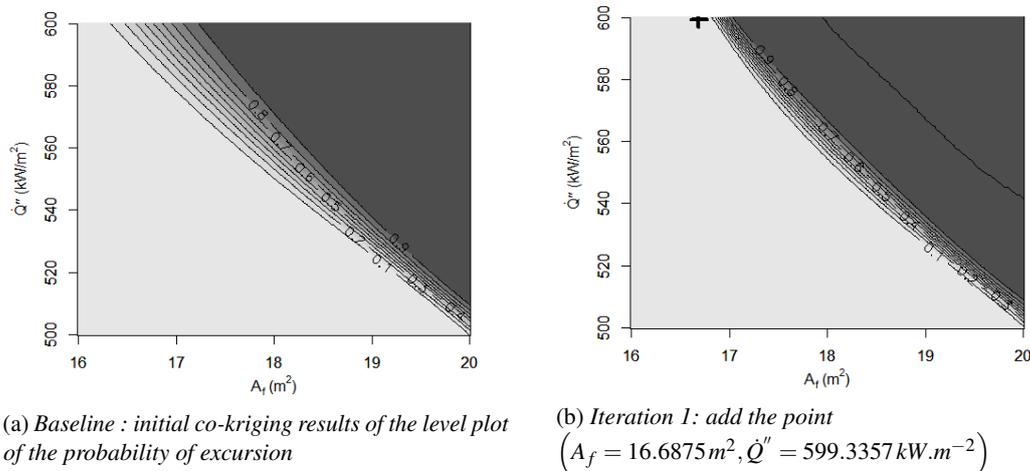


FIGURE 6: Contour plot of the initial probability of excursion (a) and the updated probability of excursion after 1 point (black cross) has been added at the first iteration of the sequential co-kriging procedure (b). Figure (a) is a zoom of figure 5b on the working area defined by the dashed rectangle and so contains no points from the initial database.

7.9. Practical interpretation of probability level plots

In the context of fire safety engineering, figure 7b could be used in order to estimate the reliability of the smoke system control used in the building. In function of the fire surface and the heat released per unit area, one can therefore determinate the critical fire scenario where tenability conditions are not met. The tenability criterion $T_U > 200^\circ\text{C}$ has an impact on the people evacuation.

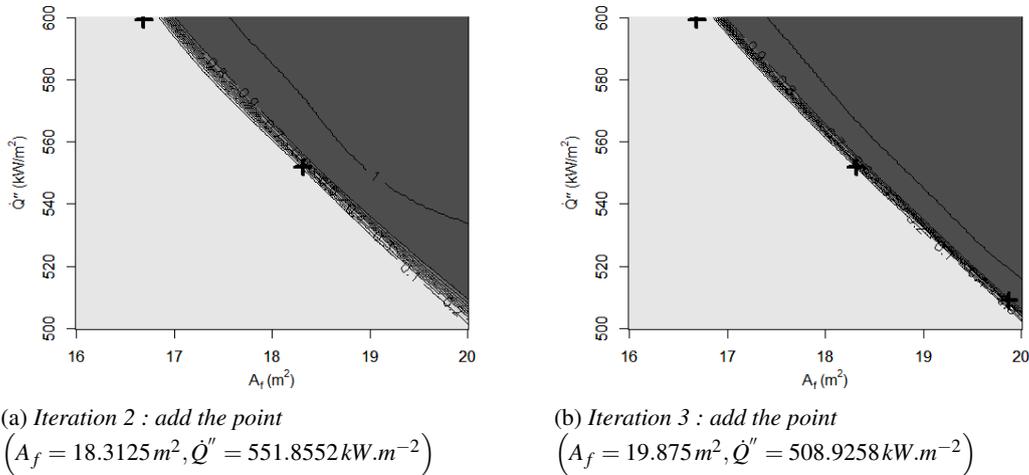


FIGURE 7: Contour plot of the updated probability of excursion at the second iteration of the sequential co-kriging procedure (a) and at the third iteration (b). At each iteration, the new point is marked with a black cross.

When $p_f \rightarrow 0$, this result induces that the smoke system controls are sufficient to avoid that fire environment in the building does not exceed the regulatory threshold. In contrast, when $p_f \rightarrow 1$, the fire conditions (heat release and fire area) can be assumed as critical because the tenability criteria are achieved and exceeded in the surrounding environment. In this condition, the smoke system control is inadequate and some modifications should be applied on it in order to respect the regulation and reduce the risk.

8. Conclusion

A multifidelity sequential co-kriging methodology has been proposed to estimate probability of failure of computationally intensive high fidelity systems. This approach combines classical sequential kriging criteria for failure evaluation with multifidelity criteria for level choice. A variety of criteria has been proposed ranging from computationally cheap to very expensive (integral criteria) depending on the quantity of information taken into account (e.g. pointwise estimates or metamodel distribution). The methodology has been applied to a real, but simplified, fire safety engineering case study with two fidelity levels where the low-fidelity level is a simplified physical model. This example displays the full methodology, exhibits the underlying hypotheses and provides a practical analysis of the results. From the fire engineering point of view, the recourse to low fidelity simulations (in addition to unbiased high fidelity simulations) allows to compute probability of conformity based on all available information regarding the environment and the likely fires, which could be an aid to conformity decisions in the future. More generally, the method could be applied to fields of the industry already using kriging methods to tackle reliability issues. Further work will consist in simultaneously optimizing the sampling points and the levels in a multifidelity setting, while addressing high dimensional issues.

Acknowledgments

This work is supported by the European Metrology Research Programm (EMRP), which is jointly funded by the EMRP participating countries within EURAMET and the European Union. The authors thank their colleague Alexandre Allard for his valuable contribution to this paper.

Appendix A: Covariance function of a Gaussian process

The choice of the covariance function depends on prior knowledge about the behaviour of the function. One can distinguish, at first, stationary and nonstationary covariance functions. The former ensures that the smoothness of the function does not depend on the location but is a function of $x - x'$. Inversely, the latter allows the smoothness to vary with the location. A stronger property is the isotropy which ensures that the covariance only depends on the euclidean distance $\|x - x'\|_2$. A variety of covariance functions can be found in [Rasmussen and Williams \(2006\)](#).

The most commonly used covariance function is the Matérn covariance function (see [Rasmussen and Williams, 2006](#)), whose hyperparameters $\psi = (\nu, \theta)$ control the smoothness and the length scale respectively. Values $\nu = 3/2$ or $\nu = 5/2$ yield the two mostly used matérn kernels that arise as the product of an exponential and a polynomial. For instance, the Matérn-5/2 is defined as a function of $r = |x - x'|$

$$k_{\nu=5/2}(r) = \left(1 + \frac{\sqrt{5}r}{\theta} + \frac{5r^2}{3\theta^2}\right) \exp\left(-\frac{\sqrt{3}r}{\theta}\right) \quad (35)$$

For $\nu = 1/2$ the correlation is able to model chaotic functions and is known as the exponential correlation function defined as $k_{\nu=1/2}(r) = \exp\left(-\frac{r}{\theta}\right)$.

The value $\nu = \infty$ yields the Gaussian correlation function defined as $k_{\infty}(r) = \exp\left(-\frac{r^2}{\theta^2}\right)$ used to model very smooth functions.

Usually, the value for ν is fixed by the user, the length scale θ is estimated on the data.

Figure 8 illustrates the effect of the prior choice of the covariance function on simulated trajectories of a Gaussian process with constant mean fitted on the same data points (black dots). The Matérn-5/2 covariance yields smooth trajectories (red curves) whereas the exponential covariance yields chaotic trajectories (black curves). In both cases, the posterior simulations interpolate the data points. Conditional trajectories are nearly confounded under the Matérn covariance whereas a large dispersion occurs under the exponential covariance, which implies that predictions under the exponential model will have larger predicted variances outside the data points.

Appendix B: Mean function of a Gaussian process

The mean function governs the trend of the simulated trajectories. Usually, the mean function is taken as constant. If additional information is available, the mean can be modelled as a function of the inputs.

Figure 9 shows that a constant mean allows to produce trended conditional trajectories. This property is usually true within the domain of variation of the inputs. After removing the rightmost

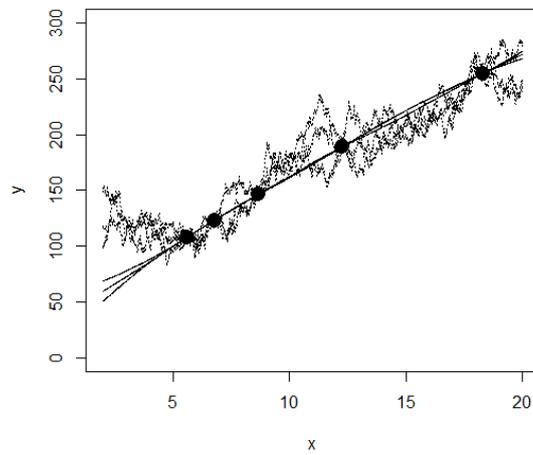


FIGURE 8: Plot of the simulated trajectories fitting the observation points under Matérn covariance (plain curves) and exponential covariance (dotted curves) models under constant mean.

point (with respect to figure 8), figure 9 shows that the effect of the mean function is mostly visible at the extremities of the domain of variation and outside. Indeed, under constant mean $m(x) = \beta_0$, predictions over $[20, 30]$ tend to be distributed around the trend uniformly on $[0, 300]$ (as $x \rightarrow 30$). Under the trended mean $m(x) = \beta_0 + \beta_1 x$, predictions outside the domain are concentrated on the trend line with no dispersion. The constant mean model can be used instead of a trended model in the interpolation domain but the predictions will follow the trend outside. Conversely, the trended model can be used in the interpolation domain and outside if the trend is justified.

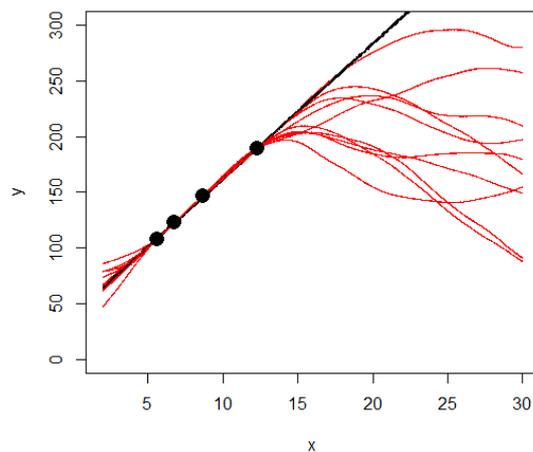


FIGURE 9: Plot of the simulated trajectories fitting the observation points obtained under matérn 5_2 covariance and a) constant mean $m(x) = \beta_0$ (10 red curves) and b) trended mean $m(x) = \beta_0 + \beta_1 x$ (10 black curves).

Appendix C: Matrix notation for universal co-kriging

Conditional on hyperparameters β_2 , σ_δ^2 and ψ_2 , $\delta(\cdot)$ is modelled as a Gaussian process with mean $m_\delta(\cdot) = h_2(\cdot)^T \beta_2$ and covariance matrix $\Sigma_\delta = (k_\delta(x, x'))_{x, x' \in \mathcal{D}}$ where $k_\delta(x, x') = \sigma_\delta^2 r_\delta(x, x'; \psi_2)$ parametrized by ψ_2 . For $x = x'$, $k_\delta(x, x') = \sigma_\delta^2$ is the variance of the process $\delta(\cdot)$.

Conditional on hyperparameters β_1 , σ_1^2 and ψ_1 , $Z_1(\cdot)$ is modelled as a Gaussian process with mean $m_1(\cdot) = h_1(\cdot)^T \beta_1$ and covariance matrix $k_1(x, x') = \sigma_1^2 r(x, x'; \psi_1)$ parametrized by ψ_1 . For $x = x'$, $k_1(x, x') = \sigma_1^2$ the variance of the process $Z_1(\cdot)$.

As a consequence of model (3), $Z_2(\cdot)$ is a Gaussian process with mean $m_2(x) = h'(x)^T \beta$ with $h'(x)^T = (\rho h_1(x)^T, h_2(x)^T)$ and $\beta^T = (\beta_1^T, \beta_2^T)^T$, and covariance function $k_2(x, x') = \rho^2 k_1(x, x') + k_\delta(x, x')$. It follows that the prior variance of the process \tilde{y}_2 is obtained as $k_2(x, x) = \rho^2 \sigma_1^2 + \sigma_\delta^2$.

Let us denote H the $(n_1 + n_2) - by - (p_1 + p_2)$ matrix of the regression functions for points in \mathcal{D}_1 and \mathcal{D}_2

$$H = \begin{pmatrix} h_1(\mathcal{D}_1) & \vec{0}_{n_1} \\ \rho h_1(\mathcal{D}_2) & h_2(\mathcal{D}_2) \end{pmatrix} \quad (36)$$

where $h_{t'}(\mathcal{D}_t) = (h_{t'}(x_{t1})^T, \dots, h_{t'}(x_{tm_t})^T)^T$ for $t, t' = \{1, 2\}$ is a $n_t - by - p_{t'}$ matrix and $\vec{0}_{n_1}$ denotes the $n_1 - by - p_2$ zero matrix where p_1 and p_2 are the number of regression functions in h_1 and h_2 respectively.

It follows that $m = H\beta$ in section 4.2.

Let us denote $k_1(\mathcal{D}_1, \mathcal{D}_2) = \{k_1(x, x'), x \in \mathcal{D}_1, x' \in \mathcal{D}_2\}$ the $n_1 - by - n_2$ data covariance matrix between points in \mathcal{D}_1 and \mathcal{D}_2 and let us use the shorthand $k_t(\mathcal{D}_t) = k_t(\mathcal{D}_t, \mathcal{D}_t)$ to denote the covariance matrix between points in \mathcal{D}_t . Note that $k_1(\mathcal{D}_2, \mathcal{D}_1) = k_1(\mathcal{D}_1, \mathcal{D}_2)^T$.

The data covariance matrix V can be expressed as the following block matrix

$$V = \begin{pmatrix} k_1(\mathcal{D}_1) & \rho k_1(\mathcal{D}_1, \mathcal{D}_2) \\ \rho k_1(\mathcal{D}_2, \mathcal{D}_1) & \rho^2 k_1(\mathcal{D}_2) + k_2(\mathcal{D}_2) \end{pmatrix} \quad (37)$$

References

- Alpert, R. L. (2008). Ceiling jet flows. In Nenno, P. D., editor, *SFPE handbook of fire protection engineering*, pages 2.21–2.36. National fire protection association, Quincy, MA, 4th edition.
- Au, S., Wang, Z.-H., and Lo, S.-M. (2007). Compartment fire risk analysis by advanced Monte Carlo simulations. *Engineering Structures*, pages 2381–2390.
- Bect, J., Ginsbourger, D., L. L., Picheny, V., and Vazquez, E. (2012). Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing*, 22:773–793.
- 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.
- Carlotti, P. and Lamalle, D. (2013). Sensitivity to boundary conditions for simulation of fire plume in enclosure. Proceedings of the 21e congrès français de mécanique, France.
- Chen, R., Hung, Y., and nad S.W. Yen, W. (2013). Contour estimation via two fidelity computer simulators under limited resources. *Computational Statistics*, 28:1813–1834.
- Chevalier, C., Bect, J., Ginsbourger, D., Vazquez, E., Picheny, V., and Richet, Y. (2014a). Fast parallel kriging-based stepwise uncertainty reduction with application to the identification of an excursion set. *Technometrics*, 56(4):455–465.

- Chevalier, C., Picheny, V., and Ginsbourger, D. (2014b). The kriginv package : An efficient and user-friendly R implementation of kriging-based inversion algorithms. *Computational Statistics and Data Analysis*, 71:1021–1034.
- Courrier, N., Boucard, P. A., and Soulier, B. (2014). The use of partially converged simulations in building surrogate models. *Advances in Engineering Software*, 67:186–197.
- Demeyer, S., Wright, L., Pendrill, L., and Bär, M. (2015). Intermediate report on case study on fire engineering. A report of the EMRP joint research project NEW04 Novel mathematical and statistical approaches to uncertainty evaluation.
- Drysdale, D. (2011). *An introduction to fire dynamics*. John Wiley, 3rd edition.
- 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.
- Fernández-Godino, G., Park, C., Kim, N., and Haftka, R. (2016). Review of multi-fidelity models. *ArXiv e-prints*.
- French regulation (2004). Instruction technique, n°246 relative au désenfumage dans les établissements recevant du public, annexe pour le calcul du taux alpha. In *Arrêté du 22 mars 2004*. Documentation BATISS (French Regulation).
- Gano, S., Renaud, J., Martin, J., and Simpson, T. (2006). Update strategies for kriging models used in variable fidelity optimization. *Structural and Multidisciplinary Optimization*, 32:287–298.
- Geraci, G., Eldred, M., and Iaccarino, G. (2015). *A multifidelity control variate approach for the multilevel Monte Carlo technique*. Center for Turbulence Research Annual Research Briefs.
- Giles, M. B. (2013). *Multilevel Monte Carlo Methods*, pages 83–103. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Huang, D., Allen, T., Notz, W., and Miller, R. (2006). Sequential kriging optimization using multiple-fidelity evaluations. *Structural and Multidisciplinary Optimization*, 32:369–382.
- ISO 13571 (2012). NF EN ISO 13571 - Guidelines for the estimation of time to compromised tenability in fires. volume ICS: 13.220.01.
- ISO 7730 (2006). NF EN ISO 7730 - Analytical determination and interpretation of thermal comfort using calculation of the pmv and ppd indices and local thermal comfort criteria. volume ICS: 13.18024.
- Jones, W., Peacock, R., Forney, G., and P.A., R. (2009). Cfast - consolidated model of fire growth and smoke transport (version 6), technical reference guide. *NIST SP - 1026*, page 125 pp.
- Kandasamy, K., Dasarathy, G., Oliva, J. B., Schneider, J., and Poczos, B. (2016). Multi-fidelity Gaussian Process Bandit Optimisation. *ArXiv e-prints*.
- Kennedy, M. C. and O'Hagan, A. (2000). Predicting the output from a complex computer code when fast approximations are available. *Biometrika*, 87:1–13.
- Kong, D., Johansson, N., van Hess, P., Lu, S., and Lo, S. (2012). A Monte Carlo analysis of the effect of heat release rate uncertainty on available safe egress time. *Journal of Fire Protection Engineering*, pages 5–29.
- Le Gratiet, L. (2012). *MuFiCokriging: Multi-Fidelity Cokriging models*. R package version 1.2.
- Le Gratiet, L. (2013). Bayesian analysis of hierarchical multifidelity codes. *SIAM/ASA Journal on Uncertainty Quantification*, 1:244–269.
- Le Gratiet, L. and Cannamela, C. (2015). Cokriging-based sequential design strategies using fast cross-validation techniques for multi-fidelity computer codes. *Technometrics*, 57(3):418–427.
- McGrattan, K., McDermott, R., Weinschenk, C., and Overholt, K. (2014). Fire dynamics simulator - technical reference guide, volume 1 mathematical model. *NIST SP - 1018*, page 175 pp.
- Myers, R. H. and Montgomery, D. C. (2015). *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. John Wiley, 3rd edition.
- Oakley, J. and A.O'Hagan (2002). Bayesian inference for the uncertainty distribution of computer model outputs. *Biometrika*, 89:769–784.
- Overholt, K. J. and Ezekoye, O. A. (2012). Characterizing heat release rates using an inverse fire modeling technique. *Fire Technology*, 48:893–909.
- Peherstorfer, B., Willcox, K., and Gunzburger, M. (2016). *Survey of multifidelity methods in uncertainty propagation, inference, and optimization*. ACDL Technical Report TR16-1.
- Peng, C.-Y. and Wu, J. (2014). On the choice of nugget in kriging modeling for deterministic computer experiments. *Journal of Computational and Graphical Statistics*, 23:151–168.
- Perdikaris, P., Venturi, D., Royset, J. O., and Karniadakis, G. E. (2015). Multi-fidelity modelling via recursive co-kriging and Gaussian-Markov random fields. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 471.
- Picheny, V., Ginsbourger, D., Roustant, O., Haftka, R., and Kim, N.-H. (2010). Adaptive designs of experiments for

- accurate approximation of a target region. *Journal of Mechanical Design*, 132.
- Ranjan, P., Bingham, D., and Michailidis, G. (2008). Sequential experiment design for contour estimation from complex computer codes. *Technometrics*, 50(4):527–541.
- Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. MIT Press.
- Roustant, O., Ginsbourger, D., and Deville, Y. (2012). DiceKriging, DiceOptim: Two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization. *Journal of Statistical Software*, 51(1):1–55.
- Rubinstein, R. Y. and Kroese, R. Y. (2008). *Simulation and the Monte Carlo Method*. Wiley.
- Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989). Design and analysis of computer experiments. *Statistical Science*, 4:409–423.
- Santner, T. J. and Notz, W. (2003). *The design and analysis of computer experiments*. Springer Verlag, Springer Series in Statistics.
- Stroh, R., Bect, J., Demeyer, S., Fischer, N., and Vazquez, E. (2017). Assessing Fire Safety using Complex Numerical Models with a Bayesian Multi-fidelity Approach. *Fire Safety Journal*. submitted.
- Xiong, S., Qian, P., and Wu, J. (2013). Sequential design and analysis of high-accuracy and low-accuracy computer codes. *Technometrics*, 55:37–46.