Quantifying uncertainty on Pareto fronts with Gaussian Process conditional simulations

Mickaël Binois, David Ginsbourger, Olivier Roustant

Quantifying uncertainty on Pareto fronts with Gaussian Process conditional simulations

M. Binois[a,b,*], D. Ginsbourger[c], O. Roustant[b]

[a] *Renault S.A.S. ,78084 Guyancourt, France*
[b] *École Nationale Supérieure des Mines de Saint-Étienne, FAYOL-EMSE, LSTI, F-42023 Saint-Étienne, France*
[c] *University of Bern, Department of Mathematics and Statistics, Alpeneggstrasse, 22 CH-3012 Bern, Switzerland*

Abstract

Multi-objective optimization algorithms aim at finding Pareto-optimal solutions. Recovering the Pareto front or the Pareto set from a limited number of function evaluations are challenging problems. A popular approach in the case of expensive-to-evaluate functions is to appeal to metamodels of the objective functions. Kriging has been shown efficient as a base for sequential multi-objective optimization, notably through infill sampling criteria balancing exploitation and exploration such as the Expected Hypervolume Improvement. Here we consider kriging metamodels not only for selecting new points, but as a tool for estimating the whole Pareto Front and quantifying how much uncertainty remains on it at any stage of kriging-based multi-objective optimization algorithms. Our approach relies on the Gaussian Process interpretation of kriging, and bases upon conditional simulations. Using concepts from random set theory, we propose to adapt the Vorob'ev expectation and deviation to capture the variability of the set of non-dominated points. Numerical experiments illustrate the potential of the proposed workflow, and it is shown on examples how Gaussian process simulations and the estimated Vorob'ev deviation can be used to monitor the ability of kriging-based multi-objective optimization algorithms to accurately learn the Pareto front.

Keywords: Kriging, Gaussian Process conditional simulations, Multi-objective black-box optimization, Uncertainty quantification, Attainment function, Vorob'ev Expectation

1. Introduction

The interest in Multi-Objective Optimization (MOO) has been growing over the last decades, resulting in the development of numerous dedicated methods, especially in Evolutionary MOO [1]. The latter are able to cope with challenging problems occurring when few information about the properties of the objective functions is available (black-box optimization). One specific difficulty is a limited budget of evaluations, because of expensive experiments or high fidelity simulation, as for example in car crash safety design [2].

In this context, see e.g. [3] for a review, a common approach is to rely on a surrogate model or metamodel to alleviate the computational costs of the optimization process. In particular, kriging metamodels have proven to be efficient because they not only give a response surface but also a quantification of prediction uncertainty. In mono-objective optimization, this property has been extensively used following the EGO algorithm [4] based on the Expected Improvement to balance between exploitation and exploration. Extensions to MOO have been developed, from scalarization approaches [5, 6] to the use of multi-objective improvement criteria such as the Expected Hypervolume Improvement [7, 8].

While results about the optimality of solutions from aggregation approaches have been reported (see e.g. [9]), things are more difficult to analyze for MOO and even more in metamodel based MOO, where an additional source of uncertainty due to surrogate modeling must be taken into account. Monitoring the convergence empirically has been

*Corresponding author. Tel.: +33 4 77 42 02 85;
E-mail addresses: binois@emse.fr (M. Binois), ginsbourger@stat.unibe.ch (D. Ginsbourger), roustant@emse.fr (O. Roustant).

proposed [10] while in EGO the values of the Expected Improvement provide a stopping rule [4].

Similarly to what has been proposed for excursion sets in Chevalier et al. [11, 12], we propose here to use notions from the theory of random sets [13] for quantifying uncertainty on Pareto fronts, through conditional simulations. We show that it is possible to obtain a metamodel-based estimation of the Pareto front, with a value for the deviation and an illustration of the remaining uncertainty. Possible applications for the practitioner include convergence assessment as well as visualization of areas where Pareto-optimal solutions could be found in the objective space.

The paper is organized as follows: Section 2 details the framework of multi-objective optimization based on Gaussian process models. In Section 3 we propose an original definition of uncertainty using the Vorob'ev expectation and deviation. Finally, Section 4 is dedicated to applications of the proposed methodology to two different test cases, where the potential of the approach to quantify uncertainty and monitor convergence within a sequential MOO algorithm is illustrated.

2. Multi-objective optimization using Gaussian Processes

2.1. Notions in MOO

Multi-objective optimizers aim at minimizing (say) several objectives at once: $f_1(\mathbf{x}), \ldots, f_m(\mathbf{x})$ with $\mathbf{x} = (x_1, \ldots, x_d)^T$ a vector of decision variable in $\mathbf{E}$ (usually $\mathbf{E} \subset \mathbb{R}^d$) and $f : \mathbf{E} \to \mathbb{R}^m$ the vector valued function whose coordinates are the f_i, $i = 1, \ldots, m$. As the objectives are usually in competition, there is no optimal solution minimizing every objective at once. This leads to the definition of a compromise solution following the Pareto dominance: a vector is said to be *dominated* if there exists another vector which is not worse in any objective and better for at least one. If a vector is not dominated by any other vector, it is optimal in the Pareto sense.

The set of optimal (or non-dominated) points in $\mathbf{E}$ is called *Pareto set* and the corresponding image by f, composed of non-dominated vectors, is called *Pareto front*. The optimization process aims at finding non-dominated objective vectors as close as possible to the true underlying Pareto front, creating a discrete approximation sometimes called a *Pareto front approximation* [14].

2.2. Kriging / Gaussian Process Regression

A common solution to perform optimization even with a tight evaluation budget is the construction of a probabilistic model relying on random fields for prediction. Originating from geostatistics and spatial statistics with a technique named *kriging* [15], it is known in the machine learning community under the term *Gaussian Process Regression* [16]. Such models have the property to interpolate the data. Furthermore, due to their probabilistic nature, they also provide a quantification of the prediction uncertainty.

Without loss of generality, here the responses are supposed to be independent (non-correlated) as advised in [17, 18] where no significant gain of using a dependent model has been shown. Following the settings of Gaussian Process Regression [19, 20], each of the objective functions f_i is supposed to be the realization of a centered Gaussian process Y_i :

$$Y_i(.) = \mathbf{g}^T(.)\boldsymbol{\beta}^{(i)} + Z_i(.) \quad (\textit{Universal Kriging})$$

where $\mathbf{g}(.)^T$ is a vector of known basis functions, $\boldsymbol{\beta}^{(i)}$ a vector of unknown coefficient and $Z_i(.)$ is a zero mean Gaussian process with given covariance function, or kernel, $k^{(i)}$. The responses are here considered to be deterministic but noise could be considered. By conditioning on n evaluations $\left\{Y_i(\mathbf{x}_1) = y_1^{(i)}, \ldots, Y_i(\mathbf{x}_n) = y_n^{(i)}, i \in \{1, \ldots, m\}\right\}$ denoted $\mathcal{A}_n$, the best predictor (or kriging mean) and the prediction covariance (also referred to as kriging covariance) are expressed as:

$$m_n^{(i)}(\mathbf{x}) = \mathbb{E}(Y_i(\mathbf{x})|\mathcal{A}_n) = \mathbf{g}(\mathbf{x})^T\hat{\boldsymbol{\beta}}^{(i)} + k_n^{(i)}(\mathbf{x})^T\mathbf{K}_n^{(i)-1}\left(\mathbf{y}_n^{(i)} - \mathbb{G}_n\hat{\boldsymbol{\beta}}^{(i)}\right),$$

$$\begin{aligned} c_n^{(i)}(\mathbf{x}, \mathbf{x'}) &= cov\left(Y_i(\mathbf{x}), Y_i(\mathbf{x'})|\mathcal{A}_n\right) \\ &= k^{(i)}(\mathbf{x}, \mathbf{x'}) - k_n^{(i)}(\mathbf{x})^T\mathbf{K}_n^{(i)-1}k_n^{(i)}(\mathbf{x'}) + \left(\mathbf{g}(\mathbf{x})^T - k_n^{(i)}(\mathbf{x})^T\mathbf{K}_n^{(i)-1}\mathbb{G}_n\right)^T\left(\mathbb{G}_n^T\mathbf{K}_n^{(i)-1}\mathbb{G}_n\right)^{-1}\left(\mathbf{g}(\mathbf{x'})^T - k_n^{(i)}(\mathbf{x'})^T K_n^{(i)-1}\mathbb{G}_n\right) \end{aligned}$$

where $\mathbf{y}_n^{(i)} = \left(y_1^{(i)}, \ldots, y_n^{(i)}\right)$, $\mathbf{K}_n^{(i)} = \left(k^{(i)}(\mathbf{x}_s, \mathbf{x}_t)\right)_{1 \leq s,t \leq n}$, $k_n^{(i)}(\mathbf{x}) = \left(k^{(i)}(\mathbf{x}, \mathbf{x}_1), \ldots, k^{(i)}(\mathbf{x}, \mathbf{x}_n)\right)^T$, $\mathbb{G}_n = \left(g(\mathbf{x}_1)^T, \ldots, g(\mathbf{x}_n)^T\right)^T$ and $\hat{\boldsymbol{\beta}}^{(i)} = \left(\mathbb{G}_n^T \mathbf{K}_n^{(i)-1} \mathbb{G}_n\right)^{-1} \mathbb{G}_n^T \mathbf{K}_n^{(i)-1} y_n^{(i)}$.

The covariance functions are chosen according to prior hypothesis about the unknown functions, such as regularity, sparsity, possible symmetries, etc. [21]. While there exists a variety of admissible covariance functions, the most commonly chosen are the stationary "Gaussian" and "Matérn" kernels [19]. Maximum likelihood is often used to estimate values for the kernel hyperparameters [16]. An example of kriging model is proposed in Figure 1, a).

2.3. Multi-objective expected improvement

Sequential approaches in MOO aim at adding new observations with a balance between exploration and exploitation. Similar to [4], several extensions of the EGO algorithm have been proposed for MOO. The main idea is to derive criteria in the vein of the *Expected Improvement* by defining a generalization of the notion of improvement for multiple objectives. Popular methods include scalarization approaches like ParEGO [5] or MOEAD-EGO [6] or truly multi-objective methods based on the definition of improvement functions over the current Pareto front $\mathcal{P}_n$ with respect to the observations. Considered improvement functions are respectively based on Euclidean distance [22], Hypervolume [7] or Maximin distance [18] i.e. the distance to the closest point of $\mathcal{P}_n$, the volume added over $\mathcal{P}_n$ and an axis-wise distance to $\mathcal{P}_n$.

In the application of Section 4, we use the Expected Hypervolume Improvement to sequentially add points. This criterion has been efficiently applied to problems with limited budget [23], enjoys some theoretical properties ([8, 24]) and furthermore is related to the concept of attainment function which is of particular importance in what follows.

2.4. Conditional simulations

Following 2.2, let us consider that the unknown objective functions are sample paths from Y_i ($i = 1, \ldots, m$) interpolating the observations, more precisely realizations of Y_i conditionally on $\mathcal{A}_n$ (i.e. realization of a conditional Gaussian process). Besides, under the Gaussian process assumptions, the kriging predictor corresponds to the theoretical expectation of those sample paths and for a given $\mathbf{x}$, the probability for a conditional simulation to be in a given interval is related to the kriging variance.

Conditional simulations can be generated using a variety of methods, from matrix decomposition to spectral method, as presented in [25, 26, 27, 28]. They have been applied in mono-objective optimization in [29] as a tool to estimate an information gain when no analytical formula is available, as opposed to the Expected Improvement. Examples of conditional simulations are displayed Figure 1, b). In the multi-objective case, conditional simulations have not been used since some analytical formulas exist in few cases. This leads to the main question of this article: how can conditional simulation apply for MOO?

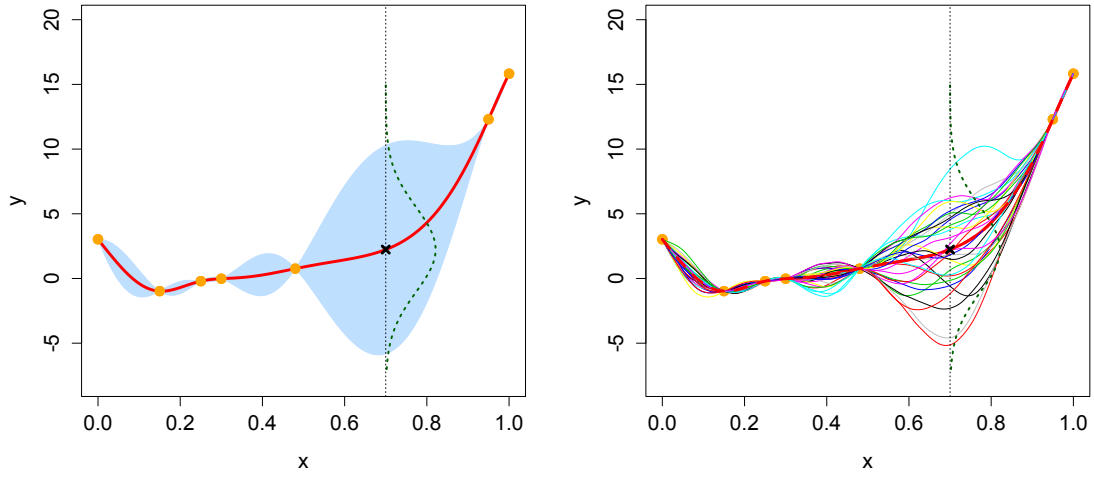
3. Quantification of uncertainty

In this section we assume that a kriging model (see Section 2.2) is fitted to each objective function, given a set of n observations $\mathcal{A}_n$. We denote $Y_1^C, \ldots, Y_m^C$ the corresponding conditional Gaussian processes. The Y_j^C ($j = 1, \ldots, m$) allow us to generate conditional Pareto front realizations and further estimate the uncertainty on the Pareto front. Indeed, they provide a convenient framework to apply concepts from random sets theory.

3.1. Conditional simulation for MOO: generation of conditional Pareto fronts and corresponding attained sets

Here we use conditional simulation to generate so-called *conditional Pareto fronts* (CPF). It is done by producing conditional GP simulations Y_j^C modeling the objectives at some simulation points in the design space, before selecting the non-dominated responses, as described in Algorithm 1. Doing this provides conditional Pareto sets and fronts, as illustrated in Figure 2.

Note that what we denote by CPF are actually approximations of theoretical conditional Pareto fronts, just like conditional simulations of Gaussian random fields are generally approximated realizations relying on a finite number



(a) (b)

Figure 1: a) Example of kriging model based on the observations (points), with kriging mean (bold line) and kriging 95% confidence intervals (shaded area). The Gaussian predictive distribution for $x = 0.7$ is represented by the vertical dashed line. b) Conditional simulations (thin lines) from the fitted Gaussian process model.

of points.

Algorithm 1 Simulation of N conditional Pareto sets and fronts

for $i = 1, 2, \ldots, N$ **do**
  Choose p simulation points $\mathbf{x}_1, \ldots, \mathbf{x}_p$ in $\mathbf{E}$ (fixed or different at each iteration).
  for $j = 1, 2, \ldots, m$ **do**
    Generate a conditional simulation at $\mathbf{x}_1, \ldots, \mathbf{x}_p$ for the j^{th} objective: $\mathbf{Y}_j^{C,(i)} = \left(Y_j^{C,(i)}(\mathbf{x}_1), \ldots, Y_j^{C,(i)}(\mathbf{x}_p)\right)$.
  end for
  Determine the Pareto set and front of $\left\{\mathbf{Y}_1^{C,(i)}, \ldots, \mathbf{Y}_m^{C,(i)}\right\}$.
end for

From now on we focus on the use of the CPFs uniquely, since the practitioner is mostly interested in visualizing results in the objective space. Each of the CPF is composed of non-dominated points in the objective space. They have been considered to assess the performance of MO optimizers [30, 14] under the term Random Non-dominated Point (RNP) sets: sets of random vectors in $\mathbb{R}^m$, non-dominated with respect to each other and with random finite cardinality (see e.g. [31]). An alternative view is to consider the set of all objective vectors dominated by a realization of a RNP set, called an *attained set*. Realizations of RNP sets and the corresponding attained sets are presented in Figure 3.

Similarly to the definition of the kriging mean and confidence intervals with conditional simulations, it would be interesting to define their equivalent for the simulated CPFs and corresponding attained sets. Nevertheless, defining an expectation and/or an index of variability for those is not straightforward and requires concepts from random sets theory [13].

3.2. Basics from random sets theory: quantifying uncertainty with the Vorob'ev deviation

Set-valued random elements, in particular *random closed sets* [13] have gained in popularity over the last decades. There exists several candidate notions to define the mean of a random closed set, see [13] (Chapter 2). We choose a rather intuitive one, based on the notion of coverage function:

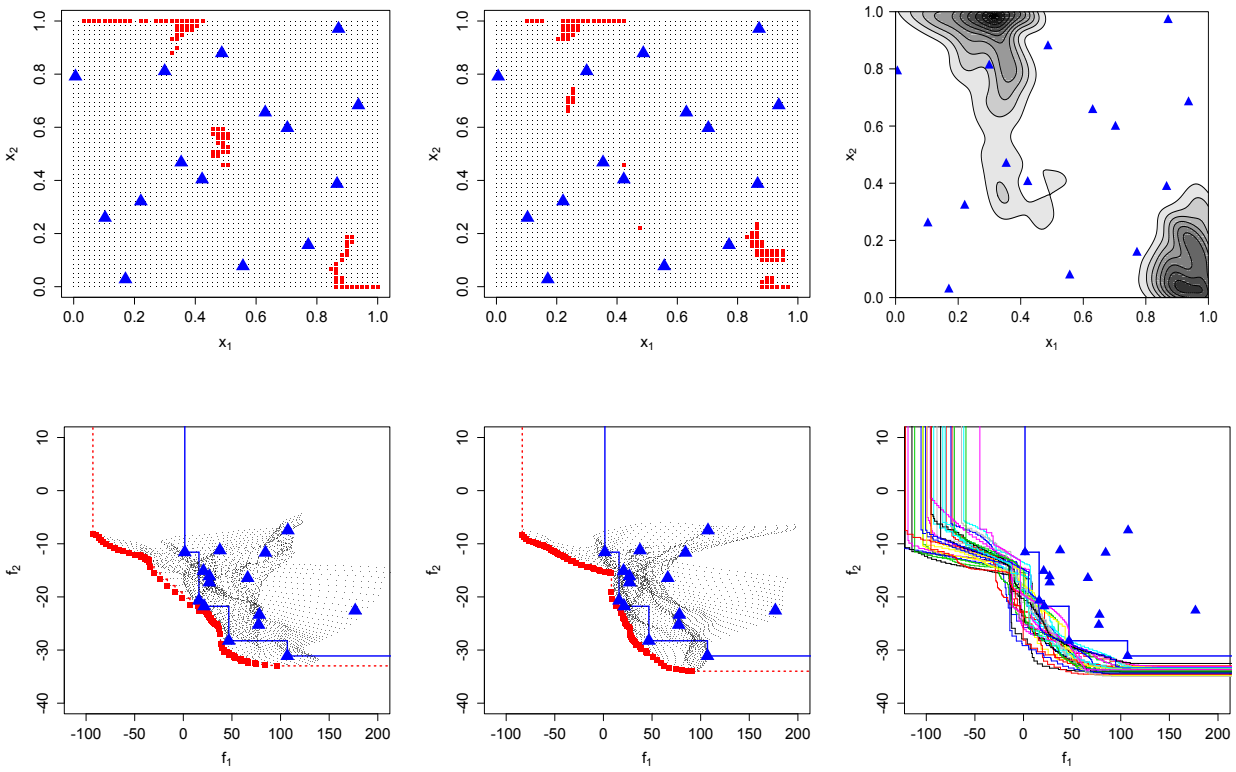


Figure 2: Conditional Pareto sets and fronts corresponding to the GP models Y_1^C, Y_2^C, based on the observations $\mathcal{A}_n$ represented by blue triangles. Left and center: examples of two conditional Pareto sets (top) and fronts (bottom) simulations, where the simulations are performed on a regular 100×100 grid. The simulation points and simulated responses are plotted with dots. The corresponding non-dominated points are represented by red squares. Right: contour plot of the probability density of optimal points in the decision space estimated from 30 conditional simulations (top) and superposition of simulated conditional Pareto fronts (bottom).

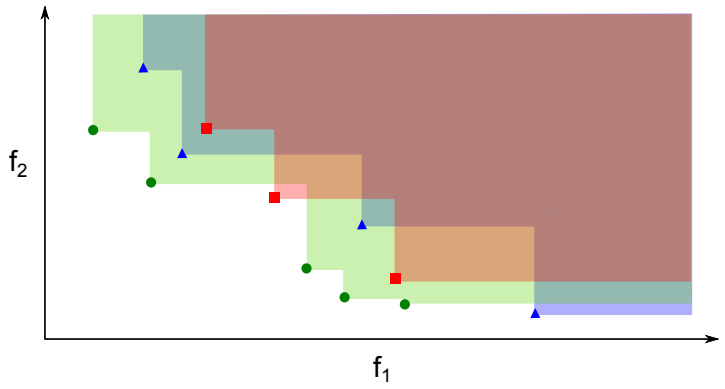


Figure 3: Example of 3 realizations of RNP sets (points, triangles and squares) and the corresponding attained sets (shaded areas).

Definition 1 (Coverage function). Let $\mathcal{Y}$ be a random closed set on a space D (here $D \subset \mathbb{R}^m$). The coverage function $p_{\mathcal{Y}}$ is defined by $p_{\mathcal{Y}} : x \in D \mapsto P(x \in \mathcal{Y})$.

This definition has been applied in the kriging framework to estimate sets of critical input values [11, 12]. It uses the Vorob'ev expectation, based on the upper level sets $Q_\beta = \{z \in \mathbb{R}^m, p_{\mathcal{Y}}(z) \geq \beta\}$, called β-quantiles.

Definition 2 (Vorob'ev expectation). Denote μ the Lebesgue measure on $\mathbb{R}^m$. Assuming that $\mathbb{E}(\mu(\mathcal{Y})) < +\infty$, the Vorob'ev expectation is the β^*-quantile Q_{β^*} such that $\mathbb{E}(\mu(\mathcal{Y})) = \mu(Q_{\beta^*})$ if this equation has a solution, if not it is defined from the condition $\mu(Q_\beta) \leq \mathbb{E}(\mu(\mathcal{Y})) \leq \mu(Q_{\beta^*}), \forall \beta > \beta^*$.

The Vorob'ev deviation is then defined by:

Definition 3 (Vorob'ev deviation). Assuming that $\mathbb{E}(\mu(\mathcal{Y})) < +\infty$, $\mathbb{E}(\mu(Q_{\beta^*} \Delta \mathcal{Y}))$ is called the *Vorob'ev deviation* of $\mathcal{Y}$, where Δ denotes the symmetric difference between sets.

The Vorob'ev expectation is a global minimizer of the deviation among all deterministic closed sets with volume equal to the average volume of $\mathcal{Y}$ (see [13] for a proof).

3.3. Application to the uncertainty quantification on Pareto fronts

Attained sets are closed[1] and unbounded subsets in $\mathbb{R}^m$. Hence, the attained sets obtained with the simulated CPFs are modeled as realizations of a random closed set and are denoted by $\mathcal{Y}_i$, $(i = 1, \ldots, N)$.

In the MOO literature, the study of distribution location and spread of an attained set $\mathcal{X}$ rely on the attainment function $\alpha_{\mathcal{X}}$ [31]: the probability for a given point in the objective space to be dominated by a RNP set, which is in fact a coverage function of $\mathcal{X}$. For proofs about the equivalence of the distribution of a RNP set and the corresponding attained set as well as for a definition of the attainment function in terms of coverage function, the interested reader is referred to [31].

In practice the attainment function is estimated by taking the mean number of RNP sets dominating a vector in the objective space:

Definition 4 (Empirical attainment function). $\hat{\alpha}_N(z) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{I}\{z \in \mathcal{Y}_i\}$

In our context this empirical attainment function is simply computed by taking the average of the number of conditional simulations dominating a given value in $\mathbb{R}^m$, as presented in Algorithm 2. An example of empirical attainment function is presented in Figure 4, showing where in the objective space there is a high probability to improve on the current Pareto front.

Algorithm 2 Empirical Attainment of the $\mathcal{Y}_i, i = 1, \ldots, N$

Select q points $z_1, \ldots, z_q$ for the computation of the attainment function (typically a grid) in the objective space
for j = 1,2,..., q **do**
  Compute the frequency of CPFs dominating z_j: $\hat{\alpha}_N(z_j)$
end for

It should be emphasized that $\mathcal{Y}$ is required to be bounded for its Vorob'ev expectation to exist. Hence it is necessary to define a reference point to bound the integration domain, similarly to the choice of the reference point for the hypervolume quality indicator [24]. Unless there is a previous knowledge about the range of the objectives, we choose the extremal non-dominated points reached by conditional simulation to bound the integration domain.

We can now determine the value of the Vorob'ev threshold β^*: the value of β corresponding to the Vorob'ev expectation, as described in Algorithm 3. The Pareto frontier of the Vorob'ev expectation provides us with an estimate of the Pareto front, as illustrated in Figure 4.

Algorithm 3 Estimation of the Vorob'ev threshold

1: Define the integration domain by finding the extremal values for the objectives over the RNP sets realizations:
$\Omega = \left[\min_{i\in(1,\dots,N)} Y_1^{(i)}, \max_{i\in(1,\dots,N)} Y_1^{(i)}\right] \times \cdots \times \left[\min_{i\in(1,\dots,N)} Y_m^{(i)}, \max_{i\in(1,\dots,N)} Y_m^{(i)}\right]$

2: Determine the average volume of the attained sets $\mathcal{Y}_i$ by calculating the average dominated hypervolume over Ω:

$$\frac{1}{N}\sum_{i=1}^{N}\int_{\Omega}\mathbf{I}\{z \in \mathcal{Y}_i\}\mu(\mathrm{d}z)$$

3: Find the value of β^* by dichotomy.

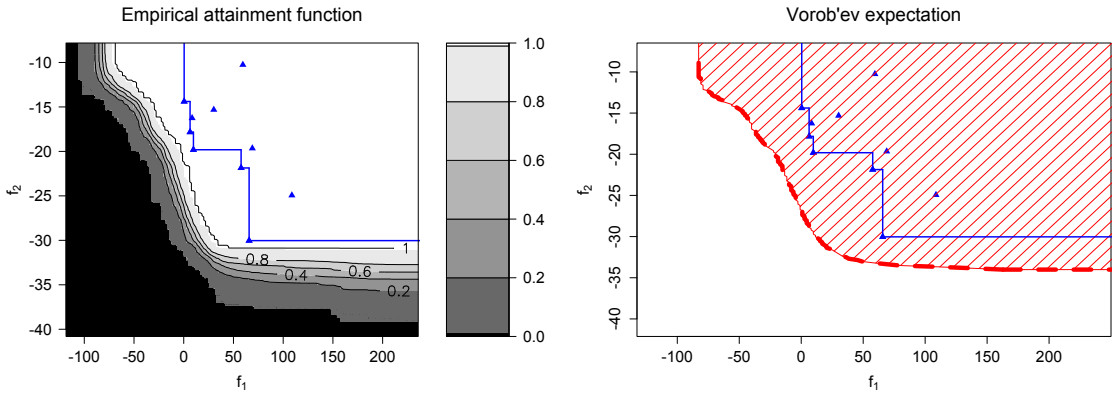


Figure 4: Example of empirical attainment function (left) and corresponding Vorob'ev expectation in the shaded area (right). The estimation of the underlying Pareto front (dashed line) is the Pareto frontier of the Vorob'ev expectation. The observations are marked by blue triangles.

Finally, given β^*, we can compute the Vorob'ev deviation. This is done by taking the average volume of the symmetric difference between the Vorob'ev expectation and each simulated CPF. The value of the Vorob'ev deviation gives an idea about the variability of the simulated CPF and can be monitored as observations are added.

From a practical point of view, it is also useful for visualization purposes to display the superposition of all the symmetric differences by defining an analogue of the attainment function:

Definition 5 (Symmetric-deviation function). The function $\delta_{\mathcal{Y}} : z \in \mathbb{R}^m \mapsto P(z \in \mathcal{Q}_{\beta^*}\Delta\mathcal{Y})$ is called the symmetric-deviation function of $\mathcal{Y}$.

$\delta_{\mathcal{Y}}$ is the coverage function of $\mathcal{Q}_{\beta^*}\Delta\mathcal{Y}$. It is estimated with the empirical symmetric-deviation function:

$$\hat{\delta}_N(z) = \frac{1}{N}\sum_{i=1}^{N}\mathbf{I}\{z \in \mathcal{Q}_{\beta^*}\Delta\mathcal{Y}_i\}.$$

Figure 5 presents an example of a symmetric difference between two sets and an empirical symmetric-deviation function. This shows the variability around the estimated Pareto front: the darker an area is, the more often it is only dominated by either the attained set or the Vorob'ev expectation. White areas denotes regions where both agree.

[1] as a finite union of closed sets (hyper quadrants)

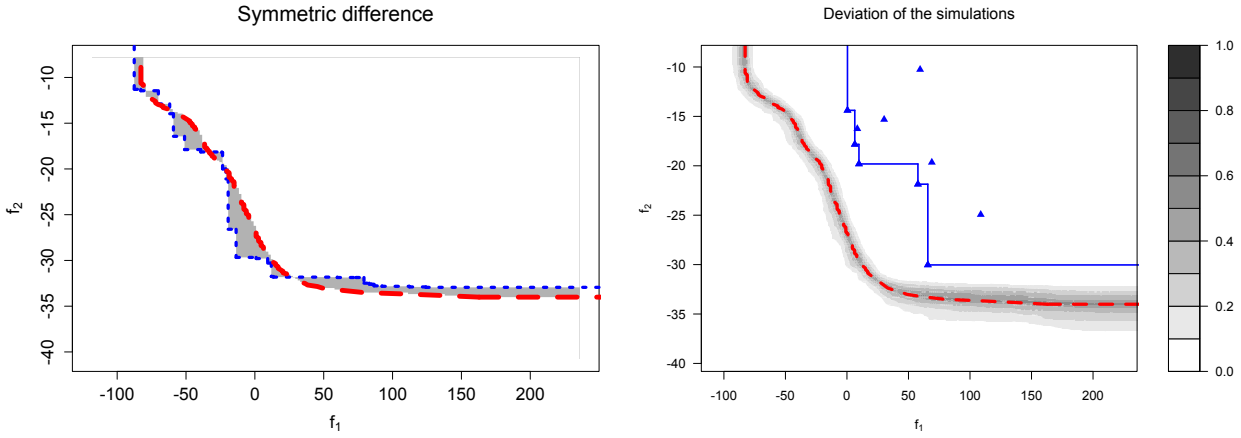


Figure 5: Left: symmetric difference between the Vorob'ev expectation (the level line of the Vorob'ev threshold is represented by the dashed line) and a simulated CPF's attained set (dotted line). Right: illustration of the deviation around the estimated Pareto front with an example of empirical symmetric-deviation function (level plot).

4. Application

In this section, we illustrate the benefits of the proposed methodology for estimating the Pareto front. We consider the following bi-objective optimization problems from the literature:

(P1) The problem presented in [32], which has a convex Pareto front:

$$f_1(\mathbf{x}) = \left(b_2 - \frac{5.1}{4\pi^2}b_1^2 + \frac{5}{\pi}b_1 - 6\right)^2 + 10\left[\left(1 - \frac{1}{8\pi}\right)\cos(b_1) + 1\right]$$

$$f_2(\mathbf{x}) = -\sqrt{(10.5 - b_1)(b_1 + 5.5)(b_2 + 0.5)} - \frac{1}{30}\left(b_2 - \frac{5.1}{4\pi^2}b_1^2 - 6\right)^2 - \frac{1}{3}\left[\left(1 - \frac{1}{8\pi}\right)\cos(b_1) + 1\right]$$

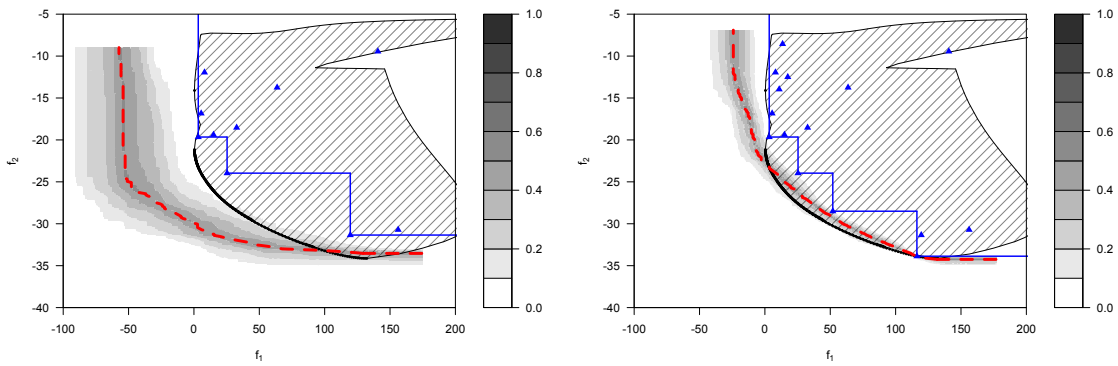
where $b_1 = 15x_1 - 5$, $b_2 = 15x_2$ and $x_1, x_2 \in [0, 1]$.

(P2) The ZDT3 problem [33] which has a disconnected Pareto front.

For each example, we start with a set of few observations that allow fitting initial Gaussian process models for the two objective functions. Then we add new points sequentially by maximizing the Expected Hypervolume Improvement, based on the formula detailed in [34]. At each step, the Gaussian process models are updated and their hyperparameters re-estimated. These models are then used to simulate CPFs, from which we compute the estimates of the Vorob'ev mean and the measures of uncertainty: Vorob'ev deviation and symmetric difference deviation. Since the integration domain varies as points are added, the values are displayed divided by the volume of this integration domain. The following test problems are fast to compute, so it is possible to compare the outcome of the proposed workflow to a reference Pareto front by using the volume of the symmetric difference.

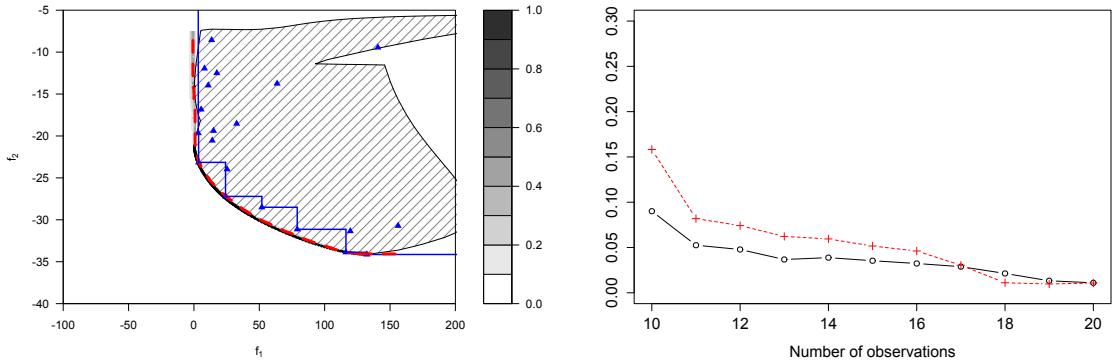
The results are presented in Figure 6 and Figure 7, showing the evolution of the estimated Pareto fronts with the corresponding uncertainty around it. For the problem (P1) the sequence is detailed, demonstrating the strength of the proposed approach for giving insights of the uncertainty on the Pareto front. In particular, the uncertainty measures are helpful for choosing a minimal number of observations for approximating the Pareto front: while 10 initial observations may not be enough (Figure 7, a) regarding the large symmetric-deviation, adding 10 observations more reduces dramatically the uncertainty (Figure 7, c).

The conclusions are similar for problem (P2), where the Pareto front is disconnected, starting this time with 20 observations and adding again ten by optimization.



(a) 10 initial observations

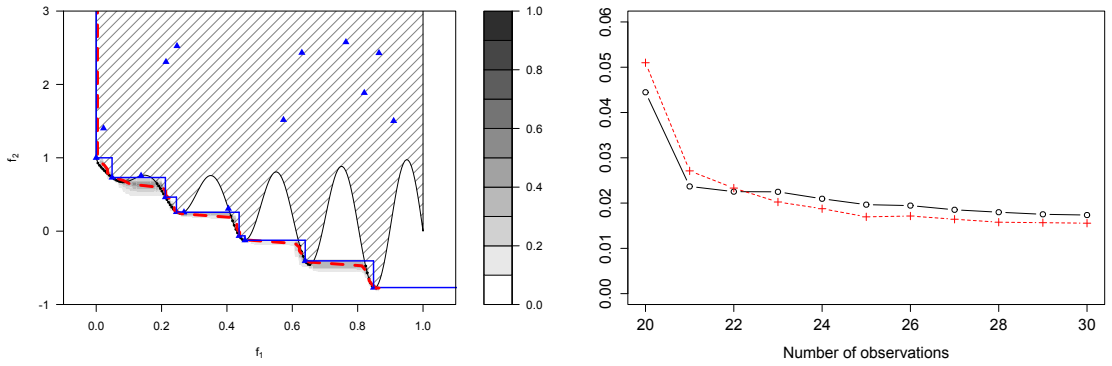
(b) 15 observations



(c) 20 observations

(d)

Figure 6: Evolution of the deviation with new observations added using Expected Hypervolume Improvement for Problem (P1). The shaded area represents the image of $\mathbf{E}$ by f with a thicker border for the Pareto front. Observations are marked with blue triangles and the blue solid line represents the current Pareto front. The dashed line is the estimated Pareto front, with the corresponding values of the symmetric-deviation in level plot. Bottom right: evolution of the scaled Vorob'ev deviation (black solid line with circles) and of the volume of the symmetric difference between the Pareto front and its estimation (red dotted line with crosses).



(a) 30 observations

(b)

Figure 7: Evolution of the deviation with new observations added using Expected Hypervolume Improvement for Problem (P2). The figure description is the same than Figure 6

Comments on the conditional simulations

- The aforementioned methodology depends on the number and location of simulation points used to obtain the CPFs from the Gaussian process models. As a first study, we have compared two sampling strategies: uniform sampling and space-filling sampling relying on a Sobol sequence. The objective functions are taken as sample paths of centered Gaussian processes with Matérn covariance kernel ($\nu = 5/2$), with parameters equal to $0.3/\sqrt{3}$ for f_1, and $0.5/\sqrt{3}$ for f_2. We compute the approximation error over the set of non-dominated points obtained from the two sampling strategies. Three error indicators are used: Hypervolume difference, epsilon and R2 quality indicators [35]. The tests are repeated one hundred times. The results presented on Figure 8 show that space-filling sampling outperforms uniform sampling to get an accurate estimation of the Pareto front.

- The error due to conditional simulation must be mitigated and put into perspective with the considered problems. Until now optimization methods based on kriging have been applied with up to six variables/objectives [36], giving the hope of getting reliable estimates with conditional simulations using smart sampling schemes.

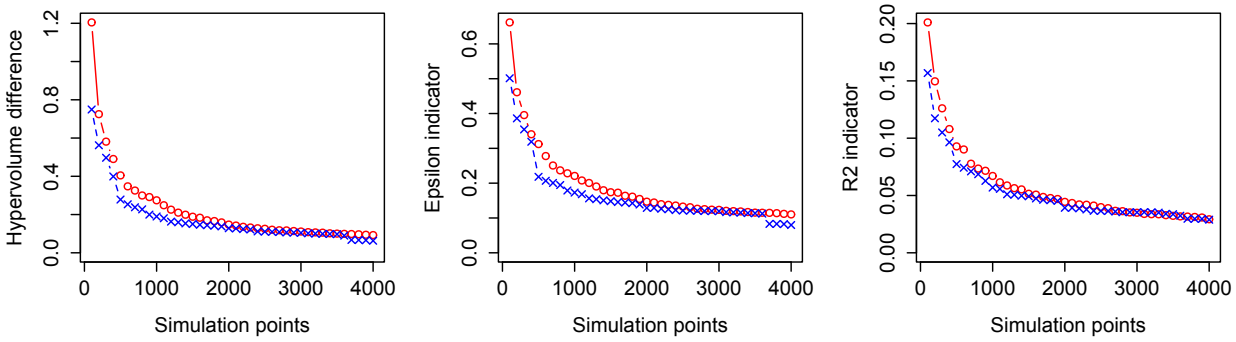


Figure 8: Hypervolume difference, epsilon and R2 quality indicators for the Pareto front, obtained by uniform sampling (points) or with a Sobol sequence (crosses) for an increasing number of points. The reference needed to compute the indicators corresponds to the best points obtained by both sampling strategies and an extra 50×50 regular grid.

5. Conclusion and perspectives

We presented an original methodology to estimate and visualize the uncertainty of Pareto front approximations, based on Gaussian process conditional simulations. Two uncertainty measures were defined relying on the theory of random sets, through the concept of Vorob'ev deviation. As illustrated on two bi-objective problems with convex or disconnected Pareto fronts, these measures can be used as a basis to define stopping criteria in a sequential framework.

Further work is needed to analyze the different kinds of uncertainty and biases that may occur when applying the proposed methodology. Perspectives also include the integration of the proposed uncertainty estimate in a SUR-strategy [37] to possibly reduce the uncertainty faster.

Acknowledgment

This work has been conducted within the frame of the ReDice Consortium, gathering industrial (CEA, EDF, IFPEN, IRSN, Renault) and academic (Ecole des Mines de Saint-Etienne, INRIA, and the University of Bern) partners around advanced methods for Computer Experiments.

References

[1] K. Deb, Introduction to evolutionary multiobjective optimization, in: J. Branke, et al. (Eds.), Multiobjective Optimization, Vol. 5252 of LNCS, Springer, 2008, pp. 59–96.

[2] X. Liao, Q. Li, X. Yang, W. Zhang, W. Li, Multiobjective optimization for crash safety design of vehicles using stepwise regression model, Structural and Multidisciplinary Optimization 35 (6) (2008) 561–569.

[3] L. Santana-Quintero, A. Montano, C. Coello, A review of techniques for handling expensive functions in evolutionary multi-objective optimization, Computational Intelligence in Expensive Optimization Problems (2010) 29–59.

[4] D. Jones, M. Schonlau, W. Welch, Efficient global optimization of expensive black-box functions, Journal of Global Optimization 13 (4) (1998) 455–492.

[5] J. Knowles, ParEGO: a hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems, IEEE Transactions on Evolutionary Computation 10 (1) (2006) 50–66.

[6] Q. Zhang, W. Liu, E. Tsang, B. Virginas, Expensive multiobjective optimization by MOEA/D with Gaussian process model, Evolutionary Computation, IEEE Transactions on 14 (3) (2010) 456–474.

[7] M. Emmerich, K. Giannakoglou, B. Naujoks, Single-and multiobjective evolutionary optimization assisted by Gaussian random field metamodels, Evolutionary Computation, IEEE Transactions on 10 (4) (2006) 421–439.

[8] T. Wagner, M. Emmerich, A. Deutz, W. Ponweiser, On expected-improvement criteria for model-based multi-objective optimization, Parallel Problem Solving from Nature–PPSN XI (2010) 718–727.

[9] K. Miettinen, Nonlinear multiobjective optimization, Vol. 12, Springer, 1999.

[10] T. Goel, N. Stander, A study on the convergence of multiobjective evolutionary algorithms, in: Preprint submitted to the 13th AIAA/ISSMO conference on Multidisciplinary Analysis Optimization, 2010, pp. 1–18.

[11] C. Chevalier, D. Ginsbourger, J. Bect, I. Molchanov, Estimating and quantifying uncertainties on level sets using the Vorobev expectation and deviation with Gaussian process models, in: D. Ucinski, A. C. Atkinson, M. Patan (Eds.), mODa 10 Advances in Model-Oriented Design and Analysis, Contributions to Statistics, Springer, 2013, pp. 35–43.

[12] C. Chevalier, Fast uncertainty reduction strategies relying on Gaussian process models, Ph.D. thesis, University of Bern (2013).

[13] I. Molchanov, Theory of random sets, Springer, 2005.

[14] E. Zitzler, J. Knowles, L. Thiele, Quality assessment of Pareto set approximations, Multiobjective Optimization (2008) 373–404.

[15] G. Matheron, Principles of geostatistics, Economic geology 58 (8) (1963) 1246–1266.

[16] C. E. Rasmussen, C. Williams, Gaussian Processes for Machine Learning, MIT Press, 2006.

[17] J. Kleijnen, E. Mehdad, Kriging in multi-response simulation, including a Monte Carlo laboratory, CentER Discussion Papers Series (2012-039).

[18] J. Svenson, T. Santner, Multiobjective optimization of expensive black-box functions via expected maximin improvement, Tech. rep., 43210, Ohio University, Columbus, Ohio (2010).

[19] M. L. Stein, Interpolation of spatial data: some theory for kriging, Springer, 1999.

[20] J. Sacks, W. J. Welch, T. J. Mitchell, H. P. Wynn, Design and analysis of computer experiments, Statistical science 4 (4) (1989) 409–423.

[21] D. Ginsbourger, O. Roustant, N. Durrande, Invariances of random fields paths, with applications in Gaussian process regression, arXiv preprint arXiv:1308.1359.

[22] A. J. Keane, Statistical improvement criteria for use in multiobjective design optimization, AIAA journal 44 (4) (2006) 879–891.

[23] W. Ponweiser, T. Wagner, D. Biermann, M. Vincze, Multiobjective optimization on a limited budget of evaluations using model-assisted S-metric selection, in: G. Rudolph, et al. (Eds.), PPSN X, Vol. 5199 of LNCS, Springer, 2008, pp. 784 – 794.

[24] A. Auger, J. Bader, D. Brockhoff, E. Zitzler, Hypervolume-based multiobjective optimization: Theoretical foundations and practical implications, Theoretical Computer Science 425 (2012) 75–103.

[25] A. G. Journel, Geostatistics for conditional simulation of ore bodies, Economic Geology 69 (5) (1974) 673–687.

[26] M. Hoshiya, Kriging and conditional simulation of Gaussian field, Journal of engineering mechanics 121 (2) (1995) 181–186.

[27] P. J. Diggle, P. J. Ribeiro, Model-based geostatistics, Springer, 2007.

[28] O. Roustant, D. Ginsbourger, Y. Deville, DiceKriging, DiceOptim: Two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization, Journal of Statistical Software 51 (1) (2012) 1–55.

[29] J. Villemonteix, E. Vazquez, E. Walter, An informational approach to the global optimization of expensive-to-evaluate functions, Journal of Global Optimization 44 (4) (2009) 509–534.

[30] C. M. Fonseca, V. G. Da Fonseca, L. Paquete, Exploring the performance of stochastic multiobjective optimisers with the second-order attainment function, in: Evolutionary Multi-Criterion Optimization, Springer, 2005, pp. 250–264.

[31] V. G. da Fonseca, C. M. Fonseca, The attainment-function approach to stochastic multiobjective optimizer assessment and comparison, in: Experimental methods for the analysis of optimization algorithms, Springer, 2010, pp. 103–130.

[32] J. M. Parr, Improvement criteria for constraint handling and multiobjective optimization, Ph.D. thesis, University of Southampton (2012).

[33] E. Zitzler, K. Deb, L. Thiele, Comparison of multiobjective evolutionary algorithms: Empirical results, Evolutionary computation 8 (2) (2000) 173–195.

[34] M. T. Emmerich, A. H. Deutz, J. W. Klinkenberg, Hypervolume-based expected improvement: Monotonicity properties and exact computation, in: Evolutionary Computation (CEC), 2011 IEEE Congress on, IEEE, 2011, pp. 2147–2154.

[35] E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, V. G. da Fonseca, Performance assessment of multiobjective optimizers: An analysis and review, Evolutionary Computation, IEEE Transactions on 7 (2) (2003) 117–132.

[36] J. D. Svenson, Computer experiments: Multiobjective optimization and sensitivity analysis, Ph.D. thesis, The Ohio State University (2011).

[37] C. Chevalier, J. Bect, D. Ginsbourger, E. Vazquez, V. Picheny, Y. Richet, Fast parallel kriging-based stepwise uncertainty reduction with application to the identification of an excursion set, To appear in Technometrics.