

## A STOCHASTIC-KRIGING-BASED MULTIOBJECTIVE SIMULATION OPTIMIZATION ALGORITHM

Sebastian Rojas-Gonzalez

Department of Decision Sciences  
and Information Management  
KU Leuven  
Naamsestraat 69  
Leuven, 3000, BELGIUM

Hamed Jalali

Department of Information Systems,  
Supply Chain and Decision Making  
Neoma Business School  
1 Rue du Marechal Juin  
Rouen, 76130, FRANCE

Inneke Van Nieuwenhuysse

Research Group Logistics  
Hasselt University  
Agoralaan, Building D  
Diepenbeek, 3590, BELGIUM

### ABSTRACT

We consider the multiobjective simulation optimization problem, where we seek to find the non-dominated set of designs evaluated using noisy simulation evaluations, in the context of numerically expensive simulators. We propose a metamodel-based scalarization approach built upon the famous ParEGO algorithm. Our approach mainly differentiates from ParEGO and similar algorithms in that we use stochastic kriging, which explicitly characterizes both the extrinsic uncertainty of the unknown response surface, and the intrinsic uncertainty inherent in a stochastic simulation. We additionally integrate the Multiobjective Optimal Computing Budget Allocation ranking and selection procedure in view of maximizing the probability of selecting systems with the true best expected performance. We evaluate the performance of the algorithm using standard benchmark test functions for multiobjective optimizers, perturbed by heterogeneous noise. The experimental results show that the proposed method outperforms its deterministic counterpart based on well-known quality indicators and the fraction of the true Pareto set found.

### 1 INTRODUCTION

Deterministic and stochastic simulation models have commonly been used to analyze and predict the behavior of complex real-world engineered systems, and to support difficult decision making processes. Researchers in the field of *design and analysis of computer experiments* (DACE), have devoted special attention to the development of *simulation optimization* techniques that seek to provide answers with limited computing budget. In general, the use of optimization in simulation comes with a series of obstacles: (a) it may require long running times, especially for high-dimensional problems, or when multiple replications are needed to achieve sufficient accuracy (as with stochastic simulation); and (b) there is a lack of analytic gradients, as the simulation model acts as a *black box* that numerically evaluates the objective vector for any given decision vector (see Figure 1).



Figure 1: Black box optimization problem.

The objective of the methodology described in this paper is to offer an efficient approach for the optimization of stochastic discrete-event simulation problems with multiple objectives. The approach employs *metamodeling*, also known as *surrogate modeling*, in order to speed up computations in settings where the simulation is computationally expensive. Evidently, in order for such metamodels to be useful, they have to be accurate and sufficiently cheap-to-evaluate. In particular, we propose a *sequential* strategy to solve stochastic multiobjective problems with box constraints.

For the multiobjective case, the optimization problem can be formulated as:  $\min[f^1(\mathbf{x}), \dots, f^m(\mathbf{x})]$  for  $m$  objectives and a vector of decision variables  $\mathbf{x} = [x_1, \dots, x_d]^T$  in the decision space  $D$  (usually  $D \subset \mathbb{R}^d$ ), with  $f : D \rightarrow \mathbb{R}^m$  the vector-valued function with coordinates  $f^1, \dots, f^m$  in the objective space  $\Theta \subset \mathbb{R}^m$ . The goal is to find a set  $F$  of all vectors  $\mathbf{x}^* = [x_1^*, \dots, x_d^*]^T$  that yields the optimum values for all objectives and satisfies all the constraints.

Given that the objectives are usually competing, finding a global optimum for all objectives is unlikely to happen. This leads to a tradeoff between objectives. The optimization task is then to find a set of points that yield a solution where one objective cannot be improved without negatively affecting the output quality of another objective. The points in this solution set are referred to as *non-dominated* or *Pareto-optimal* points, and form the *Pareto set*. The evaluation of these solutions in the objective space corresponds to the *Pareto front*. Assuming minimization of all objectives, for  $\mathbf{x}_1$  and  $\mathbf{x}_2$  two vectors in  $D$ ,  $\mathbf{x}_1 \prec \mathbf{x}_2$  means  $\mathbf{x}_1$  dominates  $\mathbf{x}_2$  iff  $f^j(\mathbf{x}_1) \leq f^j(\mathbf{x}_2) \forall j \in \Theta$  and  $\exists j \in \Theta$  such that  $f^j(\mathbf{x}_1) < f^j(\mathbf{x}_2)$ .

A wide variety of approaches have been proposed in the field of multiobjective optimization, where each of them aims to find a representative set of Pareto optimal solutions, or the most preferred solution by the decision-maker. Embedding a metamodel into the optimization framework speeds up the process through the evaluation of fast mathematical approximations of some aspect of the system performance, instead of resorting to the expensive-to-evaluate primary model. This is especially relevant in stochastic simulation optimization, where multiple replications per design are required to achieve sufficient accuracy.

The method proposed in this paper is built upon the famous ParEGO framework of Knowles (2006), a scalarization algorithm for multiobjective optimization, which extends the seminal Efficient Global Optimization (EGO) work of Jones et al. (1998) to multiobjective settings. Both EGO and ParEGO use *deterministic kriging* metamodels during the search phase to sample candidate points. Our work is different from ParEGO and similar algorithms as:

1. We use *stochastic kriging*, proposed by Ankenman et al. (2010), which explicitly characterizes both the *extrinsic uncertainty* of the unknown response surface, and the *intrinsic uncertainty* inherent in a stochastic simulation.
2. We integrate the Multiobjective Optimal Computing Budget Allocation (MOCBA, Lee et al. 2010) procedure for *ranking and selection* (R&S) of points with the best expected performance.

We vary the structure of the noise (the noise is thus heterogeneous), and the level of the noise (implementing high and low noise perturbations on the objective outcomes). Most of the work in the literature so far considers only homogeneous noise (see e.g., Knowles et al. 2009; Horn et al. 2017; Koch et al. 2015). To the best of our knowledge, this algorithm is the first to combine metamodeling and MOCBA to solve multiobjective stochastic optimization problems. We test the algorithm using standard benchmark functions for multiobjective optimizers (Huband et al. 2006), and evaluate its performance by means of well-known quality indicators found in the literature (Zitzler et al. 2008).

The remainder of this article is structured as follows: In section 2 we provide the basic theory of kriging metamodeling, for both deterministic and stochastic settings. Section 3 explains the sequential nature of kriging-based optimization algorithms, and describes the steps followed by the proposed algorithm. Section 4 discusses the performance evaluation of stochastic multiobjective optimizers, and depicts the numerical results of the experiments with the proposed algorithm and its deterministic counterpart. We conclude the article in section 5 and identify promising directions for further research.

## 2 KRIGING METAMODELING

*Deterministic kriging*, or *Gaussian process regression* (GPR, Sacks et al. 1989; Rasmussen and Williams 2006), is an approximation method that can give predictions of unknown values of a random function, random field, or random process (Van Beers and Kleijnen 2003). Kriging assumes that the closer the input data, the more positively correlated the prediction errors (Kleijnen 2015). This assumption is modeled through a second-order stationary covariance process.

Kriging-based optimization algorithms are increasingly popular in simulation optimization, as they alleviate the computational cost of optimization, especially when the simulation itself is time-consuming. In particular, they perform efficiently because they not only approximate outputs over the entire search space (i.e., the response surface), but a quantification of the prediction uncertainty is also provided through the mean square error (MSE), also known as *kriging variance*. More formally, let  $f(\mathbf{x})$  be an unknown function, where  $\mathbf{x} = (x_1, \dots, x_d)^T$  is a vector of design variables of dimension  $d$ . In the interest of fitting a metamodel for the response  $f(\mathbf{x})$  at  $n$  design points, **deterministic kriging** assumes that the unknown response surface can be represented as (Santner et al. 2013):

$$f(\mathbf{x}) = \mathbf{f}(\mathbf{x})^T \boldsymbol{\beta} + M(\mathbf{x}), \quad (1)$$

where  $\mathbf{f}(\mathbf{x})$  is a vector of known trend functions (i.e., a prior trend model that can be defined as a smoothly varying deterministic function or a constant value),  $\boldsymbol{\beta}$  is a vector of unknown parameters of compatible dimension and  $M(\mathbf{x})$  is a realization of a mean zero and covariance-stationary Gaussian random field, that represents the uncertainty imposed on the problem (Kleijnen 2015).

What relates one observation to another is the *covariance function*, denoted  $k$ , also referred to as *kernel*. Multiple covariance functions exist in the field of GP, the choice depends on prior hypothesis about the unknown functions. The most commonly used kernels in kriging literature are the stationary squared exponential or Gaussian, and Matérn kernels. The hyperparameters of these kernels are usually estimated using Maximum Likelihood Estimation (MLE); we refer to Rasmussen and Williams (2006) for more details on covariance functions. As deterministic kriging is not suitable for stochastic settings, we are interested in methods capable of accommodating noisy evaluations in the optimization framework. We thus focus on *stochastic kriging* (Ankenman et al. 2010).

### 2.1 Stochastic Kriging

Ankenman et al. (2010) extended deterministic kriging by fully accounting for the sampling variability inherent in a stochastic simulation (i.e., the *intrinsic uncertainty* of the simulation outcome). They refer to  $M(\mathbf{x})$  in Eq. (1) as the *extrinsic uncertainty* of the unknown response surface; that is, the noise imposed to the problem to aid the construction of the surrogate. They show that correctly accounting for both, extrinsic and intrinsic uncertainty, has an impact on the experiment design, the response-surface estimation and inference. They refer to the resulting model as **stochastic kriging** for simulation metamodeling.

The *intrinsic* noise is (a) additive, because it is added to any noise that is extrinsic to the system; (b) heterogeneous, because its variance  $\tau^2(\mathbf{x})$  typically depends on  $\mathbf{x}$  (the opposite of homogeneous noise); and (c) independent of the extrinsic noise  $M$  and identically distributed across replications. The value of  $f(\mathbf{x})$  is estimated by averaging the value of  $r^i$  replications at a given input vector  $\mathbf{x}^i$  as  $\bar{f}(\mathbf{x}^i) = \sum_{k=1}^{r^i} \tilde{f}_k(\mathbf{x}^i) / r^i$  (Law 2015). Thus, at an arbitrary point  $\mathbf{x}^i$ , Eq. (2) depicts the observed goal value  $\tilde{f}_k(\mathbf{x}^i)$  in the  $k^{th}$  replication,

and Eq. (3) the **stochastic kriging** prediction  $\hat{f}(\mathbf{x}^i)$ :

$$\tilde{f}_k(\mathbf{x}^i) = \beta + M(\mathbf{x}^i) + \varepsilon_k(\mathbf{x}^i), \quad (2)$$

$$\hat{f}(\mathbf{x}^i) = \beta + \Sigma_M(\mathbf{x}^i, \cdot)^T [\Sigma_M + \Sigma_\varepsilon]^{-1} (\bar{\mathbf{f}} - \beta \mathbf{1}_p), \quad (3)$$

where  $\varepsilon_k(\mathbf{x}^i)$  represents the intrinsic noise inherent to the stochastic simulation model with mean zero and variance  $\tau^2(\mathbf{x}^i)$  at any arbitrary point  $\mathbf{x}^i$ . In (3),  $\bar{\mathbf{f}} = [\bar{f}(\mathbf{x}^1), \dots, \bar{f}(\mathbf{x}^p)]^T$  is a vector of estimated function values at  $p$  design points, each one with estimated variance  $\widehat{\text{Var}}[\bar{f}(\mathbf{x}^i)]$ . These estimates are sample means of the simulation output across  $r^1, \dots, r^p$  replications. Moreover,  $\Sigma_M(\mathbf{x}^i, \cdot) = \{k[M(\mathbf{x}^i), M(\mathbf{x}^1)], \dots, k[M(\mathbf{x}^i), M(\mathbf{x}^p)]\}^T$  is the spatial variance-covariance matrix of size  $p \times p$ ,  $\Sigma_\varepsilon = \text{diag}[\tau^2(\mathbf{x}^1)/r^1, \dots, \tau^2(\mathbf{x}^p)/r^p]$  is a diagonal variance-covariance matrix implied by the average simulation noise, and  $\mathbf{1}_p$  is a  $p \times 1$  vector of ones. Note that a constant term  $\beta$  is considered to represent the overall surface mean, instead of the trend term  $\mathbf{f}(\mathbf{x})^T \boldsymbol{\beta}$ , because this model has shown to be the most useful in practice (Ankenman et al. 2010). Moreover, the MSE of the stochastic kriging predictor (i.e., the stochastic kriging variance), denoted by  $\hat{s}^2(\mathbf{x}^i)$ , is estimated as (Chen and Kim 2014):

$$\hat{s}^2(\mathbf{x}^i) = \hat{\Sigma}_M(\mathbf{x}^i, \mathbf{x}^i) - \hat{\Sigma}_M(\mathbf{x}^i, \cdot)^T [\hat{\Sigma}_M + \hat{\Sigma}_\varepsilon]^{-1} \hat{\Sigma}_M(\mathbf{x}^i, \cdot) + \frac{\gamma^T \gamma}{\mathbf{1}_p^T [\hat{\Sigma}_M + \hat{\Sigma}_\varepsilon]^{-1} \mathbf{1}_p},$$

with  $\gamma = \mathbf{1}_p^T [\hat{\Sigma}_M + \hat{\Sigma}_\varepsilon]^{-1} \hat{\Sigma}_M(\mathbf{x}^i, \cdot)$ .

The values of  $\beta$ ,  $\Sigma_M$  and  $\Sigma_\varepsilon$  are unknown and have to be estimated (estimates are denoted  $\hat{\beta}$ ,  $\hat{\Sigma}_M$  and  $\hat{\Sigma}_\varepsilon$  respectively). We refer to Ankenman et al. (2010) for the detailed derivation of these maximum-likelihood estimators.

### 3 KRIGING-BASED STOCHASTIC MULTIOBJECTIVE OPTIMIZATION

In general, kriging-based optimization algorithms contain *sequential* steps, as depicted in Figure 2. The metamodel is sequentially updated with new sampled points (i.e., infill points). When using scalarization, the multiobjective problem is decomposed into a set of scalar single-objective subproblems, and these are optimized in a collaborative manner (Miettinen and Mäkelä 2002). The solutions to these subproblems yield a good approximation of the Pareto front, not a single optimal solution. At each iteration, scalarization methods might seek a single infill point (as done in ParEGO), or multiple points (as done in MOEA/D-EGO, Zhang et al. 2010). The selection process is done entirely using the metamodel of the combined objectives. Once a point is selected, it is evaluated using the expensive simulator to yield the (noisy) objective values for a candidate member of the *observed* Pareto front. Finally, with the newly added response (and respective noise), the kriging metamodel is updated and the procedure is repeated until a convergence criterion or quality indicator goal is met, or the exploration budget is depleted.

During this search phase, an *infill criterion* facilitates the choice of candidate solutions. The main goal of the infill criterion is to balance between the local refinement of the Pareto front approximation and the improvement of the global model quality (Wagner et al. 2010), by evaluating analytical expressions and facilitating the choice of candidate solutions (i.e., points located in encouraging areas).

The well-known EGO algorithm of Jones et al. (1998) for deterministic and single-objective black-box optimization problems, popularized the *Expected Improvement* (EI) criterion. The EI measures the expected value of improvement relative to the currently found optimum  $f_{min}$ , based on the predicted response and standard deviation at a certain point  $\mathbf{x}$ , denoted  $\hat{f}(\mathbf{x})$  and  $s$  respectively:  $E[I(\mathbf{x})] = (f_{min} - \hat{f}(\mathbf{x})) \Phi\left(\frac{f_{min} - \hat{f}(\mathbf{x})}{s}\right) + s \phi\left(\frac{f_{min} - \hat{f}(\mathbf{x})}{s}\right)$ , where  $\Phi$  denotes the normal cumulative distribution and  $\phi$  denotes the normal probability density function.

EI is, however, inappropriate for stochastic simulation settings since it ignores the noise in the observations, assuming they were sampled with infinite precision (Quan et al. 2013). Being aware of this

deficiency, important advances have been made to extend EGO for stochastic simulation. An important number of these approaches assume homogeneous simulation noise; Picheny et al. (2013) carry out a comprehensive review of kriging-based methods for the optimization of single-objective functions with this type of noise.

In practice, however, the noise is heterogeneous (Kim and Nelson 2006). More recently, a number of kriging-based single-objective optimization algorithms have been proposed that can handle heterogeneous noise, based on stochastic kriging models. A comprehensive study to evaluate the effectiveness and the relative performance of these algorithms is carried out in Jalali et al. (2017). Their results showed that the performance of the selected algorithms depends on the magnitude and the structure of the noise. They also emphasize that the implementation of a smart ranking and selection procedure could likely improve the performance of the studied algorithms. In the algorithm we propose, we solve a single objective problem at each iteration, thus we can implement a single-objective infill criterion developed for stochastic simulation, as explained in the following section.

### 3.1 Proposed Algorithm

We focus on solving the following multiobjective optimization problem:  $\min_{\mathbf{x} \in D} [f^1(\mathbf{x}), f^2(\mathbf{x}), \dots, f^m(\mathbf{x})]$ , for  $m$  objectives in the objective space  $\Theta$  and decision vectors  $\mathbf{x} = [x_1, \dots, x_d]^T$  in the decision space  $D$ . The objectives are analytically intractable, so we rely on (expensive) simulation estimates. The solution to this problem in the decision space yields the Pareto set. Figure 2 shows the sequential steps performed by the algorithm.

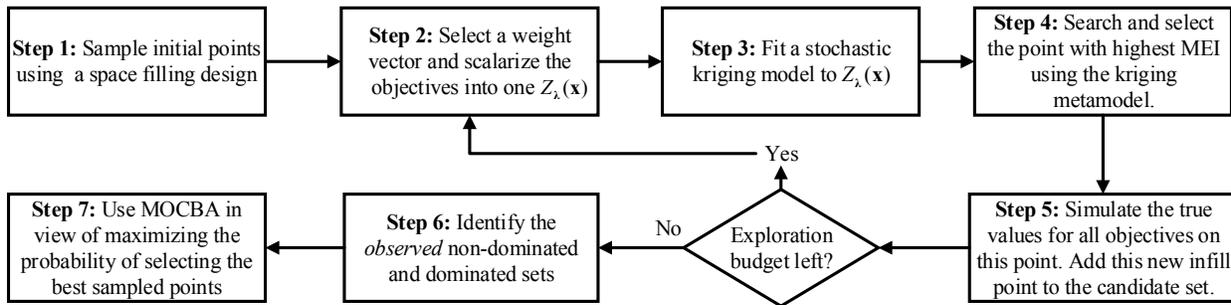


Figure 2: Sequence of steps in the proposed algorithm.

As is common in kriging-based optimization (Kleijnen 2015), in the **first step** we use a maximin latin hypercube sample (LHS), a space filling design to sample the initial points  $\mathbf{x}^i$ , with  $i = 1, \dots, n_0$ . As suggested in Jones et al. (1998), the number of initial design points is set to  $11d - 1$ , where  $d$  is the number of dimensions of  $\mathbf{x}$ . We compute the expensive simulation values for each objective on these designs, using a fixed number of replications  $B$  per point. We then distribute the computational budget between the *exploratory* and *accuracy* stages. The former refers to the exploration of the design space (i.e., the total number of infill points to sample, and the replications performed on each of them to obtain an initial estimate of the response). The latter refers to the budget assigned to the R&S procedure in order to achieve sufficient accuracy.

Very little research discusses the choice of  $B$ ; the same is true for the distribution of the budget between the exploratory and accuracy phases. In Quan et al. (2013) an allocation heuristic is proposed, but in the experiments of Jalali et al. (2017) it is proven not to be beneficial. We plan to improve on the efficient allocation of replications in further research; in this article, the choices made are arbitrary. For the initial design, we consider a fixed value of  $B = 10$ ; the total replication budget is determined by the number of infill points we wish to sample, and the choice of  $B$ . This total budget is then equally split between the exploration phase and accuracy phase. For instance, if the total number of infill points to sample is 300 points, as we take  $B = 10$ , we have a total of 3000 replications to distribute between both stages. We first

spend 1500 replications in the exploratory stage, and 1500 replications left to distribute among promising candidates according to MOCBA, in view of maximizing the probability of selecting the “true” members of the Pareto front.

In **step 2** the objectives are first normalized with respect to their known (or estimated) ranges so that each objective function lies between  $[0, 1]$ . Then the objectives are combined into one using a scalarization function. There are numerous scalarization functions/methods in the literature, the choice depends mainly on the geometrical properties of the problem (see Miettinen and Mäkelä 2002 for further details on scalarization). In this work we implement the *augmented Tchebycheff* scalarization function (4), as done in the original ParEGO algorithm (Knowles 2006):

$$Z_{\lambda}(\mathbf{x}) = \max_{j=1,\dots,m} (\lambda^j \bar{f}^j(\mathbf{x})) + \rho \sum_{j=1}^m \lambda^j \bar{f}^j(\mathbf{x}), \quad (4)$$

where  $\bar{f}^j(\mathbf{x}^i)$  are the simulation estimates of the objectives’ values,  $0 \leq \lambda^j \leq 1$ ,  $\sum_{j=1}^m \lambda^j = 1$ ,  $\forall j \in \{1, \dots, m\}$ , and  $\rho$  is a small positive value (e.g.,  $\rho = 0.05$ ). Each set of  $\lambda$  values is defined as a weight vector  $\boldsymbol{\lambda} = \lambda^1, \dots, \lambda^m$ . By varying the set of  $\lambda$  values, we assign each objective a different weight at each iteration, such that we obtain points that fall between the objectives’ extremes to construct the Pareto front (Knowles 2006). The observed scalarized single-objective subproblem  $Z_{\lambda}(\mathbf{x})$  is thus noisy. Moreover, it is common to assume that the responses are independent (Wagner et al. 2010). Though it is possible to account for correlation between the multiple objectives, for instance by using co-kriging models (Kleijnen and Mehdad 2014), recent research shows that dependent models are more complex and don’t significantly outperform independent models (Fricker et al. 2013).

In **step 3**, we use stochastic kriging to approximate  $Z_{\lambda}(\mathbf{x})$  based on the simulation estimates at all sampled solutions so far. In the **fourth step** the algorithm uses the *Modified Expected Improvement* (MEI) of Quan et al. (2013) as an infill criterion to *search* and *select* new infill points iteratively, given its strong performance in the simulation optimization experiments of Jalali et al. (2017).

In an arbitrary iteration  $p + 1$ , the algorithm looks at all unvisited alternatives and selects the solution that maximizes  $\text{MEI}(\mathbf{x}) = E\{\max[\hat{Z}(\mathbf{x}_{\min}) - Z_N^*(\mathbf{x}), 0]\}$ , where  $\hat{Z}(\mathbf{x}_{\min})$  is the stochastic kriging prediction at  $\mathbf{x}_{\min} = \arg \min_{\mathbf{x} \in D^p} \bar{Z}(\mathbf{x})$  (i.e. the alternative with the lowest sample mean among the already simulated points); and  $Z_N^*(\mathbf{x})$  is a normal random variable:  $Z_N^*(\mathbf{x}) \sim N[\hat{Z}(\mathbf{x}), s_D(\mathbf{x})]$ , where the mean  $\hat{Z}(\mathbf{x})$  is the stochastic kriging prediction at solution  $\mathbf{x}$ , and  $s_D^2(\mathbf{x})$  the estimate of the deterministic kriging prediction error. In **step 5** the algorithm selects the point that maximizes MEI and the solution is updated by performing  $B/2$  (expensive) simulation replications at the new sampled point per objective, yielding  $\bar{f}_j(\mathbf{x}_{\text{new}}), \forall j = 1, \dots, m$ . Then the algorithm updates the metamodel re-computing the kriging parameters with the new design point included.

The algorithm then checks the exploration budget. If budget allows, another iteration is performed: we go back to step 2, randomly choose another weight vector  $\boldsymbol{\lambda}$  and repeat steps 3 – 5. Depending on the number of iterations, the same objective  $Z_{\lambda}(\mathbf{x})$  (i.e., a weight vector is repeated) can be minimized several times during the run of the algorithm. When the exploration budget is depleted, the algorithm proceeds to *identify* the *observed* non-dominated set to the optimization problem by tracing back all the sampled points and comparing them. Note that the observed objective values contain noise; the correct selection of the truly non-dominated points is, thus, not trivial. This problem is referred to as *multiobjective ranking and selection* (MORS), yet very little research has been done in this regard. The setting depicted in Figure 3 exemplifies the MORS problem, for 5 designs and 3 objectives; the variances are represented by the vertical bars (95% confidence intervals).

Relying only on the *observed* objectives’ values, we may incur two possible errors due to sampling variability. A *Type I* error occurs when designs that actually belong to the non-dominated set, are considered dominated. A *Type II* error occurs when designs that are actually dominated, fall in the observed Pareto front. To ensure a high probability of *correctly* selecting a non-dominated design, we should smartly

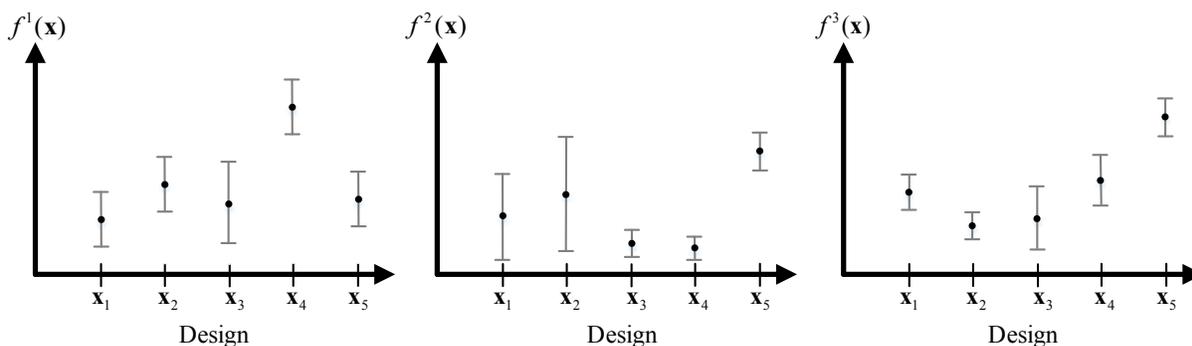


Figure 3: Based on the observed responses and variances, it is non-trivial to differentiate between superior and inferior designs. Are designs 1, 2 and 3 critical? Should we discard design 5? How to optimally allocate the accuracy budget such that we maximize the probability of selecting the designs with the best expected performance? Clearly, the more designs we sample and the more objectives we have, the more complex the problem becomes. Some designs may perform very well on some objectives, and badly in others (e.g., design 4).

allocate the *accuracy* budget to those critically competitive designs, based on their observed performance and variance. Analogously, we should not spend budget on those designs that are clearly dominated.

Ranking and selection procedures have been extensively studied for single-objective settings (Boesel et al. 2003; Chen et al. 2000). Multiple *screening*, *indifference zone* and *optimal allocation* procedures have been developed to ensure fast convergence to the true optimum by efficiently allocating extra replications among competitive points. Among the few works on MORS, MOCBA (Lee et al. 2010) is a Bayesian allocation framework that aims to minimize both type I and II errors. It is built upon its famous single-objective analogous OCBA (Optimal Computing Budget Allocation) of Chen et al. (2000).

In single-objective sequential simulation optimization algorithms, it is common to employ the R&S procedure embedded in the optimization loop (i.e., immediately after step 5 in Figure 2, before checking the budget left). Yet, we suspect that this may waste budget in the early iterations of the algorithm, when the emphasis should be on fine-tuning the kriging model, rather than identifying the final best points. In the proposed algorithm, we first sample all the candidate points, compute the observed Pareto front, and then we run MOCBA to allocate extra replications to competitive designs. Using MOCBA, designs with low objective values and high variance will be prioritized, whereas designs with large objective values and low variance will be discarded. We refer to Chen et al. (2000) and Lee et al. (2010) for the detailed derivation of the MOCBA allocation rules.

#### 4 PERFORMANCE OF THE PROPOSED ALGORITHM

To assess the performance of the algorithm, we run it on two benchmark functions with different geometries obtained from the literature (Huband et al. 2006): DTLZ2 (concave front) and WFG3 (linear front). We evaluate 4 different scenarios: DTLZ2a (2 objectives and 5 dimensions); DTLZ2b (4 objectives and 4 dimensions); WFG3a (2 objectives and 3 dimensions); and WFG3b (3 objectives and 4 dimensions). We refer to the analytical expressions and detailed characteristics of these functions in Huband et al. (2006). For all the experiments we fix a budget of 300 infill points in addition to the LHS design, and we set  $B = 10$ .

These test functions are continuous. In order to avoid the difficult problem of maximizing MEI in a continuous domain, we discretize the design space using large but finite sets of solutions. For this purpose we use low-discrepancy sequences such as Faure and Sobol (i.e., quasi-Monte Carlo sampling methods with desirable space-filling properties). The number of points depends on the dimensionality  $d$  of the problem;

we take  $d \times 1500$  points (see Lemieux 2009 for further details on quasi-Monte Carlo sampling). This set of points is the design space  $D$  and our objective is to find the Pareto set among these alternatives using the proposed algorithm.

All objectives are perturbed with heterogeneous Gaussian noise. Hence, we obtain noisy observations  $\tilde{f}_k^j(\mathbf{x}^i) = f^j(\mathbf{x}^i) + \varepsilon_k(\mathbf{x}^i)$ , with  $\varepsilon_k(\mathbf{x}^i) \sim \mathcal{N}(0, \tau(\mathbf{x}^i))$  for  $j = 1, \dots, m$  objectives at the  $k^{\text{th}}$  replication. In practice, the noise can follow any type of structure. In our experiments, we assume that the standard deviation of the noise (i.e.,  $\tau(\mathbf{x})$ ) varies linearly with respect to the objective values, analogous to the single-objective experiments of Jalali et al. (2017):  $\tau(\mathbf{x}) = af^j(\mathbf{x}) + b, \forall j \in \{1, \dots, m\}$ .

The maximum and minimum values of  $\tau(\mathbf{x})$  are linked to the range of each objective value in the region of interest (i.e.,  $\max_{\mathbf{x} \in D} f^j(\mathbf{x}) - \min_{\mathbf{x} \in D} f^j(\mathbf{x}), \forall j \in \{1, \dots, m\}$ ), denoted  $R_f^j$  (Huang et al. 2006; Picheny and Ginsbourger 2014). For the initial LHS design, the noise varies between  $0.2R_f^j$  and  $0.8R_f^j$ , as in Jalali et al. (2017). For the exploratory phase, we consider a heavy level of noise perturbing the responses:  $\tau(\mathbf{x})$  varies between 80% and 200% of  $R_f^j$  (i.e.,  $\min_{\mathbf{x} \in D} \tau(\mathbf{x}) = 0.8R_f^j$  and  $\max_{\mathbf{x} \in D} \tau(\mathbf{x}) = 2R_f^j, \forall j \in \{1, \dots, m\}$ ). This way we can easily compute the  $a$  and  $b$  constants for the linear function. We consider the case where the noise decreases linearly with the objective values; we have minimum noise at the global minimum of each individual objective (i.e.,  $\min_{\mathbf{x} \in D} \tau(\mathbf{x})$  at  $\min_{\mathbf{x} \in D} f^j(\mathbf{x})$  and  $\max_{\mathbf{x} \in D} \tau(\mathbf{x})$  at  $\max_{\mathbf{x} \in D} f^j(\mathbf{x})$ ).

#### 4.1 Experiments

Measuring the quality of a given Pareto front approximation is a non-trivial task, as in practice, the “true” Pareto front is unknown. Two widely used quality indicators are the *hypervolume* and *inverted generational distance* (IGD). The hypervolume is the portion (volume) of the objective space covered by a particular Pareto front; whereas the IGD measures the Euclidean distance from a member of the approximated Pareto front and its closest member of a reference front (e.g., the true Pareto front). Thus, the former is to be maximized, and the latter minimized. These and other indicators, however, have been developed for deterministic settings (Zitzler et al. 2008). In addition, they only look into the performance of the algorithm in the objective space, rather than in the design space.

In stochastic settings, we are also concerned about how well the observed Pareto *set* matches the true Pareto set. By only measuring the quality of the observed Pareto *front* we may incur errors type I or II. Some recent research suggests to combine different existing deterministic indicators for noisy multiobjective optimization (see e.g., Fieldsend and Everson 2015), as well as using the deterministic responses to infer on the quality of the stochastic Pareto front approximation. In this work we first compare the original (deterministic) ParEGO with the proposed algorithm by means of the hypervolume and IGD indicators. For ease of visualization, Figures 4 and 5 depict the results for the biobjective DTLZ2a and WFG3a functions only, but the results for the 3- and 4-objective functions are comparable. From left to right, the first plot shows the observed Pareto front, the middle one the hypervolume and the third one the IGD.

Based on the hypervolume and IGD indicator values, it is clear that the proposed algorithm significantly outperforms ParEGO in terms of the quality of the Pareto front (i.e., the portion of the objective space covered by the front). This is also visually clear in terms of shape, richness and uniformity of the front. The hypervolume is roughly 25% and 40% better for the proposed algorithm; whereas the IGD is 3.4 and 5.5 times better for the WFG3a and DTLZ2a scenarios respectively. However, based only on the hypervolume and IGD indicator values, we cannot determine which of the members of the observed Pareto fronts actually belong to the true Pareto front. Some of the observed members might be wrongly considered non-dominated, while some of the discarded candidates might actually belong to the true non-dominated set.

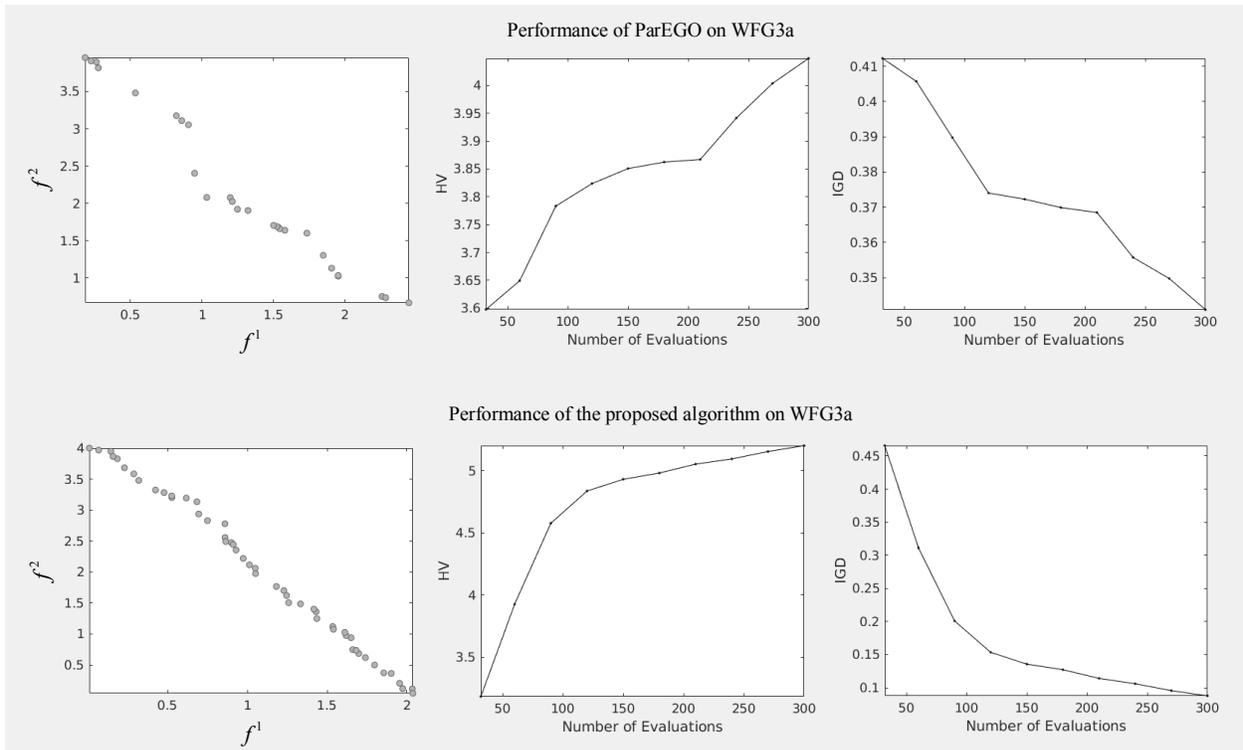


Figure 4: Performance of ParEGO and the proposed algorithm on WFG3a on a budget of 300 evaluations.

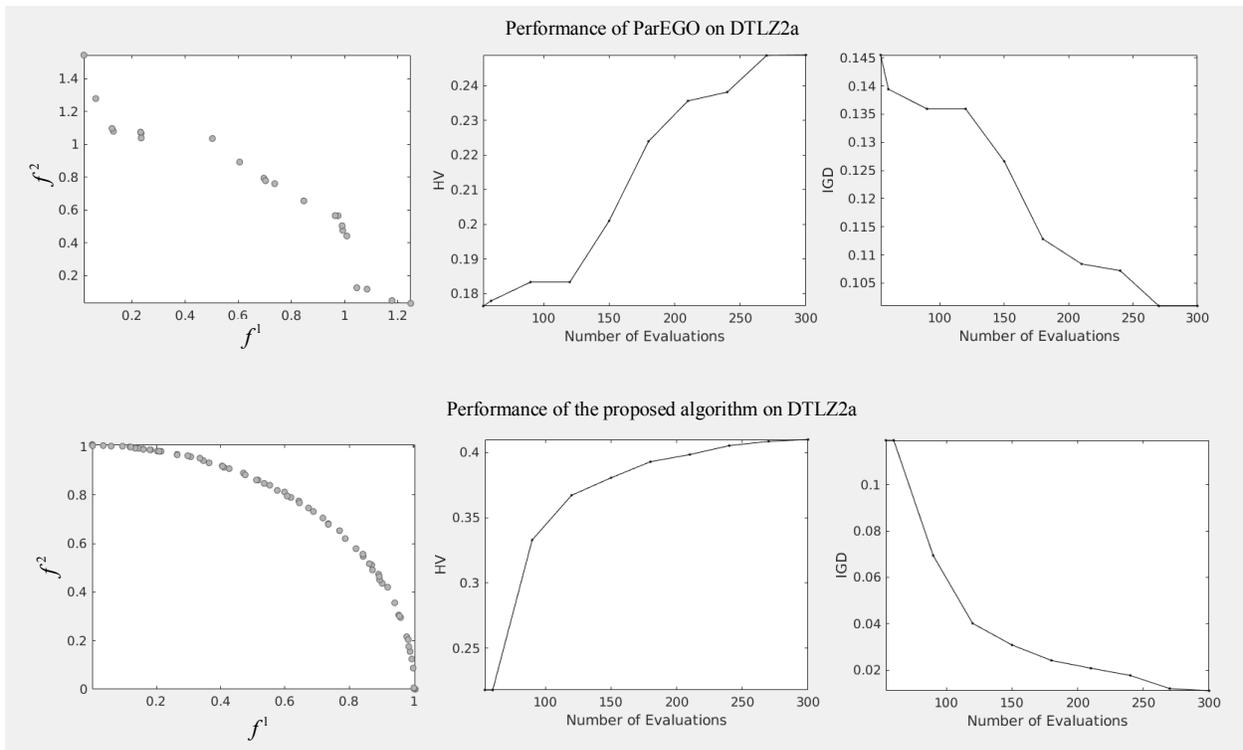


Figure 5: Performance of ParEGO and the proposed algorithm on DTLZ2a on a budget of 300 evaluations.

As we know the true objective values for the entire finite set of points, we can determine the members of the true Pareto set, and compute the percentage of the true set that has been found by the algorithms, as shown in Table 1. It is clear that the proposed algorithm is able to find better solutions than its deterministic counterpart. By employing MOCBA, we minimize both error types, but at the cost of a high replication budget. On the other hand, given the high levels of noise in the objectives, it is not surprising that ParEGO struggles in the search of non-dominated designs. Moreover, it wastes plenty of replications on non-critical designs, as it distributes the entire budget evenly among all 300 infill points.

Table 1: Percentage of the true Pareto set found by ParEGO and the proposed algorithm (average of 20 runs).

Approach	DTLZ2a	DTLZ2b	WFG3a	WFG3b
ParEGO	65%	58%	66%	56%
Proposed algorithm	91%	79%	85%	76%

## 5 CONCLUDING REMARKS

In this work we propose a preliminary stochastic-kriging-based algorithm for multiobjective simulation optimization. It builds on the ParEGO algorithm of Knowles (2006) and implements the MOCBA procedure of Lee et al. (2010) to identify the Pareto-optimal points at the end of the search loop, by discarding obviously inferior sampled points and allocating (extra) replications only to critical designs. Additionally, in contrast to the common assumption in the literature of homogeneous noise, we consider heterogeneous noise. Since stochastic kriging can tolerate more noise than deterministic kriging, it allows for less effort per design point and thus more design points can be used in the analysis. However, for large sets of design points, stochastic kriging can be slower than deterministic kriging (Staum 2009). Experimental results show that the proposed algorithm significantly outperforms ParEGO in terms of quality of the observed Pareto front and the fraction of the true Pareto set found.

There are multiple opportunities for further work. One of the main challenges for multiobjective simulation optimization algorithms is the ranking and selection problem. Apart from MOCBA, other recent efforts include Feldman et al. (2015), Branke et al. (2016) and Bonnel and Collonge (2015). Given that MOCBA requires a large replication budget to converge, running the proposed algorithm with different ranking and selection procedures would allow for inference on the efficiency of these approaches. Furthermore, *indifference zone* procedures have proven to be useful in single-objective ranking and selection. Integrating the indifference zone procedure proposed by Teng et al. (2010) into our proposed algorithm may also enhance its performance. Finally, a major problem that deserves more attention is how to distribute the computational budget between the exploratory and accuracy phases. Further experiments with different allocation rules will likely be valuable in this regard.

## REFERENCES

- Ankenman, B., B. L. Nelson, and J. Staum. 2010. “Stochastic Kriging for Simulation Metamodeling”. *Operations Research* 58(2):371–382.
- Boesel, J., B. L. Nelson, and S.-H. Kim. 2003. “Using Ranking and Selection to “Clean Up” After Simulation Optimization”. *Operations Research* 51(5):814–825.
- Bonnel, H., and J. Collonge. 2015. “Optimization Over the Pareto Outcome Set Associated with a Convex Bi-Objective Optimization Problem: Theoretical Results, Deterministic Algorithm and Application to the Stochastic Case”. *Journal of Global Optimization* 62(3):481–505.
- Branke, J., W. Zhang, and Y. Tao. 2016. “Multi-Objective Ranking and Selection Based on Hypervolume”. In *Proceedings of the 2016 Winter Simulation Conference*, edited by T. M. K. Roeder et al., 859–870. Piscataway, New Jersey: IEEE.

- Chen, C.-H., J. Lin, E. Yücesan, and S. E. Chick. 2000. "Simulation Budget Allocation for Further Enhancing the Efficiency of Ordinal Optimization". *Discrete Event Dynamic Systems* 10(3):251–270.
- Chen, X., and K.-K. Kim. 2014. "Stochastic Kriging with Biased Sample Estimates". *ACM Transactions on Modeling and Computer Simulation* 24(2):8.
- Feldman, G., S. R. Hunter, and R. Pasupathy. 2015. "Multi-Objective Simulation Optimization on Finite Sets: Optimal Allocation via Scalarization". In *Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz et al., 3610–3621. Piscataway, New Jersey: IEEE.
- Fieldsend, J. E., and R. M. Everson. 2015. "The Rolling Tide Evolutionary Algorithm: A Multiobjective Optimizer for Noisy Optimization Problems". *IEEE Transactions on Evolutionary Computation* 19(1):103–117.
- Fricker, T. E., J. E. Oakley, and N. M. Urban. 2013. "Multivariate Gaussian Process Emulators with Nonseparable Covariance Structures". *Technometrics* 55(1):47–56.
- Horn, D., M. Dagge, X. Sun, and B. Bischl. 2017. "First Investigations on Noisy Model-Based Multi-Objective Optimization". In *International Conference on Evolutionary Multi-Criterion Optimization*, edited by H. T. et al., 298–313. Cham: Springer.
- Huang, D., T. T. Allen, W. I. Notz, and N. Zeng. 2006. "Global Optimization of Stochastic Black-Box Systems via Sequential Kriging Meta-Models". *Journal of Global Optimization* 34(3):441–466.
- Huband, S., P. Hingston, L. Barone, and L. While. 2006. "A Review of Multiobjective Test Problems and a Scalable Test Problem Toolkit". *IEEE Transactions on Evolutionary Computation* 10(5):477–506.
- Jalali, H., I. Van Nieuwenhuysse, and V. Picheny. 2017. "Comparison of Kriging-based Algorithms for Simulation Optimization with Heterogeneous Noise". *European Journal of Operational Research* 261(1):279–301.
- Jones, D. R., M. Schonlau, and W. J. Welch. 1998. "Efficient Global Optimization of Expensive Black-box Functions". *Journal of Global Optimization* 13(4):455–492.
- Kim, S.-H., and B. L. Nelson. 2006. "Selecting the Best System". In *Handbooks in Operations Research and Management Science: Simulation*, edited by S. G. Henderson and B. L. Nelson, 501–534. Amsterdam: Elsevier.
- Kleijnen, J. P., and E. Mehdad. 2014. "Multivariate versus Univariate Kriging Metamodels for Multi-Response Simulation Models". *European Journal of Operational Research* 236(2):573–582.
- Kleijnen, J. P. C. 2015. *Design and Analysis of Simulation Experiments*. New York: Springer.
- Knowles, J. 2006. "ParEGO: a Hybrid Algorithm with On-line Landscape Approximation for Expensive Multiobjective Optimization Problems". *IEEE Transactions on Evolutionary Computation* 10(1):50–66.
- Knowles, J. D., D. Corne, and A. P. Reynolds. 2009. "Noisy Multiobjective Optimization on a Budget of 250 Evaluations.". In *International Conference on Evolutionary Multi-Criterion Optimization*, edited by M. E. et al., 36–50. Cham: Springer.
- Koch, P., T. Wagner, M. T. Emmerich, T. Bäck, and W. Konen. 2015. "Efficient Multi-Criteria Optimization on Noisy Machine Learning Problems". *Applied Soft Computing* 29(4):357–370.
- Law, A. M. 2015. *Simulation Modeling and Analysis*. New York: McGraw-Hill.
- Lee, L. H., E. P. Chew, S. Teng, and D. Goldsman. 2010. "Finding the Non-Dominated Pareto Set for Multi-Objective Simulation Models". *IIE Transactions* 42(9):656–674.
- Lemieux, C. 2009. *Monte Carlo and Quasi-Monte Carlo Sampling*. New York: Springer.
- Miettinen, K., and M. M. Mäkelä. 2002. "On Scalarizing Functions in Multiobjective Optimization". *OR Spectrum* 24(2):193–213.
- Picheny, V., and D. Ginsbourger. 2014. "Noisy Kriging-Based Optimization Methods: A Unified Implementation within the DiceOptim Package". *Computational Statistics & Data Analysis* 71(3):1035–1053.
- Picheny, V., T. Wagner, and D. Ginsbourger. 2013. "A Benchmark of Kriging-Based Infill Criteria for Noisy Optimization". *Structural and Multidisciplinary Optimization* 48(3):607–626.
- Quan, N., J. Yin, S. H. Ng, and L. H. Lee. 2013. "Simulation Optimization via Kriging: A Sequential Search Using Expected Improvement with Computing Budget Constraints". *IIE Transactions* 45(7):763–780.

- Rasmussen, C. E., and C. K. I. Williams. 2006. *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press.
- Sacks, J., W. J. Welch, T. J. Mitchell, and H. P. Wynn. 1989. “Design and Analysis of Computer Experiments”. *Statistical Science* 4(4):409–423.
- Santner, T. J., B. J. Williams, and W. I. Notz. 2013. *The Design and Analysis of Computer Experiments*. New York: Springer.
- Staum, J. 2009. “Better Simulation Metamodeling: The Why, What, and How of Stochastic Kriging”. In *Proceedings of the 2009 Winter Simulation Conference*, edited by M. D. Rossetti et al., 119–133. Piscataway, New Jersey: IEEE.
- Teng, S., L. H. Lee, and E. P. Chew. 2010. “Integration of Indifference-zone with Multi-Objective Computing Budget Allocation”. *European Journal of Operational Research* 203(2):419–429.
- Van Beers, W. C., and J. P. C. Kleijnen. 2003. “Kriging for Interpolation in Random Simulation”. *Journal of the Operational Research Society* 54(3):255–262.
- Wagner, T., M. Emmerich, A. Deutz, and W. Ponweiser. 2010. “On Expected-Improvement Criteria for Model-Based Multi-Objective Optimization”. In *International Conference on Parallel Problem Solving from Nature*, edited by R. S. et al., 718–727. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Zhang, Q., W. Liu, E. Tsang, and B. Virginas. 2010. “Expensive Multiobjective Optimization by MOEA/D with Gaussian Process Model”. *IEEE Transactions on Evolutionary Computation* 14(3):456–474.
- Zitzler, E., J. Knowles, and L. Thiele. 2008. “Quality Assessment of Pareto Set Approximations”. In *Multiobjective Optimization: Interactive and Evolutionary Approaches*, edited by J. Branke, K. Deb, K. Miettinen, and R. Słowiński, 373–404. Berlin, Heidelberg: Springer Berlin Heidelberg.

## AUTHOR BIOGRAPHIES

**SEBASTIAN ROJAS-GONZALEZ** is a Ph.D. student in the Department of Decision Sciences and Information Management at KU Leuven, Belgium. His doctoral work is centered on methods for multi-criteria decision making in settings with uncertainty. His website is <https://www.kuleuven.be/person/00111953> and his email address is [sebastian.rojasgonzalez@kuleuven.be](mailto:sebastian.rojasgonzalez@kuleuven.be).

**HAMED JALALI** is an Assistant Professor of Supply Chain Management at NEOMA Business School in Rouen, France. He holds a Master’s degree in Statistics and a PhD degree in Business Economics (Operations Management) from KU Leuven, Belgium. His research include simulation optimization (especially using Kriging-based algorithms) and joint assortment and pricing optimization. His research appears in IIE Transactions and European Journal of Operational Research. His website address is <http://www.neoma-bs.com/en/jalali-hamed> and his email address is [Hamed.JALALI@neoma-bs.fr](mailto:Hamed.JALALI@neoma-bs.fr).

**INNEKE VAN NIEUWENHUYSE** is a Full Professor at the Research Group Logistics, Hasselt University (Belgium), and guest professor at KU Leuven (Belgium). She holds M.Sc. and PhD degrees in Applied Economics-Business Engineering from the University of Antwerp (Belgium). She teaches courses in Supply Chain Management, Facilities Design and Operations Strategy. Her research focuses on the design and analysis of stochastic manufacturing and supply chain systems. Her email address is [inneke.vannieuwenhuyse@uhasselt.be](mailto:inneke.vannieuwenhuyse@uhasselt.be).