

# Bayesian multi-objective optimization of process design parameters in constrained settings with noise: an engineering design application

Alejandro Morales-Hernández<sup>1,2\*</sup>, Sebastian Rojas Gonzalez<sup>1,2,3</sup>, Inneke Van Nieuwenhuyse<sup>1,2</sup>, Ivo Couckuyt<sup>3</sup>, Jeroen Jordens<sup>4</sup>, Maarten Witters<sup>4</sup> and Bart Van Doninck<sup>4</sup>

<sup>1</sup>\*Flanders Make@UHasselt and Data Science Institute, Hasselt University, Agoralaan, Diepenbeek, 3590, Belgium.

<sup>2</sup>Computational Mathematics, Faculty of Sciences, Hasselt University, Agoralaan, Diepenbeek, 3590, Belgium.

<sup>3</sup>Surrogate Modeling Lab, Ghent University-IMEC, Technologiepark-Zwijnaarde, Ghent, 9052, Belgium.

<sup>4</sup>CoDesignS, Flanders Make, Oude Diestersebaan, Lommel, 3920, Belgium.

\*Corresponding author(s). E-mail(s):

[alejandromoraleshernandez@uhasselt.be](mailto:alejandromoraleshernandez@uhasselt.be);

Contributing authors: [sebastian.rojasgonzalez@ugent.be](mailto:sebastian.rojasgonzalez@ugent.be);  
[inneke.vannieuwenhuyse@uhasselt.be](mailto:inneke.vannieuwenhuyse@uhasselt.be); [ivo.couckuyt@ugent.be](mailto:ivo.couckuyt@ugent.be);

[jeroen.jordens@flandersmake.be](mailto:jeroen.jordens@flandersmake.be);

[maarten.witters@flandersmake.be](mailto:maarten.witters@flandersmake.be);

[bart.vandoninck@flandersmake.be](mailto:bart.vandoninck@flandersmake.be);

## Abstract

The use of adhesive joints in various industrial applications has become increasingly popular due to their beneficial characteristics, including their high strength-to-weight ratio, design flexibility, limited stress concentrations, planar force transfer, good damage tolerance, and fatigue resistance. However, finding the best process parameters for adhesive

bonding can be challenging. This optimization problem is inherently multi-objective, aiming to maximize break strength while minimizing cost and constrained to avoid any visual damage to the materials and ensure that stress tests do not result in adhesion-related failures. Additionally, testing the same process parameters several times may yield different break strengths, making the optimization process uncertain. Conducting physical experiments in a laboratory setting is costly, and traditional evolutionary approaches like genetic algorithms are not suitable due to the large number of experiments required for evaluation. Bayesian optimization is suitable in this context, but few methods simultaneously consider the optimization of multiple noisy objectives and constraints. This study successfully applies advanced learning techniques to emulate the objective and constraint functions based on *limited* experimental data. These are incorporated into a Bayesian optimization framework, which efficiently detects Pareto-optimal process configurations under strict budget constraints.

**Keywords:** Bayesian optimization, multi-objective optimization, constrained optimization, machine learning, adhesive bonding

## 1 Introduction

Adhesive bonding is the engineering process of joining two surfaces together by a non-metallic substance [1]. This process occurs frequently in many engineering design contexts, such as the automotive industry [2], electronics [3], and aeronautics [4]. It is a complex process, in which several physical and chemical processes occur simultaneously [5, 6], with outcomes that are influenced by many factors, such as environmental conditions, material specifications, and specific process settings. Process optimization is therefore traditionally performed by experts, based on acquired knowledge and extensive experimental campaigns [7, 8]. Yet, even for experts, optimizing adhesive bonding parameters is a difficult task, due to the complex (and, essentially, black box) relationships between the process parameters, and the case-specific nature of the bonding process. As a consequence, real physical experiments are required to detect the optimal settings for each specific case. These tend to be costly in terms of time (time to prepare the parts to be joined, time to prepare the adhesive, to join both materials, and to allow the adhesive to gain its final strength), and in terms of manual labor. Moreover, data from one industrial bonding process cannot be used to optimize another process, as not only materials and adhesives may differ, but also production process specifications. As a result, optimization of an adhesive bonding process is challenging and costly, even with expert knowledge available. Moreover, the experimental approach may easily yield suboptimal results with respect to other relevant performance metrics, such as production costs. Thus, the problem is also inherently *multi-objective*.

As observed in [9], evolutionary approaches (in particular, genetic algorithms) are currently the most common data science technique in engineering design optimization. This also holds for multi-objective engineering design problems (a.o., bonding problems). Evolutionary multi-objective algorithms (EMOAs) are applicable to black-box optimization problems and have proven to be effective derivative-free optimizers [10]. The *Non-dominated Sorting Genetic Algorithm (NSGA-II)* [11], stands out as one of the most widely used algorithms in industrial process cases. For instance, Corbett *et. al.* (2017) [12] used NSGA-II in combination with a Gaussian Process surrogate model to simultaneously optimize two objectives (force sustained by the joint and area beneath the load-displacement curve) for a novel concept for joining materials. Labbé and Drouet (2012) [13] studied designs for optimal bonding of TSL (tubular single-lap) joints using NSGA-II, highlighting the conflicting nature of mass-related and stress-related objectives.

Relative to the unconstrained case, the literature on algorithms for solving *constrained* multi-objective problems is scarce. Some of the well-known algorithms are the Constrained Non-dominated Sorting Genetic Algorithm II (C-NSGA-II [11, 14]), the Constrained Multi-Objective Evolutionary Algorithm based on Decomposition (C-MOEA/D [15]), the Two-Archive Evolutionary Algorithm for Constrained multi-objective optimization (C-TAEA [16]), and the Constrained Particle Swarm Optimization algorithm for Multi-Objective problems (C-MOPSO [17]). These algorithms differ in the way in which they handle constraints: while the first three algorithms handle the constraints during the non-dominated sort, the latter adopts a penalization function approach. We refer the reader to the original publications for further details about the implementation of these algorithms.

EMOAs are indeed an effective tool for black-box optimization, especially if the optimization problem is difficult to model; however, they are *data-inefficient* (i.e., they require numerous function evaluations), which makes them ill-suited for optimizing design problems that require (computational or experimental) *expensive* data. Even when an emulator or surrogate model is used to mitigate this issue [18], the search process in this type of algorithm remains largely based on random alterations of (promising) solutions, with convergence speeds that are sensitive to the choice of user-defined (and often problem-specific) parameters. Moreover, the measurement of the objectives is usually affected by noise and this is often neglected during the optimization, leading to suboptimal solutions.

For the sub-class of low-dimensional black-box problems that are expensive to evaluate, Bayesian optimization (BO) has emerged as a powerful alternative, with applications ranging from hyperparameter tuning of deep learning models, to design optimization in engineering, and stochastic optimization in operational research (see [19] for a comprehensive review). Several Bayesian Multiobjective Optimization (BMO) methods have been proposed to solve complex decision-making problems involving multiple objectives that are expensive to evaluate [20]. Only a few methods have considered the constrained

case [21–23], and even less literature has considered noisy objectives and/or constraints [24, 25]. The main contribution of this paper is the implementation of a full BMO algorithm for the noisy, expensive, and constrained optimization of a novel adhesive bonding process. The approach builds on the recent work of Rojas *et. al.* (2019) [25]. More specifically, the main contributions include:

- The formulation of an acquisition function that combines the expected improvement over the objectives and the constraint feasibility (based on discrete values (FEASIBLE or NOT FEASIBLE) rather in numeric values), where both the objective and constraints are black-box, noisy, and expensive to evaluate. This provides an explainable approach, as opposed to the upfront parameter choices that guide the generation of successive populations in evolutionary algorithms. The contribution in this sense is given by the fact that EAs are still traditionally used to solve such complex engineering problems and do not consider the knowledge gathered from past evaluations beyond the heuristics defined by the algorithm.
- The use of a Gaussian Process (GP) surrogate that explicitly accounts for the noise that is typically present in the outcomes of the experiments. Here, outcomes cannot be observed with perfect accuracy, and the magnitude of the noise may vary between different configurations; thus, we have *input-dependent* (i.e., heterogeneous) noise. This issue is commonly overlooked in engineering optimization problems, by simply assuming that noise is non-existent (or, at best, homogeneous). Hence the contribution of the proposed algorithm.
- A full BMO algorithm that is able to obtain *better* configurations than state-of-the-art EMOAs and surrogate-assisted EMOAs. The approach is particularly meant for settings where the analyst can only afford a very limited number of observations (as is the case with costly physical experiments in a lab).

The remainder of this article is organized as follows. Section 2 discusses the main concepts in multi-objective optimization, followed by a description of the bonding process problem under study. Section 3 details the relevant background in Bayesian optimization, and the proposed algorithms. Section 4 discusses the design of experiments, while Section 5 discusses the results. Finally, Section 6 summarizes the findings and highlights some future research directions.

## 2 Multi-objective optimization: main concepts and problem description

### 2.1 Multi-objective optimization: main concepts

In general, a multi-objective optimization problem can be defined as (assuming minimization of all the objectives):

$$\begin{aligned}
 \min \quad & [f_1(\mathbf{x}), \dots, f_m(\mathbf{x})] \\
 \text{s.t.} \quad & g_i(\mathbf{x}) \leq 0, i = 1, 2, \dots, p
 \end{aligned} \tag{1}$$

where  $\mathbf{x} = [x_1, \dots, x_d]^T$  is a vector of  $d$  decision variables in the decision space  $D$ ,  $f : D \rightarrow \mathbb{R}^m$  is a vector-valued function yielding the  $m$  objectives, and the function  $g_i$  ( $i = 1, \dots, p$ ) defines inequality constraints. The goal of a multi-objective problem is to find the set of *non-dominated* or *Pareto-optimal* solutions. A solution  $\mathbf{x}^{(1)}$  is said to dominate another solution  $\mathbf{x}^{(2)}$  when it performs better than the latter on at least one objective, while not performing worse on any of the other objectives [26]. More formally, for  $\mathbf{x}^{(1)}$  and  $\mathbf{x}^{(2)}$  two vectors in  $D$ :

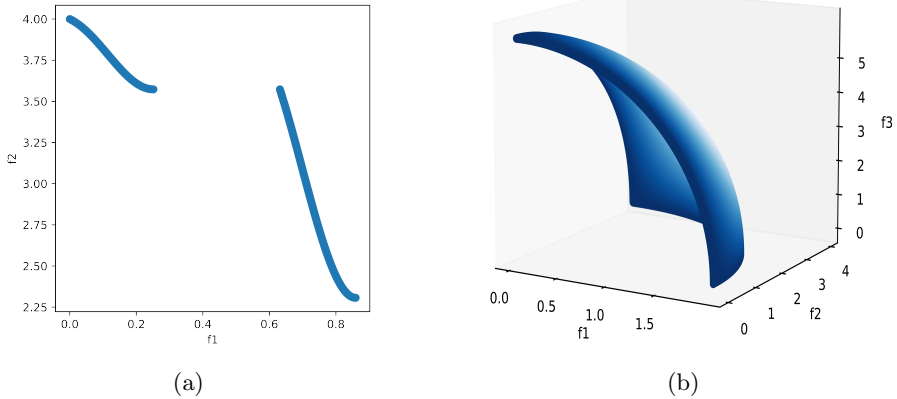
- $\mathbf{x}^{(1)} \prec \mathbf{x}^{(2)}$  means  $\mathbf{x}^{(1)}$  dominates  $\mathbf{x}^{(2)}$  iff  $f_j(\mathbf{x}^{(1)}) \leq f_j(\mathbf{x}^{(2)})$ ,  $\forall j \in \{1, \dots, m\}$ , and  $\exists j \in \{1, \dots, m\}$  such that  $f_j(\mathbf{x}^{(1)}) < f_j(\mathbf{x}^{(2)})$
- $\mathbf{x}^{(1)} \prec\prec \mathbf{x}^{(2)}$  means  $\mathbf{x}^{(1)}$  strictly dominates  $\mathbf{x}^{(2)}$  iff  $f_j(\mathbf{x}^{(1)}) < f_j(\mathbf{x}^{(2)})$ ,  $\forall j \in \{1, \dots, m\}$

The optimal solutions for the individual objectives are usually not of interest to the decision-maker, since these only reflect the extremes of the so-called *Pareto front* (i.e., the evaluation of the Pareto set in the objective space). The geometry of the Pareto front can be very diverse (see, e.g., Figure 1 for two examples generated from artificial test problems), and is usually unknown upfront. Often, it contains infinitely many solutions; the multi-objective optimization algorithm then tries to approximate this front with a finite discrete set. The solution to be implemented in practice will then depend on the preferences of the decision-maker. The reader is referred to [18] for a comprehensive survey on *deterministic* multi-objective methods for engineering problems. In such methods, it is usually assumed that the objective and, if applicable, constraint functions can be observed without noise (e.g., through a deterministic simulator, or noise-free experiments). However, in the adhesive bonding problem discussed here (and in practical settings in general), the objective and constraint observations are *noisy* [27, 28].

## 2.2 Adhesive bonding process: problem setting

The bonding process we focus on joins two PolyPhenylene Sulfide (PPS) substrates using Araldite 2011 adhesive. Figure 2 shows the general procedure of the adhesive bonding process.

The optimization focuses on the Plasma treatment step. The plasma treatment chemically modifies the top surface layer of the PPS substrate so that the surface energy increases, which impacts the adhesion strength (i.e., the strength of the connection between the adhesive and the substrate). In this process, six parameters play a role: (1) whether the surface is pre-processed or not (cleaning to remove dust and grease, which may prevent a good connection between the adhesive and the substrate), (2) the power setting of the plasma torch, (3) the speed at which the plasma torch moves across the samples, (4)



**Fig. 1:** Illustration of (a) a 2-objective problem with a disconnected Pareto front, (b) a 3-objective problem with a concave Pareto front

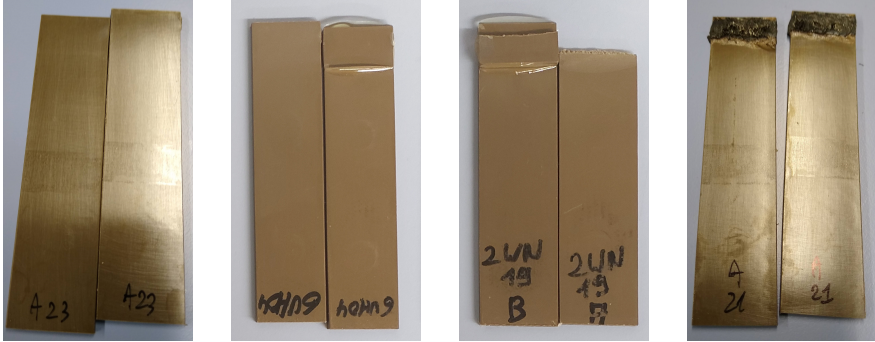


**Fig. 2:** Schematic overview of the adhesive bonding process

the distance between the plasma torch nozzle and the sample, (5) the number of passes of the plasma torch over the sample, and (6) the time between the plasma treatment and the application of the glue (as the plasma effect reverses over time). The adhesion strength is very sensitive to the configuration of these parameters.

Using lab experiments, stress tests can be performed to check the outcomes of samples that have been treated with any particular plasma parameter configuration: the lap shear strength of the sample (MPa), the failure mode (adhesive, cohesive, or substrate failure), the production cost of the sample (in euros), and the potential occurrence of visual damage (the substrates may burn when heated above their maximum allowable temperature during plasma treatment). Such physical experiments are expensive, as they require the whole process in Figure 2 to be performed, involving a human operator. Figure 3 shows different failure modes and an example of visual damage.

The goal of the optimization is to set the plasma process parameters in such a way that (1) the tensile strength (TS) is maximized, (2) the production cost (PC) is minimized, and (3) adhesive failures and visual damage are avoided. As objectives (1) and (2) are in conflict, this is a *multi-objective* optimization



(a) Initial substrate (b) Adhesion failure (c) Substrate failure (d) Burned sample

**Fig. 3:** Illustration of failure modes that may result from the stress test applied to a sample. The adhesive is applied on the substrates (a) and the failure mode can be either adhesion failure (b), substrate failure (c), or cohesive failure (not shown). In addition, visual damage might be observed after the experiment; e.g., burned sample (d)

problem. The goal is to find a set of solutions that reveal the essential trade-offs between these objectives (i.e., those solutions for which no objective can be improved without negatively affecting the performance of any other objective) while meeting the constraints (3). Equation 2 formally defines this optimization problem as

$$\begin{aligned} \min \quad & [-TS(\mathbf{x}), PC(\mathbf{x})] \\ \text{s.t.} \quad & 0.5 - \text{PoF}(\mathbf{x}) \leq 0 \end{aligned} \quad (2)$$

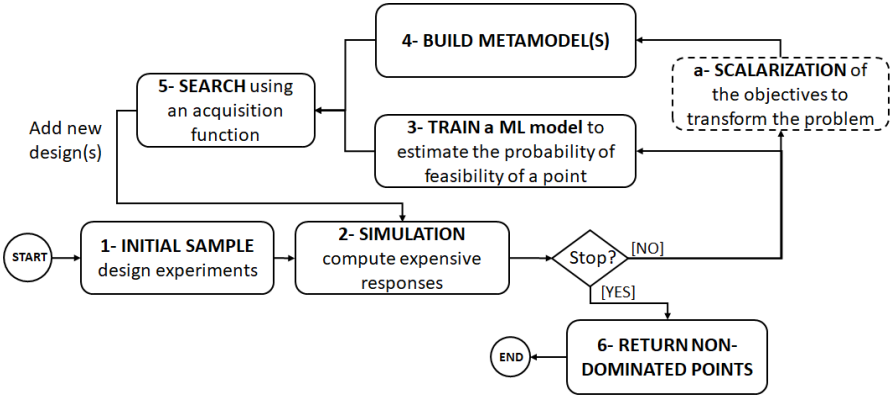
where the notation  $\text{PoF}(\mathbf{x})$  refers to the *probability of feasibility* of any given process configuration  $\mathbf{x}$  (estimated as the fraction of replications in which the configuration resulted in a feasible outcome). As the performance evaluation is expensive, the optimization algorithm should be able to detect (nearly) Pareto-optimal solutions within a small number of experiments required; collecting large amounts of experimental data is simply financially infeasible. In the following section, we discuss how the use of Bayesian optimization allows us to develop such a data-efficient optimization approach.

### 3 Constrained Bayesian multiobjective optimization: proposed algorithms

BO is a supervised learning technique, that starts with the evaluation of an initial set of input points with good space-filling properties (obtained, e.g., by means of Latin hypercube sampling or quasi-random sequences). The BO literature suggests setting the number of initial design points equal to  $k = 10d$  (with  $d$  the number of dimensions of the input space), but smaller

designs have also been advocated [29, 30]. Next, it constructs a *surrogate model* (also referred to as *metamodel*), describing our belief about the response functions under study, based on the samples observed so far. Further observations are then added sequentially using an *acquisition function* that chooses the most promising input point (referred to as *infill point*) as the one to be evaluated next. This acquisition function should balance exploration (in sparsely observed regions of the input space), and exploitation (in regions that are already known to contain good values). The metamodels are then updated, and the algorithm continues until a stopping criterion is met (e.g., the computational budget is depleted).

Each BO algorithm thus has two key elements: the type of metamodel used, and the type of acquisition function. Several acquisition functions exist (see [19] and [20] for single and multiobjective surveys respectively); allegedly, the *expected improvement (EI)* remains one of the most commonly used ones in practice [31], and GP regressors are standard metamodels in the BO literature [32]. In this work, we propose two different algorithms for constrained multi-objective optimization using EI: cMEI-SK, and cEHVI-SK. Both methods follow the same general steps depicted in Figure 4, but differ in the acquisition function used: while cMEI-SK uses a scalarization approach to transform the problem into multiple single-objective ones [25, 33], cEHVI-SK uses the *Expected Hypervolume Improvement (EHVI)* acquisition function [34, 35].



**Fig. 4:** Constrained Bayesian multi-objective optimization algorithms: general steps.

These two acquisition functions are well-known in the BO literature. However, their application has mostly been reported in deterministic unconstrained settings [36]. A few extensions have been proposed in the literature to handle constraints (e.g., [24]), and to handle observational noise (e.g., [25, 37]). Given that in our problem setting the feasibility of a process configuration is evaluated with physical tests, where both the objectives and constraints not



only are noisy but also of different nature (see Section 2.2), such methods cannot be used out-of-the-box. Further details on the proposed algorithms (and thus our main contributions) are given in sections 3.2, 3.3, and 3.4. Moreover, we explicitly differentiate two types of GP models: *ordinary kriging (OK)* [31, 38] and *stochastic kriging (SK)* [39]. While any GP model can accommodate noisy evaluations [32], OK metamodels are limited to homogeneous noise. The seminal work of [39] extended OK metamodels to handle heterogeneous noise (referred to as SK metamodels); as explained further in subsection 3.1, we exploit the information extracted from both types of metamodels.

### 3.1 Ordinary versus stochastic kriging metamodels

In ordinary kriging, it is assumed that the observations of the response function under study are deterministic (i.e., they do not exhibit uncertainty). Assume that we observed this function at  $n$  input locations  $\mathbf{x}^{(i)}, i = \{1, \dots, n\}$  (contained in matrix  $\mathbf{X}$ ), yielding function values  $\mathbf{y}^{(i)}, i = \{1, \dots, n\}$  (contained in matrix  $\mathbf{Y}$ ). This unknown function  $f^{(i)}$  for a point  $\mathbf{x}^{(i)}$  is then modeled as

$$f^{(i)} = m(\mathbf{x}^{(i)}) + M(\mathbf{x}^{(i)}) \quad (3)$$

In this expression,  $m(\mathbf{x})$  represents the mean of the process; it is often supposed to be just a constant ( $\beta_0$ ). The notation  $M(\mathbf{x})$  denotes a realization of a Gaussian random field with mean zero, and a covariance structure that exhibits spatial correlation (according to a *covariance function* or *kernel*  $k(\cdot, \cdot)$ ):

$$\text{Cov}(\mathbf{y}^{(i)}, \mathbf{y}^{(j)}) = k(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}), \quad (4)$$

Often, this covariance function is assumed to be stationary, meaning that its outcome only depends on the *distance* between the input locations. Multiple kernel functions exist in the literature [32]; a popular choice is the *squared exponential* or Gaussian kernel:

$$k(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = \sigma^2 \exp \left[ - \sum_{k=1}^d \frac{[x_k^{(i)} - x_k^{(j)}]^2}{2l_k^2} \right] \quad (5)$$

where  $\sigma^2$  is the process variance, and  $l_k$  is the length scale of the process along dimension  $k$ . The parameters  $\beta_0$ ,  $\sigma^2$ , and  $l_k$  are estimated from the already observed data, usually by means of maximum likelihood estimation.

The ordinary kriging prediction at an arbitrary unobserved location  $\mathbf{x}^{(*)}$  is then given by:

$$\hat{f}_{OK}(\mathbf{x}^{(*)}) = \beta_0 + k_* [K_n]^{-1} (\mathbf{Y} - \mathbf{1}_n \beta_0) \quad (6)$$

where  $\mathbf{1}_n$  is a  $n \times 1$  vector of ones. The mean squared error on the prediction (MSE, also referred to as *kriging variance*) is given by [31]:

$$\hat{s}_{OK}^2(\mathbf{x}^{(*)}) = k_{**} - k_* K_n^{-1} k_*^T \quad (7)$$

where

$$K_n = [k(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})], \quad i, j \in \{1, \dots, n\} \quad (8)$$

$$k_* = [k(\mathbf{x}^{(*)}, \mathbf{x}^{(i)})], \quad i \in \{1, \dots, n\} \quad (9)$$

$$k_{**} = k(\mathbf{x}^{(*)}, \mathbf{x}^{(*)}) \quad (10)$$

Note that the uncertainty involved in the model of Equation 3 is *imposed* on the problem by assumption, to aid the construction of the predictive model. For this reason, it is also referred to as the *extrinsic uncertainty* [39].

In many problem settings, however, the observations of the response function are *not* deterministic: evaluating the same input location  $\mathbf{x}^{(i)}$  multiple times (i.e., taking multiple replications) results in a *different* observation  $\mathbf{y}^{(r,i)}$  in each replication  $r$ . The observations thus exhibit *noise*, which stems from the experiments (this is also referred to as *intrinsic* noise). In the literature, this noise is often assumed to be *homogeneous* and modeled using OK meta-models [28]. In practice, however, the structure and level of the noise often depend on the input variables [38]. *Stochastic kriging* meta-models [39] explicitly account for this intrinsic (input-dependent) noise. It models the observed response value in the  $r$ -th replication at design point  $\mathbf{x}^{(i)}$  as:

$$f^{(r,i)} = m(\mathbf{x}^{(i)}) + M(\mathbf{x}^{(i)}) + \epsilon_r(\mathbf{x}^{(i)}) \quad (11)$$

where  $m$  and  $M$  are defined as in Equation 3, and  $\epsilon_r(\mathbf{x}^{(i)})$  is the intrinsic noise observed in replication  $r$ . The SK prediction at an unobserved location  $\mathbf{x}^{(*)}$  is then given by

$$\hat{f}_{SK}(\mathbf{x}^{(*)}) = \beta_0 + k_* [K_n + \Sigma_\epsilon]^{-1} (\mathbf{Y} - \mathbf{1}_n \beta_0), \quad (12)$$

and the MSE on this prediction is given by

$$\hat{s}_{SK}^2(\mathbf{x}^{(*)}) = k_{**} - k_* [K_n^{-1} + \Sigma_\epsilon] k_*^T. \quad (13)$$

In these expressions,  $K_n$ ,  $k_*$  and  $k_{**}$  are as defined in Equations 8, 9 and 10 respectively, and

$$\Sigma_\epsilon = \text{diag} \left[ \frac{\text{Var}(\mathbf{y}^{(1)})}{r^{(1)}}, \dots, \frac{\text{Var}(\mathbf{y}^{(n)})}{r^{(n)}} \right], \quad (14)$$

where  $\text{Var}(\mathbf{y}^{(i)})$  refers to the sample variance of the replication outcomes at location  $\mathbf{x}^{(i)}$ . The values on the diagonal divide this sample variance by the number of replications  $r^{(i)}$  taken at each location, and thus reflect the estimated variance of the mean outcome at each observed location.

### 3.2 Probability of feasibility

The *probability of feasibility* (PoF) has been well-studied in the BO literature when the constraints are deterministic [27, 35, 36]. In our problem setting, the constraint feasibility is checked physically (meaning that it will not entail visual

damage or adhesive failure), which results in a classification problem with two classes: YES for feasible outcomes and NO for infeasible outcomes. Thus, for each process configuration evaluated, we have a discrete target variable  $t \in \{0, 1\}$ , where 1 denotes YES and 0 denotes NO. We are interested in the probability  $p(t = 1 \mid \mathbf{X})$ ; that is, the probability of a process configuration being truly feasible conditioned on the data collected so far.

To do this, we can train a *classification* model in Step 3 of Algorithm 4. However, training a binary classifier here is not very helpful, since the *noise* on the constraints will affect the performance for the same input configuration (i.e., replicating the same process configuration may yield different outcomes for constraint violation). For instance, if the same input is replicated 5 times, yielding 2 YES and 3 NO, it is unlikely that training the classifier with e.g., the mode of the outcomes will yield accurate predictions. Therefore, we opt for training a regression model on the proportion of successful outcomes instead: for a given process configuration  $\mathbf{x}^{(*)}$ , we take the proportion of successful outcomes of the binary target  $t$  as

$$\bar{t}^{(*)} = \frac{\sum_{j=1}^{r^{(*)}} t_j}{r^{(*)}} \quad (15)$$

where  $r$  is the number of replications. We then approximate the *probability of feasibility* (PoF) of the unobserved locations by fitting an ordinary kriging model (Equation 6) on the target  $\bar{t}$  (denoted  $\hat{f}_c(\mathbf{x})$ ). That is, given the observed targets  $\bar{t} \in \mathbf{T}$  at points  $\mathbf{X}$ , the predicted distribution of an unobserved target  $t^{(*)}$  at point  $\mathbf{x}^{(*)}$  is given by

$$\text{PoF}(\mathbf{x}^{(*)}) \simeq \hat{f}_c(\mathbf{x}^{(*)}) = p(t^{(*)} = 1 \mid \mathbf{x}^{(*)}, \mathbf{X}, \mathbf{T}). \quad (16)$$

### 3.3 cMEI-SK acquisition function

When a scalarization function is used, then only one metamodel is trained on the scalarized objective at each BO iteration. In this work, we use the augmented Tchebycheff scalarization function, which is popular in general multi-objective optimization problems due to the theoretical guarantees it provides [26, 33]. The scalarized objective is given by:

$$Z_{\boldsymbol{\lambda}}(\mathbf{x}) = \max_{j \in \{1, \dots, m\}} \lambda_j f_j(\mathbf{x}) + \rho \sum_{j=1}^m \lambda_j f_j(\mathbf{x}) \quad (17)$$

where  $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_m]$ ,  $\sum_{j=1}^m \lambda_j = 1, \forall j \in \{1, \dots, m\}$ , and  $\rho$  is a small positive value (e.g.,  $\rho = 0.05$ ). We then fit an SK metamodel on  $Z_{\boldsymbol{\lambda}}(\mathbf{x})$ , explicitly accounting for the intrinsic noise on the scalarized objective (see Equation 12). The metamodel information is then used in the acquisition function as

$$\begin{aligned} \text{MEI-SK}(\mathbf{x}) &= \left[ \widehat{Z}_{SK}(\mathbf{x}_{\min}) - \widehat{Z}_{SK}(\mathbf{x}) \right] \Phi \left[ \frac{\widehat{Z}_{SK}(\mathbf{x}_{\min}) - \widehat{Z}_{SK}(\mathbf{x})}{\widehat{s}_{OK}(\mathbf{x})} \right] \\ &\quad + \widehat{s}_{OK}(\mathbf{x}) \phi \left[ \frac{\widehat{Z}_{SK}(\mathbf{x}_{\min}) - \widehat{Z}_{SK}(\mathbf{x})}{\widehat{s}_{OK}(\mathbf{x})} \right] \end{aligned} \quad (18)$$

where  $\widehat{Z}_{SK}(\mathbf{x}_{\min})$  is the SK prediction for the scalarized function ( $Z_{\lambda}$ ) at  $\mathbf{x}_{\min}$  (i.e., the point having the lowest sample mean for the scalarized objective among all feasible points already sampled), and  $\phi(\cdot)$  and  $\Phi(\cdot)$  are the standard normal density and standard normal distribution function, respectively [25, 40]. For an arbitrary input location, the corresponding constrained MEI (denoted cMEI-SK) is given by

$$\text{cMEI-SK}(\mathbf{x}) = \widehat{f}_c(\mathbf{x}) \times \text{MEI}(\mathbf{x}) \quad (19)$$

Note that at each BO iteration, a new weight vector  $\lambda$  is selected from a set of weights distributed uniformly, allowing the algorithm to sample points across the entire Pareto front [33].

### 3.4 cEHVI-SK acquisition function

When scalarization is not used, an independent SK metamodel is trained for each of the  $m$  objectives. The Expected Hypervolume Improvement (EHVI) is a popular acquisition function in *unconstrained* and *deterministic* settings [36, 41]. The hypervolume is the size of the space dominated by a Pareto front  $\mathbf{P}$  given a reference point [42]. Therefore, the hypervolume improvement of an objective vector  $\mathbf{y} \in \mathbb{R}^m$  is defined as the increment of the hypervolume indicator after  $\mathbf{y}$  is added to the current approximation of  $\mathbf{P}$  [35, 43]

$$I(\mathbf{y}, \mathbf{P}) = \mathcal{H}(\mathbf{P} \cup \{\mathbf{y}\}) - \mathcal{H}(\mathbf{P}) \quad (20)$$

where  $\mathcal{H}$  is the hypervolume calculation function. The EHVI can be defined as the integration of the hypervolume improvement function over the non-dominated area using the metamodel prediction [34]

$$\text{EHVI-SK}(\mathbf{x}) = \int_{\mathbf{y}(\mathbf{x}) \in A} I(\mathbf{y}(\mathbf{x}), \mathbf{P}) \prod_{i=1}^m \frac{1}{\widehat{s}_{SK_i}(\mathbf{x})} \phi \left( \frac{y_i(\mathbf{x}) - \widehat{f}_{SK_i}(\mathbf{x})}{\widehat{s}_{SK_i}(\mathbf{x})} \right) dy_i(\mathbf{x}) \quad (21)$$

where  $A$  stands for the non-dominated area. The terms  $\widehat{f}_{SK_i}(\mathbf{x})$  and  $\widehat{s}_{SK_i}(\mathbf{x})$  represent the objective and uncertainty estimators of the stochastic GP model respectively. Previous studies have already used EHVI, often assuming noiseless objectives [35, 41, 44], or with homogeneous noise at best [37, 45]. Finally, for a novel input configuration  $\mathbf{x}^{(*)}$ , the corresponding constrained EHVI (denoted cEHVI-SK) is given by

$$\text{cEHVI-SK}(\mathbf{x}^{(*)}) = \widehat{f}_c(\mathbf{x}^{(*)}) \times \text{EHVI-SK}(\mathbf{x}^{(*)}) \quad (22)$$

## 4 Design of experiments

To test the proposed optimization approach, we benchmark its performance against five state-of-the-art constrained EMOAs: C-NSGA-II [14], C-MOEA/D [15], C-TAEA [16], C-MOPSO [17], an adaptation of C-NSGA-II [46] to use the OK metamodel prediction to generate new populations (OK-C-NSGA-II), and a modification to the surrogate-assisted evolutionary algorithm K-RVEA [47], which we refer to as C-K-RVEA. The latter uses GP surrogates to approximate the objective functions, and it was modified to handle infeasible configurations with a penalization factor given by PoF (see Equation 2). To evaluate the impact of SK, we also implemented our proposed approaches using OK models; these are denoted with an ‘OK’ in the algorithm name. Furthermore, Zhan and Xing [36] present a different formulation for constrained EI by using the definition of PoF presented by Sobester *et. al.* [48]. However, we observed in our experiments that the formulation we propose in Equation 19 got, on average, superior results compared to the results obtained with the formulation suggested by Zhan and Xing [36]. Appendix A details these results.

The optimization of the acquisition function in Bayesian optimization tends to be non-trivial, as the function is often non-linear, non-convex, and multimodal [49]. Here we use a *Particle Swarm Optimization* (PSO) algorithm to find the infill point that maximizes the proposed acquisition functions (i.e., the fitness function of this inner optimization). Our choice is motivated by the good performance and low computational time observed in other studies with high-dimensional search space [50]. With PSO, the position of the particle represents the values of each variable to optimize.

Since the physical experiments are very expensive, a Matlab process simulator was provided by the Joining & Materials Lab<sup>1</sup>. This simulator predicts the lap shear strength of the sample (MPa), failure mode (adhesive, substrate, or cohesive failure), sample production cost (in euros), and visual quality outcome (YES in a matter of seconds. It is critical to note that this simulator is *not* meant to be a digital twin of the true process, but rather a tool for the *relative* comparison of the performance of the algorithms under different conditions, at almost zero cost. Table 1 shows the range of each process parameter considered in the optimization problem, and Table 2 summarizes the parameters of the optimization algorithms. It is worth mentioning that the Kolmogorov-Smirnov test showed that the distributions of the objective replications (for break strength) do not follow a Normal distribution. While it is common to use the mean of repeat evaluations as a way of handling noisy functions, this does still assume that the noise is normally distributed. However, we can assume normality according to the Central Limit Theorem (CLT).

---

<sup>1</sup><https://www.flandersmake.be>

In probability theory, CLT establishes that, in many situations, for independent and identically distributed random variables, the sampling distribution of the standardized sample mean tends towards the standard normal distribution even if the original variables themselves are not normally distributed.

**Table 1:** Range of the process settings (input variables) considered in the optimization.

ID	Variable	Min	Max
v1	Pre-processing	Yes or No	
v2	Power setting (W)	300	500
v3	Torch speed (mm/s)	5	250
v4	Distance between the torch and the sample (cm)	0.2	2
v5	Number of passes	1	50
v6	Time between plasma treatment and glue application (min)	1	120

The simulator allows the analyst to experiment with different levels of noise. In real life, multiple factors cause noise to occur. One of these is the so-called *contact angle*<sup>2</sup>: this measure reflects the extent to which the adhesive can maintain good contact with the material. This is important to achieve a strong adhesive bond. The contact angle depends on the type of material but also on impurities or contaminants such as wax, oil, plasticizers, etc. present on the material surface. Even though all samples in our setting are made of the same material, variations in the degree of these contaminants occur across samples, implying variations in contact angle. These result in noisy measurements of the final break strength of the bonded joints. A realistic value for the *standard deviation of the contact angle* is  $\gamma = 30\%$  of the mean, which is what we use in our experiments.

Given the experimental setting in Table 2, each algorithm evaluates exactly 60 process configurations in an expensive way, with 5 replications per configuration (i.e., 300 expensive evaluations in total). Although the number of replications performed in one input location should be in correspondence to the level of noise observed at that point, we kept this value constant for each input given the small budget we have for the optimization (300 expensive evaluations). The BO algorithms start with an initial design of 20 process configurations (note that this is smaller than the usual choice of  $k = 11d - 1$ ); exactly one infill point is then added in each of the following 40 iterations. The EMOAs, by contrast, use an initial population of 20 process configurations (the same initial set used by BO algorithms) and in each of the 2 successive generations, a novel population is generated. As common in the literature, the fitness of the configuration outcomes is evaluated based on their sample means over the 5 replications (note that, by doing so, these algorithms implicitly ignore the fact that this sample mean is in itself uncertain). Our approach

---

<sup>2</sup>Other noise factors could not be controlled in the simulator, so they are not further discussed.

**Table 2:** Summary of the parameters of the optimization algorithms

Setting	BO algorithms	Evolutionary algorithms
Size of initial design/population	LHS: $N = 20$	
Crossover probability	-	0.5 (C-MOEA/D); 0.9 o.w
Mutation probability	-	0.1
Reference directions	-	19 (C-TAEA); 3 (C-MOEA/D)
Inertia weight	-	0.4 (C-MOPSO)
C1 factor	-	2 (C-MOPSO)
C2 factor	-	2 (C-MOPSO)
Max velocity (%)	-	5 (C-MOPSO)
Reference vectors	-	151 (C-K-RVEA)
Replications	$r = 5$	
Iterations	40	2
Acquisition function	cMEI-SK/cEHVI-SK	-
Acquisition function optimization	PSO*	-
Kernel (for all GPR models)	Gaussian	-

\* PSO configuration: swarm size = 50, max iterations=1800, max stall iterations = 10, tolerance =  $1e^{-6}$

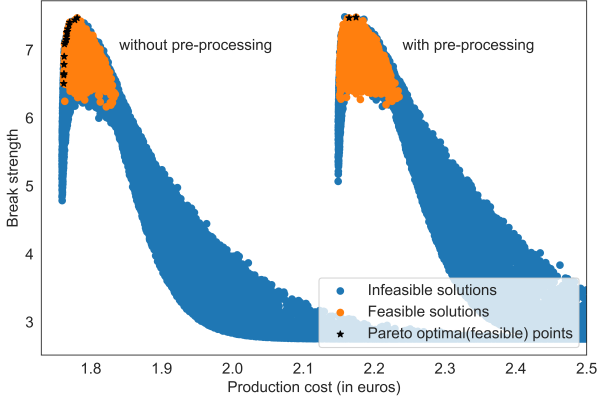
takes into account both the sample mean and the sample variance though, as explained in Section 3. While a total budget of 300 evaluations may seem high, it allows us to also study the progress the algorithms would have obtained at lower budgets, as illustrated below.

We evaluate the quality of the resulting fronts using the well-known hypervolume (HV) and IGD+ indicators [51]. The hypervolume is the volume of the objective space that is dominated by the front obtained, w.r.t. a reference point, and the IGD+ is the average smallest inverted generational distance to the closest member of the true Pareto front. The larger the hypervolume and the lower the IGD+, the better the quality of the front obtained. We also include performance analysis of the proposed methods using the *empirical attainment function* [52]. As the front obtained by the algorithms may depend on the initial design, we performed 50 macro-replications, each one starting with a different initial design.

## 5 Results

To gain some insight into the objective space to be explored by the algorithms under idealized conditions (i.e., if the contact angle could be perfectly controlled), Figure 5 shows the mean responses of the simulator on a set of 60 000 process configurations (with  $\gamma = 0\%$ ). These configurations were determined through Halton sampling, and each configuration was replicated five times. Interestingly, the feasible solutions seem to be clustered in areas with high break strength. Moreover, the use of pre-processing seems to merely lead to a cost increase, while the resulting gains in break strength are very scarce and only minor. The Pareto optimal (feasible) points estimated by means of

these Halton results, under these idealized conditions ( $\gamma = 0\%$ ), serve as the benchmark Pareto front to judge the quality of the competing optimization algorithms.

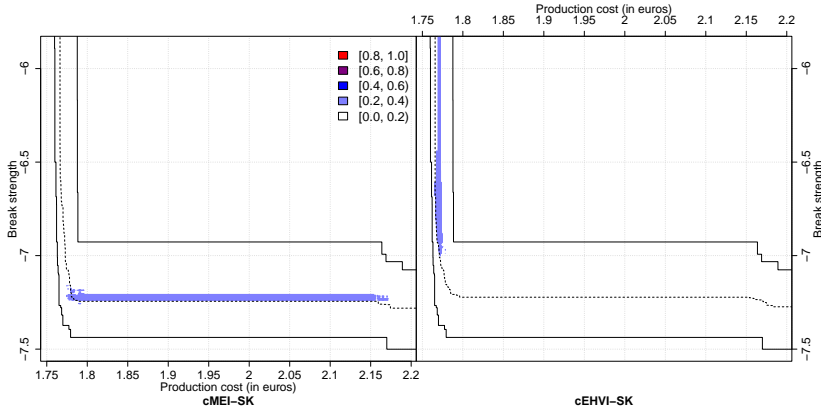


**Fig. 5:** Sample mean of break strength versus production cost, estimated by the simulator for 60 000 random process configurations ( $\gamma = 0\%$ )

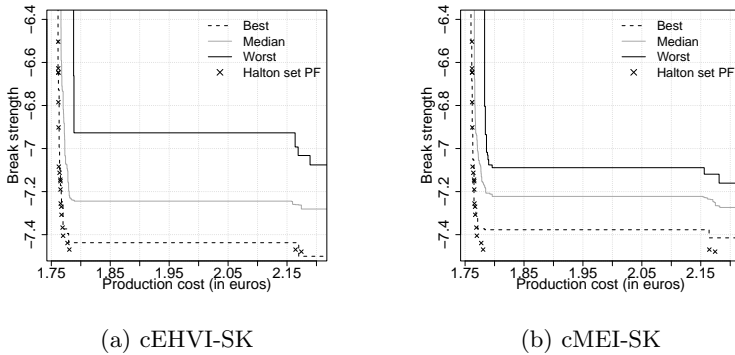
Figure 6 shows an exploratory analysis of the Pareto fronts obtained by each BO algorithm in the different macro-replications. The analysis is performed using the concept of the empirical attainment function (EAF, [52]). For each point, the EAF gives an estimate of the probability that this point is dominated (attained) by the Pareto front put forward by the given algorithm. Connecting points with the same given EAF value yields an *attainment surface* that separates the objective space into two regions: those objective vectors that are attained by the resulting Pareto fronts with (at least) that probability, and those that are not. The attainment surfaces allow us to summarize the location of the objective vectors obtained by a stochastic algorithm. The median attainment surface, for instance, consists of the objective vectors that are attained by half of the runs (representing a probability of 50%). Similarly, the worst-case results of an algorithm are reflected in the worst attainment surface, whereas the best results are given by the best attainment surface. The shaded areas in the right-hand side of Figure 6 show the objective areas where the proposed cMEI-SK approach reaches better attainment surfaces minimizing the objective **cost**. On the other hand, using cEHVI-SK means a better attainment surface minimizing the objective **break strength**.

Figure 7 shows the final *best* Pareto front obtained by the proposed approaches over 50 macro-replications (along with the *median* and *worst* fronts), for  $\gamma = 30\%$  and Figure 9 shows the percent of Pareto feasible solutions found at the end of the optimization. Although the number of feasible solutions in the Pareto front remained similar for each algorithm, BO-based approaches were able to find better fronts w.r.t. the IGD+ and hypervolume





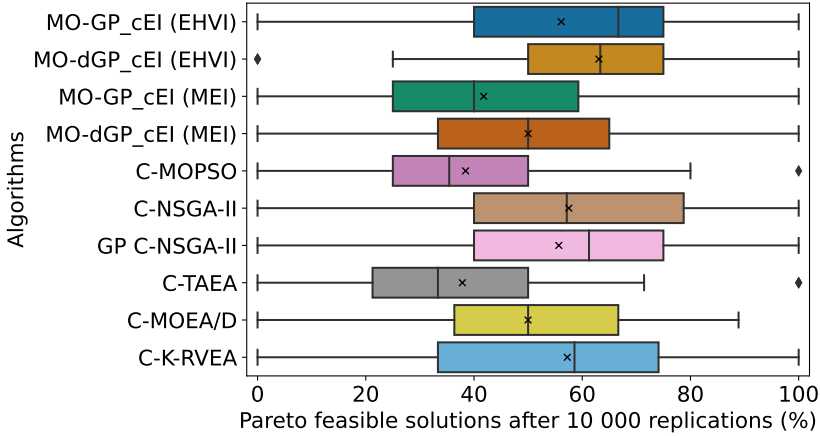
**Fig. 6:** Visualization of the differences between the EAFs using different acquisition functions in the BO methods. The Y-axis is inverted to allow for the minimization of both objectives. Each plot shows the median surface (dashed line), along with the best and worst surfaces (full lines). The areas where a better attainment surface is obtained are indicated by a shaded area (the colors indicate the improvement in the probabilities).



**Fig. 7:** Best, median, and worst Pareto front obtained by the BO methods. The Y-axis is inverted to allow minimization of both objectives

indicators. Each of these Pareto feasible solutions was considered to add 10,000 extra replications to analyze the percentage of those solutions that remained feasible. Figure 8 shows that approximately 60% of the process configurations remained feasible for most of the optimization algorithms, with C-TAEA as the worst and cEHVI-SK the best (w.r.t. the median percentage).

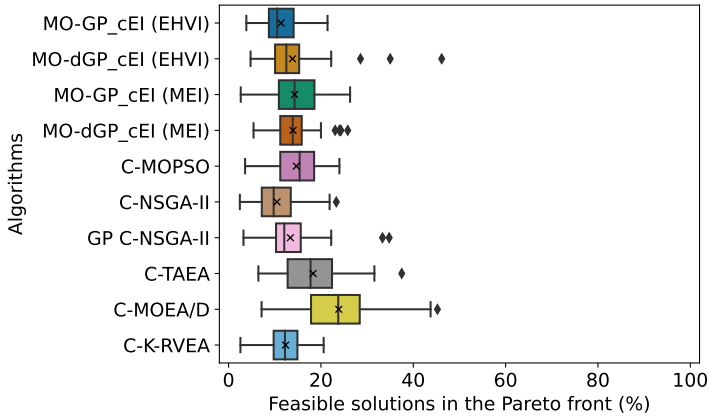
The best Pareto front obtained by cEHVI-SK is very close to the ideal Pareto front estimated by means of the Halton set exploration. It also leads



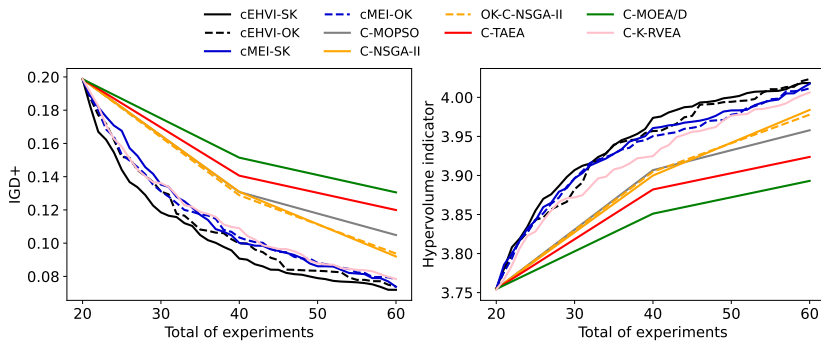
**Fig. 8:** Percent of Pareto feasible solutions (found by the algorithms during the optimization of each macro-replication) that remained as feasible when 10,000 replications were performed. The star and vertical line inside the box represent the mean and median, respectively, of the percentage computed for each macro-replication.

to a *faster* increase in HV, in terms of the number of expensive evaluations performed, than all other algorithms. This is evident from Figure 10, which shows the evolution of the *average* hypervolume and IGD+ obtained (across macro-replications) during the optimization process (again, for  $\gamma = 30\%$ ). This superior performance is also evident from Table 3, which shows the results for the average HV, along with those of the average IGD+ (the latter uses the ideal front obtained in the Halton experiment of Figure 5 as the true front). These results highlight that BO methods are able to obtain better quality results for the Pareto front than evolutionary algorithms (including surrogate-assisted ones), particularly at very limited budgets. Appendix B shows that statistical differences (Wilcoxon test,  $\alpha = 5\%$ ), both for HV and IGD+, were mainly focused amongst C-MOPSO, C-MOEA/D, and C-TAEA; and metamodel-based optimization algorithms (including C-K-RVEA).

Overall, ignoring the input-dependent noise associated with the objective *break strength* (thus using the mean value of the replications) negatively influences the performance of optimization algorithms. As shown in Figure 10, the curves of the *deterministic* BO methods (i.e., the proposed algorithms using OK metamodels) are inferior to the one when input-dependency is taken into account. Yet, they remain superior to the evolutionary algorithms, which reinforces the advantages of the exploration/exploitation performed by the optimization of the acquisition function and thus obtaining the best possible trade-off between expected performance and model error. A steep increase in the hypervolume (or decrease in IGD+) indicates that the algorithm is finding solutions that increment the area of the search space dominated by the



**Fig. 9:** Percent of Pareto feasible solutions found by the algorithms during the optimization of each macro-replication. The star and vertical line inside the box represent the mean and median, respectively, of the percentage computed for each macro-replication.



**Fig. 10:** Evolution of the IGD+ metric (Left) and hypervolume indicator (Right) of the Pareto-optimal solutions throughout the search, for  $\gamma = 30\%$  (average of 50 macro-replications). The Pareto front obtained from the Halton set is considered to compute the IGD+ metric and the reference point with production cost=3, break strength=4 is used to compute the hypervolume indicator

obtained Pareto front (and the Pareto optimal solutions are close to an ideal Pareto front). For instance, we can see that BO-based optimization quickly suggests (feasible) Pareto optimal solutions during the first 20 iterations (the first generation of the EAs). As for the number of simulations needed for convergence, from Figure 10 we can also see that EAs need more iterations (more expensive evaluations to perform) to reach the Pareto optimal solutions that BO-based optimization already found with fewer iterations (fewer expensive

evaluations). Since the type of problem addressed in this research is characterized by the expansiveness of the objective evaluation, performing more iterations can be unfeasible in practice. Hence, the superiority of BO-based optimization. Additionally, we note that surrogate-assisted evolutionary algorithms (such as C-K-RVEA) can indeed benefit from the approximation of the objectives and obtain better configurations than standard EMOAs. However, this performance remains inferior to BO approaches when data-efficiency is needed on top of effective black-box optimization.

Finally, by looking closer at the solutions (in input space) suggested by cEHVI-SK (the one with the highest HV value), we found that  $\pm 63\%$  of the solutions skip pre-processing, meaning that the production costs are reduced. Thus in Figure 11 we show the distribution of the remaining input variables. In general, cEHVI-SK suggests that the power should be between 480 W and 500 W, the speed should move at a speed between 127.5 mm/s and 152 mm/s, the distance between the torch and the sample should be in the range of 0.2cm and 0.38 cm, between 11 and 16 passes should be performed, and a time difference between 1 and 13 minutes should be considered before the glue application.

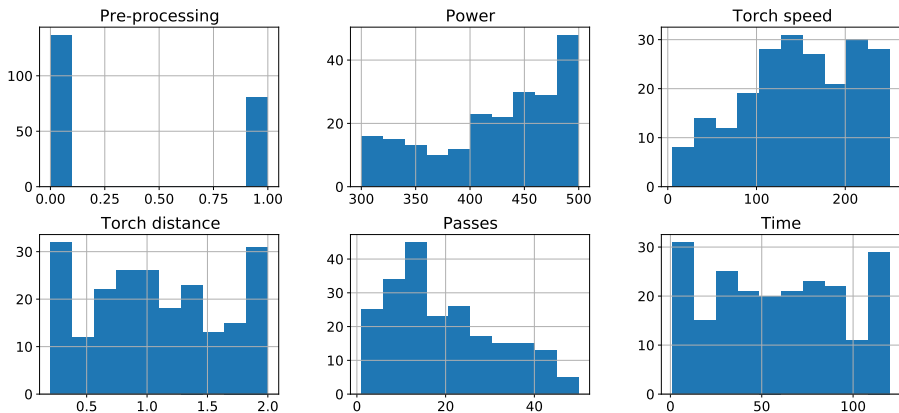
**Table 3:** Average IGD+ and HV of the fronts obtained over 50 macro-replications, for  $\gamma = 30\%$

	IGD+	HV
<b>cEHVI-SK</b>	0.0719	4.0184
<b>cEHVI-OK</b>	0.0739	4.0235
<b>cMEI-SK</b>	0.0737	4.0174
<b>cMEI-OK</b>	0.0782	4.0116
<b>C-K-RVEA</b>	0.0785	4.0062
<b>C-NSGA-II</b>	0.092	3.9839
<b>OK-C-NSGA-II</b>	0.0937	3.979
<b>C-MOPSO</b>	0.1049	3.9579
<b>C-TAEA</b>	0.1199	3.9237
<b>C-MOEA/D</b>	0.1306	3.893

## 6 Conclusions

This paper presented two constrained Bayesian optimization algorithms to solve a bi-objective problem related to the adhesive bonding process of materials (maximizing break strength while minimizing production costs). As the real experiments are carried out physically in a lab are costly, the budget for evaluations is very limited. The proposed Bayesian approach is shown to clearly outperform state-of-the-art evolutionary algorithms, which are commonly used in engineering design when solving general multi-objective, constrained problems. The difference lies in the way the experimental design is guided throughout the search: the Bayesian approach selects infill points based on an (explainable) acquisition function, which is related to the expected merit of the new infill point for optimization. The BO model ensures that the search

focuses on infill points that have a high probability of being feasible. Moreover, the GP model used to approximate the objective(s) accounts for the output (heterogenous) noise, whereas the evolutionary algorithms rely simply on the (uncertain) sample means as performance approximations. Moreover, the search in EMOAs (as it is generally with metaheuristic approaches) is guided by hard-to-tune evolutionary operators. The success of evolutionary processes is largely dependent on the availability of a sufficient experimentation budget, which is not always the case in practice.



**Fig. 11:** Distribution of the Pareto-optimal input values obtained by cEHV1-SK, across 50 macro-replications.

Our research highlights the superiority of BO methods, particularly using an EHVI-based infill criterion, over traditional EMOAs in solving the optimization problem of analysis. Additionally, the optimization algorithms considered in this research were applied to solve two other engineering problems and the results are summarized in an electronic companion of this paper. We are convinced that the use of Bayesian approaches holds great promise in solving noisy and expensive engineering problems, in terms of both search efficiency (i.e., finding solutions within a limited budget) and search effectiveness (i.e., yielding high-quality solutions). Future research will focus on the further development of an interactive software tool (a first release has been provided and is being tested), allowing lab experts to validate the results generated by the algorithm in a real-life test environment, and on the inclusion of a third objective (minimization of the *debonding* break strength). Another potential research direction could focus on a dynamic allocation of the replications applied to each input configuration to reduce the uncertainty observed in the objectives. This can lead to more optimal use of the evaluation budget and potentially greater exploration/exploitation of the search space.

## Acknowledgments

This work was supported by the Flanders Artificial Intelligence Research Program (FLAIR), the Research Foundation Flanders (FWO Grant 1216021N), and Flanders Make vzw.

## Declarations

**Conflict of interest** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix A Constrained Expected Improvement (CEI)

The Constrained Expected Improvement (CEI [36]) uses one OK metamodel to approximate the expensive objective and one metamodel to approximate each expensive constraint independently. Then, for a constrained optimization problem with  $c$  constraints

$$\begin{aligned} \min \quad & y(\mathbf{x}), \mathbf{x} \in \mathbb{R} \\ \text{s.t.} \quad & g_i(\mathbf{x}) \leq 0, i = 1, 2, \dots, c \end{aligned} \quad (\text{A1})$$

the objective value of point  $\mathbf{x}$  can be treated as a Gaussian random variable  $\mathcal{N}(\hat{y}(\mathbf{x}), \hat{s}(x))$  and the  $i$ -th constraint value of  $\mathbf{x}$  can be also treated as a Gaussian random variable  $\mathcal{N}(\hat{g}_i(\mathbf{x}), \hat{e}_i(x))$ ,  $i = 1, 2, \dots, c$ . For this,  $\hat{y}$  and  $\hat{s}$  are the GP prediction and standard error of the objective function respectively, and  $\hat{g}_i$  and  $\hat{e}_i$  are the GP prediction and standard error of the  $i$ -th constraint function respectively.

Then, from Equation 2 we can transform the constraint and derive the Probability of Feasibility as:

$$g(\mathbf{x}) = 0.5 - Pf(\mathbf{x}) \leq 0 = Pf(\mathbf{x}) \geq 0.5 = \frac{Pf(\mathbf{x}) - \hat{g}(\mathbf{x})}{\hat{e}(\mathbf{x})} \geq \frac{0.5 - \hat{g}(\mathbf{x})}{\hat{e}(\mathbf{x})} \quad (\text{A2})$$

$$\begin{aligned} PoF(\mathbf{x}) &= Prob\left(\frac{Pf(\mathbf{x}) - \hat{g}(\mathbf{x})}{\hat{e}(\mathbf{x})} \geq \frac{0.5 - \hat{g}(\mathbf{x})}{\hat{e}(\mathbf{x})}\right) \\ &= 1 - Prob\left(\frac{Pf(\mathbf{x}) - \hat{g}(\mathbf{x})}{\hat{e}(\mathbf{x})} \leq \frac{0.5 - \hat{g}(\mathbf{x})}{\hat{e}(\mathbf{x})}\right) \\ &= 1 - \Phi\left(\frac{0.5 - \hat{g}(\mathbf{x})}{\hat{e}(\mathbf{x})}\right) \end{aligned} \quad (\text{A3})$$

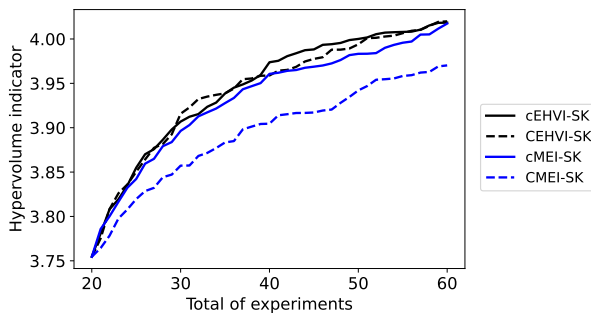
In case the GP standing for the objective and constraint function are mutually independent, the CEI can be obtained by combining the EI (to be

consistent with our work we used MEI instead and the predictors defined in Equation 12 and Equation 13) and PoF as

$$\begin{aligned} \text{CMEI-SK}(\mathbf{x}) &= \text{MEI}(\mathbf{x}) \times \text{PoF}(\mathbf{x}) \\ &= \left[ \left( \hat{f}_{SK}(\mathbf{x}_{min}) - \hat{f}_{SK}(\mathbf{x}) \right) \Phi(\mathcal{D}) + \hat{s}_{OK}(\mathbf{x}) \phi(\mathcal{D}) \right] \\ &\quad \times \left[ 1 - \Phi \left( \frac{0.5 - \hat{g}(\mathbf{x})}{\hat{e}(\mathbf{x})} \right) \right] \end{aligned} \quad (\text{A4})$$

and

$$\mathcal{D} = \frac{\hat{f}_{SK}(\mathbf{x}_{min}) - \hat{f}_{SK}(\mathbf{x})}{\hat{s}_{OK}(\mathbf{x})} \quad (\text{A5})$$



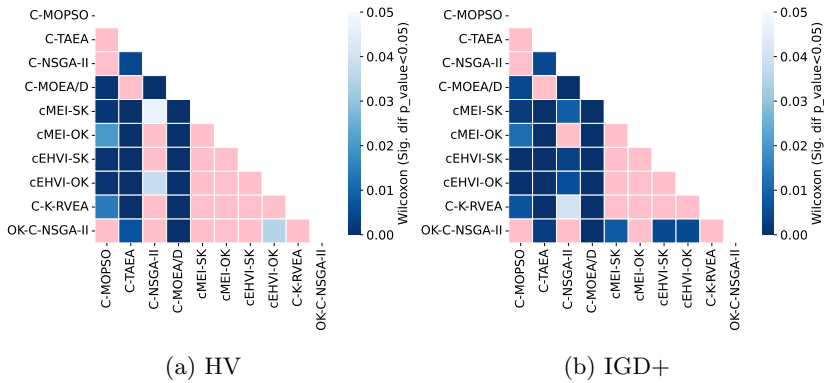
**Fig. A1:** Evolution of the mean hypervolume throughout the optimization (of 50 macro-replications). The reference point [production cost=3, break strength=4] is used to compute this metric

Note that this equation is different from our proposed Equation 19 in the derivation of the PoF. Figure A1 shows that our proposed cMEI-SK got on average better Pareto fronts (higher hypervolume values) than that of CEI. This suggests that considering the uncertainty predicted by the GP may lead the optimization to points that do not cause an increase in the hypervolume. This was more evident when the improvement was measured with MEI (and fitting only one GP to the scalarized objectives).

## Appendix B Wilcoxon test results

### References

- [1] Brockmann, W., Geiß, P.L., Klingen, J., Schröder, K.B.: Adhesive Bonding: Materials, Applications and Technology. John Wiley and Sons, - (2008)



**Fig. B2:** Wilcoxon test results for significant differences between algorithms, using (a) Hypervolume and (b) IGD+ metrics. Pink color indicate no significant differences ( $p\_value \geq 5\%$ )

- [2] Cavezza, F., Boehm, M., Terryn, H., Hauffman, T.: A review on adhesively bonded aluminium joints in the automotive industry. *Metals* **10**(6) (2020). <https://doi.org/10.3390/met10060730>
- [3] Licari, J.J., Swanson, D.W.: *Adhesives Technology for Electronic Applications: Materials, Processing, Reliability*. William Andrew, - (2011)
- [4] Correia, S., Anes, V., Reis, L.: Effect of surface treatment on adhesively bonded aluminium-aluminium joints regarding aeronautical structures. *Engineering Failure Analysis* **84**, 34–45 (2018). <https://doi.org/10.1016/j.engfailanal.2017.10.010>
- [5] da Silva, L.F.M., Ochsner, A., Adams, R.D., Spelt, J.K.: *Handbook of Adhesion Technology*. Springer, - (2011)
- [6] Pocius, A.V.: *Adhesion and Adhesives Technology: an Introduction*. Carl Hanser Verlag GmbH Co KG, - (2021)
- [7] Budhe, S., Banea, M.D., De Barros, S., Da Silva, L.F.M.: An updated review of adhesively bonded joints in composite materials. *International Journal of Adhesion and Adhesives* **72**, 30–42 (2017)
- [8] Habenicht, G.: *Applied Adhesive Bonding: a Practical Guide for Flawless Results*. John Wiley and Sons, West Sussex, UK (2008)
- [9] Chiarello, F., Belingheri, P., Fantoni, G.: Data science for engineering design: State of the art and future directions. *Computers in Industry* **129**, 103447 (2021). <https://doi.org/10.1016/j.compind.2021.103447>



- [10] Zhou, A., Qu, B.-Y., Li, H., Zhao, S.-Z., Suganthan, P.N., Zhang, Q.: Multiobjective evolutionary algorithms: A survey of the state of the art. *Swarm and Evolutionary Computation* **1**(1), 32–49 (2011). <https://doi.org/10.1016/j.swevo.2011.03.001>
- [11] Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE transactions on evolutionary computation* **6**(2), 182–197 (2002). <https://doi.org/10.1109/4235.996017>
- [12] Corbett, M., Sharos, P.A., Hardiman, M., McCarthy, C.T.: Numerical design and multi-objective optimisation of novel adhesively bonded joints employing interlocking surface morphology. *International Journal of Adhesion and Adhesives* **78**, 111–120 (2017). <https://doi.org/10.1016/j.ijadhadh.2017.06.002>
- [13] Labbé, S., Drouet, J.-M.: A multi-objective optimization procedure for bonded tubular-lap joints subjected to axial loading. *International journal of adhesion and adhesives* **33**, 26–35 (2012). <https://doi.org/10.1016/j.ijadhadh.2011.09.005>
- [14] Brownlee, A.E., Wright, J.A.: Constrained, mixed-integer and multi-objective optimisation of building designs by nsga-ii with fitness approximation. *Applied Soft Computing* **33**, 114–126 (2015). <https://doi.org/10.1016/j.asoc.2015.04.010>
- [15] Jain, H., Deb, K.: An evolutionary many-objective optimization algorithm using reference-point based nondominated sorting approach, part ii: Handling constraints and extending to an adaptive approach. *IEEE Transactions on evolutionary computation* **18**(4), 602–622 (2013)
- [16] Li, K., Chen, R., Fu, G., Yao, X.: Two-archive evolutionary algorithm for constrained multiobjective optimization. *IEEE Transactions on Evolutionary Computation* **23**(2), 303–315 (2018)
- [17] Gao, Y.-l., Qu, M.: Constrained multi-objective particle swarm optimization algorithm. In: *International Conference on Intelligent Computing*, pp. 47–55 (2012). Springer
- [18] Chugh, T., Sindhya, K., Hakanen, J., Miettinen, K.: A survey on handling computationally expensive multiobjective optimization problems with evolutionary algorithms. *Soft Computing*, 1–30 (2017). <https://doi.org/10.1007/s00500-017-2965-0>
- [19] Frazier, P.I.: Bayesian optimization. In: *Recent Advances in Optimization and Modeling of Contemporary Problems*, pp. 255–278. *Informs*, ??? (2018)

- [20] Rojas Gonzalez, S., van Nieuwenhuyse, I.: A survey on kriging-based infill algorithms for multiobjective simulation optimization. *Computers and Operations Research* **116**, 104869 (2020). <https://doi.org/10.1016/j.cor.2019.104869>
- [21] Hernández-Lobato, J.M., Gelbart, M., Hoffman, M., Adams, R., Ghahramani, Z.: Predictive entropy search for bayesian optimization with unknown constraints. In: *International Conference on Machine Learning*, pp. 1699–1707 (2015). PMLR
- [22] Gelbart, M.A., Adams, R.P., Hoffman, M.W., Ghahramani, Z., *et al.*: A general framework for constrained bayesian optimization using information-based search. *Journal of Machine Learning Research* **17**(160), 1–53 (2016)
- [23] Garrido-Merchán, E.C., Hernández-Lobato, D.: Predictive entropy search for multi-objective bayesian optimization with constraints. *Neurocomputing* **361**, 50–68 (2019)
- [24] Feliot, P., Bect, J., Vazquez, E.: A bayesian approach to constrained single-and multi-objective optimization. *Journal of Global Optimization* **67**(1), 97–133 (2017)
- [25] Rojas Gonzalez, S., Jalali, H., Nieuwenhuyse, I.V.: A multiobjective stochastic simulation optimization algorithm. *European Journal of Operational Research* **284**(1), 212–226 (2020). <https://doi.org/10.1016/j.ejor.2019.12.014>
- [26] Miettinen, K.: *Nonlinear Multiobjective Optimization* vol. 12. Springer, - (1999)
- [27] Forrester, A., Sobester, A., Keane, A.: *Engineering Design Via Surrogate Modelling (A Practical Guide)*, 1st edn. John Wiley and Sons, West Sussex, UK (2008)
- [28] Horn, D., Dagge, M., Sun, X., Bischl, B.: First investigations on noisy model-based multi-objective optimization. In: *International Conference on Evolutionary Multi-Criterion Optimization*, pp. 298–313 (2017). Springer. [https://doi.org/10.1007/978-3-319-54157-0\\_21](https://doi.org/10.1007/978-3-319-54157-0_21)
- [29] Loepky, J.L., Sacks, J., Welch, W.: Choosing the sample size of a computer experiment: a practical guide. *Technometrics* **51**(4), 366–376 (2009)
- [30] Zeng, Y., Cheng, Y., Liu, J.: An efficient global optimization algorithm for expensive constrained black-box problems by reducing candidate infilling region. *Information Sciences* **609**, 1641–1669 (2022)

- [31] Jones, D.R., Schonlau, M., Welch, W.J.: Efficient global optimization of expensive black-box functions. *Journal of Global optimization* **13**(4), 455–492 (1998). <https://doi.org/10.1023/A:1008306431147>
- [32] Williams, C.K., Rasmussen, C.E.: *Gaussian Processes for Machine Learning* vol. 2. MIT press Cambridge, MA, - (2006)
- [33] Knowles, J.: Parego: A hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems. *IEEE Transactions on Evolutionary Computation* **10**(1), 50–66 (2006). <https://doi.org/10.1109/TEVC.2005.851274>
- [34] Emmerich, M.T., Deutz, A.H., Klinkenberg, J.W.: Hypervolume-based expected improvement: Monotonicity properties and exact computation. In: 2011 IEEE Congress of Evolutionary Computation (CEC), pp. 2147–2154 (2011). IEEE
- [35] Couckuyt, I., Deschrijver, D., Dhaene, T.: Fast calculation of multiobjective probability of improvement and expected improvement criteria for pareto optimization. *Journal of Global Optimization* **60**(3), 575–594 (2014)
- [36] Zhan, D., Xing, H.: Expected improvement for expensive optimization: a review. *Journal of Global Optimization* **78**(3), 507–544 (2020). <https://doi.org/10.1007/s10898-020-00923-x>
- [37] Daulton, S., Balandat, M., Bakshy, E.: Parallel bayesian optimization of multiple noisy objectives with expected hypervolume improvement. *Advances in Neural Information Processing Systems* **34**, 2187–2200 (2021)
- [38] Kleijnen, J.P.: *Design and Analysis of Simulation Experiments*. Springer, - (2018)
- [39] Ankenman, B., Nelson, B.L., Staum, J.: Stochastic kriging for simulation metamodeling. *Operations Research* **58**(2), 371–382 (2010). <https://doi.org/10.1109/WSC.2008.4736089>
- [40] Quan, N., Yin, J., Ng, S.H., Lee, L.H.: Simulation optimization via kriging: a sequential search using expected improvement with computing budget constraints. *Iie Transactions* **45**(7), 763–780 (2013). <https://doi.org/10.1080/0740817X.2012.706377>
- [41] Loka, N., Couckuyt, I., Garbuglia, F., Spina, D., Van Nieuwenhuyse, I., Dhaene, T.: Bi-objective bayesian optimization of engineering problems with cheap and expensive cost functions. *Engineering with Computers*, 1–11 (2022)

- [42] Zitzler, E., Thiele, L.: Multiobjective optimization using evolutionary algorithms—a comparative case study. In: *Parallel Problem Solving from Nature—PPSN V: 5th International Conference Amsterdam, The Netherlands September 27–30, 1998 Proceedings* 5, pp. 292–301 (1998). Springer
- [43] Emmerich, M.T., Giannakoglou, K.C., Naujoks, B.: Single-and multi-objective evolutionary optimization assisted by gaussian random field metamodels. *IEEE Transactions on Evolutionary Computation* **10**(4), 421–439 (2006)
- [44] Daulton, S., Balandat, M., Bakshy, E.: Differentiable expected hypervolume improvement for parallel multi-objective bayesian optimization. *Advances in Neural Information Processing Systems* **33**, 9851–9864 (2020)
- [45] Koch, P., Wagner, T., Emmerich, M.T., Bäck, T., Konen, W.: Efficient multi-criteria optimization on noisy machine learning problems. *Applied Soft Computing* **29**, 357–370 (2015)
- [46] Qin, S., Sun, C., Jin, Y., Zhang, G.: Bayesian approaches to surrogate-assisted evolutionary multi-objective optimization: A comparative study. In: *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 2074–2080 (2019)
- [47] Chugh, T., Jin, Y., Miettinen, K., Hakanen, J., Sindhya, K.: A surrogate-assisted reference vector guided evolutionary algorithm for computationally expensive many-objective optimization. *IEEE Transactions on Evolutionary Computation* **22**(1), 129–142 (2016)
- [48] Sobester, A., Forrester, A., Keane, A.: *Engineering Design Via Surrogate Modelling: a Practical Guide*. John Wiley & Sons, West Sussex, UK (2008)
- [49] Wilson, J., Hutter, F., Deisenroth, M.: Maximizing acquisition functions for bayesian optimization. In: *Advances in Neural Information Processing Systems*, vol. 31. Curran Associates, Inc., ??? (2018)
- [50] Yarat, S., Senan, S., Orman, Z.: A Comparative Study on PSO with Other Metaheuristic Methods, pp. 49–72. Springer, Cham (2021)
- [51] Li, M., Yao, X.: Quality evaluation of solution sets in multiobjective optimisation: A survey. *ACM Computing Surveys* **52**(2), 26 (2019)
- [52] López-Ibáñez, M., Paquete, L., Stützle, T.: Exploratory analysis of

stochastic local search algorithms in biobjective optimization. In: *Experimental Methods for the Analysis of Optimization Algorithms*, pp. 209–222. Springer, ??? (2010). [https://doi.org/10.1007/978-3-642-02538-9\\_9](https://doi.org/10.1007/978-3-642-02538-9_9)