



Evolutionary Robust Optimization in Production Planning - Interactions between Number of Objectives, Sample Size and Choice of Robustness Measure

DOI:
[10.1016/j.cor.2016.06.020](https://doi.org/10.1016/j.cor.2016.06.020)

Document Version
Accepted author manuscript

[Link to publication record in Manchester Research Explorer](#)

Citation for published version (APA):

Diaz, J. E., Handl, J., & Xu, D. (2017). Evolutionary Robust Optimization in Production Planning - Interactions between Number of Objectives, Sample Size and Choice of Robustness Measure. *Computers and Operations Research*, 79, 266-278. <https://doi.org/10.1016/j.cor.2016.06.020>

Published in:
Computers and Operations Research

Citing this paper

Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.



Author's Accepted Manuscript

Evolutionary Robust Optimization in Production Planning - Interactions between Number of Objectives, Sample Size and Choice of Robustness Measure

Juan Esteban Diaz, Julia Handl, Dong-Ling Xu



www.elsevier.com/locate/caor

PII: S0305-0548(16)30152-6
DOI: <http://dx.doi.org/10.1016/j.cor.2016.06.020>
Reference: CAOR4032

To appear in: *Computers and Operation Research*

Received date: 30 September 2015
Revised date: 23 June 2016
Accepted date: 29 June 2016

Cite this article as: Juan Esteban Diaz, Julia Handl and Dong-Ling Xu
Evolutionary Robust Optimization in Production Planning - Interactions between
Number of Objectives, Sample Size and Choice of Robustness Measure
Computers and Operation Research, <http://dx.doi.org/10.1016/j.cor.2016.06.020>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Evolutionary Robust Optimization in Production Planning - Interactions between Number of Objectives, Sample Size and Choice of Robustness Measure

Juan Esteban Diaz^{a,*}, Julia Handl^a, Dong-Ling Xu^a

^a*Alliance Manchester Business School, The University of Manchester*

Abstract

We aim to find robust solutions in optimization settings where there is uncertainty associated with the operating/environmental conditions, and the fitness of a solution is hence best described by a distribution of outcomes. In such settings, the nature of the fitness distribution (reflecting the performance of a particular solution across a set of operating scenarios) is of potential interest in deciding solution quality, and previous work has suggested the inclusion of robustness as an additional optimization objective. However, there has been limited investigation of different robustness criteria, and the impact this choice may have on the sample size needed to obtain reliable fitness estimates. Here, we investigate different single and multi-objective formulations for robust optimization, in the context of a real-world problem addressed via simulation-based optimization. For the (limited evaluation) setting considered, our results highlight the value of an explicit robustness criterion in steering an optimizer towards solutions that are not only robust (as may be expected), but also associated with a profit that is, on average, higher than that identified by standard single-objective approaches. We also observe significant interactions between the choice of robustness measure and the sample size employed during fitness evaluation, an effect that is more pronounced for our multi-objective models.

Keywords: Evolutionary Multi-objective optimization, Production planning, Robust optimization, Simulation-based optimization, Uncertainty modelling

*Corresponding author.

Email address: juanestebandiazleiva@gmail.com (Juan Esteban Diaz)

1. Introduction

Real-world problems often require the consideration of multiple and often conflicting criteria [13, 49, 39]. When faced with problems that incorporate uncertainty, one may arguably need to take into account the robustness of solutions [4] as an additional criterion of solution quality. Ideally, robust optimization would identify a solution that simultaneously offers the best and most robust performance; in reality, these two criteria are usually conflicting [44], and a single optimal solution is unlikely to exist. Due to this expected trade-off between the quality of a solution and its robustness [30, 41], methodologies that are capable of identifying a desirable trade-off between these conflicting criteria are of significant interest in a robust optimization setting.

A complicating factor in robust optimization is the lack of a unique definition of the robustness of a solution. In some cases a solution is considered to be robust when, under certain levels of variation in the decision space or operating conditions, it performs reasonably well in terms of quality (e.g. expected objective value) [26, p. 189], feasibility or optimality [24]. Perhaps a more common idea is to consider as robust a solution that offers the best performance under the worst-case scenario [24] (e.g. maximizing the minimum value of a performance measure).

Independent of the exact definition of robustness, it has been argued that it may be advantageous to explicitly consider robustness as a separate objective [31], instead of optimizing for a single measure that implicitly considers robustness (e.g. the average of a performance measure). This is because, in doing so (using e.g. a Pareto optimization approach) more comprehensive information about the robustness of the solution can potentially be taken into account (such as the variation in the fitness values seen), and the trade-off between the main objective and its robustness measure may become more evident and facilitate decision making. On the other hand, more implicit considerations of robustness may have advantages in terms of the associated computational expense and their ease of implementation. The calculation of any measure of robustness implies additional modelling choices regarding the set of scenarios considered (i.e. how to model the way in which uncertainty presents itself, which may be in the form of variation in the design vector or the operating conditions [34]), and the number of scenarios to consider (e.g.

exhaustive enumeration versus sampling).

Some of the interactions between the above modelling choices are poorly understood, e.g. it is unclear how the choice of sample size impacts on the performance of different robustness criteria, and whether this may differ for multi-objective versus single-objective optimizers. Here, we aim to investigate this aspect in the context of a real-world robust optimization setting. Specifically, we study a production planning problem in which an optimal production plan is derived, but the profit achievable under this plan is subject to uncertainty in the factory’s operating conditions (due to the possibility of machine failures). In this setting, a simulation model of the production system can be used to translate a (deterministic) production plan into a distribution of fitness values obtained under different scenarios of operating conditions, i.e the simulation model serves the purpose of modelling the uncertainty intrinsic to the operating conditions.

1.1. Robust Optimization

To some extent, robust solutions to a problem may be obtained by replacing the original objective(s) by its/their “robust” equivalent(s) [30], e.g. a robustness measure such as the average value instead of a single evaluation. Alternatively, measures of robustness may be integrated into the optimization as hard constraints [26, p. 216]. A third possibility is to consider robustness as part of a composite function that is the weighted sum of different objectives (Deb, 2014, p. 408; Guo et al., 2008).

This are, of course, standard approaches to transform a multi-objective optimization problem into a single-objective optimization problem, with a number of known limitations: the choice of appropriate values for the weight vector can be difficult and highly subjective, as it requires the translation of qualitative information into quantitative indicators [13, p. 409]. Furthermore, the (single) solution obtained through this approach can be very sensitive to the relative weight assigned to each objective, and the method may thus be highly-user dependent. Although different solutions can be found by varying the weights, this approach is unable to find Pareto optimal solutions that lie on non-convex regions of the Pareto optimal front [37].

Pareto optimization overcomes these limitations by exploring different trade-off solutions between a set of objectives [29], without prior knowledge of a preference vector. Concretely, Pareto optimization employs the principle of Pareto dominance: A solution dominates another solution if it is strictly

better in at least one objective and not worse in any objective [14, 35]. Different to single-objective optimization approaches, Pareto optimization does not aim to find a single optimal or near-optimal solution, but the set of optimal trade-off solutions. This so-called Pareto-optimal set consists of all those solutions that are not dominated by any other solution in the search space. The reader is referred to Deb [13] and Coello [8] for more details about Pareto optimization.

In the context of robust optimization, we may use Pareto optimization to simultaneously optimize the original objective (performance measure) as well as a robustness measure. The Pareto front then serves to illustrate the trade-off between these objectives, and the decision maker may choose one solution from the set of Pareto-optimal solutions based on his experience, specific requirements, or an in-depth analysis of the quality of individual solutions.

1.2. Evolutionary Robust Optimization

Several previous works have searched for robust solutions via evolutionary single-objective (ESO) optimization. Tsutsui and Ghosh [48] proposed the single-evaluation mode (SEM) for genetic algorithms (GAs) with a robust solution searching scheme, which consists of adding perturbations to the values of the decision variables, and then a single fitness evaluation is computed per individual. Under the SEM approach robust individuals are more likely to survive across generations than less robust ones. SEM can be seen as an equivalent to an implicit averaging strategy when uncertainty is present in the decision space. Also a multi-evaluation mode (MEM), which can be seen as an equivalent to an explicit averaging strategy, was later proposed by Tsutsui [47] to search for robust solutions with GAs. The author also proposed a MEM where the fitness of an individual corresponds to the worst fitness value (MEM-W) for situations where robustness is a crucial consideration. Branke [2] analysed the potential of evolutionary algorithms (EAs) to find robust solutions via different explicit averaging strategies. In Branke [2], uncertainty is also modelled by adding random perturbations to the decision space before fitness evaluation. After analysing several strategies, the author concluded that both, re-evaluating fitter individuals and using previous fitness values to weight the fitness of those individuals are two promising ways to help EAs in the search for robust solutions. The MEM-W approach proposed by Tsutsui [47] is later applied by Ong et al. [40] within a max-min surrogate-assisted EA to generate robust solutions for engineering design, since MEM-W was

thought to be more conservative in terms of solution robustness than MEM (explicit averaging).

Considerable research in robust optimization can also be found in the existing evolutionary multi-objective (EMO) optimization literature. The explicit averaging strategy used in ESO optimization was extended to EMO optimization by Deb and Gupta [14]. Two approaches to obtain robust solutions are proposed in Deb and Gupta [14]. In the first approach, fitness evaluations for the different objectives are computed by averaging a set of neighbouring solutions in order to create perturbations in the decision space, which are later reflected in the objective space. In the second approach the optimization is performed based on fitness evaluations of the actual individuals (without perturbations) present in the population and a constraint is added to the problem for every objective considered. This constraint imposes a user-defined limit on the normalized, absolute or average normalized difference between the perturbed fitness (which may be represented by average fitness or worst fitness) and the original fitness (fitness of the actual solution without perturbations). In this second approach the desired level of robustness can be specified in terms of the maximum level of fitness variation associated with local perturbations. In both approaches multiple fitness evaluations need to be computed (using neighbouring solutions) to calculate the average fitness and the perturbed fitness in the first and second case, respectively.

In Jin and Sendhoff [31], trade-off solutions between robustness and performance were obtained by simultaneously optimizing both criteria via EMO optimization. This approach was motivated by the observation that expectation measures such as the sample mean cannot sufficiently capture fluctuations in performance [31], whereas variance-based measures fail to take into account the absolute performance of a solution. This idea of simultaneously optimizing a performance and a robustness measure via EMO optimization has been applied in several studies to search for robust solutions. For instance, Ray [43] searched for robust solutions for engineering design problems by simultaneously optimizing the average performance of neighbouring solutions, and the standard deviation of the performance of neighbouring solutions. Similarly, Goh and Tan [25] generated trade-off solutions between performance and robustness by optimizing measures of expected and worst case performance.

Lim et al. [34, 33] describe an inverse technique that avoids the need for defining an appropriate neighbourhood in the design space (as this may

be difficult to achieve a priori for some design problems). Starting from a threshold on acceptable fitness variation alone, this approach performs a multi-objective search for solutions with the highest fitness and the largest neighbourhood size consistent with this fitness threshold.

While many single and multi-objective studies have been published in the field of robust optimization, studies that directly compare ESO and EMO approaches to find robust solutions for the same problem are rare in the existing literature. Pereira et al. [42] compared the ability of evolutionary single and multi-objective algorithms to find robust solutions (link weight configurations for traffic routing processes). However, in Pereira et al. [42] a solution was considered to be robust when it performed well under two different network conditions; in this sense, two performance measures (rather than one performance measure and one robustness measure) were optimized. The weighted sum of two congestion functions (two performance measures) was applied as fitness function in the ESO algorithm, whereas in both EMO algorithms, the non-dominated sorting genetic algorithm II (NSGA II) [15] and strength Pareto evolutionary algorithm 2, both congestion functions were optimized separately. The performance advantage of NSGA II was evident under more complex network topologies and more demanding traffic requirements.

1.3. Contributions

In this study, we investigate the extent to which ESO and EMO optimization are able to generate high-quality robust solutions for a real-world production planning problem with uncertainty in its operating conditions. Specifically, we consider the sets of solutions obtained for different single and multi-objective formulations, given a limited budget of fitness evaluations (each of which requires a simulation replication).

Unique features of our experiments are as follows: (i) We contrast the use of several possible robustness measures, as well as different choices of sample size; (ii) The final quality of solutions is assessed using a large number of simulations, which provide a comprehensive picture of the fitness distribution associated with each solution (such picture is not available during optimization); (iii) The final solution sets are compared along a range of criteria, including measures of variability, average and worst case performance.

The remainder of this paper is organized as follows: In Section 2, we introduce the problem analysed and describe the ESO and EMO models. We present the details of the benchmark analysis and performance assessment

in Section 3 and 4, respectively, and the results obtained are presented in Section 5. Finally, conclusions derived from this study are given in Section 6.

2. Simulation-Based Optimization Model

In production planning, profitability is not the only criterion that needs to be considered during the specification of a production plan, but the robustness of that plan is also relevant [41].

In this study we address a production planning problem of a real manufacturing system. Detailed features of this manufacturing system are presented in Diaz [19, p. 29-42]. The robustness of production plans is a criterion that needs to be considered here because the system analysed is subject to failures of its production lines, which may cause deviations between a plan and its actual outcome. Consequently, we aim to identify a set of production plans that makes evident the trade-off between profitability and robustness.

Here, a production plan $\mathbf{x}$ is a vector of 41 decision variables, where each decision variable $x_{l,j}$ indicates the number of lots of product j to be manufactured in production line l during a finite planning horizon of one month. This batch manufacturing system has a set $M = \{1, 2, \dots, 7\}$ of independent production lines and is able to manufacture 31 different products. A feature of this system is that only a subset $A_j \subseteq M$ of production lines can manufacture a specific product j . Additionally, this manufacturing system has insufficient capacity to fully cover demand requirements and is subject to serious and non-serious failures of its production lines. We also consider that near-perfect and imperfect repairs can be undertaken after the occurrence of a failure as well as the deterioration caused by previous failures.

Simulation-based optimization (SBO) is an attractive approach to address this problem, for it can accurately capture the inherent uncertainty and complexity of the manufacturing system, while searching for near-optimal solutions for the problem analysed [21, 46, 20]. The SBO model consists of a simulation model and an EA coupled in a black-box fashion. Production plans specified by the optimizer are simulated by the simulation model, which returns the realization of that production plan. Those simulated responses are then used by the optimizer to compute the fitness values of a given production plan.

The simulation model used in this study is the one presented in Diaz [19, p. 126-136]. This simulation model developed in SimEvents[®] (The MathWorks, Inc., 2014) models the manufacturing system analysed. It employs a

combination of discrete-event simulation (DES) and Markov chains to model realistic features of manufacturing systems such as different types of production line failure, near-perfect and imperfect repairs as well as the deterioration of production lines due to previous failures.

The reader is referred to Diaz [19, p. 126-136] for a detailed description of the simulation model, whereas further details about the optimization model are provided in Section 2.1.

2.1. Optimization Model

Here, we address an extended version of the stochastic multidimensional multiple bounded knapsack problem with assignment restrictions (d -MBKARS) problem [19, p. 91-92] that considers a robustness measure as a second objective. In this problem we do not penalize deviations from a plan and we refer to it as the multi-objective d -MBKARS (M- d -MBKARS) problem without penalty. Its integer linear programming (ILP) formulation, in the context of the problem analysed, is as follows:

$$\text{maximize } f_1(\mathbf{x}) = \frac{1}{\gamma} \sum_{r=1}^{\gamma} \sum_{j=1}^{31} \sum_{l \in A_j} v_j \times s_{l,j,r} \quad (1)$$

and either

$$\text{minimize } f_2(\mathbf{x}) = \sqrt{\frac{1}{\gamma} \sum_{r=1}^{\gamma} \left(\sum_{j=1}^{31} \sum_{l \in A_j} v_j \times s_{l,j,r} - f_1(\mathbf{x}) \right)^2} \quad (2)$$

or

$$\text{maximize } f_3(\mathbf{x}) = \min_{r=1,2,\dots,\gamma} \sum_{j=1}^{31} \sum_{l \in A_j} v_j \times s_{l,j,r} \quad (3)$$

subject to

$$\sum_{j=1}^{31} \sum_{l \in A_j} w_{i,j} \times x_{l,j} \leq c_i \quad (i = 1, 2, \dots, 13), \quad (4)$$

$$\sum_{l \in A_j} x_{l,j} \leq b_j \quad (j = 1, 2, \dots, 31). \quad (5)$$

where $s_{l,j,r}$ is obtained by simulating a production plan $\mathbf{x}$ (vector that contains all decision variables $x_{l,j}$) using the simulation model. In this sense, $s_{l,j,r}$ corresponds to the number of lots of product j manufactured in production line l during the r -th simulation replication given a production plan $\mathbf{x}$. γ is the total number of simulation replications performed to compute fitness of a production plan, please see Section 2.1.1 for details about γ . The simulation model can be seen here as the function $g(\mathbf{x})$ able to map a production plan $\mathbf{x}$ onto the actual production $\mathbf{s}_r$ (vector that contains all simulated responses $s_{l,j,r}$) without the requirement of a closed form formulation. In other words $\mathbf{s}_r$ is the realization of a production plan $\mathbf{x}$ during the r -th simulation replication. v_j is the marginal (per lot) contribution to profit of product j , $w_{i,j}$ is the amount of resource i deployed by manufacturing one lot of product j and c_i is the amount of resource i available at the beginning of the manufacturing process.

We represent in the form of Equation 4 the design capacities of production lines, the amount of labour needed to manufacture one lot of product j in production line l and the situation where a product manufactured in this system is subsequently used as raw material during the manufacturing of another product. We also impose additional constraints in the form of Equation 5 to avoid that the production level of product j exceeds its maximum demand b_j . Please note that in this study, a production plan $\mathbf{x}$ is feasible if it satisfies the set of constraints imposed as Equations 4 and 5.

2.1.1. Evolutionary Optimizers

For the multi-objective SBO (MSBO) approach, we require an optimizer able to find a set of solutions that are non-dominated and are widely spread over the approximated Pareto front. EAs are attractive optimizers for Pareto optimization not only due to their wide applicability, flexibility [8] and ease of use [9], but also because multiple non-dominated solutions can be found in a single iteration [3, 13, 8]. The optimizer used in the MSBO approach is NSGA II, a well-known EMO algorithm commonly employed as benchmark in multi-objective optimization [8], that has been successfully used to solve SBO problems (see Brownlee and Wright [5] and Sanchez et al. [45] for recent examples).

In the MSBO approach, in addition to optimizing a performance measure (maximize average profit), we simultaneously optimize a robustness measure

in order to explicitly capture as a second objective the robustness of a production plan. Two different measures of robustness are evaluated in the multi-objective approach, the standard deviation of profit and the minimum profit. The minimization of the standard deviation tries to reduce the variation in profitability, whereas the maximization of the minimum profit tries to improve the worst case scenario derived from a production plan. In this sense the first MSBO model (MSBO-1) tries to simultaneously maximize average profit (Equation 1) and minimize the standard deviation of profit (Equation 2), subject to the set of constraints in the form of Equations 4 and 5. The second MSBO model (MSBO-2) differs from MSBO-1 in that, instead of minimizing the standard deviation, MSBO-2 tries to maximize minimum profit (Equation 3). Both multi-objective models, MSBO-1 and MSBO-2, use NSGA II as optimizer.

Two different single-objective SBO (SSBO) models are also evaluated, the first SSBO model (SSBO-1) tries to maximize average profit (Equation 1) subject to the set of constraints in the form of Equations 4 and 5, whereas the second SSBO model (SSBO-2) tries to maximize minimum profit (Equation 3), subject to the same set of constraints. Both single-objective models employ as optimizer the MI-LXPM algorithm proposed by Deep et al. [16].

In order to make a fair comparison between MSBO and SSBO approaches, we use the same parameters and operators for both optimizers, except for unavoidable discrepancies that arise due to the different mechanics of both GAs. Both optimizers use a population size of 100 individuals and employ binary tournament selection, Laplace crossover (crossover probability: 0.8) and power mutation (mutation probability: 0.005) as operators. See Deep et al. [16] for more details about the last two operators and the parameter values used. Also the constraint-handling method proposed by Deb [12] and the truncation procedure described in Deep et al. [16], which ensures compliance with integer constraints after crossover and mutation, are implemented in both optimizers.

It is important to mention that the selection criterion in NSGA II is based on the crowded-comparison operator [15], although it employs a binary tournament selection operator. Moreover, the way elitism is incorporated in both optimizers differs. In MI-LXPM, elitism is incorporated by having an elite set with a default size of one [17, 18]. We do not adjust this parameter, as we are aiming to maintain good levels of diversity throughout the optimization process. In NSGA II elitism is incorporated by combining the entire parent and offspring populations before extracting the new population.

Given the stochastic nature of simulation responses ($s_{l,j,r}$) and due to the nature of the objectives optimized (a sample of independent fitness values is required to compute the fitness of an individual), all fitness values are computed across a number γ of independent fitness evaluations. Two explicit averaging strategies are implemented in every single and multi-objective model. In strategy A, fitness is computed across 10 independent fitness evaluations ($\gamma = 10$), whereas in strategy B, fitness is computed across 30 independent fitness evaluations ($\gamma = 30$). The letter A or B is added to the notation of the different models according to the explicit averaging strategy that they employ, e.g. MSBO-1A and MSBO-1B.

Explicit averaging is a very expensive approach in terms of computational effort. This computational expense becomes even more critical when explicit averaging is applied in the context of SBO, where expensive simulation replications are required to compute multiple fitness values of entire populations. Therefore, in order to boost the optimization performance of NSGA II and MI-LXPM, and reduce their computational effort, we devise the specialized initialization operator presented in the following subsection. We also use a parallel implementation in all models to speed up computations.

2.2. Initialization Operator

This initialization operator creates up to 22 individuals of the GA's initial population and the rest of individuals are randomly initialized. In order to support both objectives, the initialization operator is implemented in two phases. The first phase creates solutions that are biased towards maximization of average profit, whereas the second phase tries to generate solutions that are biased towards minimization of the standard deviation of profit.

In the first phase, 200 numerical values of c'_l are generated. c'_l is the number of hours that production line l was available during a working month (24 days) and is computed as follows:

$$c'_l = \sum_{j=1}^{31} s_{l,j} \times t \quad (l = 1, 2, \dots, 7), \quad (6)$$

where $s_{l,j}$ is obtained by simulating in the simulation model a production plan that fully utilizes the theoretical capacity of all production lines and t corresponds to the manufacturing time per lot (in hours) when no failure occurs, which according to the company is 8 hours for all product lots. Subsequently, for each set of c'_l an ILP solution is found by solving the following

ILP problem:

$$\text{maximize } f(\mathbf{x}) = \sum_{j=1}^{31} \sum_{l \in A_j} v_j \times x_{l,j} \quad (7)$$

subject to

$$\sum_{j=1}^{31} \sum_{l \in A_j} w_{i,j} \times x_{l,j} \leq c_i \quad (i = 1, 2, \dots, 6), \quad (8)$$

$$\sum_{j=1}^{31} w_{l,j} \times x_{l,j} \leq c'_l \quad (x_{l,j} \in \mathbb{Z}_{\geq 0}; l = 1, 2, \dots, 7). \quad (9)$$

Here, capacities of production lines are not represented in the form of Equation 8 as other resource constraints, but in the form of Equation 9, where c'_l is used as the right hand side (RHS).

Duplicates among those 200 ILP solutions are eliminated in order to maintain diversity and only 10 ILP solutions are chosen, as described bellow, to be part of the set W , which is incorporated into the GA's initial population. If the number of ILP solutions left is less than or equal to 10, all of them are included into W , whereas if the number of solutions left is greater than 10, average fitness is computed for every ILP solution across 30 independent fitness evaluations by solving Equation 1 with $\gamma = 30$. The 10 solutions with the highest average profits are then incorporated into W .

The second phase of this initialization operator simulates (using the simulation model) every ILP solution contained in W and incorporates into the GA's initial population one simulated response $\mathbf{s}$ per each ILP solution contained in W . Those simulated responses ($\mathbf{s}$) might evidence a poor performance in terms of average profit, since they may not utilize efficiently the resources of the manufacturing system, but their standard deviations of profit should be lower than or equal to the standard deviations of the corresponding ILP solutions from which the simulated responses were obtained. This is because $s_{l,j} \leq x_{l,j}$, and thus $s_{l,j}$ is more likely to be fully realized than $x_{l,j}$.

Finally, we also incorporate into the GA's initial population the solution $\mathbf{x}^*$ obtained by optimizing Equation 7 subject to the set of constraints in the form of Equations 4 and 5, as well as the solution $\mathbf{x}'^*$ obtained by optimizing Equation 7, subject to the set of constraints in the form of Equation 5,

Equation 8 and the following:

$$\sum_{j=1}^{31} w_{l,j} \times x_{l,j} \leq \mathbb{E}[C_l] \quad (l = 1, 2, \dots, 7), \quad (10)$$

where the expected effective capacity $\mathbb{E}[C_l]$ of production line l is used as the RHS of the corresponding constraint. $\mathbb{E}[C_l]$ is computed as follows:

$$\mathbb{E}[C_l] = c_l - \frac{c_l}{t} \times \sum_{T=1}^2 \sum_{s=0}^3 p_l^{F_{T,s}} \times \mathbb{E}[\Lambda_{l,T}] \quad (l = 1, 2, \dots, 7), \quad (11)$$

where $\mathbb{E}[\Lambda_{l,T}]$ is the expected value (given in hours) of the random variable $\Lambda_{l,T}$, which models the delay caused by a repair service. $p_l^{F_{T,s}}$ is the steady state probability [7, p. 13–18] of state $F_{T,s}$, please see Diaz [19] for details about state $F_{T,s}$. $\mathbf{x}^*$ and $\mathbf{x}'^*$ are incorporated into the GA's initial population because they might contain useful alleles, especially under low levels of fitness variability.

3. Benchmark Analysis

The benchmark analysis undertaken here aims to investigate the following points:

First, we want to analyse the impact that the common practice in robust optimization of maximizing the worst case scenario [24] has on the optimization performance of the MSBO and SSBO approaches studied here. To do this we compare the solutions returned by multi-objective models that use the sample standard deviation as robustness measure (MSBO-1A, MSBO-1B) to the solutions obtained with multi-objective models that use the sample minimum instead (MSBO-2A, MSBO-2B). We also compare the solutions returned by single-objective models that maximize average profit (SSBO-1A, SSBO-1B) to the solutions obtained with single-objective models that maximize minimum profit (SSBO-2A, SSBO-2B).

Second, we intend to investigate how the allocation of limited computations between refinement of fitness values and longer optimization runs affects the optimization performance of the single and multi-objective models analysed, and whether the impact of this differs between the different formulations. To do this we compare (for every single and multi-objective

model) two sets of solutions: the ones obtained when fitness values are computed across 10 (explicit averaging strategy A) and 30 (explicit averaging strategy B) fitness evaluations.

Third, we want to investigate to what extent ESO optimization is able to generate, from a diverse final population, trade-off solutions between performance and robustness by optimizing a single measure that implicitly considers robustness, and how this compares to EMO optimization, where robustness is considered explicitly as an additional objective. To do this, we compare the solutions returned by the different single and multi-objective models.

Finally, in order to understand how performance differences evolve with increasing uncertainty levels across the different models analysed, we address the points mentioned above in two different problem instances, presented in Table 1. Table 1 shows the values assigned to p_l , ζ_l , φ_l and $\mu_{l,T}$, which are necessary to compute (as shown in Diaz [19] the transition probability matrix of each production line for both problem instances analysed in this study. p_l is the probability that a production line failure occurs (either serious or non-serious) during the manufacturing of the next product lot, given that the production line is in state O_0 (no non-serious failure has occurred after a perfect repair and the production line was operative during the manufacturing of the current product lot). ζ_l determines the probability that a serious failure occurs during the manufacturing of the next product lot, given that the production line is in state O_0 and given that a production line failure occurs during the manufacturing of the next product lot. The deterioration of a production line caused by a non-serious failure is modelled by raising p_l and ζ_l to the power of φ_l . The random variable $\Lambda_{l,T}$ is modelled by an exponential probability density function (PDF) with known mean $\mu_{l,T}$. The value of $\mu_{l,T}$ depends on the failure type of the production line, where $T = 1$ denotes a serious failure and $T = 2$ a non-serious one. Please see Diaz [19] for a detailed description of the simulation model. Please note that artificial values have been assigned to the different p_l , ζ_l , φ_l and $\mu_{l,T}$, due to the difficulty of obtaining real data for those parameters. In this manufacturing system, some products can be manufactured in several production lines (production lines 2, 3 and 4), and according to historical information provided by the company most of the failures occur on production lines 2, 3 and 4, which makes them crucial for the robustness of a production plan. For this reason only those production lines have been assigned with a positive p_l .

In order to investigate the points mentioned above, 30 independent runs

	Instance 1	Instance 2	Instance 1 and 2			
Production line (l)	p_l	p_l	$\mu_{l,1}$ (d)	$\mu_{l,2}$ (d)	ζ_l	φ_l
1	0.00	0.00	—	—	—	—
2	0.05	0.10	1.00	0.33	0.02	0.80
3	0.15	0.20	1.50	0.60	0.05	0.50
4	0.10	0.15	1.00	0.33	0.04	0.70
5	0.00	0.00	—	—	—	—
6	0.00	0.00	—	—	—	—
7	0.00	0.00	—	—	—	—

Table 1: Values of p_l , $\mu_{l,T}$, ζ_l and φ_l per production line and problem instance.

of every model are executed on both problem instances. To make a fair comparison, all models are evaluated upon a limited computational budget of 45000 independent simulation replications per run, excluding the fitness evaluations needed by the specialized initialization operator. In this sense, a run of models MSBO-1A, MSBO-2A, SSBO-1A and SSBO-2A terminates after 45 generations, whereas a run of models MSBO-1B, MSBO-2B, SSBO-1B and SSBO-2B terminates after 15 generations. The main features of every model are presented in Table 2, those parameters have been set based on preliminary experiments. All those models are implemented in MATLAB® R2014a (The MathWorks, Inc., 2014) and all computations are executed in parallel on a 12 core Intel(R) Xeon(R) CPU L5640 @ 2.27GHz with 24 GB of RAM running Scientific Linux, release 6.2.

4. Performance Assessment

To assess the “real” quality of the final population returned by the different optimizers (as compared to the “fitness estimates” used during the actual optimization process), we use a large number of simulation runs (100) to determine the profit distribution for each solution. These results are then used to identify the subsets of solutions that are non-dominated with respect to their average profit and standard deviation/minimum profit (measured in United States Dollar (USD)).

The performance assessment of different optimizers in terms of the quality of solutions is a complex issue when multiple objectives are involved [38].

	Objective 1	Objective 2	γ	Generations
MSBO-1A	max mean	min std. dev.	10	45
MSBO-2A	max mean	max minimum	10	45
MSBO-1B	max mean	min std. dev.	30	15
MSBO-2B	max mean	max minimum	30	15
SSBO-1A	max mean	—	10	45
SSBO-2A	max minimum	—	10	45
SSBO-1B	max mean	—	30	15
SSBO-2B	max minimum	—	30	15

Table 2: Objective/s, γ and number of generations across the different models analysed.

This assessment becomes even more complicated when EAs are used as optimizers, since a different set of solutions may be returned in every run of each EA (due to the randomized search mechanism of EAs) [32].

Different approaches to assess performance of multi-objective optimizers have been proposed in the existing literature, the reader is referred to Knowles et al. [32] for a comprehensive review of existing techniques. Here, in order to evaluate the points of interest mentioned in Section 3, we use the multiple sets of non-dominated solutions obtained with every model (MSBO-1A, MSBO-1B, MSBO-2A, MSBO-2B, SSBO-1A, SSBO-1B, SSBO-2A and SSBO-2B) to compute samples of values for the hypervolume indicator [22]. We then conduct Mann-Whitney U tests [36] on the samples of hypervolume values in order to derive statistical inferences about the relative performance of two models expressed in the form of the following hypotheses:

H_o : stochastic homogeneity of hypervolume values

H_a : stochastic heterogeneity of hypervolume values

In this sense, Mann-Whitney U tests tell us if one model outperforms other model in terms of the quality of approximation sets generated, under the assumption that the hypervolume indicator reflects the preference of the decision maker [32]. A non-parametric test (Mann-Whitney U test) is employed because distributions of the samples analysed do not fulfil the normality assumption.

The use of the hypervolume indicator might favour multi-objective formulations over single-objective formulations given that diversity is explicitly considered in the former models, but not in the latter ones. In order to resolve this potential issue and make a fair comparison between single and multi-

objective models, we apply the attainment function approach [10, 11]. This approach summarizes the outcome of multiple runs of the different models in the form of empirical attainment functions ¹ (EAF) [1], and thus enable us to identify the region/s where performance differences arise between single and multi-objective models, in both objective spaces analysed (*i* average profit vs. standard deviation of profit and *ii* average profit vs. minimum profit). We employ the eaf package presented in López-Ibáñez et al. [35] to create plots of the median (objective vectors attained by at least half of the runs) and best (best objective vector ever achieved) attainment surfaces as well as to illustrate the differences between EAFs obtained with different models.

5. Results

While many measures of robustness have been proposed in the existing literature [27], the sample minimum (worst case scenario for maximization problems) is particularly popular [24], see Chen et al. [6], Zhang et al. [50], Goh and Tan [25], Ong et al. [40] and Tsutsui [47] for some examples. Yet, given that the sample minimum is a biased estimator for the population [35], it is possible that its use as a robustness measure may mislead selection operators and steer an EA towards undesirable regions of the objective space. Our results provide evidence supporting the intuition that the use of less biased statistics as robustness measures reduces the computational effort needed to obtain reliable fitness estimates, thus reducing the likelihood of choosing bad solutions over good ones during the optimization and making it possible to allocate more computational effort to the optimization procedure. This can be seen in results from Mann-Whitney U tests (presented in Tables 3 and 4 for problem instance 1 and Tables A.5 and A.6 for problem instance 2 located in Appendix 8.A), which statistically show that in both objective spaces analysed (*i* average profit vs. standard deviation of profit and *ii* average profit vs. minimum profit) multi-objective models that used the sample standard deviation as robustness measure (MSBO-1A and MSBO-1B) achieved stochastically larger ($p < .01$) hypervolume values than their

¹An attainment function describes the probability of an algorithm finding at least one solution whose objective vector dominates or is equal to a specific objective vector in a single run [11, 23]. The attainment function can be approximated from the outcomes of several independent runs of an algorithm, such approximation is known as the empirical attainment function [35].

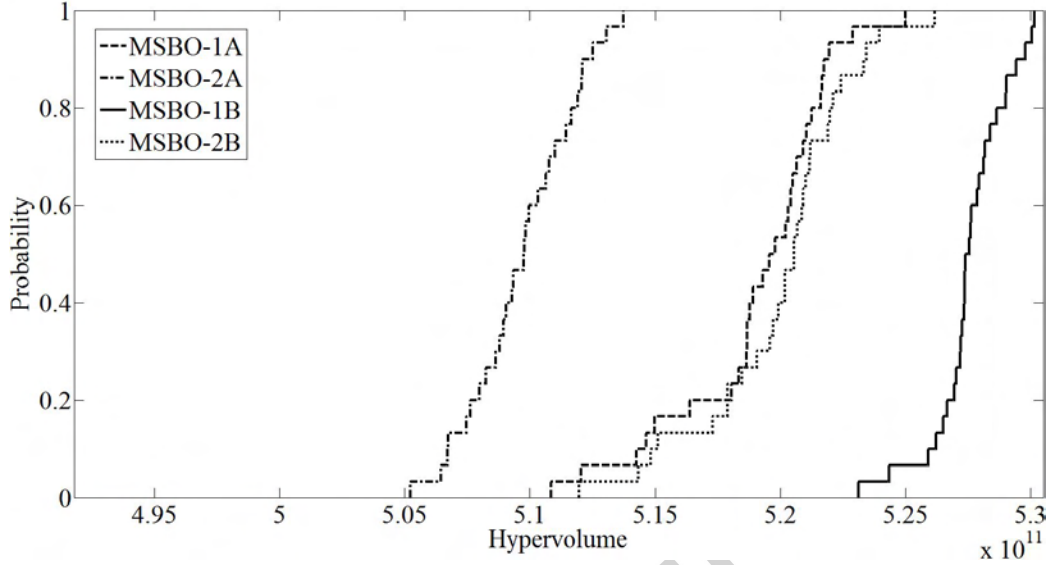


Figure 1: CDFs of hypervolume values obtained with multi-objective models (MSBO-1A, MSBO-2A, MSBO-1B, MSBO-2B) in the objective space of average profit vs. standard deviation of profit in problem instance 1.

counterparts (MSBO-2A and MSBO-2B) that used the sample minimum as robustness measure. Figures 1 and 2 also support this finding in problem instance 1, by showing that the cumulative distribution functions (CDFs) of hypervolume values obtained with MSBO-2A and MSBO-2B are dominated (first and second-order stochastic dominance) by the CDFs of hypervolume values obtained with MSBO-1A and MSBO-1B, respectively.

For the multi-objective models, our results also reveal that, given the number of fitness evaluations considered here, the allocation of more computations to obtain more reliable fitness estimates (explicit averaging strategy B) is a better strategy than optimizing across a higher number of generations using less reliable fitness estimates (explicit averaging strategy A). This finding is supported by Figures 1 and 2 for problem instance 1, which show that in both objective spaces (*i* average profit vs. standard deviation of profit and *ii* average profit vs. minimum profit) CDFs of hypervolume values obtained with multi-objective models implementing explicit averaging strategy B (MSBO-1B and MSBO-2B) dominate (first-order stochastic dominance) the CDFs of hypervolume values obtained with their correspond-

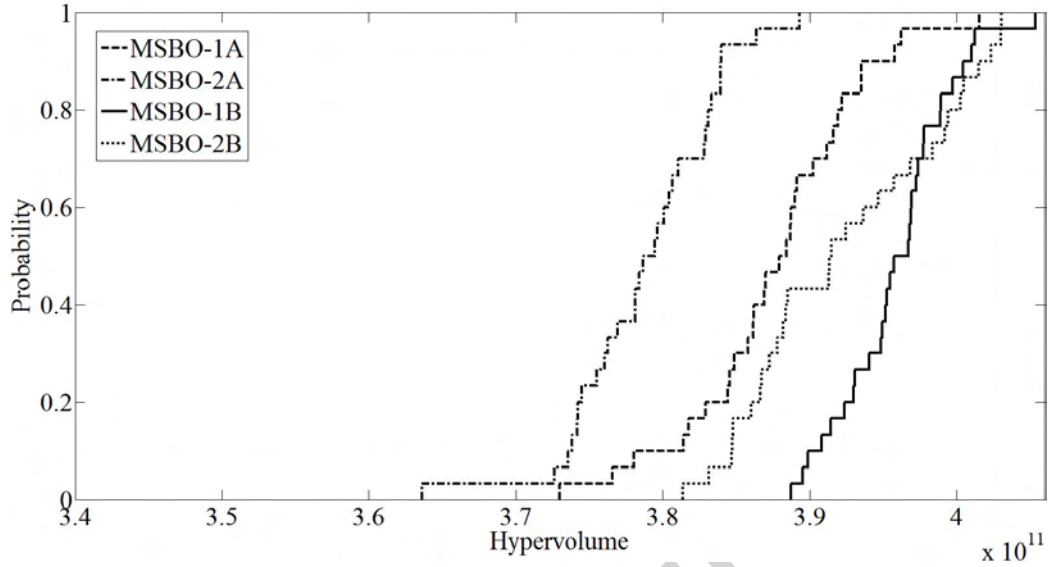
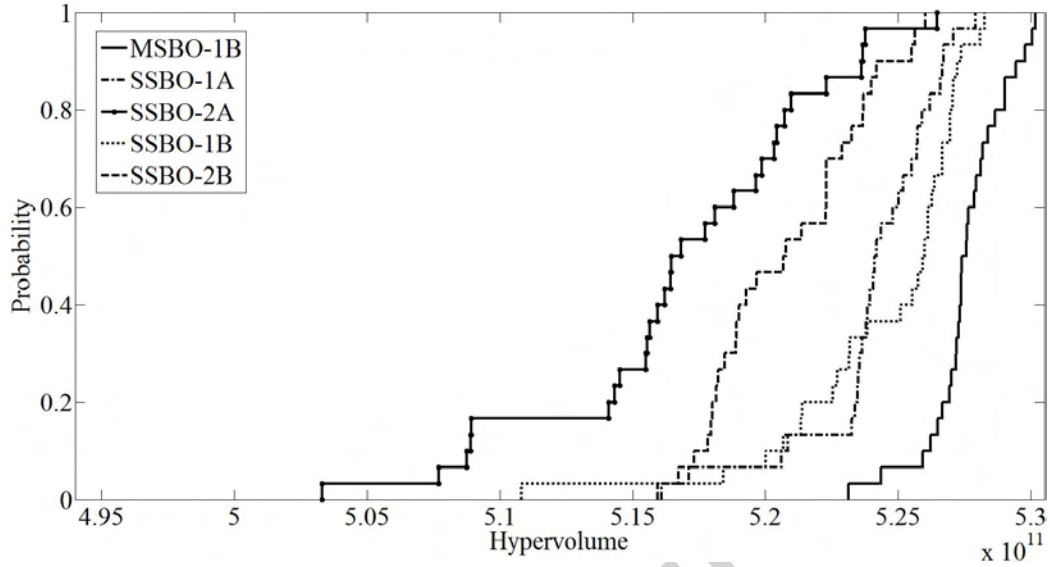


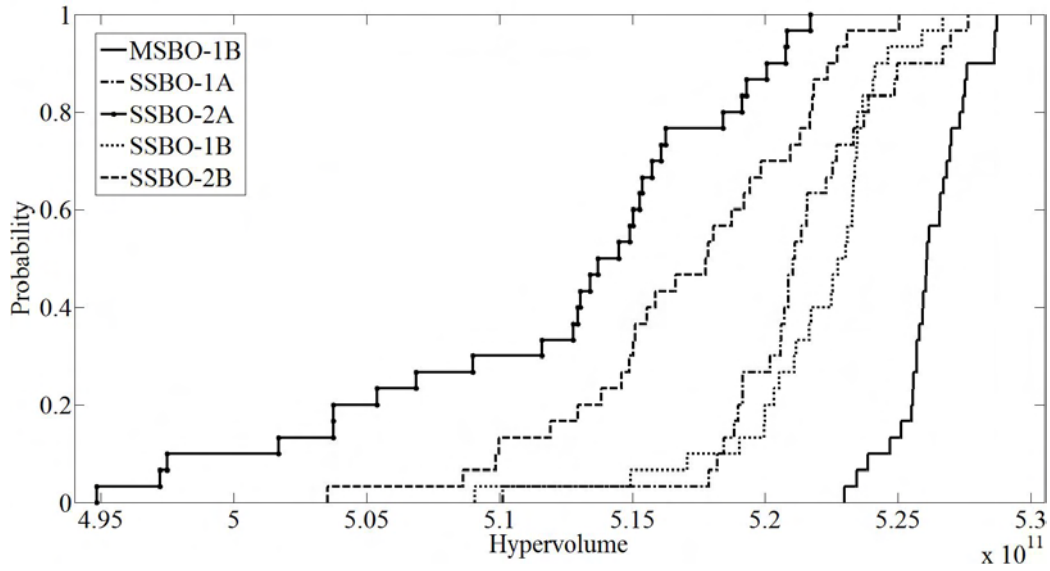
Figure 2: CDFs of hypervolume values obtained with multi-objective models (MSBO-1A, MSBO-2A, MSBO-1B, MSBO-2B) in the objective space of average profit vs. minimum profit in problem instance 1.

ing counterparts implementing strategy A (MSBO-1A and MSBO-2A). This finding is also corroborated by results from Mann-Whitney U tests, presented in Tables 3 and 4, which reveal that hypervolume values obtained with both multi-objective approaches implementing strategy B (MSBO-1B and MSBO-2B) are stochastically larger ($p < .01$) than the ones obtained with their corresponding counterparts implementing strategy A (MSBO-1A and MSBO-2A), in both objective spaces. Supportive results for problem instance 2 can be found in Figures A.9 and A.10 and in Tables A.5 and A.6 presented in Appendix 8.A.

It is important to note that in both problem instances, having more reliable fitness estimates significantly improved ($p < .01$) the optimization performance of all multi-objective models in both objective spaces analysed (*i* average profit vs. standard deviation of profit and *ii* average profit vs. minimum profit). For single-objective models in problem instance 1, the implementation of strategy B significantly improved ($p < .01$) the optimization performance of the single-objective model trying to maximize the sample minimum, but only in the solution space of average profit vs. standard de-

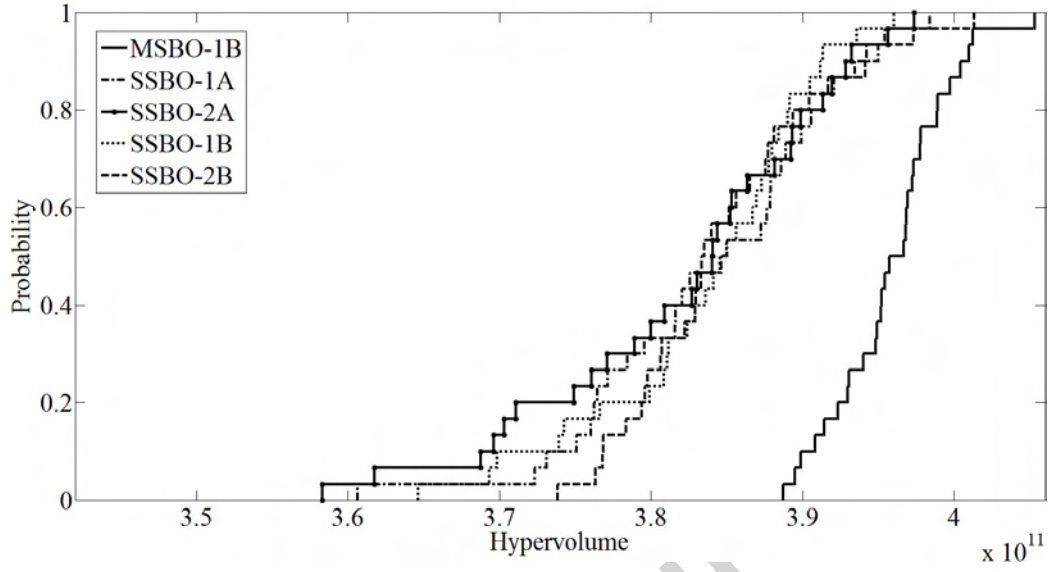


(a) Instance 1

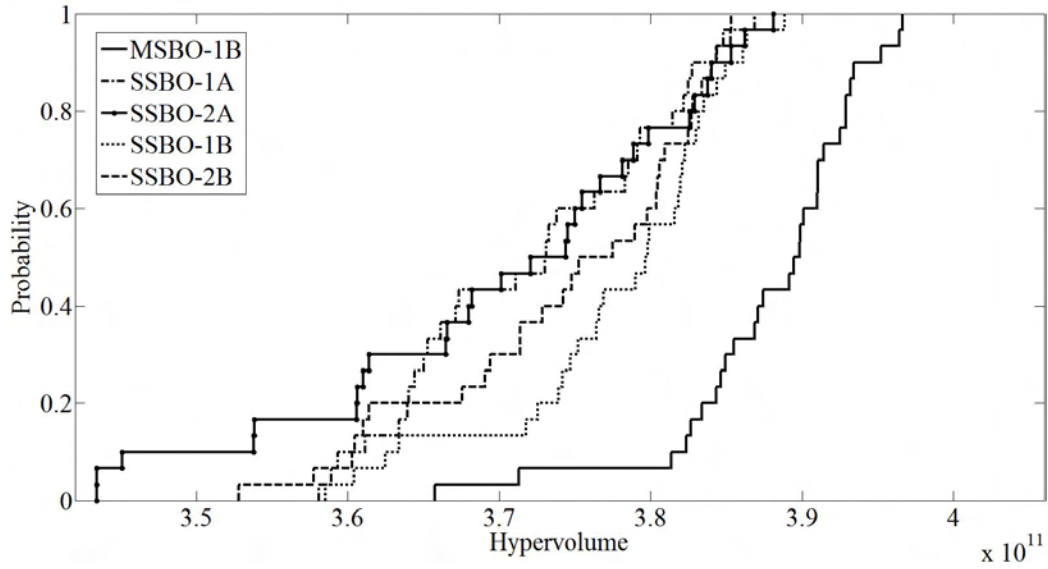


(b) Instance 2

Figure 3: CDFs of hypervolume values obtained with the best multi-objective model (MSBO-1B) and all single-objective models (SSBO-1A, SSBO-2A, SSBO-1B and SSBO-2B) in the objective space of average profit vs. standard deviation of profit.



(a) Instance 1



(b) Instance 2

Figure 4: CDFs of hypervolume values obtained with the best multi-objective model (MSBO-1B) and all single-objective models (SSBO-1A, SSBO-2A, SSBO-1B and SSBO-2B) in the objective space of average profit vs. minimum profit.

viation (see Table 3 and A.5, where SSBO-2B achieved stochastically larger hypervolume values ($p < .01$) than SSBO-2A), whereas under higher uncertainty level (problem instance 2), strategy B improved the optimization performance of the model mentioned above not only in the solution space of average profit vs. standard deviation, but also in the solution space of average profit vs. minimum profit (see Figures 3b and 4b, where the CDF of hypervolume values obtained with SSBO-2B dominate the CDFs of hypervolume values obtained with SSBO-2A - first and second-order stochastic dominance, respectively). These results are interesting because they provide evidence that using less biased statistics as objective/s might reduce the computational effort (number of samples) needed to obtain reliable fitness estimates, and thus demonstrate that not only the level of uncertainty, but also the choice of objective/s have an impact on the noise sensitivity of the optimization approach. Our results also seem to suggest that the multi-objective approach is more sensitive to noise than the single-objective approach, but further investigation is still required to support this claim. This may occur because in the single-objective case the solution selection during the optimization procedure only considers one noisy fitness estimate, whereas in the multi-objective case the noise affecting the optimization comes from two different noisy fitness estimates, which makes the solution selection more difficult. Although we cannot claim that multi-objective approaches are in general more noise sensitive than single-objective approaches, as this will highly depend on the choice of objective/s, we can say that our experiments demonstrate that for the multi-objective models considered here, allocating more computational effort towards the generation of more reliable fitness estimates seems to be more beneficial than optimizing across more generations.

Our results identified MSBO-1B as the best multi-objective model in both objective spaces analysed in problem instance 1 and 2. In the solution space of average profit vs. standard deviation of profit, this finding is confirmed by results from Mann-Whitney U tests (Tables 3 and A.5) and by analysing Figures 1 and A.9, which indicate that MSBO-1B outperformed all its multi-objective contestants (MSBO-1A, MSBO-2A and MSBO-2B). Furthermore, hypervolume values from non-dominated solutions found with MSBO-1B in the solution space of average profit vs. minimum profit were stochastically larger ($p < .01$) (see Tables 4 and A.6) than hypervolume values obtained with any other multi-objective model (MSBO-1A, MSBO-2A and MSBO-2B), in both problem instances. For problem instance 1, these

results are corroborated by Figure 2, where the CDF of hypervolume values obtained with MSBO-1B dominates the CDFs obtained with MSBO-1A (first-order stochastic dominance), MSBO-2A (first-order stochastic dominance) and MSBO-2B (second-order stochastic dominance).

It is not surprising that MSBO-1B outperforms both, MSBO-2A and MSBO-2B, in the solution space of average profit vs. standard deviation of profit, since MSBO-1B explicitly optimizes both criteria. However, it is surprising that MSBO-1B also outperforms MSBO-2A and MSBO-2B in the solution space of average profit vs. minimum profit, considering that maximization of minimum profit is an objective in MSBO-2A and MSBO-2B, but not in MSBO-1B. This counter-intuitive result further supports our finding that the use of the sample minimum as robustness measure may undermine the optimization performance of the EAs analysed, instead of returning more robust solutions, as reported in Ong et al. [40] and Tsutsui [47], compared to less biased statistics such as the sample mean or the sample standard deviation. Having a more biased statistic (sample minimum) as one of the two noisy fitness estimates may be the reason why MSBO-2B was outperformed in both objective spaces (based on CDFs and results from Mann-Whitney U tests) by MSBO-1B, which uses a less biased statistic (sample standard deviation) as fitness estimate. It is probably due to the less reliable fitness estimates that MSBO-1A performed badly (based on CDFs and results from Mann-Whitney U tests) in both objective spaces in terms of hypervolume values compared to MSBO-1B, which has more refined fitness estimates.

Our results also demonstrate that the simultaneous optimization of average profit as a performance measure and standard deviation of profit as a robustness measure (MSBO-1B) consistently generated better sets of trade-off solutions not only between average profit and standard deviation of profit, but also between average profit and minimum profit, than any other model analysed here. The supporting evidence for this finding is provided by results from Mann-Whitney U tests (Tables 3 and 4 for problem instance 1 and Tables A.5 and A.6 for problem instance 2), which determined that MSBO-1B achieved stochastically larger ($p < .01$) hypervolume values than any of the single-objective models in both objective spaces analysed. The superiority of MSBO-1B is also illustrated in Figures 3 and 4, where the CDFs of hypervolume values obtained with MSBO-1B dominate (first-order stochastic dominance) all the CDFs obtained with the single-objective models (SSBO-1A, SSBO-2A, SSBO-1B and SSBO-2B).

Finally, the median and best attainment surfaces, presented in Figures 5, 6, 7

	MSBO-1A	MSBO-1B	MSBO-2A	MSBO-2B	SSBO-1A	SSBO-1B	SSBO-2A	SSBO-2B
MSBO-1A	—	2**	14**	370	82**	94**	331	337
MSBO-1B	—	—	0**	6**	74**	135**	7**	12**
MSBO-2A	—	—	—	5**	0**	9**	115**	0**
MSBO-2B	—	—	—	—	111**	125**	280*	395
SSBO-1A	—	—	—	—	—	364 ^a	83**	174**
SSBO-1B	—	—	—	—	—	—	89**	157**
SSBO-2A	—	—	—	—	—	—	—	233**
SSBO-2B	—	—	—	—	—	—	—	—

* $p < .05$; ** $p < .01$; ^a heteroscedasticity according to non-parametric Levene test ($p < .05$) (Nordstokke and Zumbo, 2010)

Table 3: Values for Mann-Whitney U statistic obtained from pairwise comparisons of hypervolume values of all single and multi-objective models in the objective space of average profit vs. standard deviation of profit in problem instance 1.

	MSBO-1A	MSBO-1B	MSBO-2A	MSBO-2B	SSBO-1A	SSBO-1B	SSBO-2A	SSBO-2B
MSBO-1A	—	99**	112**	294*	334	308*	299*	314*
MSBO-1B	—	—	1**	310 ^a *	71**	36**	62**	76**
MSBO-2A	—	—	—	30**	267 ^a **	235**	307 ^a *	215**
MSBO-2B	—	—	—	—	209**	178**	178**	180**
SSBO-1A	—	—	—	—	—	434	413	437
SSBO-1B	—	—	—	—	—	—	426	448
SSBO-2A	—	—	—	—	—	—	—	412
SSBO-2B	—	—	—	—	—	—	—	—

** $p < .01$; ^a heteroscedasticity according to non-parametric Levene test ($p < .05$)

Table 4: Values for Mann-Whitney U statistic obtained from pairwise comparisons of hypervolume values of all single and multi-objective models in the objective space of average profit vs. minimum profit in problem instance 1.

and 8, illustrate that in problem instance 1 and 2, respectively, the regions attained in both objective spaces with the best multi-objective model (MSBO-1B) are more attractive than the ones attained with the different single-objective models (SSBO-1A, SSBO-2A, SSBO-1B and SSBO-2B). The advantage of MSBO-1B over single-objective models is also confirmed by looking at the shaded areas in those figures. Shaded areas point out the regions of the objective space where the EAF of one model is larger by at least 20% than the EAF of the other model, and thus dark regions indicate the location in the objective space where one model outperforms the other. Darker areas indicate larger differences between the estimated probability values of two models.

This finding is somewhat surprising when considering performance for one of the extremes of the Pareto front (e.g. high average profit, high minimum profit): as the single-objective optimizer is focusing on the optimization of a specific criterion, we may expect it to outperform the multi-objective optimizer in the corresponding region of the front. However, this does not occur. This appears to indicate that the simultaneous consideration of multiple objectives allows for a more comprehensive description (and recognition) of good solutions during the optimization: the solutions identified by the multi-objective approach “stand up” to the scrutiny of 100 simulations during final performance assessment, while the single-objective solutions “drop” in performance.

6. Conclusion

We have considered interactions between the choice of optimization paradigm (multi-objective versus single-objective), robustness criterion and sample size in the context of a production planning problem that is subject to machine failures. This problem is an example of a much larger set of logistics and engineering design problems in which the performance of a solution is subject to uncertainty in the operating or environmental conditions [34, 33, 30].

As may be expected, our results suggest that a dedicated multi-objective approach is able to consistently generate sets of solutions that present better trade-offs between performance and robustness than a single-objective approach. More surprisingly though, we observe that the single-objective models do not outperform the multi-objective ones in extreme regions of the Pareto front, indicating that the multi-objective approach is able to benefit from a more comprehensive description of solution quality.

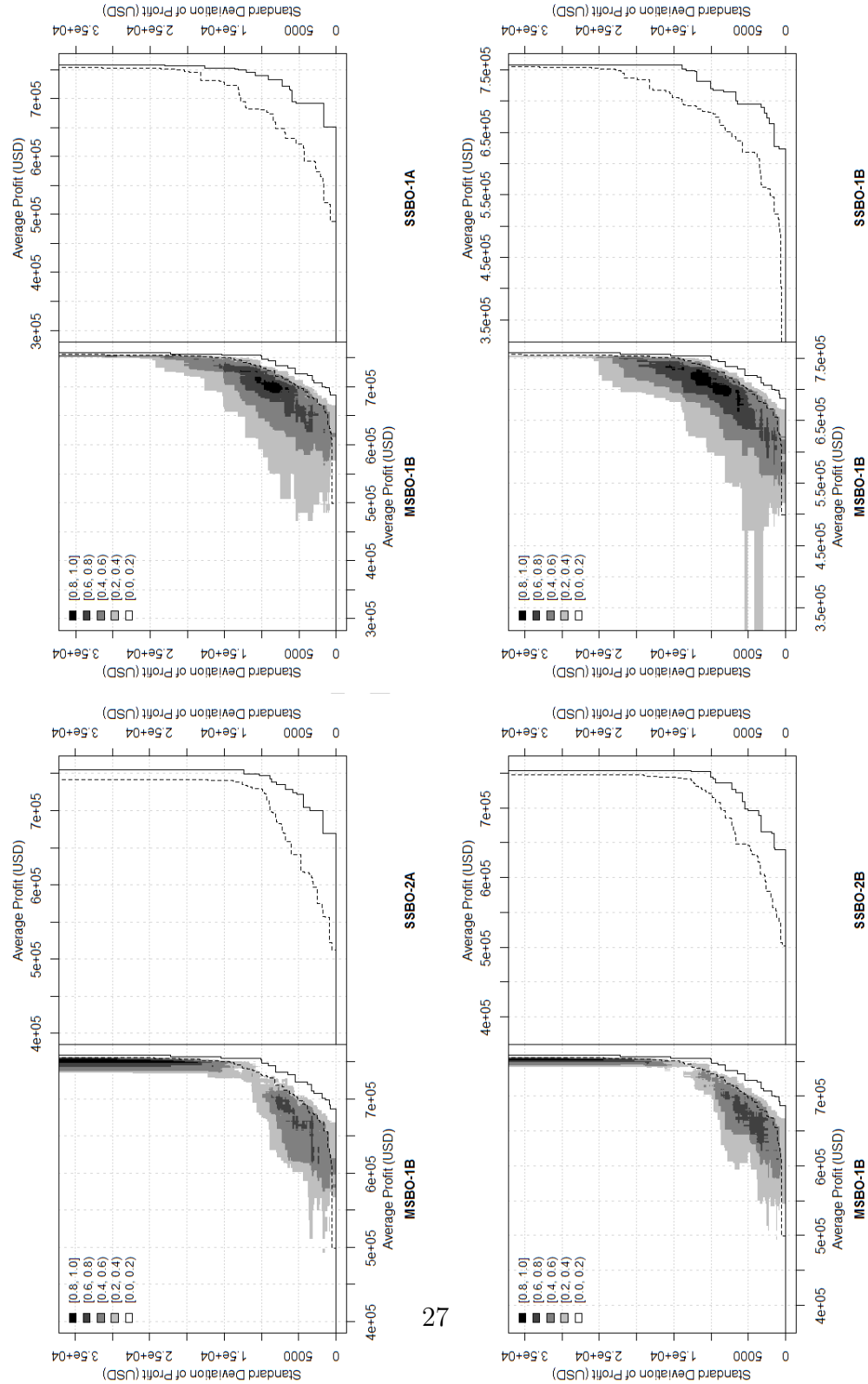


Figure 5: Median (dashed line) and best (continuous line) attainment surfaces of the best multi-objective model and all single-objective models in the objective space of average profit vs. standard deviation of profit in problem instance 1.

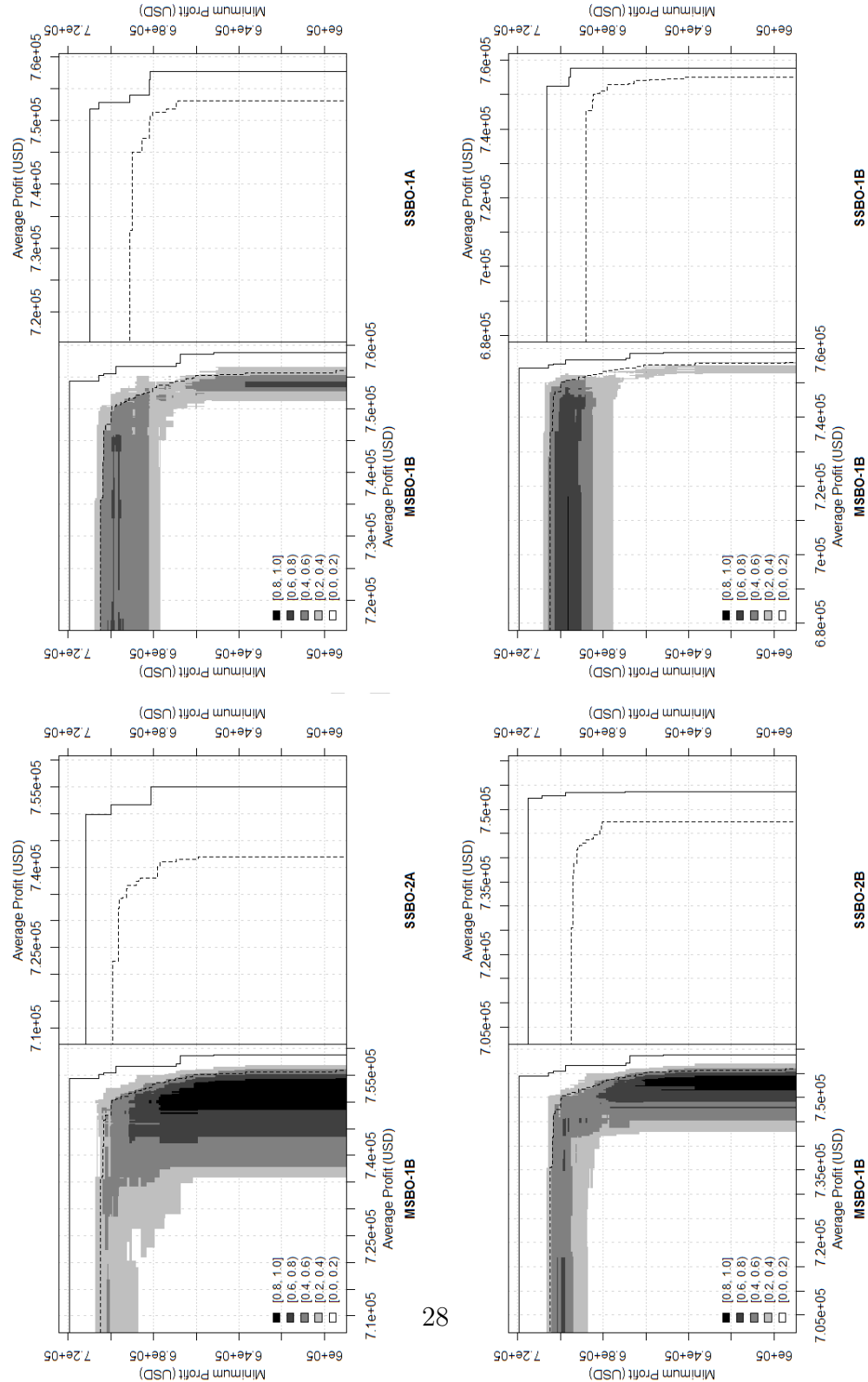


Figure 6: Median (dashed line) and best (continuous line) attainment surfaces of the best multi-objective model and all single-objective models in the objective space of average profit vs. minimum profit in problem instance 1.

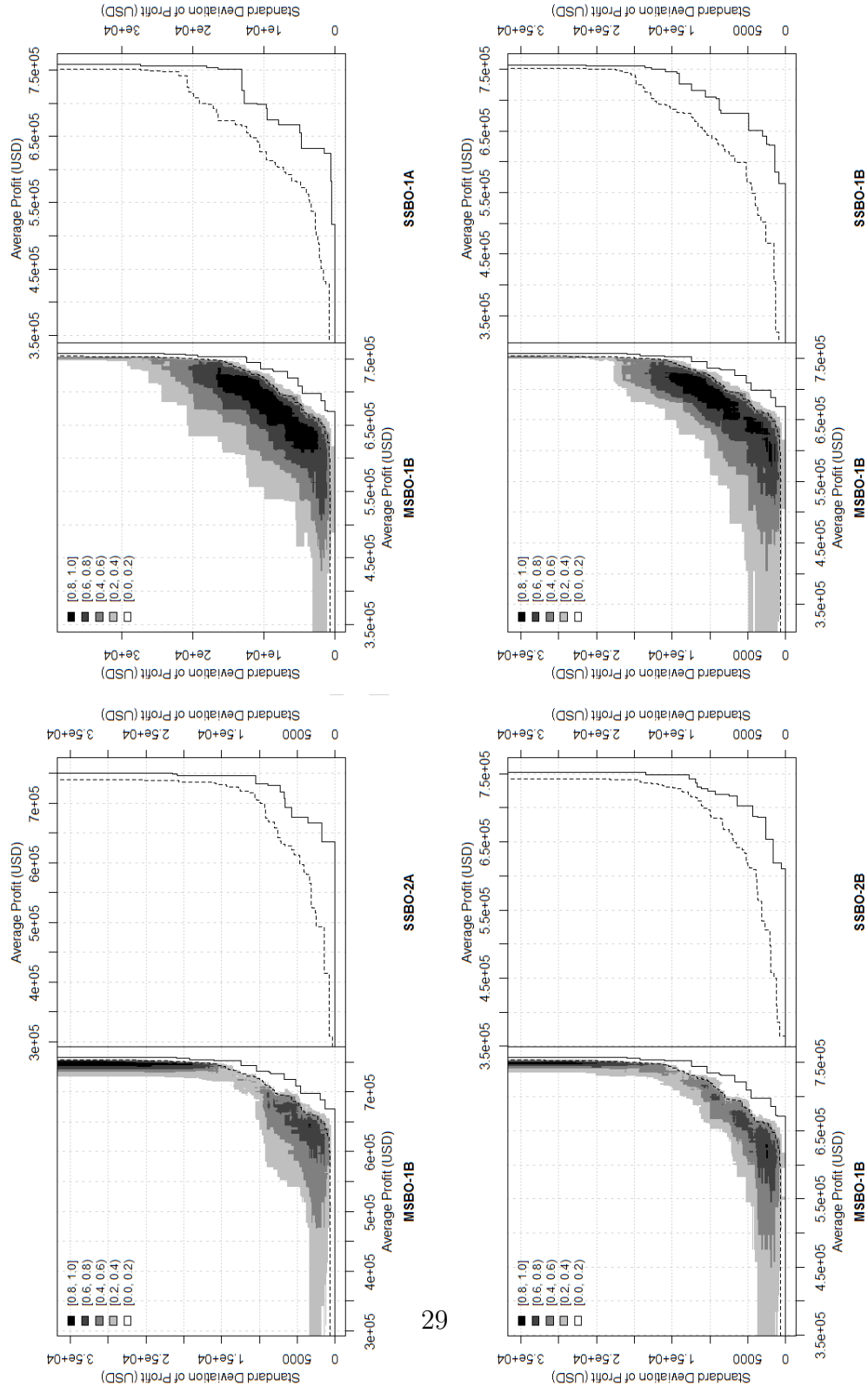


Figure 7: Median (dashed line) and best (continuous line) attainment surfaces of the best multi-objective model and all single-objective models in the objective space of average profit vs. standard deviation of profit in problem instance 2.

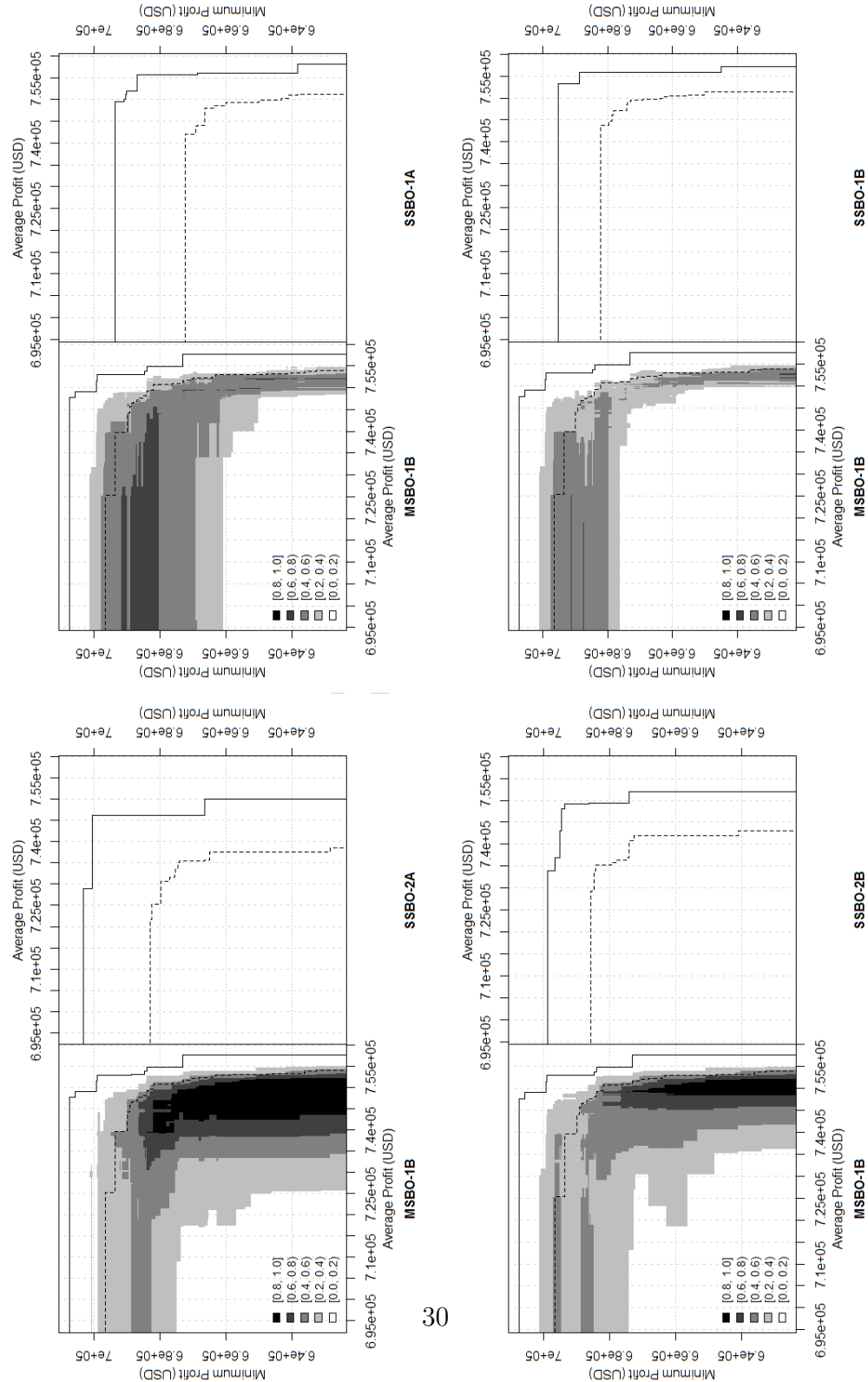


Figure 8: Median (dashed line) and best (continuous line) attainment surfaces of the best multi-objective model and all single-objective models in the objective space of average profit vs. minimum profit in problem instance 2.

We further establish that optimization performance for this problem is significantly affected by the choice of robustness measure. In particular, the use of the sample minimum as the robustness measure negatively impacts on the quality of the final solutions, for both single-objective and multi-objective optimizers. In conjunction with our experiments with sample size, we find that a larger number of evaluation replicates (i.e. more computational effort) is needed to obtain reliable estimates for this measure, compared to less biased statistics such as the sample mean or the sample standard deviation.

While our results for the multi-objective optimizers are generally positive, we do identify a sensitivity of this approach to the number of evaluation replicates employed in estimating fitness values. In particular, our results show that an increase in sample size (from 10 to 30) leads to a distinct performance improvement for all multi-objective models, while the effect is much less pronounced for the comparable single-objective models. These results highlight the importance of carefully adjusting this parameter in multi-objective formulations for similar problems, and we speculate that this may be due to the simultaneous impact of noise in two objectives.

7. Acknowledgement

Juan Esteban Diaz expresses his gratitude to the Secretariat for Higher Education, Science and Technology of Ecuador, who has supported him with a doctoral scholarship.

8. References

- [1] Binois, M., Ginsbourger, D., Roustant, O., 2015. Quantifying uncertainty on pareto fronts with gaussian process conditional simulations. *European Journal of Operational Research* 243 (2), 386–394.
- [2] Branke, J., 1998. Creating robust solutions by means of evolutionary algorithms. In: *Parallel Problem Solving from Nature PPSN V*. Springer, pp. 119–128.
- [3] Branke, J., Corrente, S., Greco, S., Słowiński, R., Zielniewicz, P., 2016. Using choquet integral as preference model in interactive evolutionary multiobjective optimization. *European Journal of Operational Research* 250 (3), 884–901.

- [4] Branke, J., Lu, K., 2015. Finding the trade-off between robustness and worst-case quality. In: Proceedings of the 2015 on Genetic and Evolutionary Computation Conference. ACM, pp. 623–630.
- [5] Brownlee, A. E., Wright, J. A., 2015. Constrained, mixed-integer and multi-objective optimisation of building designs by nsga-ii with fitness approximation. *Applied Soft Computing* 33 (0), 114–126.
- [6] Chen, L., Gendreau, M., Hà, M. H., Langevin, A., 2016. A robust optimization approach for the road network daily maintenance routing problem with uncertain service time. *Transportation Research Part E: Logistics and Transportation Review* 85, 40–51.
- [7] Ching, W. K., Huang, X., Ng, M. K., Siu, T.-K., 2006. Markov chains. Springer.
- [8] Coello, C. A. C., 2006. Evolutionary multi-objective optimization: a historical view of the field. *Computational Intelligence Magazine, IEEE* 1 (1), 28–36.
- [9] Coello, C. A. C., 2015. Multi-objective evolutionary algorithms in real-world applications: Some recent results and current challenges. In: *Advances in Evolutionary and Deterministic Methods for Design, Optimization and Control in Engineering and Sciences*. Springer, pp. 3–18.
- [10] da Fonseca, V. G., Fonseca, C. M., 2010. The attainment-function approach to stochastic multiobjective optimizer assessment and comparison. In: *Experimental methods for the analysis of optimization algorithms*. Springer, pp. 103–130.
- [11] da Fonseca, V. G., Fonseca, C. M., Hall, A. O., 2001. Inferential performance assessment of stochastic optimisers and the attainment function. In: *Evolutionary multi-criterion optimization*. Springer, pp. 213–225.
- [12] Deb, K., 2000. An efficient constraint handling method for genetic algorithms. *Computer methods in applied mechanics and engineering* 186 (2), 311–338.
- [13] Deb, K., 2014. Multi-objective optimization. In: *Search methodologies*. Springer, pp. 403–449.

- [14] Deb, K., Gupta, H., 2006. Introducing robustness in multi-objective optimization. *Evolutionary Computation* 14 (4), 463–494.
- [15] Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., 2002. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *Evolutionary Computation, IEEE Transactions on* 6 (2), 182–197.
- [16] Deep, K., Singh, K. P., Kansal, M., Mohan, C., 2009. A real coded genetic algorithm for solving integer and mixed integer optimization problems. *Applied Mathematics and Computation* 212 (2), 505–518.
- [17] Deep, K., Thakur, M., 2007. A new crossover operator for real coded genetic algorithms. *Applied Mathematics and Computation* 188 (1), 895–911.
- [18] Deep, K., Thakur, M., 2007. A new mutation operator for real coded genetic algorithms. *Applied mathematics and Computation* 193 (1), 211–230.
- [19] Diaz, J. E., 2016. Simulation-based optimization for production planning: Integrating meta-heuristics, simulation and exact techniques to address the uncertainty and complexity of manufacturing systems. Ph.D. thesis, The University of Manchester.
- [20] Ehrenberg, C., Zimmermann, J., 2012. Simulation-based optimization in make-to-order production: scheduling for a special-purpose glass manufacturer. In: *Simulation Conference (WSC), Proceedings of the 2012 Winter*. IEEE, pp. 1–12.
- [21] Figueira, G., Almada-Lobo, B., 2014. Hybrid simulation–optimization methods: A taxonomy and discussion. *Simulation Modelling Practice and Theory* 46, 118–134.
- [22] Fleischer, M., 2003. The measure of pareto optima applications to multi-objective metaheuristics. In: *Evolutionary multi-criterion optimization*. Springer, pp. 519–533.
- [23] Fonseca, C. M., Fleming, P. J., 1996. On the performance assessment and comparison of stochastic multiobjective optimizers. In: *Parallel problem solving from nature ppsn iv*. Springer, pp. 584–593.

- [24] Gabrel, V., Murat, C., Thiele, A., 2014. Recent advances in robust optimization: An overview. *European Journal of Operational Research* 235 (3), 471–483.
- [25] Goh, C. K., Tan, K. C., 2007. Evolving the tradeoffs between pareto-optimality and robustness in multi-objective evolutionary algorithms. In: *Evolutionary computation in dynamic and uncertain environments*. Springer, pp. 457–478.
- [26] Goh, C.-K., Tan, K. C., 2009. Evolutionary multi-objective optimization in uncertain environments: issues and algorithms. Vol. 186. Springer.
- [27] Goh, C. K., Tan, K. C., Cheong, C. Y., Ong, Y.-S., 2010. An investigation on noise-induced features in robust evolutionary multi-objective optimization. *Expert Systems with Applications* 37 (8), 5960–5980.
- [28] Guo, Z., Wong, W., Leung, S., Fan, J., Chan, S., March 2008. A genetic-algorithm-based optimization model for solving the flexible assembly line balancing problem with work sharing and workstation revisiting. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on* 38 (2), 218–228.
- [29] Handl, J., Kell, D. B., Knowles, J., 2007. Multiobjective optimization in bioinformatics and computational biology. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)* 4 (2), 279–292.
- [30] Jin, Y., Branke, J., 2005. Evolutionary optimization in uncertain environments-a survey. *Evolutionary Computation, IEEE Transactions on* 9 (3), 303–317.
- [31] Jin, Y., Sendhoff, B., 2003. Trade-off between performance and robustness: an evolutionary multiobjective approach. In: *Evolutionary Multi-Criterion Optimization*. Springer, pp. 237–251.
- [32] Knowles, J., Thiele, L., Zitzler, E., 2006. A tutorial on the performance assessment of stochastic multiobjective optimizers. *Tik report* 214, 327–332.
- [33] Lim, D., Ong, Y.-S., Jin, Y., Sendhoff, B., Lee, B. S., 2006. Inverse multi-objective robust evolutionary design. *Genetic Programming and Evolvable Machines* 7 (4), 383–404.

- [34] Lim, D., Ong, Y.-S., Lim, M.-H., Jin, Y., 2007. Single/multi-objective inverse robust evolutionary design methodology in the presence of uncertainty. In: *Evolutionary Computation in Dynamic and Uncertain Environments*. Springer, pp. 437–456.
- [35] López-Ibáñez, M., Paquete, L., Stützle, T., 2010. Exploratory analysis of stochastic local search algorithms in biobjective optimization. In: *Experimental methods for the analysis of optimization algorithms*. Springer, pp. 209–222.
- [36] Mann, H. B., Whitney, D. R., 1947. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, 50–60.
- [37] Marler, R. T., Arora, J. S., 2010. The weighted sum method for multi-objective optimization: new insights. *Structural and multidisciplinary optimization* 41 (6), 853–862.
- [38] Minella, G., Ruiz, R., Ciavotta, M., 2011. Restarted iterated pareto greedy algorithm for multi-objective flowshop scheduling problems. *Computers & Operations Research* 38 (11), 1521–1533.
- [39] Mlakar, M., Petelin, D., Tuar, T., Filipi, B., 2015. Gp-demo: Differential evolution for multiobjective optimization based on gaussian process models. *European Journal of Operational Research* 243 (2), 347–361.
- [40] Ong, Y.-S., Nair, P. B., Lum, K., 2006. Max-min surrogate-assisted evolutionary algorithm for robust design. *Evolutionary Computation, IEEE Transactions on* 10 (4), 392–404.
- [41] Paenke, I., Branke, J., Jin, Y., 2006. Efficient search for robust solutions by means of evolutionary algorithms and fitness approximation. *Evolutionary Computation, IEEE Transactions on* 10 (4), 405–420.
- [42] Pereira, V., Sousa, P., Cortez, P., Rio, M., Rocha, M., 2015. Comparison of single and multi-objective evolutionary algorithms for robust link-state routing. In: *Evolutionary Multi-Criterion Optimization*. Springer, pp. 573–587.

- [43] Ray, T., 2002. Constrained robust optimal design using a multiobjective evolutionary algorithm. In: Proceedings of the 2002 Congress on Evolutionary Computation. Vol. 1. IEEE, pp. 419–424.
- [44] Roederkerk, R. P., Van Heerde, H. J., 2016. Robust optimization of the 0–1 knapsack problem: Balancing risk and return in assortment optimization. *European Journal of Operational Research* 250 (3), 842–854.
- [45] Sanchez, D., Amodeo, L., Prins, C., 2010. Meta-heuristic approaches for multi-objective simulation-based optimization in supply chain inventory management. In: *Artificial Intelligence Techniques for Networked Manufacturing Enterprises Management*. Springer, pp. 249–269.
- [46] Syberfeldt, A., Ng, A., John, R. I., Moore, P., 2010. Evolutionary optimisation of noisy multi-objective problems using confidence-based dynamic resampling. *European Journal of Operational Research* 204 (3), 533–544.
- [47] Tsutsui, S., 1999. A comparative study on the effects of adding perturbations to phenotypic parameters in genetic algorithms with a robust solution searching scheme. In: *Systems, Man, and Cybernetics, 1999. IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on*. Vol. 3. IEEE, pp. 585–591.
- [48] Tsutsui, S., Ghosh, A., 1997. Genetic algorithms with a robust solution searching scheme. *Evolutionary Computation, IEEE Transactions on* 1 (3), 201–208.
- [49] Wang, Z., Guo, J., Zheng, M., Wang, Y., 2015. Uncertain multiobjective traveling salesman problem. *European Journal of Operational Research* 241 (2), 478–489.
- [50] Zhang, Q., Morari, M. F., Grossmann, I. E., Sundaramoorthy, A., Pinto, J. M., 2016. An adjustable robust optimization approach to scheduling of continuous industrial processes providing interruptible load. *Computers & Chemical Engineering* 86, 106–119.

Appendix A. Results from Problem Instance 2

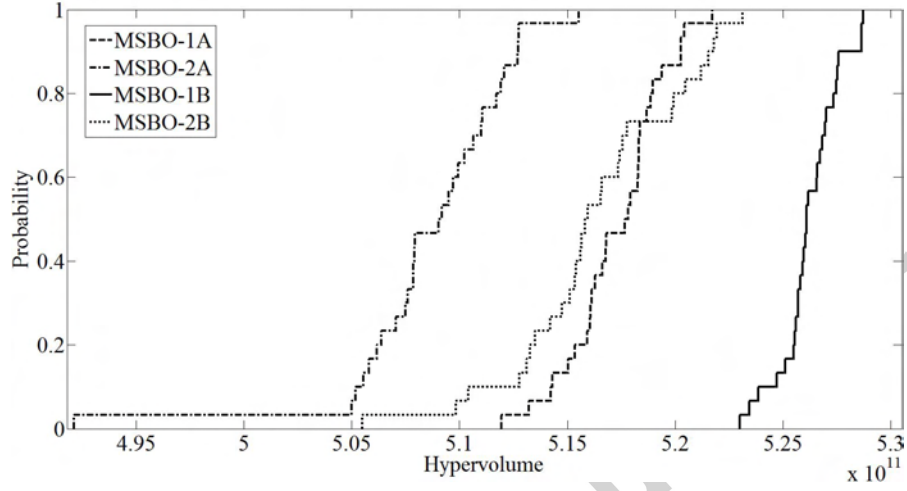


Figure A.9: CDFs of hypervolume values obtained with multi-objective models (MSBO-1A, MSBO-2A, MSBO-1B, MSBO-2B) in the objective space of average profit vs. standard deviation of profit in problem instance 2.

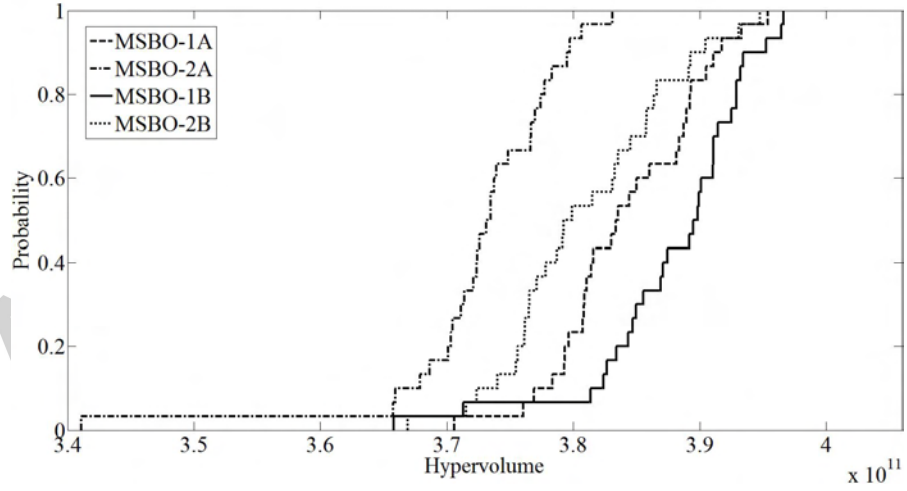


Figure A.10: CDFs of hypervolume values obtained with multi-objective models (MSBO-1A, MSBO-2A, MSBO-1B, MSBO-2B) in the objective space of average profit vs. minimum profit in problem instance 2.

	MSBO-1A	MSBO-1B	MSBO-2A	MSBO-2B	SSBO-1A	SSBO-1B	SSBO-2A	SSBO-2B
MSBO-1A	—	0**	10**	362	101**	91**	240 ^{a**}	440 ^a
MSBO-1B	—	—	0**	1**	82**	53**	0**	5**
MSBO-2A	—	—	—	58**	11**	16**	257 ^{a**}	80**
MSBO-2B	—	—	—	—	138**	114**	290**	397
SSBO-1A	—	—	—	—	—	372	82**	220**
SSBO-1B	—	—	—	—	—	—	64**	164**
SSBO-2A	—	—	—	—	—	—	—	265**
SSBO-2B	—	—	—	—	—	—	—	—

* $p < .05$; ** $p < .01$; ^a heteroscedasticity according to non-parametric Levene test ($p < .05$)

Table A.5: Values for Mann-Whitney U statistic obtained from pairwise comparisons of hypervolume values of all single and multi-objective models in the objective space of average profit vs. standard deviation of profit in problem instance 2.

	MSBO-1A	MSBO-1B	MSBO-2A	MSBO-2B	SSBO-1A	SSBO-1B	SSBO-2A	SSBO-2B
MSBO-1A	—	264**	62**	313**	127**	238**	135**	173**
MSBO-1B	—	—	52**	194**	62**	105**	76**	79**
MSBO-2A	—	—	—	164**	419 ^a	241**	427 ^a	361 ^a
MSBO-2B	—	—	—	—	209**	360	215**	275**
SSBO-1A	—	—	—	—	—	295**	437	380
SSBO-1B	—	—	—	—	—	—	300**	363
SSBO-2A	—	—	—	—	—	—	—	376
SSBO-2B	—	—	—	—	—	—	—	—

** $p < .01$; ^a heteroscedasticity according to non-parametric Levene test ($p < .05$)

Table A.6: Values for Mann-Whitney U statistic obtained from pairwise comparisons of hypervolume values of all single and multi-objective models in the objective space of average profit vs. minimum profit in problem instance 2.