

# Implicit and Explicit Averaging Strategies for Simulation-Based Optimization of a Real-World Production Planning Problem

Juan Esteban Diaz and Julia Handl

Manchester Business School, The University of Manchester, Booth Street West

M15 6PB Manchester, United Kingdom

E-mail: juan.diaz@postgrad.mbs.ac.uk, j.handl@manchester.ac.uk

**Keywords:** discrete event simulation, failure-prone manufacturing, genetic algorithms, noise handling, production planning, simulation-based optimization, uncertainty

**Received:** December 1, 2014

*In this study, we explore the impact of noise handling strategies on optimization performance in the context of a real-world production planning problem. Uncertainties intrinsic to the production system are captured using a discrete event simulation (DES) model, and the production plan is optimized using an evolutionary algorithm. The stochastic nature of the fitness values (as returned by the DES simulation) may impact on optimization performance, and we explore explicit and implicit averaging strategies to address this issue. Specifically, we evaluate the effectiveness of different strategies, when a limited budget of evaluations is available. Our results indicate a general advantage of implicit averaging in this setting, and a good degree of robustness with regard to population size. On the other hand, explicit averaging is found to be non-competitive, due to the cost of repeat-evaluations of the same solution. Finally, we explore a hybrid approach that uses explicit averaging to refine fitness estimates during final solution selection. Under increasing levels of fitness variability, this hybrid strategy starts to outperform pure implicit and explicit averaging strategies.*

*Povzetek: V študiji smo raziskali vpliv strategij za ravnanje s šumom na uspešnost optimizacije v okviru realnega problema načrtovanja proizvodnje. Negotovosti, ki se pojavljajo v proizvodnem sistemu so bile zajete z modelom simulacije diskretnih dogodkov (DES), proizvodni načrt pa je bil optimiran z uporabo evolucijskega algoritma. Ker stohastična narava vrednosti kriterijske funkcije (kot jo vrača DES) lahko vpliva na uspešnost optimizacije, smo raziskali eksplicitne in implicitne strategije povprečenja za reševanje tega problema. Natančneje, oceniti smo učinkovitost različnih strategij v primerih, ko je na voljo omejeno število ocenitev kriterijske funkcije. Rezultati v splošnem kažejo na prednost implicitnega povprečenja in dobro stopnjo robustnosti glede na velikost populacije. Po drugi strani pa smo za eksplicitno povprečenje ugotovili, da ni konkurenčno zaradi stroškov večkratnih ovrednotenj iste rešitve. Končno, raziskali smo hibridni pristop, ki uporablja eksplicitno povprečenje za izpopolnitev ocene kriterijske funkcije pri končni izbiri rešitev. S povečano stopnjo spremenljivosti kriterijske funkcije začne hibridna strategija prekašati čisti, eksplicitni in implicitni strategiji.*

## 1 Introduction

Optimization problems that include uncertainty pose challenges that are difficult to address using standard optimization methodologies. While a portion of the optimization literature is concerned with the development of methodologies capable of identifying optimal solutions to problems with uncertainty, the application of these methods often requires stringent assumptions and / or simplifications that are necessary to satisfy relevant optimality conditions. Those methods are often insufficiently powerful to accurately incorporate the full complexity and uncertainty intrinsic to real-world problems into the problem formulation, even when their consideration is essential for the generation of reliable and feasible solutions. For this reason, solutions obtained from traditional approaches to optimization under uncertainty (such as fuzzy, stochastic and

stochastic dynamic programming) may often be of limited value in producing realistic solutions for real-world problems.

Simulation-based optimization constitutes an interesting alternative in situations where the high level of complexity precludes a complete analytic formulation of a problem [7] and where uncertainty needs to be considered [8]. Simulation-based optimization involves the development of a detailed simulation model, which is then coupled with an optimizer in a black-box fashion. In other words, the optimizer operates on a (sub-)set of model parameters and the optimization process is based exclusively on the (usually stochastic) simulation responses. Evolutionary algorithms (EAs) are well-suited to black-box optimization settings, as highlighted by their wide application to real-world optimization problems that cannot be handled by analytical

approaches [14]. The feasibility and reliability of solutions become the primary consideration in such settings [8], and the EAs' flexibility in this respect typically offsets its disadvantages (specifically, the lack of guaranteed optimality of its identified solutions).

When EAs are employed as optimizers of simulation-based optimization models, fitness values become subject to the variability arising from the stochastic responses within the simulation model. The resulting noisy nature of the fitness values poses a challenge to the evolutionary optimizer, for it may mislead selection procedures [1] and lead to the propagation of inferior individuals or to the elimination of superior ones, thereby undermining algorithm performance [17]. Under these circumstances, noise handling strategies can play an important role in compensating for the impact of noise on the optimizer, and, specifically, in helping the optimizer to identify solutions that exhibit low fitness variability and give rise to high average fitness. Multiple studies have analysed situations in which noise causes perturbations during fitness evaluation, thus generating discrepancies between the observed and "true" fitness [14]. We refer the reader to [10] for a comprehensive survey of noise handling strategies proposed in the existing literature.

Implicit and explicit averaging are the two strategies most commonly employed to reduce the influence of noise in evolutionary optimization under noise. Implicit averaging relies on the EA mechanism itself to compensate for the impact of noise. Specifically, it assumes that the use of sufficiently large populations will ensure that individuals from promising regions of the search space are sampled repeatedly [10, 17], and the impact of noise can be reduced in this manner. On the other hand, explicit averaging strategies ensure that individuals are evaluated using average fitness values obtained across a specific number ( $n$ ) of fitness evaluations (replicates). Statistically, this approach ensures that the expected error of fitness estimates (i.e. the difference between the observed and the "true" fitness mean) reduces with a factor of  $\sqrt{n}$  [10].

Both implicit and explicit averaging strategies incur additional fitness evaluations due to (i) the increase in population size and due to (ii) the increase in the number of trials, respectively. Fitness evaluations present an important consideration in simulation-based optimization, as each replication of a simulation is time-consuming and the number of these replications may be limited by available computational time. Here, we investigate the efficiency and effectiveness of different noise-handling strategies in a realistic simulation-based optimization setting, in which the computational budget available for the optimization (and, therefore, the overall number of simulation replicates) is limited. Specifically, we compare explicit averaging against implicit averaging strategies for two different population sizes. Finally, we investigate a hybrid scheme that aims to combine the strengths of both approaches.

The remainder of this paper is organized as follows. Section 2 introduces the real-world optimization problem considered and the corresponding simulation-based optimization

model developed in this study. Explicit averaging, implicit averaging and a hybrid strategy combining both approaches are described in Section 3. Section 4 presents details about the comparative analysis, and empirical findings are presented in Section 5. Overall conclusions, as well as the limitations of this study and future research directions, are discussed in Sections 6 and 7, respectively.

## 2 Simulation-Based Optimization Model

In this study, a simulation-based optimization approach based on the integration of discrete event simulation (DES) and a genetic algorithm (GA) is applied to address the production planning problem of a real manufacturing company presented by Diaz Leiva and Handel [6] with the difference that, here, the objective is to achieve profit maximization. Additional modifications made to the original DES and optimization models presented in [6] are stated in Sections 2.1 and 2.2, respectively.

This problem corresponds to a big bucket, multi-product, multi-level (sub-products), capacitated (constraints are considered) production planning problem of a failure-prone manufacturing system, consisting of multiple work centres with insufficient capacity to fully cover demand requirements.

The DES model was developed in SimEvents<sup>®</sup> (The MathWorks, Inc., 2014) and Matlab's GA (MI-LXPM) implementation [5] was used as the optimizer. This is the default MATLAB<sup>®</sup> R2014a's (The MathWorks, Inc., 2013) implementation for solving integer and mixed integer problems with GA. The GA employs Laplace crossover, power mutation and binary tournament selection as operators. The truncation procedure, which ensures compliance with integer constraints after crossover and mutation is described in [5]. The inbuilt constraint-handling method is the parameter free penalty function approach proposed by Deb [4].

All computations were 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.

### 2.1 Simulation Model

The DES model employed in this study is a modified version of the model presented by Diaz Leiva and Handel [6].

The DES model represents the production of 31 products  $k$  within 7 work centres  $l$ . A work centre corresponds to the set of resources (e.g., machines, people, etc.) needed to manufacture certain products. Given that some products can be manufactured in several work centres a total of 41 processes  $j$  are considered in the DES model. A process  $j$  includes all series of events involved in the initialization of orders of a product  $k$ , its manufacturing in a specific work centre  $l$  and its storage in an specific sink  $s$  (with  $s = 1, 2, \dots, 41$ ).

This model intends to capture the delays ( $\alpha_l$ ) caused by work centre failures and provides the stock of products manufactured during a production period of one month (24 days composed of 3 shifts of 8 hours each). The total stock of a specific product ( $S_k$ ) corresponds to the sum of lots manufactured across the different work centres as shown by the following equation:

$$S_k = \sum_{l=1}^7 S_{k,l}, \tag{1}$$

where  $S_{k,l}$  is the stock of product  $k$  manufactured in work centre  $l$ .

The first modification to the original model is that instead of using probability distribution functions (PDFs) to represent the time required to manufacture a specific production lot ( $T_j$ ), constant values are assigned to those attributes according to specifications provided by the company.

Moreover, unlike the original DES model, the probability of occurrence of a work centre failure during the manufacturing of a product lot is here denoted as  $P_l$  and is modelled as attributes assigned to each work centre (rather than to each process). The probabilities used for  $P_l$ , as well as the PDFs employed to represent the delays caused by those failures ( $\alpha_l$ ), are summarized in Table 2.

Finally, an additional server was added to each process, so that the first lot of every process is processed by this server during the entire duration of each simulation replication (24 days). This modification was made to allow decision variables to take values equal to zero, a possibility not accounted for in the original model [6].

### 2.2 Optimization Model

The objective here is to generate production plans that try to maximize the expected sum of contributions to profit generated from processes undertaken by a failure-prone manufacturing system. The expected sum of contributions to profit is later referred to as “profit” for simplification purposes.

A total of 41 decision variables ( $x_j$ ) are considered, which correspond to the number of lots to be produced in each process  $j$ , and are constrained to be non-negative integers. Those decision variables, specified by the GA, constitute the input to the DES model and the responses  $S_k$  obtained from the DES model are used for computing the value of the fitness function.

The fitness value  $f$  is calculated across  $n$  independent simulation replicates for each individual  $x$  as follows:

$$\text{maximize } f(x) = \bar{c} = \frac{1}{n} \sum_{m=1}^n c_m, \tag{2}$$

where the value of  $n$  varies depending on the strategy applied (see Section 4 for details about  $n$ ).

For each replication  $m$ , the responses ( $S_k$ ) of the DES model are used to calculate  $c_m$  as follows:

$$c_m = \sum_{k=1}^{31} P_k, \tag{3}$$

where the total profit derived from product  $k$  is defined as:

$$P_k = S_k \times \rho_k, \tag{4}$$

where  $\rho_k$  denotes the contribution margin per lot of product  $k$ .

Additional constraints are imposed in the form of Equation 5 to avoid production levels greater than the maximum demand, to represent the requirement of sub-products and labour needed to undertake each process.

$$\sum_{j=1}^{41} a_{i,j} \times x_j \leq b_i \quad (i = 1, 2, \dots, 44), \tag{5}$$

where  $b_i$  denotes the magnitude of constraint  $i$  and  $a_{i,j}$  corresponds to the amount deployed from  $b_i$  by manufacturing one lot in process  $j$ .

### 3 Noise Handling Strategies

In this study we focus exclusively on implicit and explicit averaging strategies, as these are straightforward to implement in any EA and present the approaches most commonly employed in practice. Other noise handling strategies, such as averaging by means of approximated models [3, 12, 16] and modifications of the selection scheme [2, 15] have been proposed in the literature, but are not considered here.

An explicit averaging strategy uses a fixed number  $n$  of simulation replicates to obtain an average fitness value for each individual, as described in Equation 2. These average fitness estimates are then used to inform selection probabilities in the evolutionary algorithm. Therefore, under the explicit averaging strategy (ES) here analysed, fitness of an individual  $x$  is computed across 10 independent fitness evaluations ( $n = 10$ ) and the population size employed corresponds to 50 individuals.  $n$  is set at this relatively small level because a limited computational budget of 25,000 fitness evaluations is available, and assigning higher values to  $n$  or using a larger population size would reduce the number of generations that can be executed under ES.

In contrast, an implicit averaging strategy uses a single simulation replicate ( $n = 1$ ) to evaluate the quality of an individual. Increased robustness towards noise is then achieved by increasing population size relative to standard settings of this parameter. Consequently, under the implicit averaging strategy (IS), a single fitness evaluation ( $n = 1$ ) is used to compute fitness and a population size of 100 individuals is applied. Additionally, a baseline strategy (BS) is also analysed in order to evaluate the performance of implicit averaging when the population size is the same as in

ES, and therefore, twice as many generations are evolved compared to IS.

Moreover, we further describe a hybrid strategy (HS) that attempts to combine aspects of implicit and explicit averaging. This strategy applies implicit averaging ( $n = 1$  and a population size of 100 individuals) throughout the evolution process, but switches behaviour towards the end of the optimization: instead of choosing the final solution based on a single fitness value, we propose to select from the final population the feasible individual with the best average fitness. Consequently, we implement a mechanism that computes the average fitness of every feasible individual of the final population across a number  $\gamma$  of fitness evaluations, generated from independent simulation replications. The value of  $\gamma$  depends on the number of feasible individuals ( $\delta$ ) in the final population and on the computational budget available for this last step ( $E$ ), which in this case corresponds to 1000 fitness evaluations.  $\gamma$  is computed as follows:

$$\gamma = \left\lfloor \frac{E}{\delta} \right\rfloor. \quad (6)$$

Therefore, having a population size of 100, if every individual in the final population is feasible, 10 fitness evaluations are used to compute the average fitness of each individual. However, if infeasible individuals are present in the final population, those 1000 fitness evaluations are distributed amongst feasible individuals only.

Parameters specified for all four noise handling strategies introduced in this section are presented in Table 1.

## 4 Comparative Analysis

In order to test the effectiveness of the strategies under different levels of fitness variability, the following comparative analysis is undertaken for two different problem instances. Table 2 shows the different levels of uncertainty incorporated into each instance of the problem.

The performance of the EA under the proposed HS is compared with the performance observed for IS, ES and BS. In order to provide a fair comparison of the four strategies analysed, the stopping criteria selected to terminate the optimization procedure is the number of fitness evaluations. As mentioned in Section 3, a total budget of 25,000 fitness evaluations is allocated for every strategy as shown in Table 1.

The simulation-based optimization model is run 60 different times for each strategy; in each run, the best solution (based on last fitness evaluations) is selected from the final population. Consequently, 240 production plans are generated per problem instance. The precise quality of each of these plans is evaluated using extensive simulation: average profit, measured in United States Dollar (USD), is computed for every production plan across 1000 profit values obtained via stochastic simulation.

Subsequently, the four sets of average profit values are depicted as cumulative distribution functions (CDFs) and stochastic dominance criterion [18] is applied to determine whether or not the optimization performance, as measured in average profit values, differs between strategies.

Furthermore, Mann-Whitney U test [11] is then conducted for paired comparisons to test whether the optimization performance achieved under the different strategies is statistically significant, expressed in the form of the following hypotheses:

- $H_o$  : stochastic homogeneity of CDFs of average profit values obtained under both strategies
- $H_a$  : average profit values obtained under one strategy are stochastically smaller than the ones obtained under the other strategy

Mann-Whitney U test is employed instead of t-test, since distributions of the samples analysed do not fulfil the normality assumption.

## 5 Results

Descriptive statistics (means, minimum values, maximum values and standard deviations) of the average profit values (as computed across 1000 independent replications) obtained under the four strategies as well as the corresponding average computation times are presented in Tables 3 and 4 for problem instance 1 and 2, respectively.

Since our intention is to test the hypothesis presented in section 4 among the four strategies, homogeneity of variances of the ranked values across the different samples is a necessary condition for Mann-Whitney U test to be a reliable test [9]. Therefore, non-parametric Levene tests [13] were performed on every combination of samples in both problem instances. In both problem instances, results from these tests indicate that variances did not differ significantly ( $p > 0.05$ ) between the samples of ranks analysed, confirming the suitability of Mann-Whitney U test to evaluate the hypothesis above mentioned (Section 4).

Figures 1 and 2 illustrate as CDFs the 60 average profit values obtained with production plans generated under each strategy in problem instance 1 and 2, respectively. Both figures clearly show that the CDFs of average profit obtained under BS, IS and HS dominate the CDF of profit obtained under ES (first-order stochastic dominance). Furthermore, results from Mann-Whitney U test statistically show that average profit values obtained under ES are stochastically smaller ( $p < 0.01$ ) than the ones obtained under BS, IS and HS in both problem instances, as shown in Tables 5 and 6.

These results demonstrate that ES is an inadequate noise handling strategy in our setting: this result is likely to be driven by the limited computational budget available, and a stronger performance of ES may potentially be achieved when considering performance upon convergence.

Table 1: Parameters used for baseline, implicit averaging, explicit averaging and hybrid strategies

	BS	IS	ES	HS
<i>n</i>	1	1	10	1
<i>PopulationSize</i>	50	100	50	100
Generations	500	250	50	240
Fitness evaluations	25,000	25,000	25,000	≤ 25,000

Table 2: Probabilities for  $P_l$  and PDFs for  $\alpha_l$

Work centre ( <i>l</i> )	Instance 1	Instance 2	Instance 1 and 2	
	$P_l$	$P_l$	PDF	$\alpha_l$ Mean (d)
1	0.00	0.00	—	—
2	0.10	0.30	Exponential	0.0847
3	0.25	0.45	Exponential	0.0935
4	0.15	0.35	Exponential	0.1338
5	0.00	0.00	—	—
6	0.00	0.00	—	—
7	0.00	0.00	—	—

Table 3: Descriptive statistics of average profits and average computation times per strategy in problem instance 1

	BS	IS	ES	HS
Mean (USD)	703,689	715,376	555,932	707,503
Minimum (USD)	550,417	484,229	460,416	550,014
Maximum (USD)	793,316	795,716	689,685	792,696
Std dev (USD)	55,539	60,416	47,322	60,754
Average computation time (s)	1256	1144	1235	1369

Table 4: Descriptive statistics of average profits and average computation times per strategy in problem instance 2

	BS	IS	ES	HS
Mean (USD)	707,600	703,420	543,140	723,460
Minimum (USD)	536,840	538,990	437,930	561,550
Maximum (USD)	785,810	783,040	656,800	790,460
Std dev (USD)	60,137	61,744	43,412	51,258
Average computation time (s)	1284	1108	1214	1265

No dominance (neither first nor second degree stochastic dominance) can be determined among the CDFs of average profit obtained under BS, IS and HS in problem instance 1, where all three strategies appear equally competitive. These results are in accordance with results from Mann-Whitney U test, which indicate stochastic homogeneity ( $p > 0.05$ ) among samples obtained under BS, IS and HS in problem instance 1. It is interesting that population size (i.e. different setups of implicit averaging) has no significant effect in this setting as evidenced under BS and IS.

In problem instance 2, results from Mann-Whitney U tests also indicate stochastic homogeneity ( $p > 0.05$ ) among samples obtained under BS, IS and HS. However, the CDFs of average profit obtained under HS dominates the CDFs of average profit obtained under IS and under BS, respectively (first and second-order stochastic dominance). These results indicate that, even in a setting with a limited evaluation budget, the accurate fitness estimates from explicit averaging can be beneficial to optimization. It is clear that the difference between IS and HS arises during the final stages of the optimization only, while IS continues optimization during additional generations, HS focuses resources on explicit averaging across its last population. When seen in combination with the poor performance of ES, our results suggest that the trade-off between improved exploration (from evaluating more individuals) and accurate fitness evaluations (through simulation replicates for the same individual) needs to be carefully balanced in this setting. This finding appears in line with previous research that has delivered contradictory results regarding the relative performance of explicit and implicit averaging.

Table 5: Values for Mann-Whitney U statistic obtained in problem instance 1

	BS	IS	HS	ES
BS	—	1525	1705	107**
IS	—	—	1662	117**
HS	—	—	—	113**
ES	—	—	—	—

\*\*  $p < 0.01$

Table 6: Values for Mann-Whitney U statistic obtained in problem instance 2

	BS	IS	HS	ES
BS	—	1719	1549	103**
IS	—	—	1485	96**
HS	—	—	—	36**
ES	—	—	—	—

\*\*  $p < 0.01$

## 6 Conclusion

Implicit averaging strategies reduce the impact of noise by having a sufficiently large population, which ensures that individuals from promising regions of the search space are sampled repeatedly [10, 17]. On the other hand, explicit averaging strategies use average fitness values, obtained across a specific number of fitness evaluations, to ensure that evaluation of individuals is based on fitness estimates that are more robust to noise [10].

One of the key findings of this study was that, in the context of our real-world problem, a noise-handling strategy based on explicit averaging did not provide a competitive performance. More generally, this points to the fact that the computational costs incurred by simulation replicates may be problematic in constrained time settings.

Furthermore, we found that implicit averaging performed robustly for both of the population sizes used. The performance of our hybrid strategy does indicate that some effort towards explicit averaging may become important with increasing variability. Under low levels of fitness variability, the hybrid strategy, implicit averaging and our baseline showed a comparable performance. This situation changed with increasing levels of fitness variability, when HS started to enhance overall performance.

Compared to a pure, implicit averaging strategy, the hybrid strategy misses out on the last few generations of optimization. Our results show, however, that this disadvantage is more than counter-balanced by the benefits from an accurate final selection step that reduces the likelihood of choosing an inferior individual (in terms of average fitness) as the final solution.

## 7 Limitations and Future Research

The relevance of obtaining more reliable fitness estimates increases with the level of variability, since there is a higher risk of choosing an inferior solution. It is, therefore, intuitive that the final selection mechanism implemented in HS would be more beneficial in such circumstances. But at the same time, the number of fitness evaluations needed to obtain reliable estimates is expected to raise with higher fitness variability, leaving to the evolutionary process a smaller share of the computational budget. Therefore, further research may focus on investigating the right trade-off between exploration (IS) and accurate fitness evaluation (ES). In this sense, the application of different sampling techniques (e.g., Latin Hypercube) during the final selection mechanism might be worthy of future investigation, as it may allow a reduction in the number of fitness evaluations required in this last step.

Our results underline issues around the computational cost of explicit averaging, but also highlight that sporadic use of this strategy may, nevertheless, be beneficial. In this study, the use of explicit averaging in the hybrid strategy was limited to the final selection step only. Future research

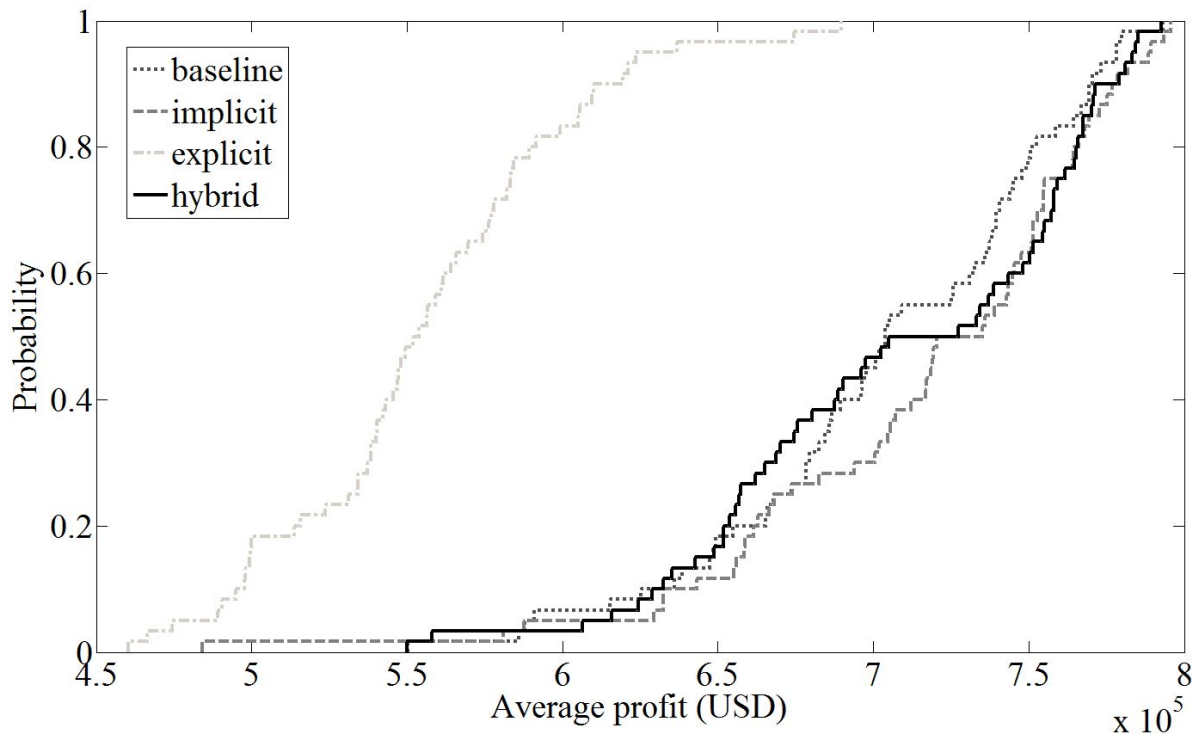


Figure 1: CDFs of average profit values obtained with production plans generated under the four different strategies in problem instance 1.

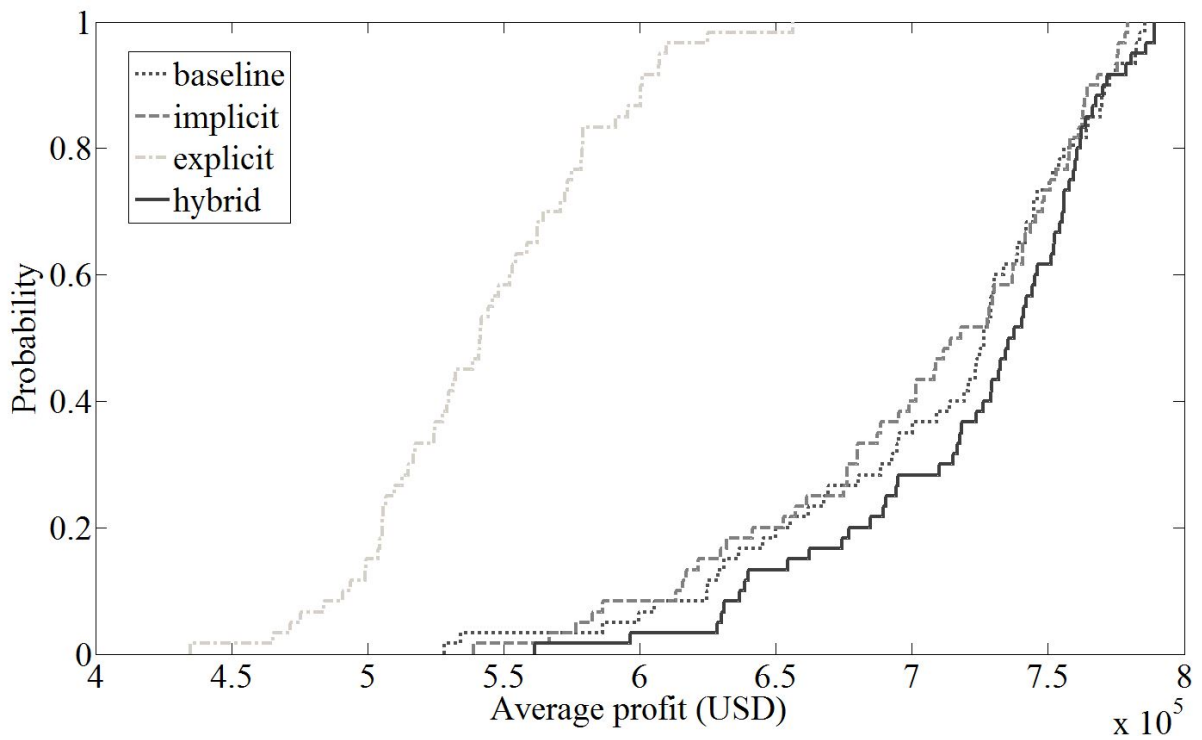


Figure 2: CDFs of average profit values obtained with production plans generated under the four different strategies in problem instance 2.

may consider the possibility of using explicit averaging at earlier points during the optimization process.

## 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.

## References

- [1] M. Bhattacharya, R. Islam, A. N. Mahmood (2014) Uncertainty and evolutionary optimization: A novel approach, *Proceedings of the 9th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, pp. 988–993.
- [2] J. Branke, C. Schmidt (2003) Selection in the presence of noise, *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pp. 766–777.
- [3] J. Branke, C. Schmidt, H. Schmeck (2001) Efficient fitness estimation in noisy environments. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pp. 243–250.
- [4] K. Deb (2000) An efficient constraint handling method for genetic algorithms, *Computer methods in applied mechanics and engineering*, vol. 186, no. 2, pp. 311–338.
- [5] K. Deep, K. P. Singh, M. Kansal, C. Mohan (2009) A real coded genetic algorithm for solving integer and mixed integer optimization problems, *Applied Mathematics and Computation*, vol. 212, no. 2, pp. 505–518.
- [6] J. E. Diaz Leiva, J. Handl (2014) Simulation-based ga optimization for production planning, *Proceedings of the Student Workshop on Bioinspired Optimization Methods and their Applications (BIOMA)*, pp. 27–39.
- [7] C. Ehrenberg, J. Zimmermann (2012) Simulation-based optimization in make-to-order production: scheduling for a special-purpose glass manufacturer, *Proceedings of the Winter Simulation Conference (WSC)*, pp. 1–12.
- [8] G. Figueira, B. Almada-Lobo (2014) Hybrid simulation-optimization methods: A taxonomy and discussion, *Simulation Modelling Practice and Theory*, vol. 46, pp. 118–134.
- [9] M.-R. Ghasemi, J. Ignatius, A. Emrouznejad (2014) A bi-objective weighted model for improving the discrimination power in MCDEA, *European Journal of Operational Research*, vol. 233, no. 3, pp. 640–650.
- [10] Y. Jin, J. Branke (2005) Evolutionary optimization in uncertain environments - a survey, *IEEE Transactions on Evolutionary Computation*, vol. 9, no. 3, pp. 303–317.
- [11] H. B. Mann, D. R. Whitney (1947) On a test of whether one of two random variables is stochastically larger than the other, *The Annals of Mathematical Statistics*, vol. 18, no. 1, pp. 50–60.
- [12] F. Neri, X. del Toro Garcia, G. L. Cascella, N. Salvatore (2008) Surrogate assisted local search in pmsm drive design, *COMPEL: The International Journal for Computation and Mathematics in Electrical and Electronic Engineering*, vol. 27, no. 3, pp. 573–592.
- [13] D. W. Nordstokke, B. D. Zumbo (2010) A new non-parametric levene test for equal variances, *Psicologica*, vol. 31, no. 2, pp. 401–430.
- [14] C. Qian, Y. Yu, Y. Jin, Z.-H. Zhou (2014) On the effectiveness of sampling for evolutionary optimization in noisy environments, *Lecture Notes in Computer Science*, vol. 8672, pp. 302–311.
- [15] G. Rudolph (2001) A partial order approach to noisy fitness functions, *Proceedings of the Congress on Evolutionary Computation (CEC)*, vol. 1, pp. 318–325.
- [16] Y. Sano, H. Kita, I. Kamihira, M. Yamaguchi (2000) Online optimization of an engine controller by means of a genetic algorithm using history of search, *Proceedings of the 26th IEEE Annual Conference of the Industrial Electronics Society (IECON)*, vol. 4, pp. 2929–2934.
- [17] A. Syberfeldt, A. Ng, R. I. John, P. Moore (2010) Evolutionary optimisation of noisy multi-objective problems using confidence-based dynamic resampling, *European Journal of Operational Research*, vol. 204, no. 3, pp. 533–544.
- [18] S. Yitzhaki (1982) Stochastic dominance, mean variance, and gini's mean difference, *The American Economic Review*, vol. 72, no. 1, pp. 178–185.