New Uncertainty Handling Strategies in Multi-objective Evolutionary Optimization

Thomas Voß[1], Heike Trautmann[2], and Christian Igel[1]

[1] Institut für Neuroinformatik, Ruhr-Universität Bochum, Germany
[2] TU Dortmund University, Dortmund, Germany
{thomas.voss,christian.igel}@ini.rub.de,
trautmann@statistik.tu-dortmund.de

Abstract. Since many real-world optimization problems are noisy, vector optimization algorithms that can cope with noise and uncertainty are required. We propose new, robust selection strategies for evolutionary multi-objective optimization in the presence of noise. We apply new measures of uncertainty for estimating the recently introduced Pareto-dominance for uncertain and noisy environments (PDU). The first measure is the inter-quartile range of the outcomes of repeated function evaluations. The second is based on axis-aligned bounding boxes around the upper and lower quantiles of the sampled fitness values in objective space. Experiments on real and artificial problems show promising results.

1 Introduction

Most real-world multi-objective optimization problems (MOPs) are inherently noisy because they rely on noisy measurements or require complex simulations that suffer from noisy conditions as well. Multi-objective evolutionary algorithms relying on indicator-based selection strategies have been successfully applied to various real-world as well as artificial MOPs. However, in their canonical form they do not work well when applied to MOPs with strong noise. By changing the selection scheme and sampling the objective functions for each candidate solution several times, the algorithms' robustness w.r.t. noise and uncertainty can be improved. Recently, an extension of the Pareto-dominance relation that addresses the problem of noisy objective function values has been presented in [1]. This extension originally relies on the convex hull of multiple objective function samples. We propose to consider axis-aligned bounding boxes and per-objective empirical quartiles instead.

In the following, we briefly introduce Pareto-dominance for uncertain and noisy environments (PDU) including our adaptation based on axis-aligned bounding boxes. Thereafter, we describe our empirical evaluation on real-world and artificial fitness functions before we finish with the conclusions and open questions.

2   A Robust Variant of Pareto Dominance

Let the noisy MOP (MNOP) be defined in a very general way without assumptions on the distribution of the objective function values nor on the error structure considered (e.g., the definition is not restricted to multiplicative or additive noise):

Definition 1 (MNOP)

$$
\begin{aligned}
&Min.\ \ \boldsymbol{y} = (\boldsymbol{f}|\boldsymbol{x}, \boldsymbol{\varepsilon}) = (f_1, \ldots, f_k|\boldsymbol{x}, \boldsymbol{\varepsilon}) && \textit{(MNOP)}\\
&with\ \ \boldsymbol{y} = (y_1, \ldots, y_k) \in \mathcal{Y}, && \textit{objectives}\\
&\quad (f_i|\boldsymbol{x}, \varepsilon_i) \sim \mathcal{F}_i, \quad i = 1, \ldots, k, && \textit{objective functions}\\
&\quad \mathcal{F}_i, \quad i = 1, \ldots, k, && \textit{distribution functions}\\
&\quad \boldsymbol{x} = (x_1, \ldots, x_n) \in \mathcal{X}, && \textit{decision variables}\\
&\quad \boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_k), && \textit{error terms.}
\end{aligned}
$$

We introduce an approach that handles the inherent uncertainty of the optimization problem by a specific variant of Pareto dominance relying on repeated evaluations of the decision vectors considered. By combining both expectation $E(f_i|\boldsymbol{x}, \varepsilon_i)$ as well as uncertainty $U(f_i|\boldsymbol{x}, \varepsilon_i)$ of the resulting objective realizations an alternative concept of Pareto-dominance (PDU) is set up [2,3]:

Definition 2 (Pareto-Dominance in Uncertain Environments (PDU)).
A solution $\boldsymbol{x}$ of MNOP dominates a solution $\boldsymbol{x}^*$ iff

$$
\exists\, i \in \{1, \ldots, k\} : E(f_i|\boldsymbol{x}, \varepsilon_i) + U(f_i|\boldsymbol{x}, \varepsilon_i)\ < E(f_i|\boldsymbol{x}^*, \varepsilon_i) - U(f_i|\boldsymbol{x}^*, \varepsilon_i);
$$
$$
\forall\, j = (1, \ldots, k), j \neq i : E(f_j|\boldsymbol{x}, \varepsilon_j) + U(f_j|\boldsymbol{x}, \varepsilon_j) \leq E(f_j|\boldsymbol{x}^*, \varepsilon_j) - U(f_j|\boldsymbol{x}^*, \varepsilon_j).
$$

The concept of PDU is based on an uncertainty zone around the expected value at a given solution. Dominance decisions are only made in case the reliability of the objective values is high enough in the sense that the uncertainty zones of two points to be compared do not overlap in any dimension (see Fig. 1). The definition of PDU above implicitly assumes symmetric noise distributions by estimating only a single $U(f_i|\boldsymbol{x}, \varepsilon_i)$ value per objective i at point $\boldsymbol{x}$.

Expectation. As the required expected value $E(f_i|\boldsymbol{x}, \varepsilon_i)$ cannot be observed directly, we estimate the location of the corresponding distribution based on a sample of m evaluations of the objective functions. To get a robust estimate, we consider the median (which is an unbiased estimator of the expectation if the expectation exists and the distribution is symmetric):

$$
\begin{aligned}
&E(f_i|\boldsymbol{x}, \varepsilon_i) \approx \operatorname{med}(f_{im}) \ \forall\, i \in \{1, ..., k\} \quad \text{with}\\
&\operatorname{med}(\boldsymbol{f}_m) = (\operatorname{med}(f_{1m}|\boldsymbol{x}, \varepsilon_1), \ldots, \operatorname{med}(f_{km}|\boldsymbol{x}, \varepsilon_k)),\\
&(f_{im}|\boldsymbol{x}, \varepsilon_i) = \left(f_i^1, \ldots, f_i^m|\boldsymbol{x}, \varepsilon_i\right), \quad i = 1, \ldots, k.
\end{aligned}
$$

Theoretically, a sample size of $m \to \infty$ is required. As in practice already moderate sample sizes are unrealistic, we suggest to use racing algorithms (see below) for dynamically adjusting the sample size depending on the amount of variability.

Uncertainty. There are several possibilities for choosing an indicator for the uncertainty of the sampled objective values for a given point in decision space. We will introduce and investigate three different approaches with increasing complexity.

1. One approach, probably the simplest one to handle noisy objectives, is to omit a special uncertainty measure and account for the variability solely by using an estimator for the expected value (i.e., $U(f_i|\boldsymbol{x}, \varepsilon_i) = 0$).
2. The inter-quartile range (IQR) of the objective realizations in each dimension suggests itself as an appropriate estimator for the inherent uncertainty. We refer to a non-parametric estimator because this is in line with using the median as the estimator for the expected value, that is

$$U(f_i|\boldsymbol{x}, \varepsilon_i) = q_{0.75}(f_{im}|\boldsymbol{x}, \varepsilon_i) - q_{0.25}(f_{im}|\boldsymbol{x}, \varepsilon_i) \ ,$$

   where q_α refers to the empirical α-quantile of $\mathcal{F}_i$.
3. Analogous to reverting to convex hulls in [2] we propose the usage of an axis-aligned bounding box (BB, [4]) based on the upper and lower quartiles of the sampled values in $\mathbb{R}^k$. Computing the box has a much lower computational complexity and has the same intuitive interpretation, namely, that with increasing volume of the bounding box the uncertainty of objectives increases simultaneously. The bounding box BBQ is defined as the cartesian product of the k intervals each of which is defined by the lower and upper quartile of the corresponding objective. The average of the distances of med($\boldsymbol{f}_m$) to the closest point on BBQ in each dimension (BBQ$_j^{\text{closest}}$) is taken as the required uncertainty measure reflecting a deviation to border points (see Fig. 1):

$$U(f_i|\boldsymbol{x}, \varepsilon_i) = \frac{1}{k} \sum_{j=1}^{k} \left| \text{med}(f_{im}) - \text{BBQ}_j^{\text{closest}} \right| \ .$$

Distribution of Function Evaluations. We apply concepts inspired by selection races [5,6,7] for dynamically controlling the sample size per individual such that the number of samples is as low as possible while the single objectives satisfy certain statistical requirements.

3   Preliminary Empirical Evaluation

This section presents the empirical evaluation of the different noise handling approaches on benchmark functions as well as on two real-life optimization problems. We considered the $(\mu + \mu)$-MO-CMA-ES (see [8,9]) and the NSGA-II (see [10]) in our experiments. All experiments have been conducted using the Shark Machine Learning Library [11]. In the following, we first outline the two real-world optimization problems from rapid manufacturing before we describe the experimental setup and the results.
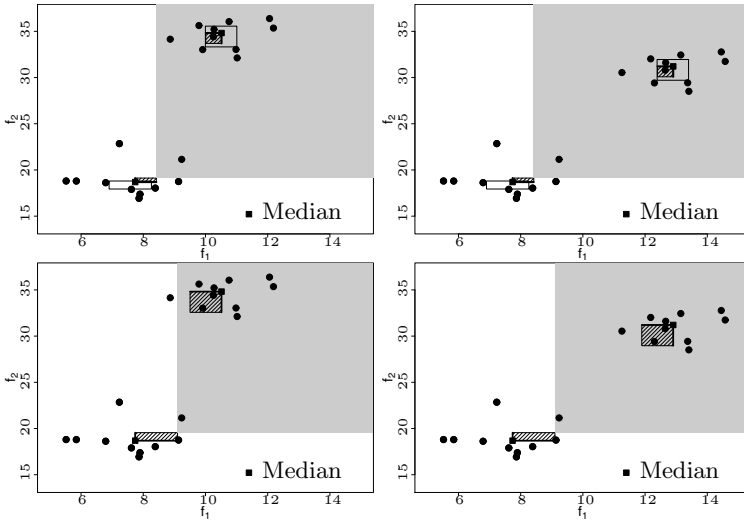
Fig. 1. Examples of PDU. Uncertainty regions are displayed by striped rectangles. The grey rectangle indicates the region dominated by the left solution. Black thick lines visualize distances from the median to the Quartile-BB in each dimension (top, approach 3) or represent the IQR (bottom, approach 2).

Rapid Manufacturing. Ready-to-use additive manufacturing (RUAM) combines welding and grinding by using a single machine while the optimization of the weld bead geometry is the crucial part of an efficient process. Structures are built in a layer-by-layer stepwise procedure which strongly relies on correct wall thickness and shape. Two different RUAM problems are addressed here.

RUAM1 [3]: Liratzis [12] identified depth of penetration P_d and sidewall penetration P_s to be key weld bead geometry quality factors being objectives to be minimized. Travel speed (T, $[900,\dots 1400]$ mm/min), wire feed speed (W, $[8.2,\dots 13.2]$ m/min), wire distance from side wall (D, $[0.3,\dots, 1.2]$ mm) and arc length correction (A_c, $[-25,\dots, 25]$ %) are the key decision variables of the process. Based on Design of Experiment (DoE) methods the following models were estimated based on coded influence factors in the interval $[0, 1]$ using the variable bounds listed above:

$$P_d = 2.55 + 0.35 \cdot W - 0.23 \cdot T + 0.21 \cdot D + 0.18 \cdot A_c + 0.14 \cdot W \cdot A_c$$
$$- 0.25 \cdot T \cdot A_c - 0.073 \cdot A_c^2 + \varepsilon_1, \ \varepsilon_1 \sim \mathcal{N}(0, 0.1661^2),$$
$$P_s = 0.60 + 0.038 \cdot W - 0.086 \cdot T - 0.074 \cdot D + 0.13 \cdot A_c + 0.025 \cdot W \cdot T$$
$$- 0.024 \cdot W \cdot D - 0.043 \cdot T \cdot D - 0.018 \cdot W^2 - 0.022 \cdot T^2 + \varepsilon_2, \ \varepsilon_2 \sim \mathcal{N}(0, 0.0331^2).$$

RUAM2 [13]: Another aspect is the maximization of the height (H) and simultaneous target optimization of the width (W) of the weld bead while considering constraints regarding the contact angle (CA) and limitations of the welding machine. The diameter of wire (X_1, $[0.8,\dots, 1.2]$ mm), wire feed speed (X_2,
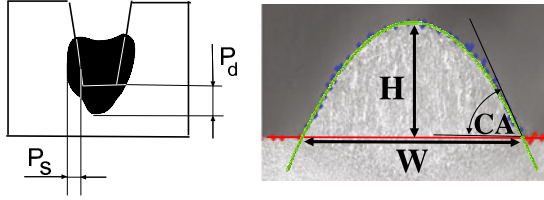
Fig. 2. Visualization of objectives (left: RUAM1, right: RUAM2)

[1.5,...,14] mm/s) and the wire feed speed divided by the travel speed (X_3, [10,...,20] mm/s) were identified as key decision variables by DoE methods:

$$H = 3.948 \cdot X_1 - 0.5386 \cdot X_2 + 0.006262 \cdot X_3^2 - 1.123 \cdot X_1^3 + 0.03353 \cdot X_2^2$$
$$- 0.0001578 \cdot X_3^3 - 0.0008612 \cdot X_2^3 + 0.1668 \cdot X_1 \cdot X_2 + \varepsilon_1, \varepsilon_1 \sim \mathcal{N}(0, 0.13^2) \ ,$$
$$W^* = |W - 5| + \varepsilon_2 = |2.5208 \cdot X_1 - 0.2699 \cdot X_2 + 0.0038 \cdot X_3^2 - 0.0004 \cdot X_2^3$$
$$+ 0.6766 \cdot X_1 \cdot X_2 - 5| + \varepsilon_2, \varepsilon_2 \sim \mathcal{N}(0, 0.6116^2) \quad \text{w.r.t.}$$
$$\text{CA} = 101.58 \cdot X_2 + 971.94 \cdot X_1 - 74.49 \cdot X_3 - 251.21 \cdot X_2 \cdot X_1 - 0.78 \cdot X_2 \cdot X_3$$
$$- 65.41 \cdot X_1 \cdot X_3 + 0.08 \cdot X_2^2 + 7.67 \cdot X_3^2 + 4.01 \cdot X_2^2 \cdot X_1 + 115.47 \cdot X_2 \cdot X_1^2$$
$$+ 0.036 \cdot X_2^2 \cdot X_3 + 28.50 \cdot X_1^2 \cdot X_3 + 0.27 \cdot X_1 \cdot X_3^2 - 0.13 \cdot X_2^3 - 306.33 \cdot X_1^3$$
$$- 0.18 \cdot X_3^3 + 0.12 \cdot X_2 \cdot X_1 \cdot X_3 \leq 90 \ ; X_1 \leq 0.0043 \cdot X_2^2 - 0.1313 \cdot X_2 + 1.7876$$

in addition to the box constraints of the decision variables.

Experimental Setup. We compare the algorithms deploying the respective combinations of noise handling methods on different classes of benchmark functions. We consider the bi-objective, constrained and non-separable WFG6 and WFG8 (see [14]). Moreover, the bi-objective, unconstrained and rotated benchmark function ELLI1 (see [8]) is used for the performance evaluation. We augment the assessment by empirically analyzing the performance of the algorithms on two optimization problems from the domain of rapid manufacturing. For all of the bi-objective functions, we considered different levels of additive, normally-distributed noise (see Table 1). For the selection races, we set δ to 0.0001 according to the suggestions given in [7] and limited the maximum number of repeated fitness function evaluations to $t_{\max} = 15$. The parent and offspring population sizes have been chosen as $\mu = \lambda = 100$. In case of the $(\mu + \mu)$-MO-CMA-ES, we adhered to the parameter setup presented in [9]. For the NSGA-II, we considered the default parameter setup (see [10]). We conducted 50 independent trials and aborted every trial after 500 generations.

Results. The performance of the different noise handling strategies is analyzed by the hypervolume (HV) and the Generalized Spread Indicator (SP) [10] over the course of generations, where the counting of the generations is done independently from the actual number of required fitness reevaluations.

Exemplary visualizations are given in Figs. 3 and 4 for the high noise levels. The objective function WFG8 is omitted as the results are very similar to WFG6.

Table 1. Empirical reference points for computing the unary hypervolume indicator and the per-objective noise levels

	Empirical reference point	Noise levels
WFG6	(2.6131, 5.47336)	$\sigma_{1,2} \in \{0.05, 0.1\}$
WFG8	(2.77644, 5.52929)	$\sigma_{1,2} \in \{0.05, 0.1\}$
ELLI1	(2.21839, 2.14453)	$\sigma_{1,2} \in \{0.1, 1\}$
RUAM1	(3.89213, 1.629089)	$\sigma_1 = 0.1661, \sigma_2 = 0.0331$
RUAM2	(1, 7.28359)	$\sigma_1 = 0.13, \sigma_2 = 0.6116$

As SP has to be minimized, the negative SP is plotted to be visually in line with the HV which has to be maximized.

The results of the algorithms without noise handling suggest that both of them are able to cope with noisy fitness functions to some degree without modifications. This may be attributed to the population-based approach as well as to the rank-based selection scheme employed in both algorithms. However, the performance of the algorithms without noise handling deteriorated quickly for low signal-to-noise ratios.

In all cases most of the variants of both the $(\mu + \mu)$-MO-CMA-ES as well as the NSGA-II employing the PDU performed better with respect to the final HV than the variants of the algorithms without noise-handling (No_NH). The same observations can be made for the lower noise levels showing slightly smaller effects. While the variant without uncertainty measurement (NUT) in some cases seems to outperform the more sophisticated noise handling approaches IQR and BBQUT with regard to HV, the analysis of SP shows that both IQR and BBQUT mostly manage to generate a better spread of solutions in the final front.

In addition, we analyzed the median age of the individuals in the final population. Exemplary results are picked up in Fig. 5. For WFG6, RUAM1 and RUAM2 the results of the NSGA-II show the same tendencies as for the MO-CMA-ES, and the WFG8 results are in line with WFG6. It becomes obvious that the median age of the individuals of both the algorithms without noise handling as well as the NUT variants is considerably higher than for IQR and BBQUT, especially for WFG6 and WFG8 with IQR being dominant for the RUAM. Thus, the approaches No_NH and NUT seem to run the risk of generating biased results with regard to randomly generated solutions which appear to be of high quality but in fact are not, for example they can be even better than the true Pareto front. The composition of the actual fronts lasts much longer than for IQR and BBQUT indicating partial stagnation.

Figure 5 as well lists the median number of nondominated solutions in the final fronts which decreases with increasing noise level. It is noticeable that the number of nondominated solutions for all problems but ELLI1 is much lower than the population size which reveals the increasing complexity of the problems due to the introduced noise.
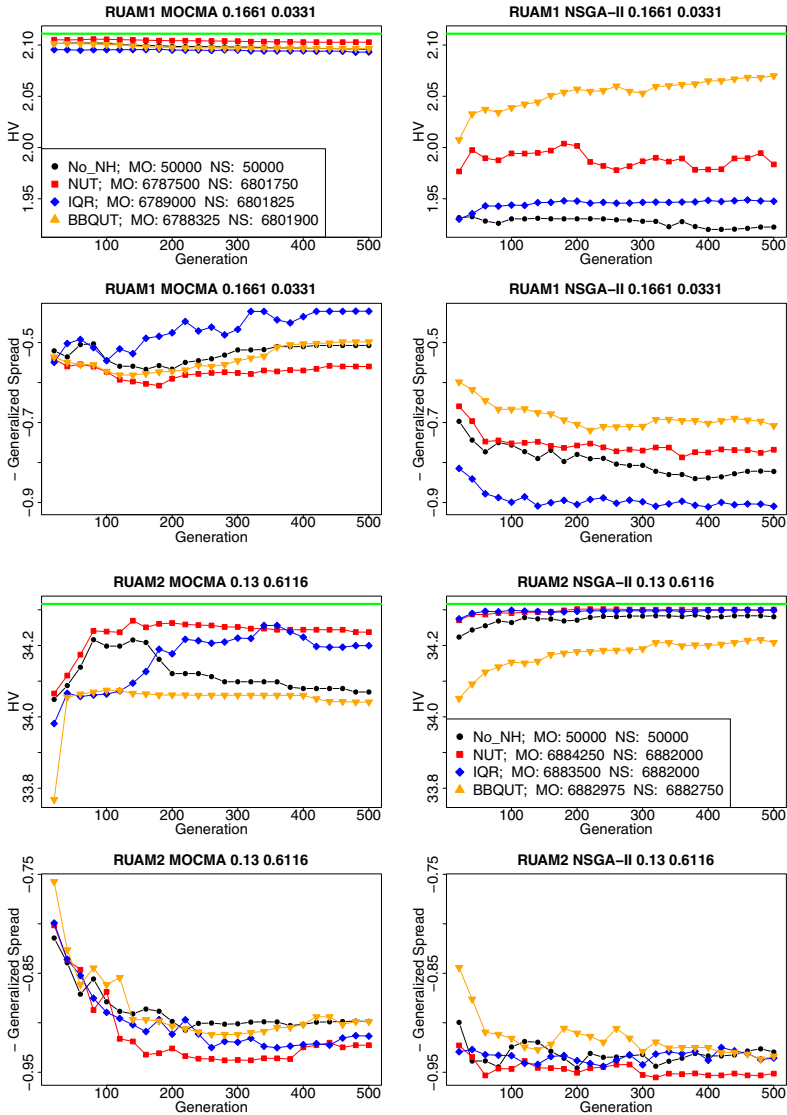
Fig. 3. Evolution of the absolute hypervolume for RUAM1 and RUAM2 over the number of generations. The upper line in the figures indicates the best results for empirically determined Pareto-approximations for the non-noisy case. For each of the noise handling approaches, the average number of fitness function evaluations spent by the $(\mu + \mu)$-MO-CMA-ES (indicated by MO) and the NSGA-II (indicated by NS) are reported.
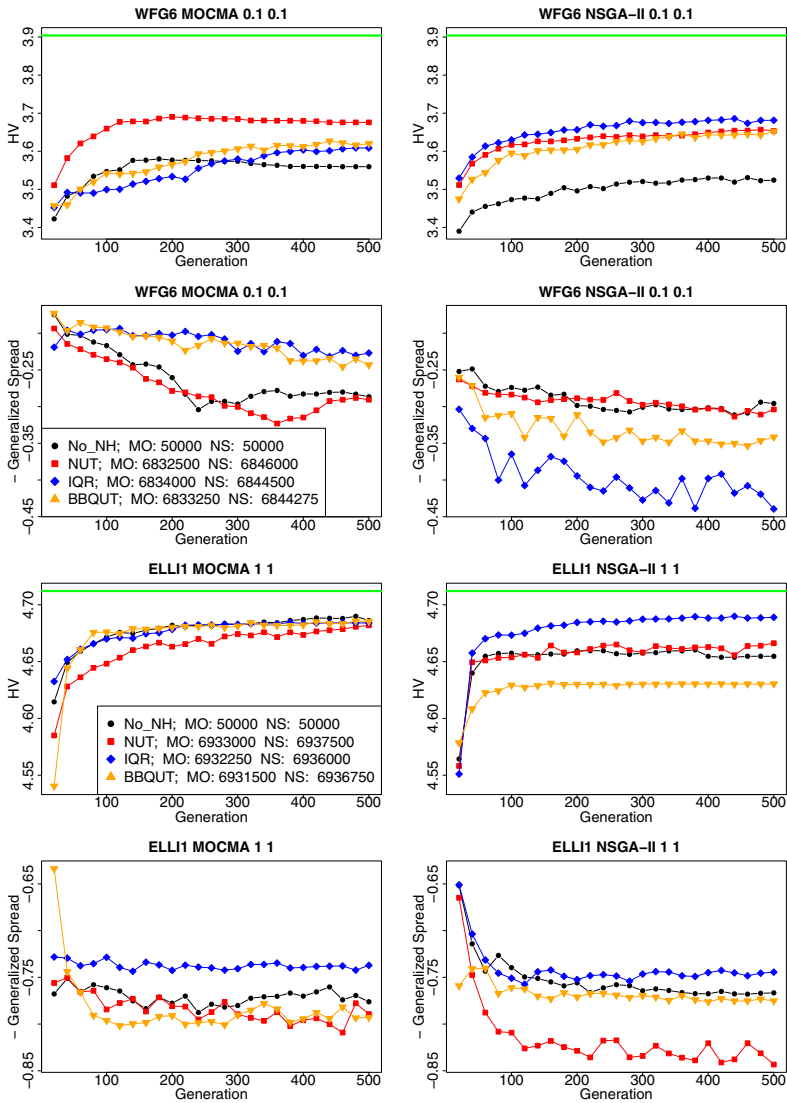
Fig. 4. Evolution of the absolute hypervolume for the fitness functions WFG6 and ELLI1 over the number of generations. All plots refer to the medians of 50 trials. For each of the noise handling approaches, the average number of fitness function evaluations spent by the $(\mu + \mu)$-MO-CMA-ES (indicated by MO) and the NSGA-II (indicated by NS) are given.

The number of non-dominated solutions in the final population is considerably lower than usually observed on standard, non-noisy MOPs. This shows the importance of the first level sorting criterion – here non-dominating sorting using different notions of dominance – for the selection process in the presence of noise and uncertainty.
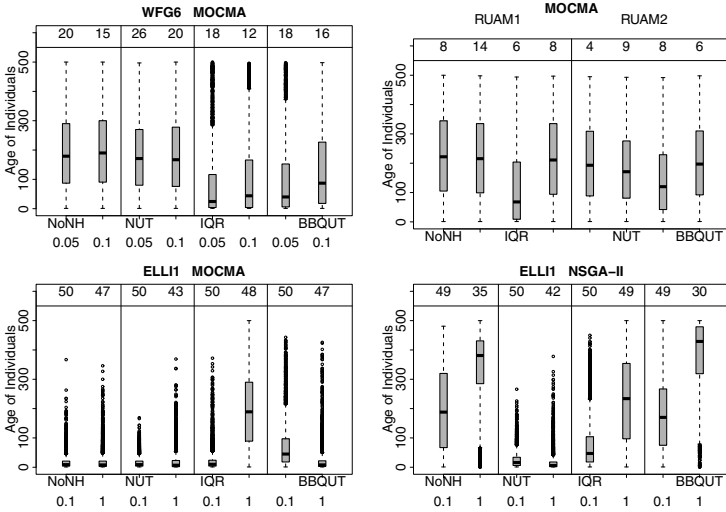
Fig. 5. Boxplots of ages of the individuals in the final population. Different noise levels are given at the bottom of the plot if relevant. The median numbers of nondominated solutions in the final front are given at the top of each boxplot.

By dynamically distributing fitness function reevaluations, we saved between 7.5% and 9% of a maximum of 750,000 (100 offspring, 500 generations, a maximum of 15 evaluations per individual) fitness function evaluations (see Figs. 3 and 4 for the total number of fitness function evaluations spent by the respective combination of algorithms and noise handling approaches).

4   Conclusions

We presented a generalization of the Pareto-dominance relation for uncertain environments that is directly applicable within any evolutionary multi-objective algorithm relying on the well-known indicator-based selection strategy. Two methods for measuring the uncertainty arising from noisy fitness function values have been introduced. We empirically investigated the performance of our noise handling approaches on both real-world optimization problems and artificial fitness functions. Our results suggest that the methods presented here improve the overall robustness of multi-objective evolutionary algorithms relying on the indicator-based selection scheme and thus, enhance the algorithms' performance in the presence of noise.

Nevertheless, the noise handling framework presented here allows for future improvements by considering more sophisticated schemes for measuring the uncertainty. Future work should include an in-depth evaluation on a broad range of fitness functions and consider different noise distributions. Additionally, we will study the integration of selection races and PDU. Further, we will consider non-elitist evolutionary multi-objective algorithms, which may be more appropriate for noisy problems (e.g., see [15]).

References

1. Trautmann, H., Mehnen, J.: Preference-Based Pareto-Optimization in Certain and Noisy Environments. Engineering Optimization 41, 23–38 (2009)
2. Trautmann, H., Mehnen, J., Naujoks, B.: Pareto-dominance in noisy environments. In: Tyrrell, A. (ed.) 2009 IEEE Congress on Evolutionary Computation, pp. 3119–3126. IEEE Press, Los Alamitos (2009)
3. Mehnen, J., Trautmann, H.: Robust Multi-objective Optimisation of Weld Bead Geometry for Additive Manufacturing. In: Teti, R. (ed.) Proceedings of the 6th CIRP International Seminar on Intelligent Computation in Manufacturing Engineering, CIRP ICME 2008 (2008)
4. de Berg, M., Cheong, O., van Kreveld, M., Overmars, M.: Computational Geometry, 3rd edn. Springer, Heidelberg (2008)
5. Maron, O., Moore, A.W.: The racing algorithm: Model selection for lazy learners. Artificial Intelligence Review 11, 193–225 (1997)
6. Audibert, J.Y., Munos, R., Szepesvári, C.: Tuning bandit algorithms in stochastic environments. In: Hutter, M., Servedio, R.A., Takimoto, E. (eds.) ALT 2007. LNCS (LNAI), vol. 4754, pp. 150–165. Springer, Heidelberg (2007)
7. Heidrich-Meisner, V., Igel, C.: Hoeffding and bernstein races for selecting policies in evolutionary direct policy search. In: Bottou, L., Littman, M. (eds.) Proceedings of the International Conference on Machine Learning (ICML 2009), pp. 401–408 (2009)
8. Igel, C., Hansen, N., Roth, S.: Covariance matrix adaptation for multi-objective optimization. Evolutionary Computation 15(1), 1–28 (2006)
9. Voß, T., Hansen, N., Igel, C.: Improved step size adaptation for the MO-CMA-ES. In: Proc. 12th Annual Conference on Genetic and Evolutionary Computation Conference (GECCO). ACM Press, New York (2010)
10. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation 6, 182–197 (2002)
11. Igel, C., Glasmachers, T., Heidrich-Meisner, V.: Shark. Journal of Machine Learning Research 9, 993–996 (2008)
12. Liratzis, T.: Tandem Gas Metal Arc Pipeline Welding. PhD thesis, Cranfield University, UK (2007)
13. Wessing, S., Ding, J., Trautmann, H., Mehnen, J., Naujoks, B.: Sequential Parameter Optimization for Multi-Objective Evolutionary Optimization of Additive Layer Manufacturing. In: Teti, R. (ed.) Proceedings of the 7th CIRP International Seminar on Intelligent Computation in Manufacturing Engineering, CIRP ICME 2010 (2010) (accepted)
14. Huband, S., Hingston, P., Barone, L., While, L.: A review of multi-objective test problems and a scalable test problem toolkit. IEEE Transactions on Evolutionary Computation 10, 477–506 (2007)
15. Büche, D., Stoll, P., Dornberger, R., Koumoutsakos, P.: Multiobjective evolutionary algorithm for the optimization of noisy combustion processes. IEEE Transactions on Systems, Man, and Cybernetics 32, 460–473 (2002)