

# An Evolutionary Approach to the Solution of Multi-Objective Min-Max Problems in Evidence-Based Robust Optimization

Simone Alicino and Massimiliano Vasile

Department of Mechanical and Aerospace Engineering

University of Strathclyde

Glasgow, G1 1XJ, United Kingdom

Email: {simone.alicino, massimiliano.vasile}@strath.ac.uk

**Abstract**—This paper presents an evolutionary approach to solve the multi-objective min-max problem (MOMMP) that derives from the maximization of the Belief in robust design optimization. In evidence-based robust optimization, the solutions that minimize the design budgets are robust under epistemic uncertainty if they maximize the Belief in the realization of the value of the design budgets. Thus robust solutions are found by minimizing, with respect to the design variables, the global maximum with respect to the uncertain variables. This paper presents an algorithm to solve MOMMP, and a computational cost reduction technique based on Kriging metamodels. The results show that the algorithm is able to accurately approximate the Pareto front for a MOMMP at a fraction of the computational cost of an exact calculation.

## I. INTRODUCTION

Given the model of a system or process, the lower expectation in the realization of the value of a particular performance index for that system or process can be defined as the degree of belief that one has in a certain proposition being true, given the available evidence. In the framework of imprecise probabilities, it can be seen as a lower bound to the cumulative distribution function of classical probability theory. Its use is therefore interesting in engineering design, as it gives the lower limit of the confidence that the design budgets under uncertainty will be below a given threshold. In this framework both epistemic and aleatory uncertainties can be treated even when no exact information on the probability distribution associated to a particular uncertain quantity is available. Stochastic variables, with associated probability, are replaced by a multivalued mapping from a collection of subsets of an uncertain space  $U$  into an envelope defined by the lower and upper expectation (Belief and Plausibility in the case of Evidence Theory) in the realization of a particular value.

In the preliminary design of an engineering system the type of uncertainty is often mainly epistemic in nature, as in a later stage of the design more information is generally available. It is therefore natural to assign belief masses to intervals of values rather than precise probabilities. The main drawback with the use of multivalued mappings is that the computation of the lower and upper expectations, Belief and Plausibility in the case of belief functions, has a complexity that is exponential with the number of uncertain variables. Recently, Vasile et al. [1] proposed some

strategies to obtain an estimation of the maximum Belief with a reduction of the computational cost. The approach starts by translating an optimization under uncertainty into a single or multi-objective min-max problem equivalent to a worst-case scenario optimization problem. Although this approach avoids the intrinsic exponential complexity in the computation of the Belief, it still requires a high number of function evaluations. This paper presents an algorithm that implements special heuristics to increase the probability to find the exact global Pareto front in the case of a min-max problem on multi-modal functions. Also, it presents an approach based on Kriging surrogate models to reduce the computational cost associated to the solution of the MOMMP. The combination of evolutionary algorithms and surrogate models has been object of considerable research efforts in the last decade, as summarized in [2], [3]. However, to the authors' knowledge, not many papers deal with surrogate-based min-max optimization. Ideas of combining evolutionary optimization and surrogate modelling to solving min-max problems were proposed in [4], [5]. More recently Marzat et al. [6] proposed a non-evolutionary approach using Kriging metamodels to reduce the computational cost of the solution of single objective min-max problem in worst-case scenario optimization.

The approach proposed in this paper addresses specifically multi-objective min-max optimization problems and differs from previous work in the way the surrogate model interfaces with the optimization algorithm. As it will be shown, the ultimate goal is to build a surrogate of the maximization process. The paper starts with a brief introduction to Evidence Theory and its use in the context of robust design optimization in Section II. Section III introduces an algorithm to compute a multi-objective optimal design solution under uncertainty. Section IV explains the use of surrogate modelling to reduce the computational cost in multi-objective optimization cases. Section V presents the results on some scalable test cases that were applied to assess the performance of the proposed algorithm.

## II. EVIDENCE-BASED ROBUST DESIGN OPTIMIZATION

Evidence Theory (also known as Dempster-Shafer Theory) [7] allows to adequately model both epistemic and aleatory uncertainty when no information on the probability distributions is available. For instance, during the prelimi-

nary design of an engineering system, experts can provide informed opinions by expressing their belief in an uncertain parameter  $u$  being within a certain set of intervals. The level of confidence an expert has in  $u$  belonging to one of the intervals is quantified by using a mass function generally known as Basic Probability Assignment (*bpa*). Note that the *bpa* is a belief rather than an actual probability. All the intervals form the so-called frame of discernment  $\Theta$ , which is a set of mutually exclusive elementary propositions. The frame of discernment can be viewed as the counterpart of the finite sample space in probability theory. The power set of  $\Theta$  is called  $U = 2^\Theta$ , or the set of all the subsets of  $\Theta$  (the uncertain space in the following). An element  $\theta$  of  $U$  that has a non-zero *bpa* is called focal element. When more than one parameter is uncertain, the focal elements are the result of the Cartesian product of all the elements of each power set associated to each uncertain parameter. The *bpa* of a given focal element is then the product of the *bpa* of all the elements in the power set associated to each parameter. All the pieces of evidence completely in support of a given proposition form the cumulative belief function *Bel*, whereas all the pieces of evidence partially in support of a given proposition form the cumulative plausibility function *Pl*. The Belief *Bel* and the Plausibility *Pl* functions are defined as follows:

$$Bel(A) = \sum_{\forall \theta_i \subseteq A} m(\theta_i) \quad (1)$$

$$Pl(A) = \sum_{\forall \theta_i \cap A \neq \emptyset} m(\theta_i) \quad (2)$$

where  $A$  is the proposition about which Belief and Plausibility need to be evaluated. For example, the proposition can be expressed as:

$$A = \{\mathbf{u} \in U \mid f(\mathbf{u}) \leq \nu\} \quad (3)$$

where  $f$  is the outcome of the system model and the threshold  $\nu$  is the desired value of a design budget (e.g. the mass). Thus, focal elements intercepting the set  $A$ , but not fully included in it, are considered in *Pl* but not in *Bel*. It is important to note that the set  $A$  can be disconnected or present holes, likewise the focal elements can be disconnected or partially overlapping.

An engineering system to be optimized can be modelled as a function  $f : D \times U \subseteq \mathbb{R}^{m+n} \rightarrow \mathbb{R}$ . The function  $f$  represents the model of the system budgets (e.g. power budget, mass budget, etc.), and depends on some uncertain parameters  $\mathbf{u} \in U$  and design parameters  $\mathbf{d} \in D$ , where  $D$  is the available design space and  $U$  the uncertain space. What is interesting for the designers is the value of the function  $f$  for which  $Bel = 1$ , i.e. it is maximum. This value of the design budget is the threshold  $\nu_{\max}$  above which the design is certainly feasible, given the current body of evidence. If  $q$  objective functions exist, then the following problem can be solved without considering all the focal elements:

$$\nu_{\max} = \min_{\mathbf{d} \in D} \mathbf{F} = \min_{\mathbf{d} \in D} [\max_{\mathbf{u} \in U} f_1(\mathbf{d}, \mathbf{u}), \dots, \max_{\mathbf{u} \in U} f_q(\mathbf{d}, \mathbf{u})]^T \quad (4)$$

Problem in 4 is a multi-objective min-max over the design space  $D$  and the uncertain space  $\bar{U}$ , where  $\bar{U}$  is a unit hypercube collecting all the focal elements in a compact set

with no overlapping or holes. The transformation between  $U$  and  $\bar{U}$  is given by:

$$\mathbf{x}_U = \frac{(\mathbf{b}_{U,i}^u - \mathbf{b}_{U,i}^l)}{(\mathbf{b}_{\bar{U},i}^u - \mathbf{b}_{\bar{U},i}^l)} \mathbf{x}_{\bar{U},i} + \mathbf{b}_{U,i}^l - \frac{(\mathbf{b}_{U,i}^u - \mathbf{b}_{U,i}^l)}{(\mathbf{b}_{\bar{U},i}^u - \mathbf{b}_{\bar{U},i}^l)} \mathbf{b}_{\bar{U},i}^l \quad (5)$$

where  $\mathbf{b}_{U,i}^u$  and  $\mathbf{b}_{U,i}^l$  (resp.  $\mathbf{b}_{\bar{U},i}^u$  and  $\mathbf{b}_{\bar{U},i}^l$ ) are the upper and lower boundaries of the  $i$ -th hypercube to which  $\mathbf{x}_{U,i}$  (resp.  $\mathbf{x}_{\bar{U},i}$ ) belongs.

### III. A MULTI-OBJECTIVE MIN-MAX EVOLUTIONARY ALGORITHM

Problem in 4 searches for the minimum of the maxima of all the functions over  $\bar{U}$  and represents an example of worst-case scenario design optimization. The maximum of every function is independent of the other functions and corresponds to a different uncertain vector. Therefore, all the maxima can be computed in parallel with  $q$  single-objective maximizations. The maximization of each function is performed by running a global optimization over  $\bar{U}$  using Inflationary Differential Evolution (IDEA). The minimization over  $D$  is performed by means of MACS $\nu$ , a modified version of MACS2.

IDEA [8] is a population-based memetic algorithm for single-objective optimization. It hybridizes Differential Evolution and Monotonic Basin Hopping paradigms in order to simultaneously improve local convergence and avoid stagnation, as demonstrated for some space trajectory optimization problems.

MACS2 [9] is a memetic algorithm for multi-objective optimization based on a combination of Pareto ranking and Tchebycheff scalarization. The search for non-dominated solutions is performed by a population of agents which combine individualistic and social actions. The initial population is randomly generated in the search domain  $D$ . Individualistic actions perform a sampling of the search space in a neighborhood  $N_p$  of each agent. Then, subsets of the population perform social actions aiming at following particular descent directions in the criteria space. Social agents implement a Differential Evolution operator and assess the new candidate solutions using Tchebycheff scalarization. Current non-dominated solutions are then stored in an archive. Both social and individualistic actions make use of a combination of the population and the archive.

MACS $\nu$  is the min-max variant of MACS2. Indeed, in a classical minimization problem two solutions  $\mathbf{d}_1$  and  $\mathbf{d}_2$  are ranked according to which one gives the lower value of the function. In the minimization loop of a min-max problem, the same can be done only if the maximization loop has returned the actual global maxima  $\tilde{\mathbf{u}}_1$  and  $\tilde{\mathbf{u}}_2$ . However, this is usually not true. Therefore a mechanism of cross-check such that also  $(\mathbf{d}_1, \mathbf{u}_2)$  and  $(\mathbf{d}_2, \mathbf{u}_1)$  are evaluated is needed in order to increase the probability that each maximization identifies the global maximum, and correctly rank two solutions. For this reason, MACS $\nu$  (see Algorithm 1) endows MACS2 with special heuristics. More in detail, Algorithm 2 is used to discriminate and archive the solutions that are Pareto dominant. A cross check is necessary to compare the values of the objective functions for a newly generated design vector against the function

values of an already archived solution. Consider, in fact, that a different design vector corresponds in general to a different landscape of the objective functions, and therefore to a different location of the maxima. Moreover, the cross-check performs a local search or a simple function evaluation in the inner maximization loop depending on whether the location of the maxima changes or not, respectively, for different design vectors.

At the end of this cross-check, Algorithm 3 is run to select the design vectors to attribute to the next generation, once a new candidate population is generated after either individualistic or social moves. The following heuristics are implemented: if  $\mathbf{d}$  (resp.  $\mathbf{u}$ ) is unchanged, the old  $\mathbf{u}$  (resp.  $\mathbf{d}$ ) is replaced with the new one, if it yields a better value of the objective function; if both  $\mathbf{d}$  and  $\mathbf{u}$  are different, the new vectors will replace the old ones.

A validation (see Algorithm 4) is run at the last iteration of MACS $\nu$ , after the individualistic and social moves have been performed. It performs a global search for the extremes in the archive, and replaces the corresponding uncertain vector and objective if the new ones give a better value of the objective function, until there is no more variation in the extremes of the archive. This step is introduced to mitigate the possibility that the cross-check operators assign the same incorrect  $\mathbf{u}$  to all  $\mathbf{d}$  vectors in the population and archive.

#### A. Optimization Performance Metrics

In order to assess the Pareto front  $f$  found by MACS $\nu$  with respect to a reference front  $g$ , we made use of convergence  $M_{conv}$  (Equation 6), and spreading  $M_{spr}$  (Equation 7), defined as:

$$M_{conv} = \frac{1}{N_p} \sum_{i=1}^{N_p} \min_{j \in M_p} 100 \left\| \frac{g_j - f_i}{g_j} \right\| \quad (6)$$

$$M_{spr} = \frac{1}{M_p} \sum_{i=1}^{M_p} \min_{j \in N_p} 100 \left\| \frac{f_j - g_i}{g_i} \right\| \quad (7)$$

where  $N_p$  and  $M_p$  are the cardinality of the fronts  $f$  and  $g$ , respectively. As one can see, the lower the value of  $M_{conv}$  and  $M_{spr}$ , the better convergence and spreading, respectively. In addition, two performance indexes  $p_{conv} = P(M_{conv} < t_{conv})$  and  $p_{spr} = P(M_{spr} < t_{spr})$  compute the success rate with respect to some thresholds  $t_{conv}$  and  $t_{spr}$ . These indexes tell how many times, over a certain number of runs, convergence and spreading are equal to or below such thresholds.

#### IV. SURROGATE MODELLING

The solution of the min-max problem involves an increase in the number of times the objective function needs to be evaluated. In this context, the use of surrogate models can play a valuable role. The surrogates are constructed using data drawn from high-fidelity models, and provide fast approximations of the objectives at new design points. In this paper we propose an approach in which a surrogate is built in the  $D$  space only, as shown in Figure 1(a). The surrogate in the design space has  $\mathbf{d}$  as design sites, and  $g$  as response, and is therefore constructed and updated in the outer loop of the optimization, i.e. the minimization over  $D$ .

---

#### Algorithm 1 MACS $\nu$

---

```

1: Initialize population  $P$ , archive  $A = P$ ,  $n_{feval} = 0$ ,
    $\epsilon = 0.7$ ,  $\delta = 10^{-6}$ 
2: while  $n_{feval} < n_{feval,max}$ 
3:   Run individualistic moves and generate trial population  $P_t$ 
4:   for all  $\mathbf{d} \in P_t$ 
5:     for all  $\mathbf{d}_{arch} \in A$ 
6:       if  $\mathbf{d} \succ \mathbf{d}_{arch}$ 
7:         CROSS-CHECK( $P_t, A$ ) ▷ Alg. 2
8:       end if
9:     end for
10:  end for
11:  MIN-MAX SELECTION( $P, P_t$ ) ▷ Alg. 3
12:  Update  $P$  and  $A$ 
13:   $Z \leftarrow \|\mathbf{F}_{arch}^{max} - \mathbf{F}_{arch}^{min}\|$ 
14:  Run social moves and generate candidate population  $P_s$ 
15:  for all  $\mathbf{d} \in P_s$ 
16:    for all  $\mathbf{d}_{arch} \in A$ 
17:      if  $\mathbf{d} \succ \mathbf{d}_{arch}$  or  $\|\mathbf{F}(\mathbf{d}) - \mathbf{F}(\mathbf{d}_{arch})\| > \epsilon Z$ 
18:        CROSS-CHECK( $P_s, A$ ) ▷ Alg. 2
19:      end if
20:    end for
21:  end for
22:  MIN-MAX SELECTION( $P, P_s$ ) ▷ Alg. 3
23:  Update  $P$  and  $A$ 
24:  VALIDATION( $A$ ) ▷ Alg. 4
25:  for all  $\mathbf{d} \in P$ 
26:    for all  $\mathbf{d}_{arch} \in A$ 
27:      if  $\mathbf{d} \succ \mathbf{d}_{arch}$  or  $\mathbf{d} \prec \mathbf{d}_{arch}$ 
28:        CROSS-CHECK( $P, A$ ) ▷ Alg. 2
29:      else if  $\|\mathbf{F}(\mathbf{d}) - \mathbf{F}(\mathbf{d}_{arch})\| < \delta$ 
30:        Replace  $\mathbf{u} \in P$  with  $\mathbf{u} \in A$ 
31:      end if
32:    end for
33:  end for
34:  Fit surrogate model in  $D$  with the elements of  $A$ 
35: end while

```

---

The approach can be formalized as follows. Let us consider, without loss of generality, Problem 4 in the single-objective case:

$$\nu_{\max} = \min_{\mathbf{d} \in D} \max_{\mathbf{u} \in \bar{U}} f(\mathbf{d}, \mathbf{u}) \quad (8)$$

This can also be written as

$$\nu_{\max} = \min_{\mathbf{d} \in D} g(\mathbf{d}, \mathbf{u}) \quad (9)$$

where

$$g(\mathbf{d}, \mathbf{u}) = \max_{\mathbf{u} \in \bar{U}} f(\mathbf{d}, \mathbf{u}) \quad (10)$$

i.e. it is the result of the inner loop of the optimization.

The aim of this approach is to separate the outer and inner loops. The idea behind this is that the inner loop can be called only if the surrogate  $(\mathbf{d}, g)$  needs to be updated. If the accuracy of the surrogate  $(\mathbf{d}, g)$  is above a certain threshold, only the outer loop is run, hence saving computational expense. In addition, such approach allows for the surrogate model to be dependent only on the relevant parameters, i.e. the design vector  $\mathbf{d}$  for the outer loop (Figure 1(b)).

**Algorithm 2** Cross-check

---

```

1: function CROSS-CHECK( $P, A$ )
2:   for all  $l \in \{1, \dots, q\}$ 
3:     if  $\|\mathbf{u}_{arch}^l - \mathbf{u}^l\| \neq 0$ 
4:       Compute local maxima  $\tilde{\mathbf{u}}_{arch}^l$  and  $\tilde{\mathbf{u}}^l$  associated to  $\mathbf{d}$  and  $\mathbf{d}_{arch}$ 
5:       if  $f^l(\mathbf{d}_{arch}, \tilde{\mathbf{u}}^l) \geq f^l(\mathbf{d}_{arch}, \mathbf{u}_{arch}^l)$ 
6:         replace  $\mathbf{u}_{arch}^l$  with  $\tilde{\mathbf{u}}^l$ 
7:       end if
8:       if  $f^l(\mathbf{d}, \tilde{\mathbf{u}}_{arch}^l) \geq f^l(\mathbf{d}, \mathbf{u}^l)$ 
9:         replace  $\mathbf{u}^l$  with  $\tilde{\mathbf{u}}_{arch}^l$ 
10:      end if
11:    end if
12:  end for
13:  return  $P, A$ 
14: end function

```

---

**Algorithm 3** Min-Max Selection

---

```

1: function MIN-MAX SELECTION( $P, P_{new}$ )
2:   for all  $l \in \{1, \dots, q\}$ 
3:     for all  $(\mathbf{d} \in P) \cup (\mathbf{d}_{new} \in P_{new})$ 
4:       if  $\|\mathbf{d}_{new} - \mathbf{d}\| = 0$ 
5:         if  $f_{new}^l(\mathbf{d}_{new}, \mathbf{u}_{new}^l) \geq f^l(\mathbf{d}, \mathbf{u}^l)$ 
6:           replace  $\mathbf{u}^l$  with  $\mathbf{u}_{new}^l$ 
7:           replace  $f_l$  with  $f_{new}^l$ 
8:         end if
9:       else if  $\|\mathbf{d}_{new} - \mathbf{d}\| > 0$ 
10:        if  $f_{new}^l(\mathbf{d}_{new}, \mathbf{u}_{new}^l) < f^l(\mathbf{d}, \mathbf{u}^l)$ 
11:          replace  $\mathbf{d}$  with  $\mathbf{d}_{new}$ 
12:          replace  $f_l$  with  $f_{new}^l$ 
13:        end if
14:      end if
15:    end for
16:  end for
17:  return  $P$ 
18: end function

```

---

The surrogate model is built by using the archived solutions of the minimization loop as design sites (see Algorithm 1, line 34). The surrogate is therefore built and updated only if there are new elements in the archive. Note that a strategy that builds the surrogate progressively as new agents are deployed into the minimization loop, therefore using all the design solutions, is not applicable because the

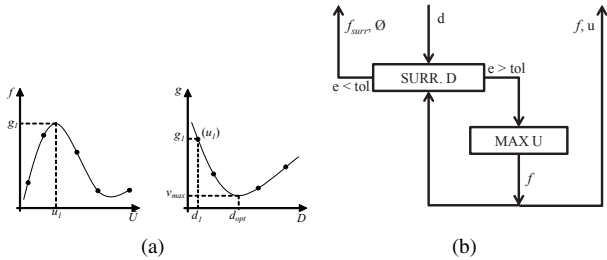


Fig. 1. (a) Conceptual example of the surrogate modelling strategy. The plot represents the surrogate in the  $D$  space. The dots are the design sites. Notice that  $g_1 = \max_u f(d_1, u)$ . (b) Schematic of the surrogate modelling strategy.

**Algorithm 4** Validation

---

```

1: function VALIDATION( $A$ )
2:   for all  $l \in \{1, \dots, q\}$ 
3:      $\Delta f_{best} = 1$ 
4:     while  $\Delta f_{best} \neq 0$ 
5:        $j = \text{argmin } f^l \in A$ 
6:       Run global optimization over  $\bar{U}$  and compute new  $\bar{f}^l$  and associated  $\bar{\mathbf{u}}^l$ 
7:       if  $\bar{f}^l > f^l$ 
8:         replace  $\mathbf{u}^l \in A$  with  $\bar{\mathbf{u}}^l$ 
9:         replace  $f^l \in A$  with  $\bar{f}^l$ 
10:       $\Delta f_{best} = \bar{f}^l - f^l$ 
11:    end if
12:  end while
13: end for
14: return  $A$ 
15: end function

```

---

design solutions need to be cross-checked. Without these steps, the surrogate would contain design solutions that, while being non-dominated, do not maximize the inner loop. A further feature of our technique is that the number of design sites is kept below a user-specified threshold of  $n_s$  points. Once the number of sample points overcomes such threshold, the sites to be retained are chosen so that the spreading of the Pareto front is preserved, and the members of the archive evenly distributed. By noting that the design sites are, as explained in the previous paragraph, the current archive of Pareto set and front, one can define a sector in the criteria space defined by the origin and the extrema of the front. Then, such sector is divided into  $n_s - 1$  smaller sectors of equal angle, and the design sites to be retained are the  $n_s$  ones which distance to the boundaries between the sectors is minimum. This allows for the surrogate to approximate accurately the current Pareto front, and also to keep the surrogate to a reasonable size.

**A. Kriging Predictor**

Starting from a set of sample points, or design sites, the Kriging predictor is an interpolation technique that makes use of a regression function and a correlation model to predict the response of a function at a desired point. Being an interpolation method, it gives an exact prediction of the response at the sample points. Moreover, it assumes that the output function values are correlated in design space, i.e. closer points are more highly correlated. A complete derivation of the Kriging model can be found in Jones [10]. If we suppose to have a set of  $n$  design sites  $(\mathbf{x}, \mathbf{y})$ , the correlation matrix  $\mathbf{R}$  of the design points can be expressed as

$$\mathbf{R} = [R_{ij}] = \exp \left[ - \sum_{l=1}^d \theta_l \|\mathbf{x}_{il} - \mathbf{x}_{jl}\|^{p_l} \right] \quad (11)$$

where  $i, j = 1, \dots, n$ . In the same way, the correlation of the new point  $\mathbf{x}^*$  at which we want to predict the response  $\hat{y}(\mathbf{x}^*)$ , with the design points can be expressed as

$$\mathbf{r} = [r_i] = \exp \left[ - \sum_{l=1}^d \theta_l \|\mathbf{x}_i^* - \mathbf{x}_{il}\|^{p_l} \right] \quad (12)$$

The correlation model depends on two parameters  $\theta$  and  $p$ . They can be found to be the ones that minimize the mean squared error between the predicted response  $\hat{y}$  and the actual response  $y$ . Therefore an optimization is to be performed during the training phase. Under the assumption that the regression function is a zero-th order polynomial, i.e. it is a  $n \times 1$  vector of ones  $\mathbf{1}$ , the prediction  $\hat{y}(\mathbf{x}^*)$  can be found to be

$$\hat{y}(\mathbf{x}^*) = \hat{\mu} + \mathbf{r}^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{1} \hat{\mu}) \quad (13)$$

where

$$\hat{\mu} = \frac{\mathbf{1}^T \mathbf{R}^{-1} \mathbf{y}}{\mathbf{1}^T \mathbf{R}^{-1} \mathbf{1}} \quad (14)$$

One of the key benefits of Kriging is the provision of an estimated error of its predictions. The estimated mean squared error (MSE) for a Kriging model is

$$s^2(\mathbf{x}^*) = \hat{\sigma}^2 \left[ 1 - \mathbf{r}^T \mathbf{R}^{-1} \mathbf{r} + \frac{(\mathbf{1} - \mathbf{r}^T \mathbf{R}^{-1} \mathbf{r})^2}{\mathbf{1}^T \mathbf{R}^{-1} \mathbf{1}} \right] \quad (15)$$

where

$$\hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{1} \hat{\mu})^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{1} \hat{\mu})}{n} \quad (16)$$

is the estimated variance.

The availability of an estimate of the prediction error is a very convenient feature, as it can be exploited in the surrogate update strategy. Letting  $y_{min}$  be the current best function value, and  $Y(\mathbf{x})$  a random variable normally distributed, with mean  $\hat{y}(\mathbf{x})$  and standard deviation  $s(\mathbf{x})$ , describing the uncertainty about the function's value, an improvement  $I$  is achieved if  $I(\mathbf{x}) = y_{min} - \hat{y}(\mathbf{x}) > 0$ . In this paper we consider a method based on the probability of improvement, that is deemed to be one of the best two-stage methods to the Kriging update [10] [11]. The main advantage of this method is that the probability of improvement is independent of magnitude and units of the function value.

The probability of an improvement is the area of the PDF, with mean  $\hat{y}(\mathbf{x})$  and standard deviation  $s(\mathbf{x})$ , calculated from  $-\infty$  to  $y_{min}$ . That is

$$P(I(\mathbf{x})) = \Phi(u(\mathbf{x})) \quad (17)$$

where  $\Phi$  is the normal cumulative distribution function, and

$$u(\mathbf{x}) = \frac{y_{min} - \hat{y}(\mathbf{x})}{s(\mathbf{x})} = \frac{I(\mathbf{x})}{s(\mathbf{x})} \quad (18)$$

One can set a threshold  $\tau$  on this probability of improvement, and update the Kriging predictor if  $P(I) \geq \tau$ . The advantage of such method is that the threshold  $\tau$  is a simple non-dimensional probability measure, therefore independent of the magnitude and units of the function value.

The probability of improvement can be extended to the case of multiple objectives. In this case, an improvement is achieved when the new point dominates at least one member of the archived non-dominated points. Considering two uncorrelated objectives, the improvement  $I(\mathbf{x})$  is then defined as  $y_{1P} - \hat{y}_1(\mathbf{x}) > 0 \cup y_{2P} - \hat{y}_2(\mathbf{x}) > 0$ , where  $y_{1P}$  and  $y_{2P}$  represent the members of the archived Pareto

TABLE I. TEST FUNCTIONS.

ID	Function
MV1	$f = \sum_{i=1}^n d_i u_i^2$
MV2	$f = \sum_{i=1}^n (d_i - u_i)^2$
MV3	$f = \sum_{i=1}^n (5 - d_i) (1 + \cos u_i) + (d_i - 1) (1 + \sin u_i)$
MV8	$f = \sum_{i=1}^n (2\pi - u_i) \cos(u_i - d_i)$
MV9	$f = \sum_{i=1}^n (d_i - u_i) \cos(-5u_i + 3d_i)$
MV10	$f = \sum_{i=1}^n (d_i + u_i) \cos(-u_i(5 d  + 5) + 3d_i)$
EM1	$f = \sum_{i=1}^n (u_i - 3d_i) \sin u_i + (d_i - 2)^2$

TABLE II. TEST CASES.

Test Case	Functions	$\mathbf{d}$	$n$
TC1	$f_1 = \text{MV1}, f_2 = \text{MV3}$	$[1, 5]^n$	2
TC2	$f_1 = \text{MV2}, f_2 = \text{MV8}$	$[0, 3]^n$	8
TC3	$f_1 = \text{MV2}, f_2 = \text{EM1}$	$[1, 5]^n$	8
TC4	$f_1 = \text{MV8}, f_2 = \text{MV9}$	$[1, 3]^n$	2
TC5	$f_1 = \text{MV8}, f_2 = \text{EM1}$	$[1, 5]^n$	4
TC6	$f_1 = \text{MV10}, f_2 = \text{MV9}$	$[-4, 2\pi]^n$	1

TABLE III. *bpa* STRUCTURE.

MV1, MV2, MV3	Interval	[-5 -4]	[-3 0]	[-1 3]
	<i>bpa</i>	0.1	0.25	0.65
MV8	Interval	[0 1]	[2 4]	[3 2\pi]
	<i>bpa</i>	0.1	0.25	0.65
MV9	Interval	[-\pi/2 -\pi/6]	[0 \pi]	[3\pi/4 3\pi/2]
	<i>bpa</i>	0.1	0.4	0.5
MV10	Interval	[\pi 4]	[5 6]	[5.5 2\pi]
	<i>bpa</i>	0.1	0.25	0.65
EM1	Interval	[0 5]	[7 14]	[12 20]
	<i>bpa</i>	0.1	0.5	0.4

front. If the Pareto front is composed of  $N_P$  members, by defining

$$u_1^i(\mathbf{x}) = \frac{y_{1P} - \hat{y}_1(\mathbf{x})}{s_1(\mathbf{x})}, \quad u_2^i(\mathbf{x}) = \frac{y_{2P} - \hat{y}_2(\mathbf{x})}{s_2(\mathbf{x})} \quad (19)$$

for  $i = 1, \dots, N_P$ , the probability of improvement is

$$P(I(\mathbf{x})) = \Phi(u_1^1(\mathbf{x})) + \sum_{i=1}^{N_P-1} [\Phi(u_1^{i+1}(\mathbf{x})) - \Phi(u_1^i(\mathbf{x}))] \times \Phi(u_2^{i+1}(\mathbf{x})) + [1 - \Phi(u_1^{N_P}(\mathbf{x}))] \Phi(u_2^{N_P}(\mathbf{x})) \quad (20)$$

This expression can be obtained by integrating the Gaussian distribution underlying the improvement  $I$  for the two objective functions between  $-\infty$  and the Pareto front [12].

## V. TEST CASE

The surrogate-based optimization algorithm described in the previous sections was tested on six bi-objective test cases. The chosen functions, reported in Table I, have the property of being easily scalable. More in detail, functions MV9, MV10, and EM1 present very challenging landscapes, with multiple maxima that can change significantly with the design vector. Function MV10, in particular, is characterized by having the maxima located on top of multiple sharp, steep peaks. Such functions can also be deceptive for the Kriging surrogate. The six test cases are composed of the functions in Table I as reported in Table II, where  $n$  is the dimension of the design vector  $\mathbf{d}$ , as well as the uncertain vector  $\mathbf{u}$ , therefore the total dimension of the test cases is  $2n$ . The uncertain vector  $\mathbf{u}$  is assigned the *bpa* structure reported in Table III. Note that test cases TC4 and TC6 are very challenging, as they have multiple maxima and TC6 also a disconnected front.

TABLE IV. SENSITIVITY OF THE ALGORITHM TO THE NUMBER OF AGENTS

$s.r.(f_1, f_2)$ ( $M_{conv}, M_{spr}$ )		Total agents		
		5	10	20
Social/total	1/3	(100, 82.4) (1.9, 3.6)	(100, 89.9) (1.1, 2.6)	(100, 97.4) (0.7, 1.9)
	1/2	(100, 84.2) (1.8, 3.5)	(100, 87.9) (1.4, 2.7)	(100, 96.8) (0.8, 2.0)
	1	(100, 77.3) (2.5, 3.4)	(100, 88.1) (1.4, 2.7)	(100, 96.6) (0.8, 2.1)

TABLE V. SENSITIVITY OF THE ALGORITHM TO  $F$  AND  $C_R$  IN THE INNER LOOP

$s.r.(f_1, f_2)$ ( $M_{conv}, M_{spr}$ )		$F$			
		0.1	0.5	1	2
$C_R$	0.1	(100, 81.7) (2.2, 3.3)	(100, 87.9) (1.4, 2.7)	(100, 91.5) (0.9, 2.7)	(100, 91.4) (1.0, 2.7)
	0.5	(100, 80.0) (2.4, 3.7)	(100, 90.1) (1.1, 2.7)	(100, 91.6) (0.9, 2.6)	(100, 88.3) (1.5, 3.0)
	0.9	(100, 78.5) (2.4, 3.5)	(100, 82.4) (1.8, 3.0)	(100, 87.8) (1.5, 2.9)	(100, 84.5) (1.6, 2.8)

A sensitivity analysis was conducted in order to assess how the several parameters of the solver affect its performance. In particular, coupled analyses were carried out for total number of agents and number of social agents for MACS $\nu$ , and  $F$  vs.  $C_R$  for IDEA. Note that, as shown in [9],  $F$  and  $C_R$  of MACS $\nu$  do not have a big impact. The sensitivity analysis was run for test case TC4 with a total of 2E5 function evaluations. The results were assessed in terms of success rate of finding the global maximum in the inner loop, as well as convergence and spreading as per Equations 6 and 7. In Tables IV - V the numbers in brackets are the success rate of the maximization for the two objectives, indicated as  $s.r.(f_1, f_2)$ , and convergence and spreading. The best results are obtained for 20 agents for MACS $\nu$ , half of which perform the social actions, and  $F = 1$  and  $C_R = 0.1$ . The total number of function evaluations was set equal to 2E5 $n$ , and 200 $n$  for the inner loop. The Kriging predictor makes use of a zero-th order polynomial regression function and a Gaussian correlation function. The number of design sites is kept below 20. Three thresholds for the surrogate update have been tested, and they were set to 0.3, 0.5 and 0.9, i.e. 30%, 50% and 90% probability of improvement. The reference solution, i.e. the real front in Figures 2 - 4, was computed by merging the results of 200 runs of the same problem without using the surrogate model. Tables VI to VIII summarize the performance metrics of the surrogate-based optimization for the six test cases at the several thresholds for the surrogate update. The results are the average performances obtained from the 200 runs needed to achieve a confidence interval of 95% on the success rate being within a  $\pm 5\%$  interval containing its estimated value [8]. The column *Inner/Outer* contains the percentage of times that the inner loop, hence the real function, was called. The columns  $M_{conv}$  and  $M_{spr}$  contain the mean value of  $M_{conv}$  and  $M_{spr}$  respectively of the surrogate-based MACS $\nu$ , and in brackets for MACS $\nu$  without the surrogate, for the sake of comparison. The columns  $p_{conv}/t_{conv}$  and  $p_{spr}/t_{spr}$  contain the success rate of computing a front which convergence and spreading are below the thresholds  $t_{conv}$  and  $t_{spr}$  also contained in the columns after the '/' symbol. For this paper the thresholds were set equal to the mean values plus the standard deviation of  $M_{conv}$  and  $M_{spr}$  of

TABLE VI. PERFORMANCE METRICS FOR SURROGATE UPDATE THRESHOLD OF 30%.

Test Case	Inner/Outer	$M_{conv}$	$M_{spr}$	$p_{conv} / t_{conv}$	$p_{spr} / t_{spr}$
TC1	63.6%	0.3(0.2)	2.0(1.7)	94(100) / 1	82(87) / 3
TC2	92.4%	0.8(0.8)	3.9(4.0)	100(98) / 1	75(77) / 5
TC3	87.9%	0.7(0.5)	3.6(1.8)	85(98) / 1	50(90) / 3
TC4	69.8%	2.1(0.3)	0.9(0.9)	44(98) / 1	96(96) / 2
TC5	92.6%	2.5(1.7)	1.9(1.3)	71(91) / 3	72(94) / 2
TC6	74.3%	2.2(0.3)	4.3(1.1)	23(100) / 1	44(99) / 3

TABLE VII. PERFORMANCE METRICS FOR SURROGATE UPDATE THRESHOLD OF 50%.

Test Case	Inner/Outer	$M_{conv}$	$M_{spr}$	$p_{conv} / t_{conv}$	$p_{spr} / t_{spr}$
TC1	44.3%	1.4(0.2)	7.7(1.7)	78(100) / 1	46(87) / 3
TC2	86.6%	0.8(0.8)	5.0(4.0)	98(98) / 1	59(77) / 5
TC3	52.1%	4.0(0.5)	12.4(1.8)	3(98) / 1	0(90) / 3
TC4	61.3%	7.2(0.3)	3.9(0.9)	32(98) / 1	66(96) / 2
TC5	57.4%	7.7(1.7)	8.6(1.3)	27(91) / 3	7(94) / 2
TC6	63.1%	4.7(0.3)	10.6(1.1)	14(100) / 1	33(99) / 3

TABLE VIII. PERFORMANCE METRICS FOR SURROGATE UPDATE THRESHOLD OF 90%.

Test Case	Inner/Outer	$M_{conv}$	$M_{spr}$	$p_{conv} / t_{conv}$	$p_{spr} / t_{spr}$
TC1	32.4%	0.7(0.2)	17.3(1.7)	88(100) / 1	6(87) / 3
TC2	86.4%	0.8(0.8)	5.6(4.0)	97(98) / 1	46(77) / 5
TC3	51.1%	4.3(0.5)	12.6(1.8)	1(98) / 1	0(90) / 3
TC4	49.0%	3.0(0.3)	4.9(0.9)	58(98) / 1	31(96) / 2
TC5	53.4%	6.3(1.7)	8.4(1.3)	19(91) / 3	2(94) / 2
TC6	40.3%	11.5(0.3)	51.1(1.1)	20(100) / 1	6(99) / 3

MACS $\nu$  without the surrogate, rounded to the higher natural number.

The Pareto fronts for the six test cases are shown in Figures 2 to 4 for the three surrogate update threshold of 30%, 50% and 90%. As a comparison, also the fronts computed by MACS $\nu$  without the surrogate, of which performance metrics are in brackets in Tables VI to VIII, are displayed.

With reference to Tables VI to VIII, some considerations can be made:

- The inner loop is called progressively less times as the surrogate update threshold increases, as expected;
- There is a variability of the computational expense among the test cases for the same thresholds: This is due to the fact that certain functions are more or less complicated to approximate by the surrogate model;
- Overall performances degrade, even substantially for some test cases, as the surrogate is updated less frequently;
- Higher surrogate update thresholds affect more the spreading than the convergence: This is an indication that the spreading-preserving technique to downsize the surrogate is effective. Indeed, the higher the threshold, the less frequent is the surrogate update, the less design sites are contained in surrogate, and the less times the surrogate downsizing is performed.

By and large, the surrogate-based MACS $\nu$  produces very good results for all the test cases when the surrogate update threshold is 30%, with a computational saving ranging from about 8% to about 35%. For instance, test case TC1 could be solved at 64% of the exact computational expense without loss of performance. No loss of performance was

also obtained for TC2, but with a computational saving of 8%. Note, however, that TC1 has a total of 4 dimensions, whereas TC2 has 16. The algorithm performed fairly well also for test cases TC3 and TC5, with a 10% reduction of the computational cost. Test case TC4 is an interesting one because the rightmost portion of the front (see Figure 2(d)) has multiple fronts above it, making very challenging the identification of the global one. Indeed, *surMACS<sub>v</sub>* could identify it, but not accurately compute it. TC6 is also interesting in that it presents a disconnected front. Nevertheless, it could be accurately computed with a 25% reduction of the computational cost. However, the success rates for TC6 were lower than for the other test cases.

Increasing the update threshold obviously causes the computational saving to be higher because less points are found that have 50% probability of improving the surrogate. By looking at the fronts in Figure 3 one can see that the surrogate-based *MACS<sub>v</sub>* is able to find, very accurately for three out of six test cases, the real fronts, at a computational cost that ranges between 40% and 60% the exact one. However, this also caused a decrease of the quality of the results in terms of both convergence and spreading. Also, note that the success rates  $p_{conv}$  and  $p_{spr}$  decrease substantially for some test cases, TC3 for example. This means that it is more difficult for the algorithm to compute a front which convergence and spreading are lower than the thresholds  $p_{conv}$  and  $p_{spr}$ . Nevertheless, the solution found is still good for all the six test cases, even for the disconnected front of TC6.

When the surrogate update threshold is pushed as high as 90% (Figure 4), the fronts remain good for four of the six test cases. At this threshold the computational saving is around 50% for four test cases, as high as 68% for TC1, and 14% for TC2. It is interesting to note that, for TC6, further increasing the threshold to 90% causes the surrogate to “interpolate” the portions of the front without finding them accurately, yet the region in the objective space where the real front lies could still be identified.

## VI. CONCLUSION

An algorithm for surrogate-based multi-objective worst-case optimization has been presented. The algorithm, called *MACS<sub>v</sub>*, has the peculiarity of performing a number of checks and cross checks to increase the probability to find the global optimum. In this paper, the optimization algorithm is coupled with surrogate modelling. The aim is to create a surrogate model that approximates the mapping between design parameters and function values, so that the expensive inner loop of the min-max algorithm is called only a fraction of the times. The algorithm was tested on six challenging bi-objective test cases made of scalable test functions, presenting multiple maxima and challenging Pareto fronts. An analysis of sensitivity of the parameters of the algorithm was carried out, showing how several combinations of total vs. social agents, and  $F$  vs.  $C_R$  affected the performance of the algorithm. Moreover, three surrogate update thresholds were tested. The results were good for four out of six test cases, with a good approximation of the real Pareto front obtained with an average computational cost of 50% of the exact one. In the case of disconnected fronts, the computed

front lays on the real one, though the disconnected portions of the front were interpolated but not completely identified. However, this study is limited to the use of the probability of improvement as surrogate update strategy. Other techniques, such as the expected improvement, were proposed in the framework of surrogate-based optimization, and their use in conjunction with *MACS<sub>v</sub>* is under investigation. A real case application of *MACS<sub>v</sub>* is the maximization of Belief function, which finds importance in engineering robust design, for example, as it gives the value of the design budgets for which the design is certainly feasible under uncertainty.

## ACKNOWLEDGMENT

The Kriging predictor used for this work is the one implemented in the DACE Toolbox by Nielsen, Lophaven and Sondergaard, and freely available at <http://www2.imm.dtu.dk/hbni/dace/>. The authors would like to thank the anonymous reviewers for helping to improve the quality of the paper. This work is partially supported through an ESA/NPI grant ‘Evidence-based Robust Design Optimization’.

## REFERENCES

- [1] M. Vasile, E. Minisci and Q. Wijnands, “Approximated Computation of Belief Functions for Robust Design Optimization,” *53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, Apr. 2012.
- [2] M. Bhattacharya, “Surrogate based Evolutionary Algorithm for Design Optimization,” *World Academy of Science, Engineering and Technology* vol. 1, 2007
- [3] Y. Jin, “Surrogate-Assisted Evolutionary Computation: Recent Advances and Future Challenges,” *Swarm and Evolutionary Computation* vol. 1, pp 61-70, 2011
- [4] A. Zhou and Q. Zhang, “A Surrogate-Assisted Evolutionary Algorithm for Minimax Optimization,” *2010 IEEE Conference on Evolutionary Computation (CEC 2010)*, Jul. 2010.
- [5] Y.-S. Ong, P. B. Nair, K. Y. Lum, “Max-Min Surrogate-Assisted Evolutionary Algorithm for Robust Design,” *IEEE Trans. Evol. Comp.* vol. 10, pp 392-404, 2006
- [6] J. Marzat, E. Walker and H. Piet-Lahanier, “Worst-case Global Optimization of Black-Box Functions through Kriging and Relaxation,” *J. Global Optim.* vol. 55, pp. 707-727, 2013.
- [7] G. Shafer, *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ, 1976.
- [8] M. Vasile, E. Minisci and M. Locatelli, “An Inflationary Differential Evolution Algorithm for Space Trajectory Optimization,” *IEEE Trans. Evol. Comp.* vol. 15, pp. 267-281, Apr. 2011.
- [9] F. Zuiani and M. Vasile, “Multi Agent Collaborative Search Based on Tchebycheff Decomposition,” *Computational Optimization and Applications*, Mar. 2013.
- [10] D. R. Jones, “A Taxonomy of Global Optimization Methods Based on Response Surface,” *J. Global Optim.* vol. 21, pp. 345-383, 2001.
- [11] A. I. J. Forrester and A. J. Keane, “Recent Advances in Surrogate-based Optimization,” *Prog Aerosp Sci.* vol. 45, pp. 50-79, 2009.
- [12] A. J. Keane, “Statistical Improvement Criteria for Use in Multiobjective Design Optimization,” *AIAA Journal*, vol. 44, no. 4, pp. 879-891, 2006.

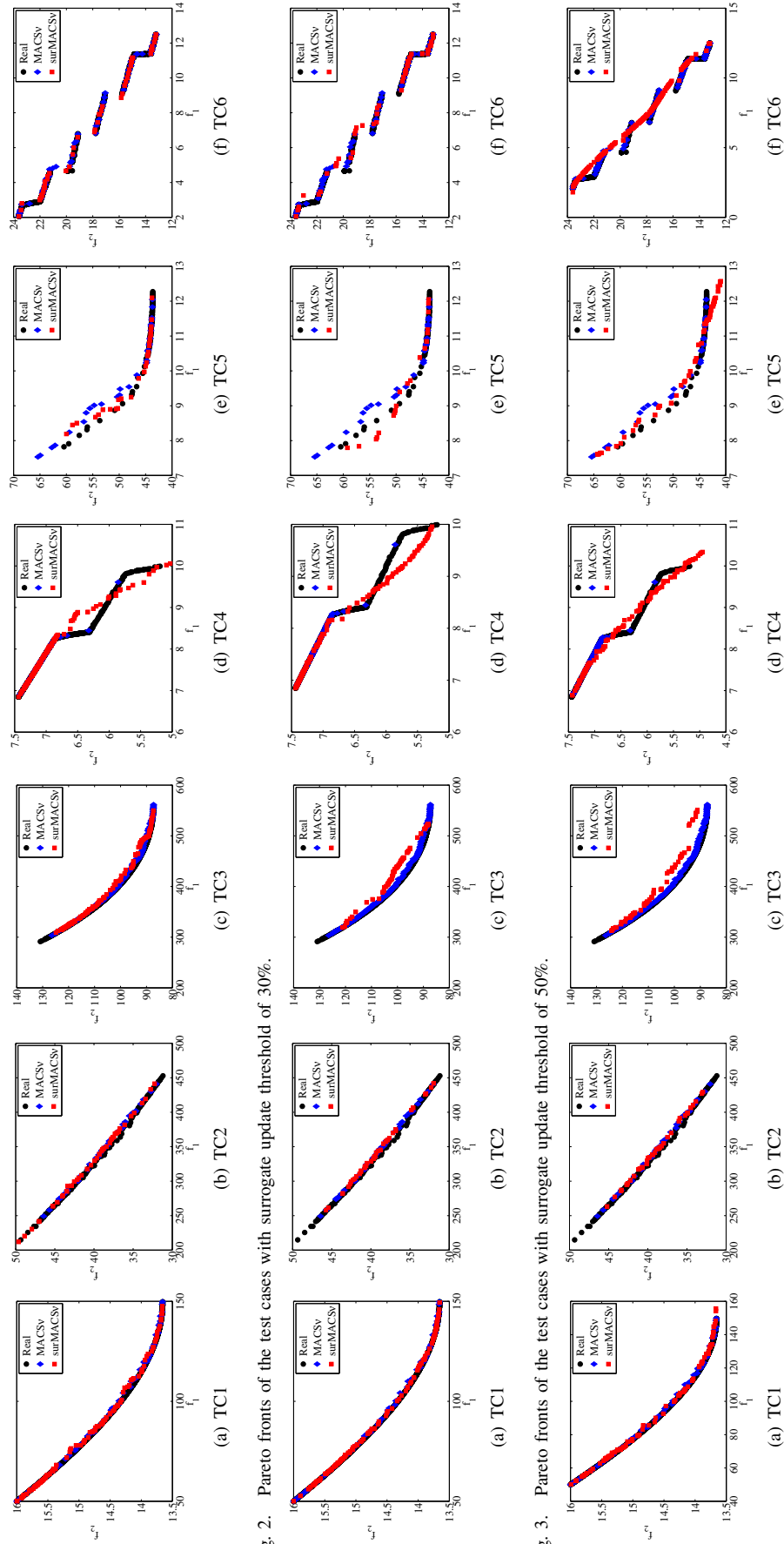


Fig. 2. Pareto fronts of the test cases with surrogate update threshold of 30%.

Fig. 3. Pareto fronts of the test cases with surrogate update threshold of 50%.

Fig. 4. Pareto fronts of the test cases with surrogate update threshold of 90%.