J. Yaghob (Ed.): ITAT 2015 pp. 159–166
Charles University in Prague, Prague, 2015

Investigation of Gaussian Processes in the Context of Black-Box Evolutionary Optimization

Andrej Kudinov[1], Lukáš Bajer[2], Zbyněk Pitra[3], and Martin Holeňa[4]

[1] Faculty of Information Technology, Czech Technical University, Prague, kudinand@fit.cvut.cz,
[2] Faculty of Mathematics and Physics, Charles University, Prague, bajeluk@gmail.com,
[3] Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University, Prague, z.pitra@gmail.com,
[4] Czech Academy of Sciences, Prague, martin@cs.cas.cz

Abstract: Minimizing the number of function evaluations became a very challenging problem in the field of black-box optimization, when one evaluation of the objective function may be very expensive or time-consuming. Gaussian processes (GPs) are one of the approaches suggested to this end, already nearly 20 years ago, in the area of general global optimization. So far, however, they received only little attention in the area of evolutionary black-box optimization.

This work investigates the performance of GPs in the context of black-box continuous optimization, using multimodal functions from the CEC 2013 competition. It shows the performance of two methods based on GPs, Model Guided Sampling Optimization (MGSO) and GPs as a surrogate model for CMA-ES. The paper compares the speed-up of both methods with respect to the number of function evaluations using different settings to CMA-ES with no surrogate model.

1 Introduction

Evolutionary computation became very successful during the past few decades in continuous black-box optimization. In such optimization, we have no mathematical definition of the optimized function available, thus we can calculate analytically neither that function itself, nor its derivatives. In such cases, there is no option but to empirically evaluate the objective function through measurements, tests or simulations.

In various real-world optimization problems, the evaluation of the objective function is very expensive or time-consuming, e.g., protein's folding stability optimization [6], computer-assisted design [1] and job allocations in a computational grid [13]. In such cases, we need to keep the number of function evaluations as low as possible, without impairing the quality of expected results.

In this paper, we employ two approaches addressing that task. The first, called Model Guided Sampling Optimization (MGSO) [2], is one of the recent implementations of GPs. The second employed approach is surrogate modeling, recalled in Subsection 2.2, which we will use in conjunction with CMA-ES.

This work investigates the performance of both methods on the set of niching functions from the CEC 2013 competition [10], characterized by a high number of local optima, which makes evolutionary search for the global optimum difficult because evolutionary algorithms (EAs) tend to get trapped in one of the local optima.

The following section describes the theoretical fundamentals of GPs and introduces the MGSO method. It also explains surrogate modeling and using GPs as a surrogate model for CMA-ES. Section 3 presents results of experimental evaluation of the considered methods. Section 5 summarizes the results and concludes the paper.

2 Gaussian Processes in Optimization

GP is a random process such that any finite sequence $X_1, \dots, X_k$ of the involved random variables has a multivariate Gaussian distribution. GP is defined by its mean value and covariance matrix described by a function with relatively small number of hyper-parameters, which are usually fitted by the maximum likelihood method. Firstly, GP is trained with N data points from the input space $\mathbb{X}$,

$$\mathbf{X}_N = \{\mathbf{x}_i | \mathbf{x}_i \in \mathbb{R}^D\}_{i=1}^N$$

with known input-output values $(\mathbf{x}_N, \mathbf{y}_N)$, then it is used for predicting the $(N+1)$-st point. The conditional density of the extended vector $\mathbf{y}_{N+1} = (y_1, \dots, y_N, y_{N+1})$, conditioned on $\mathbf{X}_{N+1} = \mathbf{X}_N \cup \{\mathbf{x}_{N+1}\}$ is

$$p(\mathbf{y}_{N+1}|\mathbf{X}_{N+1}) = \frac{\exp(-\frac{1}{2}\mathbf{y}_{N+1}^{\mathrm{T}} \mathbf{C}_{N+1}^{-1} \mathbf{y}_{N+1})}{\sqrt{(2\pi)^{N+1} \det(\mathbf{C}_{N+1})}},$$

where $\mathbf{C}_{N+1}$ is the covariance matrix of a $(N+1)$-dimensional Gaussian distribution. The covariance matrix can be expressed as

$$\mathbf{C}_{N+1} = \begin{pmatrix} \mathbf{C}_N & \mathbf{k} \\ \mathbf{k}^{\mathrm{T}} & \kappa \end{pmatrix},$$

where κ is the variance of the new point itself, $\mathbf{k}$ is is the vector of covariances between the new point and training data and $\mathbf{C}_N$ is the covariance of the Gaussian distribution corresponding to the N training data points [5].

Covariance functions provide prior information about the objective function and express the covariance between the function values of each two data points $\mathbf{x}_i$, $\mathbf{x}_j$ as $cov(f(\mathbf{x}_i), f(\mathbf{x}_j)) = k(\mathbf{x}_i, \mathbf{x}_j)$. Because the matrix $x_1, \dots, x_N$ serves as a covariance matrix, it has to positive semidefinite.

2.1 Model Guided Sampling Optimization

MGSO has the ability to use a regression model for prediction and error estimation in order to get the probability of obtaining a better solution. It was inspired by two previously proposed methods in the field of black-box optimization. The first method, Estimation of Distribution Algorithms [9], creates a new set of solutions for the next generation using estimated probability distribution from previously selected candidate solutions. The second approach is surrogate modeling, described in Section 2.2.

MGSO was proposed as an alternative method for Jones' Efficient Global Optimization (EGO) [8]. Unlike EGO, MGSO produces not a single solution, but a whole population of solutions. The selection of candidate solutions is performed by sampling the probability of improvement (PoI) of the GP model, which serves as a measure of how promising the chosen point is for locating the optimum. PoI is determined by means of a chosen threshold T and the estimation of the objective function shape by the current GP model.

The crucial step in the MGSO algorithm is the sampling of PoI, which is determined by the predicted mean $\hat{f}(\mathbf{x}) = \hat{y}$ and the standard deviation $\hat{s}(\mathbf{x}) = s_y$ of the GP model $\hat{f}$ in any point $\mathbf{x}$ of the input space

$$\mathrm{PoI}_T(\mathbf{x}) = \Phi\left(\frac{T - \hat{f}(\mathbf{x})}{\hat{s}(\mathbf{x})}\right) = \mathrm{P}(\hat{f}(\mathbf{x} \leq T)),$$

which corresponds to the value of cumulative distribution function of the Gaussian for the value of threshold T. Although all the variables are sampled from Gaussian distribution, $\mathrm{PoI}(\mathbf{x})$ is not Gaussian-shaped as it depends on the threshold T and the modelled function f.

2.2 GP as Surrogate Model for CMA-ES

Surrogate modeling is a technique used in optimization in order to decrease the number of expensive function evaluations. A surrogate model, which is a regression model of suitable kind (in our case a GP), is constructed by training with known values of the objective function for some inputs first, and then it is used by the employed evolutionary optimization algorithm instead of the original objective function (in evolutionary optimization usually called fitness) during the search for the global optimum. Although, creating and training a model increases time complexity of the optimization algorithm, using a model instead of the original fitness function decreases the number of its evaluations, which is a crucial objective in expensive optimization.

Every regression model approximates the original fitness function with some error. To prevent the optimization from being mislead from such an erroneous approximation, it is necessary to use the original fitness function for some subset of evaluations. That subset is determined by the evolution control (EC) strategy [5].

An *individual-based* EC strategy consists in determining the subset of individuals evaluated by the original fitness function in each generation. The following description is illustrated by Figure 1. Denote λ to be the size of the population provided by CMA-ES. First, $\lambda'' = \alpha\lambda$ points are sampled from $N(m, \sigma^2\mathbf{C})$, where $\alpha \in [0,1]$, m is the mean, σ is the step-size and $\mathbf{C}$ stands for the covariance matrix (1). These λ'' points are evaluated by the original fitness function and included in training the model. Then, the extended population $\lambda' = \beta(\lambda - \lambda'')$, where $\beta \in [1,\infty)$, is sampled by a model using the same distribution (2). The extended population is required by the model for choosing promising points for re-evaluation by the original fitness function. Subsequently, $\gamma(\lambda - \lambda'')$ points, where $\gamma \in [0,1]$, are chosen according to some criterion from among the extended population, e.g. fitness value, and used in the evaluation by the original fitness function (3). The complement to λ points is gathered from the rest of the extended population by dividing it into $k = (1-\gamma)(\lambda - \lambda'')$ clusters and selecting the best point from each cluster, which are also evaluated by the original fitness function and added to the final population (4) [3].

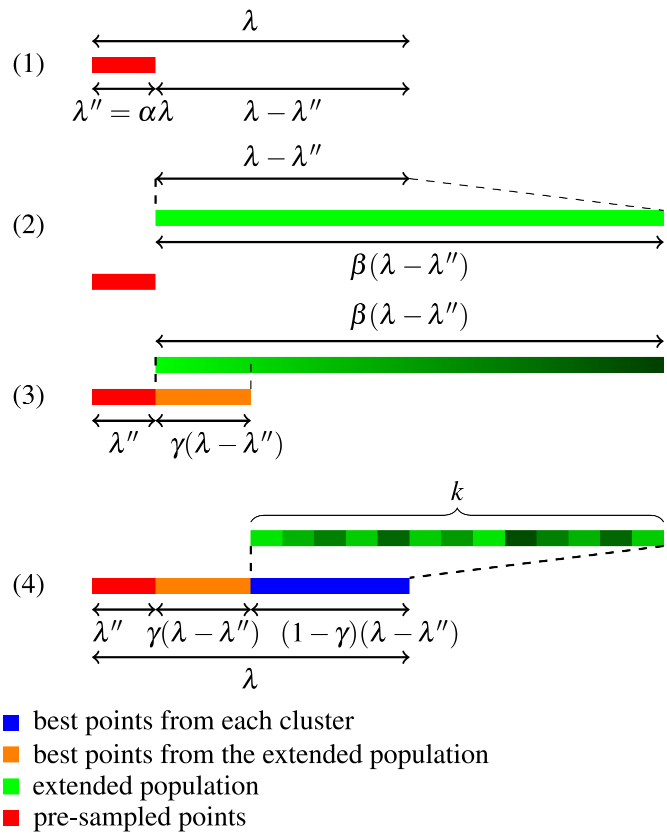

Figure 1: Individual-based EC

A *generation-based* EC strategy determines the number of model-evaluated generations between two generations evaluated by the original function. After a generation is evaluated by the original fitness function, the model is trained using the obtained values. The number of consecutive model-evaluated generations can be determined also dynamically, as introduced in so-called *adaptive* EC strategy [11], when the deviation between the original and the model fitness function is assessed and then it is decided whether to evaluate using the original fitness or the model.

Determining the most suitable EC parameters, however, is an open problem, which depends on the properties of the fitness function and the current performance of the surrogate model. Moreover, the most suitable parameters change during the optimization process.

3 Experimental Evaluation

Previous investigations compared the performance of MGSO [2] and CMA-ES with GP surrogate model [4] (denoted hereafter S-CMA-ES) with CMA-ES without a model, using the standard black-box optimization benchmarks [7]. In this work, we compare those methods using an additional set of 12 multimodal fitness functions from the CEC 2013 competition on niching methods for multimodal function optimization [10]:

- f1: Five-Uneven-Peak Trap (1D),
- f2: Equal Maxima (1D),
- f3: Uneven Decreasing Maxima (1D),
- f4: Himmelblau (2D),
- f5: Six-Hump Camel Back (2D),
- f6: Shubert (2D, 3D),
- f7: Vincent (2D, 3D),
- f8: Modified Rastrigin - All Global Optima (2D),
- f9: Composition Function 1 (2D),
- f10: Composition Function 2 (2D),
- f11: Composition Function 3 (2D, 3D, 5D, 10D),
- f12: Composition Function 4 (3D, 5D, 10D, 20D).

3.1 MGSO Performance

MGSO performance was examined using two covariance functions, isotropic squared exponential ($\mathbf{K}_{\mathrm{SE}}^{\mathrm{iso}}$) and squared exponential with automatic relevance determination ($\mathbf{K}_{\mathrm{SE}}^{\mathrm{ard}}$), with parameters shown in Table 1. The results in Tables 2 and 3 show the speed-up of MGSO with respect to CMA-ES. As can be seen, the $\mathbf{K}_{\mathrm{SE}}^{\mathrm{iso}}$ covariance function performed better among these two in more than the half of cases. Table 1 shows used parameter settings in our evaluations.

3.2 S-CMA-ES Performance

The speed-up results are shown in Tables 2 and 3. In performed evaluations, four covariance functions in the GP surrogate model were used, two types of the squared exponential covariance function, the isotropic version

$$\mathbf{K}_{\mathrm{SE}}^{\mathrm{iso}}(\mathbf{x}_i,\mathbf{x}_j)=\sigma_f^2\exp\left(-\frac{1}{2\ell^2}(\mathbf{x}_i-\mathbf{x}_j)^\top(\mathbf{x}_i-\mathbf{x}_j)\right), \quad (1)$$

S-CMA-ES	
covariance functions	$cov \in \{\mathbf{K}_{\mathrm{Matérn}}^{\nu=\frac{5}{2}}, \mathbf{K}_{\exp}, \mathbf{K}_{\mathrm{SE}}^{\mathrm{iso}}, \mathbf{K}_{\mathrm{SE}}^{\mathrm{ard}}\}$
starting values of (σ_f^2,ℓ)	$(0.1,10)$ for $\mathbf{K}_{\mathrm{SE}}^{\mathrm{iso}}$ $(0.05\times\boldsymbol{J}_{1,D},0.1)$ for $\mathbf{K}_{\mathrm{SE}}^{\mathrm{ard}}$ $(0.5,2)$ otherwise
starting values of σ_n^2	0.01
MGSO	
covariance functions	$cov \in \{\mathbf{K}_{\mathrm{SE}}^{\mathrm{iso}}, \mathbf{K}_{\mathrm{SE}}^{\mathrm{ard}}\}$
starting values of (σ_f^2,ℓ)	$(0.1,10)$ for $\mathbf{K}_{\mathrm{SE}}^{\mathrm{iso}}$ $(0.05\times(\boldsymbol{J}_{1,D}),0.1)$ for $\mathbf{K}_{\mathrm{SE}}^{\mathrm{ard}}$
starting values of σ_n^2	0.01

Table 1: Model parameter settings for S-CMA-ES and MGSO performance testing. The symbols $\mathbf{K}_{\mathrm{SE}}^{\mathrm{iso}}$, $\mathbf{K}_{\mathrm{SE}}^{\mathrm{ard}}$, $\mathbf{K}_{\exp}$, $\mathbf{K}_{\mathrm{Matérn}}^{\nu=\frac{5}{2}}$, denote, respectively, the isotropic squared exponential, squared exponential with automatic relevance determination, exponential and Matérn with parameter $\nu=\frac{5}{2}$ covariance functions. $\boldsymbol{J}_{1,D}$ denotes the vector of ones of length equal to the dimension D of the input space.

and the version using automatic relevance determination

$$\mathbf{K}_{\mathrm{SE}}^{\mathrm{ard}}(\mathbf{x}_i,\mathbf{x}_j)=\sigma_f^2\exp\left(-\frac{1}{2}(\mathbf{x}_i-\mathbf{x}_j)^\top\lambda^{-2}(\mathbf{x}_i-\mathbf{x}_j)\right), \quad (2)$$

where λ stands for the characteristic length scale which measures the distance for being uncorrelated along x_i. The covariance matrices 1 and 2 differ only when λ is a diagonal matrix instead of a scalar. Two types of the Matérn covariance function were used,

$$\mathbf{K}_{\mathrm{Matérn}}^{\nu=\frac{1}{2}}(r)=\exp\left(-\frac{r}{\ell}\right), \quad (3)$$

which is better known as exponential covariance function ($\mathbf{K}_{\exp}$), and

$$\mathbf{K}_{\mathrm{Matérn}}^{\nu=\frac{5}{2}}(r)=\sigma_f^2\left(1+\frac{\sqrt{5}r}{\ell}+\frac{5r^2}{3\ell^2}\right)\exp\left(-\frac{\sqrt{5}r}{\ell}\right), \quad (4)$$

where $r=(\mathbf{x}_i-\mathbf{x}_j)$ and the parameter ℓ is the characteristic length-scale with which the distance of two considered data points is compared and σ_f^2 is the signal variance. The description of the listed covariance functions can be found in [12]. The considered covariance functions parameters are shown in Table 1.

In the performed experiments, different variants of the chosen EC strategies, described in Section 2.2, were examined, *generation-based* and *individual-based*. The result are discussed in the following sections.

3.3 Generation-Based EC Strategy

Apart from covariance function selection, *generation-based* EC strategy was determined by two other parameters, the number of model-evaluated generations and the multiplication factor of CMA-ES' step size σ, which is used in the original-evaluated generations in order to provide points for model training from a broader region of the input space. In the implementation, the first parameter was varied among the values 1, 2, 4 and 8 consecutive model-evaluated generations and the second parameter was varied among the values 1 and 2.

3.4 Individual-Based EC Strategy

Apart from covariance function selection, three other parameters, described in Section 2.2, were examined in the case of *individual-based* EC strategy. The first parameter $\alpha \in [0,1]$ determines the proportion of the original population to be pre-sampled and evaluated by the original fitness function. The second parameter $\beta \in [1,\infty)$ is a multiplicator determining the size of extended population. The third parameter $\gamma \in [0,1]$ determines the amount of points with the best model-fitness chosen from the extended population to be re-evaluated by the original fitness function to become a part of the final population. This parameter also determines the number of clusters, where the best point is chosen from each cluster and added to the final population.

In performed evaluations, the parameter α was varied among the values 0, 0.0625, 0.125, and 0.25, the parameter β was varied among the values 5 and 10 and γ was varied among the values 0, 0.1 and 0.2.

4 Results and Their Assessment

4.1 Result Tables

Tables 2 and 3 show the speed-up of S-CMA-ES and MGSO, compared to CMA-ES without a surrogate model. For the respective targets (distances to the true optimum Δf_{opt}), the speed-up of the expected running time (ERT) is shown. ERT is the number of function evaluations needed to reach the target divided by the ratio of the successful runs, which reached the target. Stopping criteria: the distance 10^{-8} to the true optimum and $100D$ original fitness function evaluations.

The first column in each box corresponds to the overall best settings (described in Section 4.2), the covariance function and EC settings of S-CMA-ES using the *generation-based* EC strategy in terms of the average speed-up. The second column corresponds to the best covariance function and *generation-based* EC settings for the respective function-dimension combination, if there was any better than the overall best observed settings. Analogously, the third and fourth columns show results for S-CMA-ES using *individual-based* EC strategy. Similarly, the last two columns in each box show the speed-up of the MGSO.

Signs "-" instead of the speed-up values mean that, unlike the CMA-ES, no run of the considered method (S-CMA-ES or MGSO) was able to reach that target. Signs "+" mean that, unlike the employed method, no CMA-ES run was able to reach the target. Signs "*" mean that neither the considered method nor CMA-ES were able to reach the target. Speed-ups written in bold mark cases where the S-CMA-ES' or MGSO's median of the ERT is significantly lower than the median of the CMA-ES according to the one-sided Wilcoxon's test on the significance level $\alpha = 0.05$.

4.2 Observations

The MGSO method brought the highest speed-up in the case of $f2$, $f4$, 3D version of $f7$, 2D version of $f11$ and 5D version of $f12$. The worst results were observed in the case of $f1$, 2D version of $f10$ and 3D and 10D versions of $f12$. The best results were achieved using $\mathbf{K}_{\mathrm{SE}}^{\mathrm{iso}}$ covariance function, however, in the case of $f4$ and 2D version of $f7$ $\mathbf{K}_{\mathrm{SE}}^{\mathrm{ard}}$ covariance function brought much better results.

In the case of the *generation-based* EC, the overall best settings with respect to the median values are $(\mathbf{K}_{\mathrm{SE}}^{\mathrm{iso}}, 8, 1)$ – 8 consecutive model-evaluated generations with unmodified step size in combination with $\mathbf{K}_{\mathrm{SE}}^{\mathrm{iso}}$ covariance function. The overall best *generation-based* EC settings showed to be also the best *generation-based* EC settings of the respective functions, except for 3D version of $f12$, where S-CMA-ES performed better using larger step size. Using different covariance functions didn't bring much better results than the overall best covariance function.

The *individual-based* EC strategy achieved the best results with the overall settings $(\mathbf{K}_{\mathrm{SE}}^{\mathrm{ard}}, 0, 5, 0.1)$ – squared exponential covariance matrix with automatic relevance determination, no pre-sampling before training the model, 5 as the multiplicator determining the size of extended population and 0.1 as a multiplicator determining the amount of best points chosen from the extended population. The best results using described parameters were achieved in the case of functions $f3$ and 2D and 3D version of $f6$. However, the overall performance of the *individual-based* EC strategy lags far behind the *generation-based* EC strategy, MGSO and even CMA-ES itself.

4.3 Best-Fitness Progress Diagrams

Figure 2 shows examples of the best-fitness progress with the best observed settings (see Table 2 for details). Medians and the first and third quartiles of the best fitness reached are shown; medians and quartiles measured for MGSO and S-CMA-ES on 15 and 10 independent runs (for both EC strategies), respectively.

The optimization progress of the *individual-based* EC strategy showed to be the slowest in comparison to other methods. MGSO outperformed CMA-ES in most cases and the highest speed-up was achieved in the later phase of the optimization process. The *generation-based* EC strategy achieved the highest speed-up in the middle phase of the optimization process. However, the *generation-based*

f1 (1–D)

Δf_{opt}	S-CMA-ES - generation EC		S-CMA-ES - individual EC		MGSO	
1e1	00.90	01.00	01.00		00.17	00.14
1e0	00.90	01.00	01.00		00.06	00.14
1e-1	00.90	01.00	01.00		00.06	00.08
1e-2	00.90	01.00	01.00		-	00.05
1e-3	00.90	01.00	01.00		-	00.05
1e-4	00.90	01.00	01.00		-	-
1e-5	00.90	01.00	01.00		-	-
1e-6	00.90	01.00	01.00		-	-
1e-7	00.90	01.00	01.00		-	-
1e-8	00.90	01.00	01.00		-	-
param:	$(\mathbf{K}_{SE}^{\text{iso}}, 8, 1)$	$(\mathbf{K}_{\text{SE}}^{\text{ard}}, 8, 1)$	$(\mathbf{K}_{\text{SE}}^{\text{ard}}, 0, 5, 0.1)$		$\mathbf{K}_{\text{SE}}^{\text{iso}}$	$\mathbf{K}_{\text{SE}}^{\text{ard}}$

f2 (1–D)

Δf_{opt}	S-CMA-ES - generation EC		S-CMA-ES - individual EC		MGSO	
1e-1	01.00	**01.44**	00.76	00.13	01.30	
1e-2	**01.41**	**02.54**	00.30	00.10	**03.30**	
1e-3	**01.85**	**02.54**	00.12	00.86	01.08	
1e-4	**03.11**	**03.75**	-	01.73	**01.73**	
1e-5	**04.04**	**04.49**	-	-	**02.25**	
1e-6	**04.34**	**04.83**	-	-	**02.37**	
1e-7	**12.95**	**13.57**	-	-	**08.05**	
1e-8	12.21	15.23	-	-	16.78	
param:	$(\mathbf{K}_{SE}^{\text{iso}}, 8, 1)$	$(\mathbf{K}_{\text{SE}}^{\text{ard}}, 8, 1)$	$(\mathbf{K}_{\text{SE}}^{\text{ard}}, 0, 5, 0.1)$	$(\mathbf{K}_{\text{Matérn}}^{\nu=\frac{5}{2}}, 2^{-2}, 10, 0.2)$	$\mathbf{K}_{\text{SE}}^{\text{iso}}$	

f3 (1–D)

Δf_{opt}	S-CMA-ES - generation EC		S-CMA-ES - individual EC		MGSO	
1e-1	**02.44**		00.52	00.45	01.83	
1e-2	05.71		01.44	01.51	02.13	
1e-3	21.36		07.12	17.80	09.49	
1e-4	18.41		07.12	-	08.63	
1e-5	20.07		07.76	-	09.41	
1e-6	20.07		07.76	-	09.41	
1e-7	*		*	*	*	
1e-8	*		*	*	*	
param:	$(\mathbf{K}_{SE}^{\text{iso}}, 8, 1)$		$(\mathbf{K}_{\text{SE}}^{\text{ard}}, 0, 5, 0.1)$	$(\mathbf{K}_{\text{exp}}, 0, 5, 0.2)$	$\mathbf{K}_{\text{SE}}^{\text{iso}}$	

f4 (2–D)

Δf_{opt}	S-CMA-ES - generation EC		S-CMA-ES - individual EC		MGSO	
1e2	00.54	01.00	00.16	00.11	00.35	00.35
1e1	01.24	**01.63**	00.17	00.44	**01.55**	01.55
1e0	**02.59**	**03.52**	-	01.26	**02.20**	04.00
1e-1	**02.74**	**03.47**	-	-	**01.97**	02.62
1e-2	**02.99**	**03.66**	-	-	**02.44**	03.25
1e-3	**03.91**	**05.00**	-	-	**03.58**	04.57
1e-4	**04.45**	**05.22**	-	-	**04.52**	04.68
1e-5	**13.44**	**14.83**	-	-	00.67	13.23
1e-6	**17.63**	**20.10**	-	-	-	18.12
1e-7	+	+	*	*	*	+
1e-8	+	+	*	*	*	+
param:	$(\mathbf{K}_{SE}^{\text{iso}}, 8, 1)$	$(\mathbf{K}_{\text{exp}}, 8, 1)$	$(\mathbf{K}_{\text{SE}}^{\text{ard}}, 0, 5, 0.1)$	$(\mathbf{K}_{\text{exp}}, 2^{-2}, 5, 0.1)$	$\mathbf{K}_{\text{SE}}^{\text{iso}}$	$\mathbf{K}_{\text{SE}}^{\text{ard}}$

f5 (2–D)

Δf_{opt}	S-CMA-ES - generation EC		S-CMA-ES - individual EC		MGSO	
1e0	01.00		00.12	00.09	00.65	00.65
1e-1	**01.72**		00.05	00.08	00.80	00.80
1e-2	**03.28**		-	00.70	**01.52**	01.52
1e-3	**02.95**		-	-	**01.51**	01.92
1e-4	**03.27**		-	-	**01.96**	02.48
1e-5	**03.74**		-	-	**02.55**	03.23
1e-6	*		*	*	*	*
1e-7	*		*	*	*	*
1e-8	*		*	*	*	*
param:	$(\mathbf{K}_{SE}^{\text{iso}}, 8, 1)$		$(\mathbf{K}_{\text{SE}}^{\text{ard}}, 0, 5, 0.1)$	$(\mathbf{K}_{\text{Matérn}}^{\nu=\frac{5}{2}}, 0, 10, 0.2)$	$\mathbf{K}_{\text{SE}}^{\text{iso}}$	$\mathbf{K}_{\text{SE}}^{\text{ard}}$

f6 (2–D)

Δf_{opt}	S-CMA-ES - generation EC		S-CMA-ES - individual EC		MGSO	
1e2	**01.18**	01.18	-	01.74	**03.04**	02.03
1e1	04.97	04.20	-	-	00.36	02.21
1e0	07.16	08.40	-	-	-	01.80
1e-1	+	+	*	*	*	+
1e-2	+	+	*	*	*	+
1e-3	+	+	*	*	*	+
1e-4	+	+	*	*	*	+
1e-5	*	*	*	*	*	*
1e-6	*	*	*	*	*	*
1e-7	*	*	*	*	*	*
1e-8	*	*	*	*	*	*
param:	$(\mathbf{K}_{SE}^{\text{iso}}, 8, 1)$	$(\mathbf{K}_{\text{SE}}^{\text{ard}}, 4, 1)$	$(\mathbf{K}_{\text{SE}}^{\text{ard}}, 0, 5, 0.1)$	$(\mathbf{K}_{\text{SE}}^{\text{ard}}, 0, 10, 0)$	$\mathbf{K}_{\text{SE}}^{\text{iso}}$	$\mathbf{K}_{\text{SE}}^{\text{ard}}$

f6 (3–D)

Δf_{opt}	S-CMA-ES - generation EC		S-CMA-ES - individual EC		MGSO	
1e4	01.00		01.00	01.00	00.27	
1e3	17.55		-	07.42	04.53	
1e2	07.72		-	-	04.06	
1e1	05.96		-	-	-	
1e0	+		*	*	*	
1e-1	+		*	*	*	
1e-2	+		*	*	*	
1e-3	+		*	*	*	
1e-4	+		*	*	*	
1e-5	+		*	*	*	
1e-6	+		*	*	*	
1e-7	+		*	*	*	
param:	$(\mathbf{K}_{SE}^{\text{iso}}, 8, 1)$		$(\mathbf{K}_{\text{SE}}^{\text{ard}}, 0, 5, 0.1)$	$(\mathbf{K}_{\text{SE}}^{\text{ard}}, 0, 10, 0.1)$	$\mathbf{K}_{\text{SE}}^{\text{iso}}$	

f7 (2–D)

Δf_{opt}	S-CMA-ES - generation EC		S-CMA-ES - individual EC		MGSO	
1e-2	**03.28**	**02.93**	00.05	00.43	00.14	01.02
1e-3	**04.42**	**04.81**	-	-	00.27	00.54
1e-4	**05.63**	**06.09**	-	-	-	00.52
1e-5	**07.91**	**07.91**	-	-	-	01.45
1e-6	**16.24**	**16.24**	-	-	-	03.39
1e-7	65.25	68.62	-	-	-	11.29
1e-8	+	+	*	*	*	+
param:	$(\mathbf{K}_{SE}^{\text{iso}}, 8, 1)$	$(\mathbf{K}_{\text{exp}}, 8, 1)$	$(\mathbf{K}_{\text{SE}}^{\text{ard}}, 0, 5, 0.1)$	$(\mathbf{K}_{\text{SE}}^{\text{ard}}, 2^{-4}, 10, 0)$	$\mathbf{K}_{\text{SE}}^{\text{iso}}$	$\mathbf{K}_{\text{SE}}^{\text{ard}}$

f7 (3–D)

Δf_{opt}	S-CMA-ES - generation EC		S-CMA-ES - individual EC		MGSO	
1e-1	**03.15**		00.03	01.28	01.14	
1e-2	**05.26**		-	-	**01.56**	
1e-3	**10.36**		-	-	**03.16**	
1e-4	**13.33**		-	-	**04.95**	
1e-5	64.47		-	-	22.83	
1e-6	+		*	*	+	
1e-7	+		*	*	+	
1e-8	+		*	*	+	
param:	$(\mathbf{K}_{SE}^{\text{iso}}, 8, 1)$		$(\mathbf{K}_{\text{SE}}^{\text{ard}}, 0, 5, 0.1)$	$(\mathbf{K}_{\text{Matérn}}^{\nu=\frac{5}{2}}, 0, 10, 0)$	$\mathbf{K}_{\text{SE}}^{\text{iso}}$	

f8 (2–D)

Δf_{opt}	S-CMA-ES - generation EC		S-CMA-ES - individual EC		MGSO	
1e0	**02.68**	**02.68**	-	00.96	**01.68**	01.68
1e-1	**03.91**	**03.91**	-	-	**02.22**	02.22
1e-2	**04.46**	**04.79**	-	-	**03.20**	03.20
1e-3	**07.15**	**07.65**	-	-	**03.29**	03.29
1e-4	**12.95**	**13.74**	-	-	**05.61**	06.73
1e-5	**22.40**	**24.85**	-	-	**11.39**	13.53
1e-6	+	+	*	*	+	+
1e-7	+	+	*	*	+	+
1e-8	+	+	*	*	+	+
param:	$(\mathbf{K}_{SE}^{\text{iso}}, 8, 1)$	$(\mathbf{K}_{\text{Matérn}}^{\nu=\frac{5}{2}}, 8, 1)$	$(\mathbf{K}_{\text{SE}}^{\text{ard}}, 0, 5, 0.1)$	$(\mathbf{K}_{\text{exp}}, 2^{-4}, 5, 0)$	$\mathbf{K}_{\text{SE}}^{\text{iso}}$	$\mathbf{K}_{\text{SE}}^{\text{ard}}$

Table 2: Speed-up of S-CMA-ES using *individual*- and *generation-based* strategies and MGSO, compared to CMA-ES without a surrogate model – functions $f1 - f10$ (see Section 4.1 for details). Empty columns signify that the best observed settings for the respective function-dimension combination are identical with the overall best observed settings.

f9 (2–D)

Δf_{opt}	S-CMA-ES - generation EC		S-CMA-ES - individual EC		MGSO
1e1	**03.40**	**02.74**	-	00.10	01.06
1e0	**03.38**	**03.11**	-	00.07	**01.15**
1e-1	**03.82**	**03.82**	-	-	**01.53**
1e-2	**04.35**	**04.35**	-	-	**01.52**
1e-3	**07.80**	**07.80**	-	-	**02.68**
1e-4	**22.40**	**23.56**	-	-	**08.54**
1e-5	+	+	*	*	+
1e-6	+	+	*	*	+
1e-7	+	+	*	*	+
1e-8	+	+	*	*	*
param:	$(\mathbf{K}_{SE}^{iso}, 8, 1)$	$(\mathbf{K}_{SE}^{ard}, 8, 1)$	$(\mathbf{K}_{SE}^{ard}, 0, 5, 0.1)$	$(\mathbf{K}_{SE}^{ard}, 2^{-4}, 5, 0)$	$\mathbf{K}_{SE}^{iso}$

f10 (2–D)

Δf_{opt}	S-CMA-ES - generation EC		S-CMA-ES - individual EC		MGSO
1e2	**01.98**	**01.76**	00.03	00.26	00.39
1e1	**02.29**	**02.62**	-	-	00.46
1e0	**03.01**	**03.21**	-	-	00.07
1e-1	**03.71**	**04.00**	-	-	-
1e-2	**06.15**	**07.02**	-	-	-
1e-3	18.28	21.88	-	-	-
1e-4	+	+	*	*	*
1e-5	+	+	*	*	*
1e-6	+	+	*	*	*
1e-7	+	+	*	*	*
1e-8	+	+	*	*	*
param:	$(\mathbf{K}_{SE}^{iso}, 8, 1)$	$(\mathbf{K}_{Matérn}^{\nu=\frac{5}{2}}, 8, 1)$	$(\mathbf{K}_{SE}^{ard}, 0, 5, 0.1)$	$(\mathbf{K}_{SE}^{ard}, 0, 10, 0)$	$\mathbf{K}_{SE}^{iso}$

f11 (2–D)

Δf_{opt}	S-CMA-ES - generation EC		S-CMA-ES - individual EC		MGSO
1e2	**03.92**	**03.38**	00.04	01.06	**01.24**
1e1	**03.69**	**03.69**	-	-	**01.43**
1e0	**04.63**	**05.01**	-	-	**02.32**
1e-1	**05.19**	**05.19**	-	-	**03.18**
1e-2	**08.70**	**09.23**	-	-	**02.51**
1e-3	**12.06**	**14.04**	-	-	**03.72**
1e-4	59.40	65.25	-	-	18.46
1e-5	+	+	*	*	+
1e-6	+	+	*	*	*
1e-7	+	+	*	*	*
1e-8	+	+	*	*	*
param:	$(\mathbf{K}_{SE}^{iso}, 8, 1)$	$(\mathbf{K}_{SE}^{ard}, 8, 1)$	$(\mathbf{K}_{SE}^{ard}, 0, 5, 0.1)$	$(\mathbf{K}_{SE}^{iso}, 2^{-4}, 10, 0)$	$\mathbf{K}_{SE}^{iso}$

f11 (3–D)

Δf_{opt}	S-CMA-ES - generation EC		S-CMA-ES - individual EC		MGSO
1e3	00.53	01.00	01.00	01.00	00.27
1e2	**02.51**	**03.04**	-	-	**01.33**
1e1	**03.04**	**03.38**	-	-	**01.41**
1e0	**03.12**	**03.47**	-	-	**01.85**
1e-1	**03.90**	**04.33**	-	-	**01.71**
1e-2	**05.95**	**07.21**	-	-	**02.19**
1e-3	27.83	32.15	-	-	11.06
1e-4	+	+	*	*	+
1e-5	+	+	*	*	+
1e-6	+	+	*	*	+
1e-7	+	+	*	*	*
1e-8	+	+	*	*	*
param:	$(\mathbf{K}_{SE}^{iso}, 8, 1)$	$(\mathbf{K}_{exp}, 8, 1)$	$(\mathbf{K}_{SE}^{ard}, 0, 5, 0.1)$	$(\mathbf{K}_{SE}^{ard}, 0, 5, 0)$	$\mathbf{K}_{SE}^{iso}$

f12 (3–D)

Δf_{opt}	S-CMA-ES - generation EC		S-CMA-ES - individual EC		MGSO	
1e2	**02.05**	01.72	00.37	00.27	00.91	00.46
1e1	**01.42**	02.27	-	00.67	00.31	00.63
1e0	-	**02.28**	-	-	00.48	00.65
1e-1	-	**02.76**	-	-	-	-
1e-2	-	**03.78**	-	-	-	-
1e-3	-	19.56	-	-	-	-
1e-4	*	+	*	*	*	*
1e-5	*	+	*	*	*	*
1e-6	*	+	*	*	*	*
1e-7	*	+	*	*	*	*
1e-8	*	+	*	*	*	*
param:	$(\mathbf{K}_{SE}^{iso}, 8, 1)$	$(\mathbf{K}_{SE}^{ard}, 8, 2)$	$(\mathbf{K}_{SE}^{ard}, 0, 5, 0.1)$	$(\mathbf{K}_{Matérn}^{\nu=\frac{5}{2}}, 0, 10, 0.1)$	$\mathbf{K}_{SE}^{iso}$	$\mathbf{K}_{SE}^{ard}$

f11 (5–D)

Δf_{opt}	S-CMA-ES - generation EC	S-CMA-ES - individual EC	MGSO
1e2	**03.02**	-	**01.36**
1e1	**03.24**	-	**01.51**
1e0	**04.37**	-	**01.96**
1e-1	**09.79**	-	03.12
1e-2	27.12	-	07.51
1e-3	47.82	-	13.17
1e-4	+	*	+
1e-5	+	*	*
1e-6	+	*	*
1e-7	+	*	*
1e-8	+	*	*
param:	$(\mathbf{K}_{SE}^{iso}, 8, 1)$	$(\mathbf{K}_{SE}^{ard}, 0, 5, 0.1)$	$\mathbf{K}_{SE}^{iso}$

f11 (10–D)

Δf_{opt}	S-CMA-ES - generation EC		S-CMA-ES - individual EC		MGSO
1e3	00.89	01.14	00.17	00.64	00.27
1e2	**03.65**	**04.11**	-	-	01.05
1e1	**08.51**	**09.35**	-	-	02.61
1e0	**09.78**	**10.59**	-	-	02.10
1e-1	**20.02**	**21.50**	-	-	-
1e-2	+	+	*	*	*
1e-3	+	+	*	*	*
1e-4	+	+	*	*	*
1e-5	+	+	*	*	*
1e-6	+	+	*	*	*
1e-7	*	+	*	*	*
1e-8	*	*	*	*	*
param:	$(\mathbf{K}_{SE}^{iso}, 8, 1)$	$(\mathbf{K}_{Matérn}^{\nu=\frac{5}{2}}, 8, 1)$	$(\mathbf{K}_{SE}^{ard}, 0, 5, 0.1)$	$(\mathbf{K}_{SE}^{iso}, 2^{-4}, 5, 0)$	$\mathbf{K}_{SE}^{iso}$

f12 (5–D)

Δf_{opt}	S-CMA-ES - generation EC		S-CMA-ES - individual EC		MGSO
1e2	**01.64**	**01.70**	-	00.04	00.78
1e1	**05.60**	**06.61**	-	-	**04.73**
1e0	19.24	33.01	-	-	28.86
1e-1	+	+	*	*	+
1e-2	+	+	*	*	+
1e-3	+	+	*	*	+
1e-4	+	+	*	*	*
1e-5	+	+	*	*	*
1e-6	+	+	*	*	*
1e-7	+	*	*	*	*
1e-8	*	*	*	*	*
param:	$(\mathbf{K}_{SE}^{iso}, 8, 1)$	$(\mathbf{K}_{SE}^{ard}, 4, 1)$	$(\mathbf{K}_{SE}^{ard}, 0, 5, 0.1)$	$(\mathbf{K}_{SE}^{iso}, 2^{-2}, 5, 0)$	$\mathbf{K}_{SE}^{iso}$

f12 (10–D)

Δf_{opt}	S-CMA-ES - generation EC	S-CMA-ES - individual EC	MGSO	
1e2	**08.98**	-	**01.88**	02.02
1e1	**06.83**	-	-	01.05
1e0	**17.61**	-	-	-
1e-1	+	*	*	*
1e-2	+	*	*	*
1e-3	+	*	*	*
1e-4	+	*	*	*
1e-5	+	*	*	*
1e-6	*	*	*	*
1e-7	*	*	*	*
1e-8	*	*	*	*
param:	$(\mathbf{K}_{SE}^{iso}, 8, 1)$	$(\mathbf{K}_{SE}^{ard}, 0, 5, 0.1)$	$\mathbf{K}_{SE}^{iso}$	$\mathbf{K}_{SE}^{ard}$

f12 (20–D)

Δf_{opt}	S-CMA-ES - generation EC	S-CMA-ES - individual EC		MGSO
1e3	01.39	00.05	00.18	00.42
1e2	**02.73**	-	-	00.03
1e1	**07.24**	-	-	-
1e0	59.06	-	-	-
1e-1	+	*	*	*
1e-2	+	*	*	*
1e-3	+	*	*	*
1e-4	+	*	*	*
1e-5	+	*	*	*
1e-6	*	*	*	*
1e-7	*	*	*	*
1e-8	*	*	*	*
param:	$(\mathbf{K}_{SE}^{iso}, 8, 1)$	$(\mathbf{K}_{SE}^{ard}, 0, 5, 0.1)$	$(\mathbf{K}_{SE}^{iso}, 0, 10, 0)$	$\mathbf{K}_{SE}^{iso}$

Table 3: Speed-up of S-CMA-ES using *individual-* and *generation-based* strategies and MGSO, compared to CMA-ES without a surrogate model – functions $f11 - f20$ (see Section 4.1 for details). Empty columns signify that the best observed settings for the respective function-dimension combination are identical with the overall best observed settings.

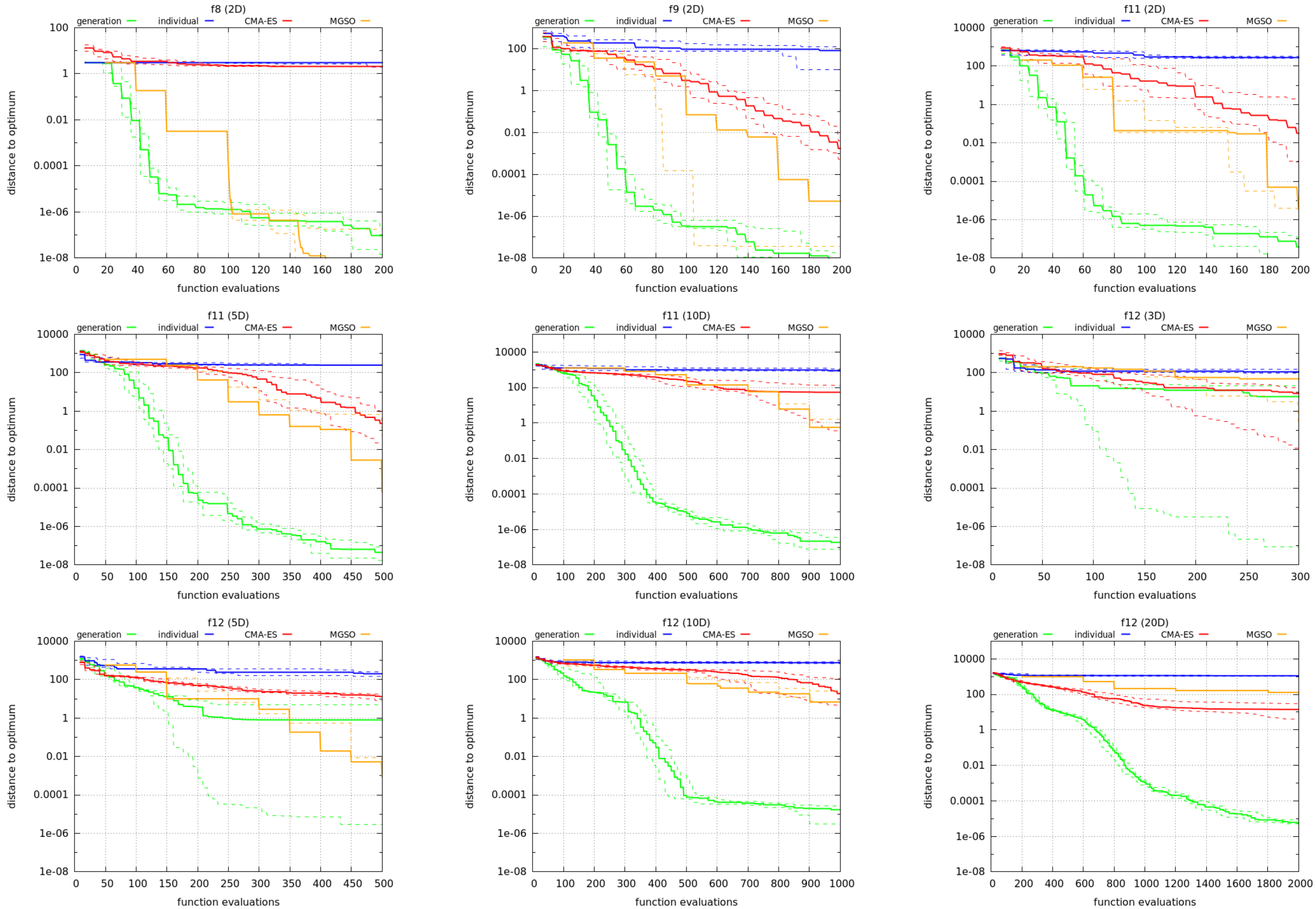

Figure 2: Examples of the best-fitness progress with the best observed settings, solid line: median, dashed lines: quartiles.

EC strategy generally brought the best results, it was outperformed by MGSO in the case of 2D version of $f8$ and 5D version of $f12$ in the later phase of the optimization process.

5 Conclusion

In this paper, two optimization approaches based on Gaussian processes were tested on the set of multimodal fitness functions from the CEC 2013 competition [10], and were compared to the state-of-the-art evolutionary approach in black-box optimization, CMA-ES. One of them is Model Guided Sampling Optimization [2], the other approach, S-CMA-ES [5], consists in using GP as a surrogate model for CMA-ES. The performance of the methods was compared with respect to the number of function evaluations.

In the case of S-CMA-ES, two evolution control strategies were used, the *individual-* and *generation-based*. Although S-CMA-ES using *generation-based* EC strategy outperformed MGSO, both methods showed the performance improvement in most cases. On the other hand, the *individual-based* EC strategy brought the worst results of all considered methods. We also observed, that S-CMA-ES performs better using *generation-based* EC setting with more consecutive model-evaluated generations. Isotropic squared exponential covariance function showed to be the most suitable for the optimization from all tested covariance functions.

This is a work in progress that is a part of a broader ongoing research. Terefore, it would be premature to draw deeper conclusions at this stage. We hope to be able to draw such conclusions after further investigations will be performed in the future.

Acknowledgement

The research reported in this paper has been supported by the Czech Science Foundation (GAČR) grant 13-17187S.

References

[1] Nik, M. A., Fayazbakhsh, K., Pasini, D., Lessard, L.: A comparative study of metamodeling methods for the design optimization of variable stiffness composites. Composite Structures **107** (2014), 494–501

[2] Bajer, L., Charypar, V., Holeňa, M.: Model guided sampling optimization with gaussian processes for expensive black-box optimization. In: Blum, C. (ed.), GECCO Companion'13, 2013, New York: ACM

[3] Bajer, L., Holeňa, M.: Two gaussian approaches to black-box optomization. CoRR, abs/1411.7806, 2014

[4] Bajer, L., Pitra, Z., Holeňa, M.: Benchmarking Gaussian processes and random forests surrogate models on the BBOB noiseless testbed. In: GECCO'15, Madrid, Spain, 2015, ACM

[5] Bajer, L., Pitra, Z., Holeňa, M.: Investigation of Gaussian processes and random forests as surrogate models for evolutionary black-box optimization. In: GECCO'15, Poster Abstracts, Madrid, Spain, 2015, ACM, TBA

[6] Chaput, J. C., Szostak, J. W.: Evolutionary optimization of a nonbiological ATP binding protein for improved folding stability. Chemistry & Biology bf 11(6) (June 2004), 865–874

[7] Hansen, N., Finck, S., Ros, R., Auger, A.: Real-parameter black-box optimization benchmarking 2012: experimental setup. Technical Report, INRIA, 2012

[8] Jones, D. R., Schonlau, M., Welch, W. J.: Efficient Global Optimization of expensive black-box functions. J. of Global Optimization **13(4)** (Dec. 1998), 455–492

[9] Larranaga, P., Lozano, J.: Estimation of distribution algorithms: a new tool for evolutionary computation. Kluwer Academic Pub, 2002

[10] Li, X., Engelbrecht, A., Epitropakis, M. G.: Benchmark functions for CEC'2013 special session and competition on niching methods for multimodal function optimization'. 2013.

[11] Loshchilov, I., Schoenauer, M., Sebag, M.: Self-adaptive surrogate-assisted covariance matrix adaptation evolution strategy. In: Proceedings of the Fourteenth International Conference On Genetic And Evolutionary Computation Conference, GECCO'12, 321–328, New York, NY, USA, 2012, ACM

[12] Rasmussen, C., Williams, C.: Gaussian processes for machine learning. Adaptive Computation and Machine Learning, MIT Press, Cambridge, MA, USA, Jan. 2006

[13] Tesauro, G., Jong, N. K., Das, R., Bennani, M. N.: On the use of hybrid reinforcement learning for autonomic resource allocation. Cluster Computing **10(3)** (2007), 287–299