

# Comparing SVM, Gaussian Process and Random Forest Surrogate Models for the CMA-ES

Zbyněk Pitra<sup>1,2</sup>, Lukáš Bajer<sup>3,4</sup>, and Martin Holeňa<sup>3</sup>

<sup>1</sup> National Institute of Mental Health  
Topolová 748, 250 67 Klecany, Czech Republic  
z.pitra@gmail.com

<sup>2</sup> Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague  
Břehová 7, 115 19 Prague 1, Czech Republic

<sup>3</sup> Institute of Computer Science, Academy of Sciences of the Czech Republic  
Pod Vodárenskou věží 2, 182 07 Prague 8, Czech Republic  
{bajer,holena}@cs.cas.cz

<sup>4</sup> Faculty of Mathematics and Physics, Charles University in Prague  
Malostranské nám. 25, 118 00 Prague 1, Czech Republic

*Abstract:* In practical optimization tasks, it is more and more frequent that the objective function is black-box which means that it cannot be described mathematically. Such functions can be evaluated only empirically, usually through some costly or time-consuming measurement, numerical simulation or experimental testing. Therefore, an important direction of research is the approximation of these objective functions with a suitable regression model, also called surrogate model of the objective functions. This paper evaluates two different approaches to the continuous black-box optimization which both integrates surrogate models with the state-of-the-art optimizer CMA-ES. The first Ranking SVM surrogate model estimates the ordering of the sampled points as the CMA-ES utilizes only the ranking of the fitness values. However, we show that continuous Gaussian processes model provides in the early states of the optimization comparable results.

## 1 Introduction

Optimization of an expensive objective or fitness function plays an important role in many engineering and research tasks. For such functions, it is sometimes difficult to find an exact analytical formula, or to obtain any derivatives or information about smoothness. Instead, values for a given input are possible to be obtained only through expensive and time-consuming measurements and experiments. Those functions are called black-box, and because of the evaluation costs, the primary criterion for assessment of the black-box optimizers is the number of fitness function evaluations necessary to achieve the optimal value.

The Covariance Matrix Adaptation Evolution Strategy (CMA-ES) [5] is considered to be the state-of-the-art of the black-box continuous optimization. The important property of the CMA-ES is that it advances through the search space only according to the ordering of the function values in current population. Hence, the search of the algorithm is rather local which predisposes it to premature convergence in local optima if not used with sufficiently large

population size. This issue resulted in development of several restart strategies [12], such as IPOP-CMA-ES [1] and BIPOP-CMA-ES [6] performing restarts with population size successively increased, or aCMA-ES [9] using also unsuccessful individuals for covariance matrix adaptation.

Furthermore, the CMA-ES often requires more fitness function evaluations to find the optimum than many real-world experiments can offer. In order to decrease the number of evaluations in evolutionary algorithms, it is convenient to periodically train a surrogate model of the fitness function and use it for evaluation of new points instead of the original function. The second option is to use the model for selection of the most promising points to be evaluated by the original fitness.

Loshchilov's surrogate-model-based algorithm <sup>s\*</sup>ACM-ES [13] utilizes the former approach: it estimates the ordering of the fitness values required by the CMA-ES using Ranking Support Vector Machines (SVM) as an ordinal regression model. Moreover, it has been shown [13] that model parameters (hyperparameters) used to construct Ranking SVM model can be optimized during the search by the pure CMA-ES algorithm. Later proposed <sup>s\*</sup>ACM-ES extensions, referred to as <sup>s\*</sup>ACM-ES-k [15] and BIPOP-<sup>s\*</sup>ACM-ES-k [14], use a more intensive exploitation of the surrogate model by increasing population size in generations evaluated by the model.

More recently, a similar algorithm based on regression surrogate model called S-CMA-ES [3] has been presented. As opposed to the former algorithm, S-CMA-ES is performing continuous regression by Gaussian processes (GP) [17] and random forests (RF) [4].

This paper compares the two mentioned surrogate CMA-ES algorithms, <sup>s\*</sup>ACM-ES-k and S-CMA-ES, and the original CMA-ES itself. We benchmark these algorithms on the BBOB/COCO testing set [7, 8] not only in their one population IPOP-CMA-ES version, but also in combination with the two-population-size BIPOP-CMA-ES.

The remainder of the paper is organized as follows. The next chapter briefly describes tested algorithms: the CMA-ES, the BIPOP-CMA-ES, the  $s^*$ ACM-ES-k, and the S-CMA-ES. Section 3 contains experimental setup and results, and Section 4 concludes the paper and suggests further research directions.

## 2 Algorithms

### 2.1 The CMA-ES

In each generation  $g$ , the CMA-ES [5] generates  $\lambda$  new candidate solutions  $\mathbf{x}_k \in \mathbb{R}^D$ , where  $k = 1, \dots, \lambda$ , from a multivariate normal distribution  $N(\mathbf{m}_{(g)}, \sigma^2_{(g)} \mathbf{C}_{(g)})$ , where  $\mathbf{m}_{(g)}$  is the mean interpretable as the current best estimate of the optimum,  $\sigma^2_{(g)}$  the step size, representing the overall standard deviation, and  $\mathbf{C}_{(g)}$  the  $D \times D$  covariance matrix. The algorithm selects the  $\mu$  points with the lowest function value from  $\lambda$  generated candidates to adjust distribution parameters for the next generation.

The CMA-ES uses restart strategies to deal with multimodal fitness landscapes and to avoid being trapped in local optima. A multi-start strategy where the population size is doubled in each restart is referred to as IPOP-CMA-ES [1].

### 2.2 BIPOP-CMA-ES

The BIPOP-CMA-ES [6], unlike IPOP-CMA-ES, considers two different restart strategies. In the first one, corresponding to the IPOP-CMA-ES, the population size is doubled in each restart  $i_{\text{restart}}$  using a constant initial step-size  $\sigma_{\text{large}}^0 = \sigma_{\text{default}}^0$ :

$$\lambda_{\text{large}} = 2^{i_{\text{restart}}} \lambda_{\text{default}}. \quad (1)$$

In the second one, the smaller population size  $\lambda_{\text{small}}$  is computed as

$$\lambda_{\text{small}} = \left\lceil \lambda_{\text{default}} \left( \frac{1}{2} \frac{\lambda_{\text{large}}}{\lambda_{\text{default}}} \right)^{U[0,1]^2} \right\rceil, \quad (2)$$

where  $U[0,1]$  denotes the uniform distribution in  $[0,1]$ . The initial step-size is also randomly drawn as

$$\sigma_{\text{small}}^0 = \sigma_{\text{default}}^0 \times 10^{-2U[0,1]}. \quad (3)$$

The BIPOP-CMA-ES performs the first run using the default population size  $\lambda_{\text{default}}$  and the initial step-size  $\sigma_{\text{default}}^0$ . In the following restarts, the strategy with less function evaluations summed over all algorithm runs is selected.

### 2.3 $s^*$ ACM-ES-k

Loshchilov's version of the CMA-ES using the ordinal regression by Ranking SVM as surrogate model in specific generations instead of the original function is referred to as  $s^*$ ACM-ES [13], and its extension using a more intensive exploitation is called  $s^*$ ACM-ES-k [15].

Before the main loop starts, the  $s^*$ ACM-ES-k evaluates  $g_{\text{start}}$  generations by the original function, then it repeats the following steps: First, the surrogate model is constructed using hyperparameters  $\theta$ , and the original function-evaluated points from previous generations. Second, the surrogate model is optimized by the CMA-ES for  $g_m$  generations with population size  $\lambda = k_\lambda \lambda_{\text{default}}$  and the number of best points  $\mu = k_\mu \mu_{\text{default}}$ , where  $k_\lambda, k_\mu \geq 1$ . Third, the following generation is evaluated by the original function using  $\lambda = \lambda_{\text{default}}$  and  $\mu = \mu_{\text{default}}$ . To avoid a potential divergence when  $g_m$  fluctuate between 0 and 1,  $k_\lambda > 1$  is used only in the case of  $g_m \geq g_{m\lambda}$ , where  $g_{m\lambda}$  denotes the number of generations suitable for effective exploitation using the model. Then the model error is calculated according to the comparison of ranking between the original and model evaluation of the last generation. After that, the  $g_m$  is adjusted in accordance with the model error. As the last step, the  $s^*$ ACM-ES-k searches a hyperparameter space by one generation of the CMA-ES minimizing the model error to find the most suitable hyperparameter settings  $\theta_{\text{new}}$  for the next model-evaluated generations.

The  $s^*$ ACM-ES-k version using BIPOP-CMA-ES proposed in [14] is called BIPOP- $s^*$ ACM-ES-k.

### 2.4 S-CMA-ES

As opposed to the former algorithms, a different approach to surrogate model usage is incorporated in the S-CMA-ES [3]. The algorithm is a modification of CMA-ES where the original evaluating and sampling phases are substituted by the Algorithm 1 at the beginning of each CMA-ES generation.

In order to avoid the false convergence of the algorithm in the BBOB benchmarking toolbox, the model-predicted values are adapted to never be lower than the so far minimum of the original function (see the step 17 in the pseudocode).

The main difference between the S-CMA-ES and the  $s^*$ ACM-ES-k is in the manner how the CMA-ES is utilized. Considering S-CMA-ES, the model prediction or training is performed within each generation of the CMA-ES. On the contrary in the  $s^*$ ACM-ES-k, individual generations of the CMA-ES are started to optimize either original fitness, surrogate fitness, or model itself.

### 3 Experimental Evaluation

The core of this paper lies in a systematic comparison of the two mentioned approaches to using surrogate models with the CMA-ES and the original CMA-ES algorithm itself. The first group of surrogate-based algorithms is formed by the S-CMA-ES algorithms using Gaussian processes and random forests models, and the other group is formed by the  $s^*$ ACM-ES algorithm. These four algorithms (CMA-ES, GP-CMA-ES, RF-CMA-ES,  $s^*$ ACM-ES) are tested in their IPOP version (based on IPOP-CMA-ES) [1] and in the bi-population restart strategy version (based on BIPOP-CMA-ES and its derivatives) [6].

#### 3.1 Experimental Setup

The experimental evaluation is performed through the noiseless part of the COCO/BBOB framework (COMparing Continuous Optimizers / Black-Box Optimization Benchmarking) [7, 8]. It is a collection of 24 benchmark functions with different degree of smoothness, uni-/multi-modality, separability, conditionality etc. Each function is

---

#### Algorithm 1 Surrogate CMA-ES Algorithm [3]

---

**Input:**  $g$  (generation),  $g_m$  (number of model generations),  $\sigma$ ,  $\lambda$ ,  $\mathbf{m}$ ,  $\mathbf{C}$  (CMA-ES internal variables),  $r$  (maximal distance between training points and  $\mathbf{m}$ ),  $n_{\text{REQ}}$  (minimal number of points for model training),  $n_{\text{MAX}}$  (maximal number of points for model training),  $A$  (archive),  $f_M$  (model),  $f$  (original fitness function)

- 1:  $\mathbf{x}_k \sim N(\mathbf{m}, \sigma^2 \mathbf{C}) \quad k = 1, \dots, \lambda \quad \{CMA-ES \text{ sampling}\}$
- 2: **if**  $g$  is original-evaluated **then**
- 3:  $y_k \leftarrow f(\mathbf{x}_k) \quad k = 1, \dots, \lambda \quad \{fitness \text{ evaluation}\}$
- 4:  $A = A \cup \{(\mathbf{x}_k, y_k)\}_{k=1}^{\lambda}$
- 5:  $(\mathbf{X}_{\text{tr}}, \mathbf{y}_{\text{tr}}) \leftarrow \{(\mathbf{x}, y) \in A \mid (\mathbf{m} - \mathbf{x})^\top \sigma \mathbf{C}^{-1/2} (\mathbf{m} - \mathbf{x}) \leq r\}$
- 6: **if**  $|\mathbf{X}_{\text{tr}}| \geq n_{\text{REQ}}$  **then**
- 7:  $(\mathbf{X}_{\text{tr}}, \mathbf{y}_{\text{tr}}) \leftarrow$  choose  $n_{\text{MAX}}$  points if  $|\mathbf{X}_{\text{tr}}| > n_{\text{MAX}}$
- 8:  $\{transformation \text{ to the eigenvector basis:}\}$   
 $\mathbf{X}_{\text{tr}} \leftarrow \{(\sigma \mathbf{C}^{-1/2})^\top \mathbf{x}_{\text{tr}} \text{ for each } \mathbf{x}_{\text{tr}} \in \mathbf{X}_{\text{tr}}\}$
- 9:  $f_M \leftarrow \text{trainModel}(\mathbf{X}_{\text{tr}}, \mathbf{y}_{\text{tr}})$
- 10: mark  $(g + 1)$  as model-evaluated
- 11: **else**
- 12: mark  $(g + 1)$  as original-evaluated
- 13: **end if**
- 14: **else**
- 15:  $\mathbf{x}_k \leftarrow (\sigma \mathbf{C}^{-1/2})^\top \mathbf{x}_k \quad k = 1, \dots, \lambda$
- 16:  $y_k \leftarrow f_M(\mathbf{x}_k) \quad k = 1, \dots, \lambda \quad \{model \text{ evaluation}\}$
- 17:  $\{shift \ y_k \text{ values if } (\min y_k) < \text{best } y \text{ from } A\}$   
 $y_k = y_k + \max\{0, \min_A y - \min y_k\} \quad k = 1, \dots, \lambda$
- 18: **if**  $g_m$  model generations passed **then**
- 19: mark  $(g + 1)$  as original-evaluated
- 20: **end if**
- 21: **end if**

**Output:**  $f_M, A, (y_k)_{k=1}^{\lambda}$

---

defined for any dimension  $D \geq 2$ ; the dimensions used for our tests are 2, 5, 10, and 20. The set of functions comprises, among others, well-known continuous optimization benchmarks like ellipsoid, Rosenbrock's, Rastrigin's, Schwefel's or Weierstrass' function.

The framework calls the optimizers on 15 different instances for each function and dimension, meaning that 1440 optimization runs were called for each of the eight considered algorithms. The graphs at the end of the paper show detailed results in a per-function and per-group-of-function manner. The following paragraphs summarize the parameters of the algorithms.

*The CMA-ES.* The original CMA-ES was used in its IPOP-CMA-ES version (Matlab code v. 3.61) with number of restarts = 4, IncPopSize = 2,  $\sigma_{start} = \frac{8}{3}$ ,  $\lambda = 4 + \lfloor 3 \log D \rfloor$ . The remainder settings were left default.

*$s^*$ ACM-ES.* We have used Loshchilov's GECCO 2013 Matlab code `xacmes.m` [14] in its  $s^*$ ACM-ES version, setting the parameters CMAactive = 1, newRestartRules = 0 and withSurr = 1, modelType = 1, withModelEnsembles = 0, withModelOptimization = 1, hyper\_lambda = 20,  $\lambda_{\text{Mult}} = 1$ ,  $\mu_{\text{Mult}} = 1$  and  $\Lambda_{\text{minIter}} = 4$ .

*S-CMA-ES: GP5-CMA-ES and RF5-CMA-ES.* The number after the GP/RF in the names of the algorithms denotes the number of model-evaluated generations  $g_m$ , which are evaluated by the model in row. All considered S-CMA-ES versions use the distance  $r = 8$  (see algorithm 1). For the GP model,  $K_{\text{Matérn}}^{v=5/2}$  covariance function with starting values  $(\sigma_n^2, l, \sigma_f^2) = \log(0.01, 2, 0.5)$  has been used (see [3] for the details). We have tested RF comprising 100 regression trees, each containing at least two training points in each leaf. The CMA-ES parameters (IPOP version,  $\sigma_{start}$ ,  $\lambda$ , IncPopSize etc.) were used the same as in the pure CMA-ES experiments. All S-CMA-ES parameter values were chosen according to preliminary testing on several functions from the COCO/BBOB framework.

*BIPOP version of the algorithms.* The bi-population versions BIPOP-CMA-ES and BIPOP- $s^*$ ACM-ES use the same Loshchilov's Matlab code `xacmes.m` with the parameter BIPOP = 1. The BIPOP-GP5-CMA-ES and BIPOP-RF5-CMA-ES algorithms are constructed in the same manner as the S-CMA-ES was transformed from the CMA-ES – by integration of the Algorithm 1 into every generation of the BIPOP- $s^*$ ACM-ES.

#### 3.2 Results

The performance of the algorithms is compared in the graphs placed in Figures 1–3. The graphs in Figure 1 depict the *expected running time (ERT)*, which depends on a given target function value  $f_t = f_{\text{opt}} + \Delta f$  – the true optimum  $f_{\text{opt}}$  of the respective benchmark function raised by a small value  $\Delta f$ . The ERT is computed over all relevant

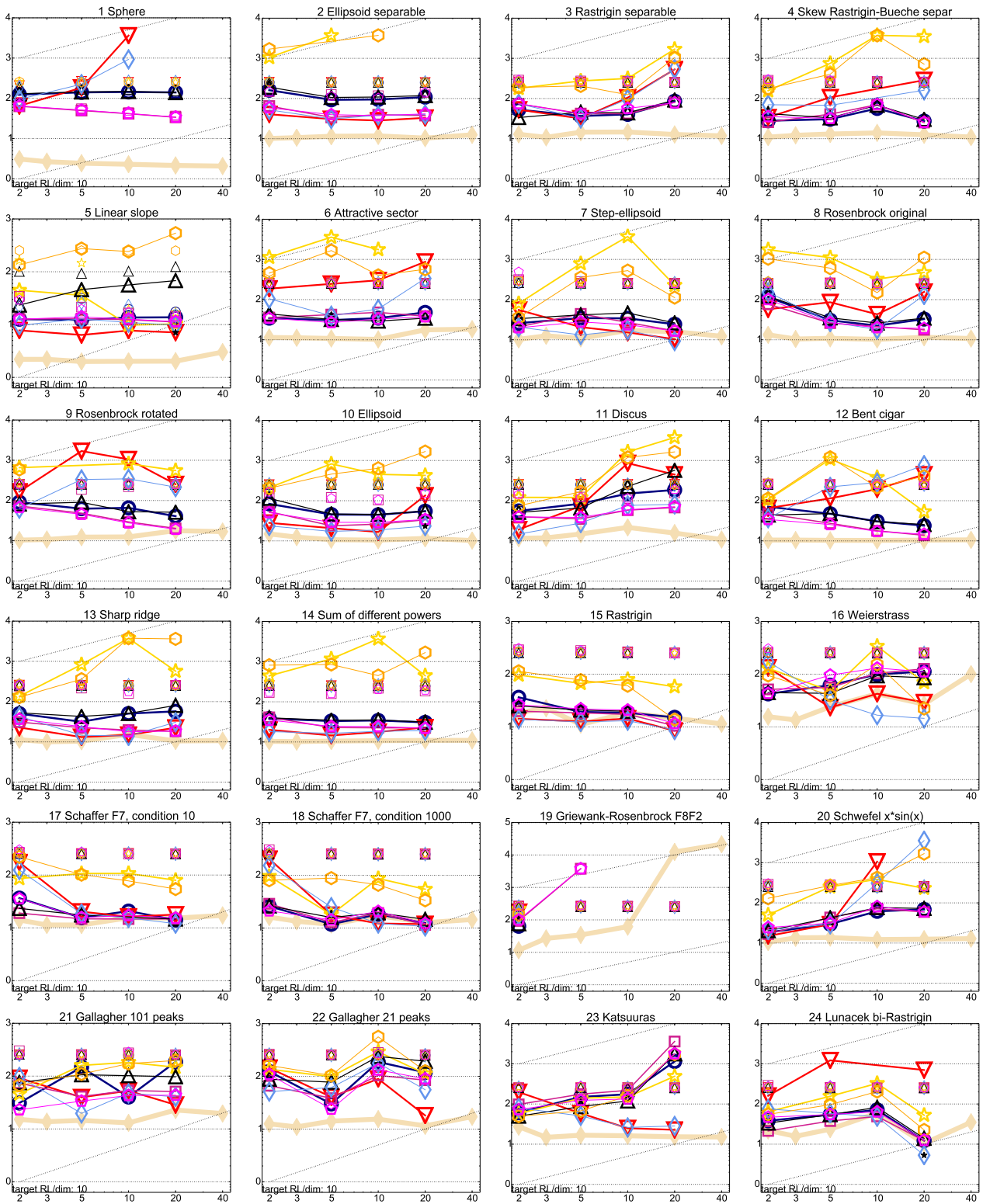


Figure 1: Expected running time (ERT in number of  $f$ -evaluations as  $\log_{10}$  value) divided by dimension versus dimension. The target function value is chosen such that the bestGECCO2009 artificial algorithm just failed to achieve an ERT of  $10 \times \text{DIM}$ . Different symbols correspond to different algorithms given in the legend of  $f_1$  and  $f_{24}$ . Light symbols give the maximum number of function evaluations from the longest trial divided by dimension. Black stars indicate a statistically better result compared to all other algorithms with  $p < 0.01$  and Bonferroni correction number of dimensions (six). Legend:  $\circ$ :BIPOP-CMAES,  $\nabla$ :BIPOP-GP5,  $\star$ :BIPOP-RF5,  $\square$ :BIPOP-saACMES,  $\triangle$ :CMA-ES,  $\diamond$ :GP5-CMAES,  $\circ$ :RF5-CMAES,  $\diamond$ :saACMES

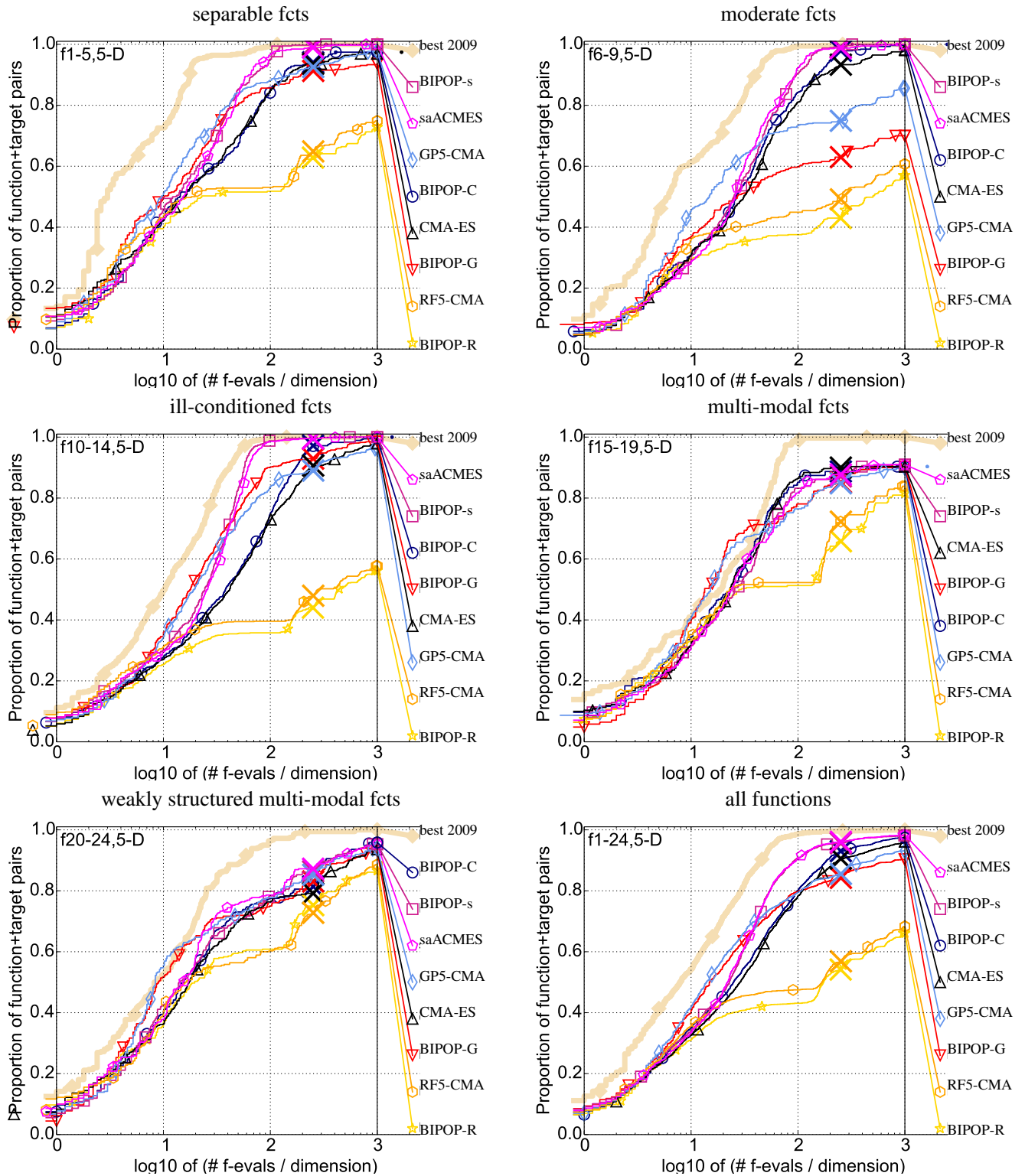


Figure 2: Bootstrapped empirical cumulative distribution of the number of objective function evaluations divided by dimension (FEvals/DIM) for all functions and subgroups in 5-D. The targets are chosen from  $10^{[-8..2]}$  such that the bestGECCO2009 artificial algorithm just not reached them within a given budget of  $k \times \text{DIM}$ , with  $k \in \{0.5, 1.2, 3, 10, 50\}$ . The “best 2009” line corresponds to the best ERT observed during BBOB 2009 for each selected target.

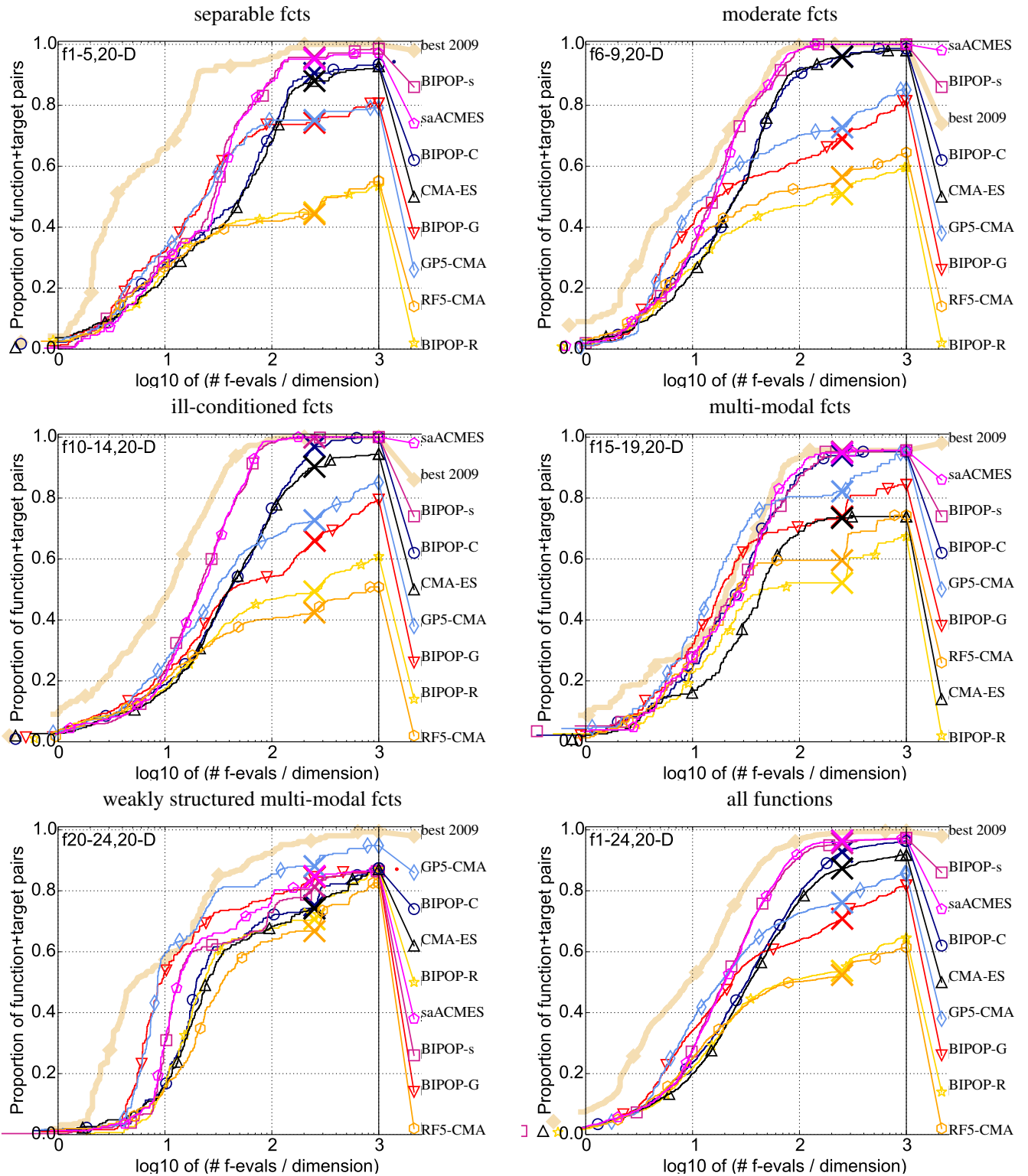


Figure 3: Bootstrapped empirical cumulative distribution of the number of objective function evaluations divided by dimension (FEvals/DIM) for all functions and subgroups in 20-D. The targets are chosen from  $10^{[-8..2]}$  such that the bestGECCO2009 artificial algorithm just not reached them within a given budget of  $k \times \text{DIM}$ , with  $k \in \{0.5, 1.2, 3, 10, 50\}$ . The “best 2009” line corresponds to the best ERT observed during BBOB 2009 for each selected target.

trials as the number of the original function evaluations (FEs) executed during each trial until the best function value reached  $f_i$ , summed over all trials and divided by the number of trials that actually reached  $f_i$  [7].

As we can see in Figure 1, the 24 functions can be roughly divided into two groups according to the algorithm which performed the best (at least in 10D and 20D). The first group of functions where the CMA-ES performed best consists of functions 1, 3, 4, 6, and 20 while on functions 2, 5, 7, 10, 11, 13–16, 18, 21, 23, and 24, GP5-CMA-ES is usually better. The usage of the BIPOP versions generally leads to no improvement or even to performance decrease.

The graphs in Figures 2 and 3 summarize the performance over subgroups of the benchmark functions and show the proportion of algorithm runs that reached the target value  $f_i \in 10^{[-8..2]}$  indeed ( $f_i$  was actually different for each respective function, see the figures captions). Roughly speaking, the higher the colored line, the better the performance of the algorithm is for the number of the original evaluations given on the horizontal axis.

Thus we can see that our GP5-CMA-ES usually outperforms the other algorithms when we consider the evaluations budget  $FEs \leq 10^{1.5}D$ , i.e.  $FEs \leq 150$  for 5D and  $FEs \leq 600$  for 20D. However, as the number of the considered original evaluations rises, the original CMA-ES or the  $s^*$ ACM-ES usually performs better. This fact can be summarized that our GP5-CMA-ES is convenient especially for the applications where a very low number of function evaluations is available, such as in [2].

## 4 Conclusions & Future Work

In this paper, we have compared the surrogate-assisted S-CMA-ES, which uses GP and RF continuous regression models, with  $s^*$ ACM-ES-k algorithm based on ordinal regression by Ranking SVM, and the original CMA-ES, all in their IPOPOP and BIPOP versions. The comparison shows that Gaussian process S-CMA-ES usually outperforms the ordinal-based  $s^*$ ACM-ES-k in early stages of the algorithm search, especially on multimodal functions (BBOB functions 15–24). However, the algorithms and surrogate models should be further analyzed and compared since, for example, the NEWUOA [16] or SMAC [10, 11] algorithms spend a considerably lower number of function evaluations than the CMA-ES in these early optimization phases. The BIPOP versions of the algorithms did not increase performances of appropriate IPOPOP versions except BBOB function 5.

A natural perspective of improving S-CMA-ES is to make the number of model-evaluated generations self-adaptive. We will additionally investigate different properties of continuous and ordinal regression in view of their applicability as regression models. Different cases and benchmarks where the ordinal regression is clearly superior to continuous regression will be further identified. For

example, hybrid surrogate models combining both kinds of regression will be attempted.

## Acknowledgements

This work was supported by the Czech Science Foundation (GAČR) grant P103/13-17187S, by the Grant Agency of the Czech Technical University in Prague with its grant No. SGS14/205/OHK4/3T/14, and by the project “National Institute of Mental Health (NIMH-CZ)”, grant number CZ.1.05/2.1.00/03.0078 (and the European Regional Development Fund.). Further, access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum, provided under the programme “Projects of Large Infrastructure for Research, Development, and Innovations” (LM2010005), is greatly appreciated.

## References

- [1] Auger, A., Hansen, N.: A restart CMA evolution strategy with increasing population size. In: The 2005 IEEE Congress on Evolutionary Computation **2**, 1769–1776, IEEE, Sept. 2005
- [2] Baerns, M., Holeňa, M.: Combinatorial development of solid catalytic materials. Design of high-throughput experiments, data analysis, data mining. Imperial College Press / World Scientific, London, 2009
- [3] Bajer, L., Pitra, Z., Holeňa, M.: Benchmarking gaussian processes and random forests surrogate models on the BBOB noiseless testbed. In: Proceedings of the 17th GECCO Conference Companion, Madrid, July 2015, ACM, New York
- [4] Breiman, L.: Classification and regression trees. Chapman & Hall/CRC, 1984
- [5] Hansen, N.: The CMA evolution strategy: A comparing review. In: J. A. Lozano, P. Larranaga, I. Inza, E. Bengoetxea, (eds), Towards a New Evolutionary Computation, 192 in Studies in Fuzziness and Soft Computing, 75–102, Springer Berlin Heidelberg, Jan. 2006
- [6] Hansen, N.: Benchmarking a BI-population CMA-ES on the BBOB-2009 function testbed. In: Proceedings of the 11th Annual GECCO Conference Companion: Late Breaking Papers, GECCO’09, 2389–2396, New York, NY, USA, 2009, ACM
- [7] Hansen, N., Auger, A., Finck, S., Ros, R.: Real-parameter black-box optimization benchmarking 2012: Experimental setup. Technical Report, INRIA, 2012
- [8] Hansen, N., Finck, S., Ros, R., Auger, A.: Real-parameter black-box optimization benchmarking 2009: Noiseless functions definitions. Technical Report RR-6829, INRIA, 2009, updated February 2010
- [9] Hansen, N., Ros, R.: Benchmarking a weighted negative covariance matrix update on the BBOB-2010 noiseless testbed. In: Proceedings of the 12th Annual Conference Companion on Genetic and Evolutionary Computation, GECCO’10, 1673–1680, New York, NY, USA, 2010, ACM

- [10] Hutter, F., Hoos, H., Leyton-Brown, K.: Sequential model-based optimization for general algorithm configuration. In: C. Coello, (ed.), *Learning and Intelligent Optimization 2011*, volume 6683 of *Lecture Notes in Computer Science*, 507–523, Springer Berlin Heidelberg, 2011
- [11] Hutter, F., Hoos, H., Leyton-Brown, K.: An evaluation of sequential model-based optimization for expensive black-box functions. In: *Proceedings of the 15th Annual Conference Companion on Genetic and Evolutionary Computation, GECCO'13 Companion*, 1209–1216, New York, NY, USA, 2013, ACM
- [12] Loshchilov, I., Schoenauer, M., Sebag, M.: Alternative restart strategies for CMA-ES. In: C. A. C. Coello, V. Cutello, K. Deb, S. Forrest, G. Nicosia, and M. Pavone, (eds), *PPSN (1)*, volume 7491 of *Lecture Notes in Computer Science*, 296–305, Springer, 2012
- [13] Loshchilov, I., Schoenauer, M., Sebag, M.: Self-adaptive surrogate-assisted covariance matrix adaptation evolution strategy. In: *Proceedings of the 14th GECCO, GECCO '12*, 321–328, New York, NY, USA, 2012, ACM
- [14] Loshchilov, I., Schoenauer, M., Sebag, M.: BI-population CMA-ES algorithms with surrogate models and line searches. In: *Genetic and Evolutionary Computation Conference (GECCO Companion)*, 1177–1184, ACM Press, July 2013
- [15] Loshchilov, I., Schoenauer, M., Sebag, M.: Intensive surrogate model exploitation in self-adaptive surrogate-assisted CMA-ES (saACM-ES). In: *Genetic and Evolutionary Computation Conference (GECCO)*, 439–446, ACM Press, July 2013
- [16] Powell, M. J. D.: The NEWUOA software for unconstrained optimization without derivatives. In: G. D. Pillo, M. Roma, (eds), *Large-Scale Nonlinear Optimization*, number 83 in *Nonconvex Optimization and Its Applications*, 255–297, Springer US, 2006
- [17] Rasmussen, C. E., Williams, C. K. I.: *Gaussian processes for machine learning*. *Adaptative Computation and Machine Learning Series*, MIT Press, 2006