Overview of Surrogate-model Versions of Covariance Matrix Adaptation Evolution Strategy

Zbyněk Pitra
National Institute of Mental Health, Czech Republic
Czech Technical University
Břehová 7, Prague 115 19, Czech Republic
z.pitra@gmail.com

Lukáš Bajer
Charles University
Malostranské náměstí 25
Prague 118 00, Czech Republic
bajeluk@gmail.com

Jakub Repický
Czech Academy of Sciences, Czech Republic
Charles University
Malostran. nám. 25, Prague 118 00, Czech Republic
j.repicky@gmail.com

Martin Holeňa
Czech Academy of Sciences
Pod Vodárenskou věží 2
Prague 182 07, Czech Republic
martin@cs.cas.cz

ABSTRACT

Evaluation of real-world black-box objective functions is in many optimization problems very time-consuming or expensive. Therefore, surrogate regression models, used instead of the expensive objective function and in that way decreasing the number of its evaluations, have received a lot of attention. Here, we briefly survey surrogate-assisted versions of the state-of-the-art algorithm for continuous black-box optimization — the CMA-ES (Covariance Matrix Adaptation Evolution Strategy). We compare five of them, together with the original CMA-ES, on the noiseless benchmarks of the Comparing-Continuous-Optimisers platform in the expensive scenario, where only a small budget of evaluations is available.

CCS CONCEPTS

•**Computing methodologies → Continuous space search;** *Model development and analysis;* Uncertainty quantification;

KEYWORDS

black-box optimization, evolutionary optimization, surrogate modelling

ACM Reference format:
Zbyněk Pitra, Lukáš Bajer, Jakub Repický, and Martin Holeňa. 2017. Overview of Surrogate-model Versions of Covariance
Matrix Adaptation Evolution Strategy. In *Proceedings of GECCO '17 Companion, Berlin, Germany, July 15-19, 2017,* 8 pages.
DOI: http://dx.doi.org/10.1145/3067695.3082539

1 INTRODUCTION

A principal challenge in many research and engineering tasks is the optimization of problems with no information about its mathematical description. Functions describing such problems are called *black-box functions*, i. e., functions for which we are only able to obtain values at specified points of the input space. Black-box optimization problems regularly appear in applications where the values of a *fitness* or an *objective* function can be obtained only empirically through measurements, expensive experiments, or via time-consuming computer simulations. The absence of derivatives turns the application of traditional methods of smooth optimization difficult. On the other hand, stochastic optimization methods such as evolutionary algorithms appear to be quite successful for continuous black-box optimization.

If the evaluation of the objective function is expensive, evolutionary optimization becomes less useful due to a large amount of evaluations necessary to achieve the optimal value. Surrogate modelling, originally from the field of smooth optimization, is an approach to reduce the number of evaluations of the expensive function through using its regression model [29]. This model, a. k. a. *surrogate model*, is trained on the already available input–output value pairs $(\mathbf{x}_i, y_i), i = 1, \ldots, N$, and is used instead of the original expensive fitness to evaluate some of the points needed by the optimization algorithm. To our knowledge, the following types of regression models have been employed so far in single-objective continuous black-box optimization:

- low degree polynomials [4, 21], which are models in the spirit of traditional *response surface models* [28];
- *artificial neural networks*, in particular multilayer perceptrons and radial basis function networks [36];
- *support vector machine* (SVM) *regression* [23, 25];
- *Gaussian processes* (GPs), a. k. a. kriging [5, 9, 22, 30, 38];
- *random forests* (RFs), i. e., ensembles of decision trees [5].

The *Covariance Matrix Adaptation Evolution Strategy* (CMA-ES) proposed by Hansen and Ostermeier [17] is considered as one of the most successful continuous black-box optimization algorithms. It will be described in some detail at the beginning of Section 2. The investigation of combining surrogate models with the CMA-ES resulted in various algorithms, several of which we will describe in more detail in the following text.

In 2001, Jin et al. [19] proposed the following two evolution control strategies for the utilization of surrogate models in the CMA-ES: *individual-based* strategy evaluates λ points using the original fitness function selected out of a larger set of points evaluated by the surrogate model. In *generation-based* strategy, the entire population is evaluated on the original fitness function for η generations and on the model for the subsequent $\kappa > \eta$ generations.

The individual evolution control was employed in 2002 by Emmerich et al. proposing the first GP based pre-selection strategy for the CMA-ES in *Metamodel-Assisted Evolution Strategy* (MA-ES) [11] (see Section 2.2 for more information).

In 2004, Least-Square minimization has been used in LS-CMA-ES [4] to train a quadratic model of the fitness function for covariance matrix adaptation in the CMA-ES. This surrogate model employed automatic detection of the model inaccuracy allowing to switch between the original CMA-ES and the surrogate model.

One year later, *Gaussian Process Optimization Procedure* (GPOP) [9] suggests a different approach to surrogate modelling: in each GPOP generation, a local GP model is constructed around the so-far-best solution and then the model is directly optimized by the CMA-ES to find its optimum which is subsequently evaluated using the original fitness function.

An effective combination of building local surrogate models and controlling changes in population ranking after fitness function evaluation is incorporated in the *local meta-model CMA-ES* (lmm-CMA-ES) proposed in 2006 by Kern et al. [21] and later improved by Auger et al. in 2013 [2] (see Section 2.3).

In 2010, GPs were combined with the CMA-ES in order to find robust solutions on noisy functions in *Kriging metamodelling based CMA-ES* [22]. The algorithm builds a local GP model for each offspring using the points from the archive evaluated with the original noisy fitness function and subsequently estimates the function values without noise. The original fitness evaluation is performed only if the archive does not contain a representative sample set for the surrogate model construction. The covariance matrix of the CMA-ES is also employed to transform the input space.

The ^{s*}ACM-ES proposed in 2012 by Loshchilov et al. [23] employs SVM ordinal regression to estimate the ordering of the fitness function values. The algorithm will be described in some detail in Section 2.4.

Another surrogate-assisted approach using an ensemble of local GP models sharing the same parameters has been proposed in 2013 by Lu et al. [26]. The algorithm selects the best points out of a larger population evaluated using the model according to one of several implemented strategies.

In 2015, the combination of the *Efficient Global Optimization* algorithm [20] using GP for direct optimization and the CMA-ES resulted in the *EGO-CMA* algorithm [27]. The algorithm runs EGO for a few iterations; subsequently, it estimates the core CMA-ES initial variables such as the covariance matrix and the step-size from the EGO Gaussian process model, and finally starts CMA-ES with the computed initial values.

In the same year, the *Surrogate CMA-ES* (S-CMA-ES) algorithm employing GPs and RFs in generation-based evolution control was introduced in [5]. It will be presented below in 2.5, and its extension proposed in [30], called *Doubly Trained Surrogate CMA-ES* (DTS-CMA-ES), in 2.6.

This paper gives an overview of several previously mentioned surrogate-assisted versions of the CMA-ES and provides a comparison among them and with the original CMA-ES. The comparison is performed on the COCO/BBOB testbed [15, 16]. Our selection of those algorithms is based primarily on their properties and published results, but partially also on the availability or at least reproducibility of their implementation.

The remainder of the paper is organized as follows. Section 2 describes the tested algorithms. Section 3 contains experimental setup and results. Section 4 summarizes the results and draws conclusions.

2 SELECTED SURROGATE MODELLING ALGORITHMS IN THE CMA-ES CONTEXT

In this section, we give some details about the CMA-ES and 5 out of the above mentioned surrogate-model versions of it that we consider the most important.

2.1 CMA-ES

The base of all considered algorithms in this article is the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) [13]. In each generation g, the CMA-ES generates λ new candidate solutions $\mathbf{x}_k \in \mathbb{R}^D$, $k = 1, \ldots, \lambda$, from a multivariate normal distribution $\mathcal{N}(\mathbf{m}, \sigma^2\mathbf{C})$ where $\mathbf{m}$ is the mean of the current estimate of the probability distribution and $\sigma^2\mathbf{C}$ its covariance matrix, decomposed in such a way that the scalar σ is used as the current step-size. After a new offspring is evaluated with the fitness function f, the algorithm selects the ordering of μ best points with the lowest function value to improve distribution parameters for the next generation.

The fact that the CMA-ES uses only the ranking of function values for the adaptation of σ and $\mathbf{C}$ makes it invariant to order-preserving transformations of the fitness function and to general linear transformations of the search space.

To be more robust on multimodal functions and to avoid premature convergence to local optima, the CMA-ES utilizes restart strategies. A multi-start strategy where the population size is doubled in each restart is referred to as IPOP-CMA-ES [3]. A BIPOP-CMA-ES [14] switches between IPOP-CMA-ES strategy and a strategy where the population size and the step-size of the restarted algorithm are rather smaller than their previous values. Each time the algorithm restarts, the strategy with the smaller number of already used function evaluations is applied.

2.2 MA-ES

In the Metamodel-Assisted Evolution Strategy (MA-ES), a Gaussian process-based surrogate model guides an evolution strategy by a mechanism of *pre-selection* [11], whereby λ most promising individuals are pre-selected from an extended population of $\lambda_{\mathrm{Pre}} > \lambda$ points according to the GP predicted mean. In [38], the *Probability of Improvement*, i. e., the probability that the point's function value will be lower than the so far achieved minimum of the original fitness function, was examined as the pre-selection criterion. The surrogate model in the MA-ES is trained on a number of most recently evaluated points.

A *Gaussian process* [34] is a collection of random variables $(f(\mathbf{x}))_{\mathbf{x}\in\mathbb{R}^D}$, such that each finite subcollection $(f(\mathbf{x}_1), \ldots, f(\mathbf{x}_N))$ has an N-dimensional normal, i. e., Gaussian distribution. Moreover, the means of those distributions have to be defined by a *mean function* $m : \mathbb{R}^D \to \mathbb{R}$ and their covariances by a *covariance function*

$k : \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}$. The values of the covariance function on all the pairs $(\mathbf{x}_i, \mathbf{x}_j)$ of some N training data form the matrix $\mathbf{K}_N \in \mathbb{R}^{N\times N}$, $\{\mathbf{K}_N\}_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j)$. If an additive i. i. d. noise is considered, the covariance matrix of $(f(\mathbf{x}_1), \ldots, f(\mathbf{x}_N))$ is $\mathbf{K}_N + \sigma_n^2 \mathbf{I}_N$, where σ_n^2 is the variance of the noise.

2.3 lmm-CMA-ES

The local meta-model CMA-ES (lmm-CMA-ES), proposed in [21] and later improved in [2, 6] to be numerically more stable, employs *locally weighted regression* [1] to build an individual quadratic surrogate model for every offspring $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^D$ using a set of k_{nn} previously evaluated points nearest to $\mathbf{x}_i$ according to the Mahalanobis distance. Each quadratic model is trained through minimizing the weighted least square error.

The original fitness evaluation phase of the CMA-ES is replaced by the following cycle. Using an *approximate ranking procedure* [35], the cycle is repeated, until the ranking difference of a prescribed fraction of the offspring in two consecutive iterations is lower then a defined threshold or the entire offspring has been evaluated using the original function:

(1) *build* models for the entire offspring;
(2) *rank* the offspring according to their model values;
(3) *evaluate* the n_b best ranked points, where an initial value n_{init} is used instead of n_b in the first iteration of the cycle;
(4) *rank* the offspring again and calculate the ranking difference.

After the cycle, adjust n_{init} according to the number of originally evaluated points from the offspring.

In a later version nlmm-CMA-ES [6], the condition used to stop evaluating new points using original fitness was improved to keep the speed-up for larger populations than the CMA-ES default.

2.4 ^{s*}ACM-ES-k

Loshchilov et al. combined the ability of Ranking SVM [18] to preserve the CMA-ES invariance properties with an adaptation of surrogate-model hyperparameter values during the search in their ^{s*}ACM-ES [23]. An extension of that algorithm using a more intensive exploitation is called ^{s*}ACM-ES-k [25].

Ranking SVM [18] is a variant of SVM for ordinal regression based on maximizing the margins between individual rank boundaries.

The ^{s*}ACM-ES-k starts with evaluating g_{start} generations using the original fitness. Then it iterates through the following steps:

(1) *train* a surrogate model with parameters θ using the points evaluated with the original fitness function;
(2) *optimize* the surrogate model by the CMA-ES for g_m generations with population size $\lambda = k_\lambda \lambda_{\text{default}}$ and the number of parents $\mu = k_\mu \mu_{\text{default}}$, where $k_\lambda, k_\mu \geq 1$;
(3) *evaluate* the original function f on the CMA-ES generated offspring using $\lambda = \lambda_{\text{default}}$ and $\mu = \mu_{\text{default}}$;
(4) *calculate* the model error and subsequently the new g_m using ranks of the original and model evaluations of the last generation;
(5) *search* the parameter space of the surrogate model by the CMA-ES to find the most convenient settings θ_{new} for the next-generation model, using the model error as a fitness function.

In [24], the ^{s*}ACM-ES-k version using BIPOP-CMA-ES, called BIPOP-^{s*}ACM-ES-k, and a hybrid of BIPOP-^{s*}ACM-ES-k, the STEP method [37] and the NEWUOA algorithm [33], called HCMA, were proposed.

2.5 S-CMA-ES

In the S-CMA-ES, introduced in [5], the generation-based evolution control [19] is employed to evaluate points sampled by the CMA-ES. At first, the population of one generation is evaluated using the original fitness. After that, a surrogate model is built using the original-evaluated data if the model has enough training points. Otherwise, the next generations are sampled and evaluated with the original fitness until the number of training points is sufficient. In a prescribed number of subsequent generations, the surrogate model is utilized to obtain function values of sampled points. The model employs GP, described in Section 2.2, or RF, described below.

Random forest is an ensemble of decision trees, in the particular case of the S-CMA-ES binary regression trees [7], where each observation $\mathbf{x} = (x_1, x_2, \ldots, x_D) \in \mathbb{R}^D$ passes through a series of binary decisions (for example $x_i \overset{?}{<} c \in \mathbb{R}$) associated with internal nodes and arrives in one of the leaf nodes containing real values utilized as the prediction of function values $\mathbf{y}$. The forest gains randomness during training by *bagging* [8]. The overall forest prediction is obtained through averaging all tree predictions.

2.6 DTS-CMA-ES

The Doubly Trained S-CMA-ES [30] is the S-CMA-ES successor replacing the generation evolution control by the *doubly trained evolution control*, which utilizes the ability of Gaussian processes to provide the distribution of predicted points.

The doubly trained evolution control proceeds in the following steps:

(1) *predict* the means and variances of the λ offspring by the GP model trained using originally-evaluated points from antecedent generations;
(2) *evaluate* the offspring using an uncertainty criterion C;
(3) *evaluate* the n_{orig} points most promising according to C using the original fitness;
(4) *retrain* the GP model including the points evaluated in (3);
(5) *evaluate* the remaining $\lambda - n_{\text{orig}}$ points using the retrained model from (4).

In [32], the DTS-CMA-ES using an ordinal-regression GP model was tested showing lower performance than the original DTS-CMA-ES using a metric-regression GP.

3 EXPERIMENTAL RESULTS

In this section, the above surveyed 5 surrogate-model versions are compared with the original CMA-ES. Because the S-CMA-ES is used both with GPs and with RFs, the comparison finally includes 7 algorithms.

3.1 Experimental Setup

The considered algorithms were compared on the set of all 24 noiseless functions from the COCO/BBOB framework [15, 16] in dimensions $D = 2, 5, 10$, and 20 on 15 different instances per function. Each algorithm had a budget of $250D$ function evaluations to reach the target distance $\Delta f_T = 10^{-8}$ from the function optimum. The parameters of the tested algorithms are summarized in the following paragraphs.

The original CMA-ES was employed in its IPOP-CMA-ES version[1] (Matlab code v. 3.61) with the following settings: the number of restarts = 4, IncPopSize = 2, $\sigma_{\text{start}} = \frac{8}{3}$, $\lambda = 4 + \lfloor 3 \log D \rfloor$. The remaining settings were left default.

The MA-ES was used with parent number $\mu = 2$ and population size $\lambda = 10$. The number of training points was set to $2\lambda = 20$ and the size of extended population λ_{Pre} was $3\lambda = 30$. The mean of GP model prediction was used as the pre-selection criterion.

The lmm-CMA-ES was used in its improved version published in [2]. The results have been downloaded from the BBOB results data archive[2] in its GECCO 2013 settings.

We have used the bipopulation version of the ^{s*}ACM-ES-k, the BIPOP-^{s*}ACM-ES-k [24]. Similarly to the lmm-CMA-ES, the algorithm results have also been downloaded from the BBOB results data archive[3] in its GECCO 2013 settings.

In connection with the S-CMA-ES, GPs were used as a surrogate model for 5 generations evaluated with the model and RFs for one generation only. The GP and RF model parameters have been taken from [31]. Similarly to [30], all the function values were normalized to zero mean and unit variance before surrogate model training. Also the remaining S-CMA-ES parameters were left the same as in [30] and the CMA-ES parameters were set identically to the original CMA-ES.

The DTS-CMA-ES was tested using the overall best settings from [30]: the GP prediction variance as the uncertainty criterion, the population size $\lambda = 8 + \lfloor 6 \log D \rfloor$, and the number of originally-evaluated points $n_{\text{orig}} = \lceil 0.05\lambda \rceil$.

3.2 Results

Results from experiments are presented in Figures 1, 2, and 3 and also in Tables 1 and 2. The graphs in Figures 1 and 2 depict the scaled logarithm $\Delta_f^{\log}$ of the median Δ_f^{med} of minimal distance from the function optimum over runs on 15 independent instances dependent on function evaluations divided by dimension (FE/D) (see [30] for details). The values are scaled to the $[-8, 0]$ interval, where -8 corresponds to the minimal and 0 to the maximal distance. This visualization was chosen due to better ability to distinguish the differences in the convergence of tested algorithms more than the default visualization used by the COCO/BBOB platform. The results averaged through the full set of all 24 benchmark functions are shown in Figure 3. More detailed results can be found on an authors' webpage[4].

[1]The results in [31] show that the performances of the IPOP- and the BIPOP-CMA-ES are almost similar in the experimental setup identical to the setup tested in this paper.
[2]http://coco.gforge.inria.fr/data-archive/2013/lmm-CMA-ES_auger_noiseless.tgz
[3]http://coco.gforge.inria.fr/data-archive/2013/BIPOP-saACM-k_loshchilov_noiseless.tgz
[4]http://uivty.cs.cas.cz/~cma/gecco2017/

Dim	5D		20D	
#FEs/#FE$_T$	1/3	1	1/3	1
CMA-ES	6.58	5.29	6.46	4.83
MA-ES	3.25	3.40	3.71	4.08
lmm-CMA-ES	2.54	2.50	3.33	2.58
^{s*}ACM-ES-k	3.88	3.33	3.17	2.54
S-CMA-ES GP	3.62	4.56	3.75	5.17
S-CMA-ES RF	5.04	5.83	4.67	5.25
DTS-CMA-ES	3.08	3.08	2.92	3.54
F_F	15.90	11.34	10.80	9.15

Table 1: Mean rank of each algorithm over the BBOB and the Iman-Davenport statistic for the 4 considered combinations of dimensionalities and evaluation budgets.

We compare all algorithms on the BBOB noiseless benchmarks using the non-parametric Friedman test [10]. To take into account different settings for the comparison, the test is conducted separately for two dimensionalities of the input space, $5D$ and $20D$, and two function evaluation budgets, a higher and a lower one. Let #FE$_\text{T}$ be the smallest number of function evaluations on which at least one algorithm reached the target, i. e., satisfied $\Delta_f^{\text{med}} \leq \Delta f_T$, or #FE$_\text{T}$ = $250D$ if no algorithm reached the target within $250D$ evaluations. We set the higher budget for the tests to #FE$_\text{T}$ and the lower budget to $\frac{\#\text{FE}_\text{T}}{3}$. The algorithms are ranked on each BBOB function with respect to Δ_f^{med} at a given budget of function evaluations.

Table 1 reports the mean ranks over all the BBOB functions and the Iman-Davenport statistic F_F. If the null hypothesis of equally distributed performance of all algorithms is valid, then F_F follows the $F_{K-1,(K-1)(N-1)}$ distribution with K the number of tested algorithms and N the number of benchmark functions. In our case, $F_{K-1,(K-1)(N-1)} = F_{6,138}$ and its critical value for the significance level $\alpha = 0.05$ is 2.16. In each test scenario, F_F exceeds the critical value, thus, we reject the null hypothesis.

We proceed to a pairwise $K \times K$ comparison of the algorithms' mean ranks by the post-hoc Friedman test with the Bergmann-Hommel correction of the family-wise error [12]. To better illustrate the algorithms' differences, we also count the number of benchmark functions at which one algorithm was ranked higher than the other. The test is again performed separately for each considered combination of dimensionalities and function evaluation budgets. The pairwise score and the statistical significance of the pairwise rank differences are reported in Table 2.

The MA-ES has clearly the best performance out of compared algorithms on f_6, which is generally very hard to regress. It also achieves very good results on multi-modal functions (f_{15-19}). The reason for the MA-ES behavior might be in the fact that the individual strategy does not mislead the CMA-ES by imprecise model evaluations. On the other hand, this strategy probably leads to saving fewer evaluations on the remaining functions.

The lmm-CMA-ES presents very balanced performance on all benchmarks (apart from the results on f_{19}, which could be erroneous). The ^{s*}ACM-ES-k usually provides one of the best performances on unimodal functions, especially in $20D$. The speed-up of the ^{s*}ACM-ES-k is not affected by the growing dimension so much as the speed-up of other algorithms, which can be caused also by the potentially lower sensitivity of the Ranking SVM model to the curse of dimensionality.

The only surrogate-assisted algorithm with significantly lower performance than the remainder was the S-CMA-ES. The results of RFs were generally worse than GPs results in combination with the S-CMA-ES. However, the RF models can be considered more robust in $20D$ than the GP models, which are known to suffer from higher dimensionalities.

The DTS-CMA-ES often converges within the budget of 100 FE$/D$; then it sometimes gets stuck in a local optimum, which can be observed more often in $20D$. The best results from the compared algorithms are achieved on smooth unimodal functions in $5D$ (f_2, f_{8-11}, and f_{14}). This success is probably caused by the constant low number of original-evaluated points per generation regardless of the model error.

4 CONCLUSIONS

In this paper, we have presented an overview of several algorithms using surrogate models to speed up the state-of-the-art black-box optimization algorithm CMA-ES. Five surrogate-model versions of the CMA-ES using four different surrogate models were compared with the original CMA-ES on the noiseless benchmarks from the COCO/BBOB framework.

We have found that all surrogate models significantly improve the CMA-ES convergence on most of the noiseless functions from the COCO/BBOB testbed. On the other hand, there is no surrogate model or algorithm using the model to improve the CMA-ES significantly better than the remaining algorithms in the expensive scenario.

Both the ^{s*}ACM-ES-k and the lmm-CMA-ES follow the CMA-ES invariance to monotonous transformations via intrinsic usage of f-values ranking instead of their values which is probably the main reason of their success on a broad spectrum of COCO/BBOB functions. This fact suggests a possible direction of research into GP-based surrogate-assisted CMA-ES algorithms. On the other hand, on multi-modal functions, the algorithms using GPs as a surrogate model become competitively successful, because such functions are difficult to approximate by simpler models such as quadratic regression or Ranking SVM. Another possible direction of development can be increasing the level of model-parameter optimization such as the online choice of the GP model's covariance function.

ACKNOWLEDGMENTS

The reported research was supported by the CSF grant No. 17-01251, the GA of the CTU No. SGS17/193/OHK4/3T/14, and by the project Nr. LO1611 with a financial support from the MEYS under the NPU I program. Further, access to the National Grid Infrastructure MetaCentrum, provided under the programme LM2010005, is greatly appreciated.

REFERENCES

[1] C. G. Atkeson, A. W. Moore, and S. Schaal. 1997. Locally Weighted Learning. *Artif. Intell. Rev.* 11, 1-5 (Feb. 1997), 11–73.

[2] A. Auger, D. Brockhoff, and N. Hansen. 2013. Benchmarking the local metamodel CMA-ES on the noiseless BBOB'2013 test bed. In *Genetic and Evolutionary Computation Conference, GECCO '13, Amsterdam, The Netherlands, July 6-10, 2013, Companion Material Proceedings*. 1225–1232.

[3] A. Auger and N. Hansen. 2005. A restart CMA evolution strategy with increasing population size. In *The 2005 IEEE Congress on Evolutionary Computation, 2005*, Vol. 2. IEEE, 1769–1776.

[4] A. Auger, M. Schoenauer, and N. Vanhaecke. 2004. LS-CMA-ES: A Second-Order Algorithm for Covariance Matrix Adaptation. In *Parallel Problem Solving from Nature - PPSN VIII (Lecture Notes in Computer Science)*, Vol. 3242. Springer Berlin Heidelberg, 182–191.

[5] L. Bajer, Z. Pitra, and M. Holeňa. 2015. Benchmarking Gaussian Processes and Random Forests Surrogate Models on the BBOB Noiseless Testbed. In *Proceedings of the 17th GECCO Conference Companion*. ACM, New York, Madrid.

[6] Z. Bouzarkouna, A. Auger, and D. Y. Ding. 2010. Investigating the Local-Meta-Model CMA-ES for Large Population Sizes. In *3rd European event on Bio-inspired algorithms for continuous parameter optimisation (Lecture Notes in Computer Science)*, Vol. 6024. Istanbul, Turkey, 402–411.

[7] L. Breiman. 1984. *Classification and regression trees.* Chapman & Hall/CRC.

[8] L. Breiman. 1996. Bagging predictors. *Machine learning* 24, 2 (1996), 123–140.

[9] D. Büche, N. N. Schraudolph, and P. Koumoutsakos. 2005. Accelerating evolutionary algorithms with Gaussian process fitness function models. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 35, 2 (May 2005), 183–194.

[10] J. Demšar. 2006. Statistical Comparisons of Classifiers over Multiple Data Sets. *Journal of Machine Learning Research* 7 (2006), 1–30.

[11] M. Emmerich, A. Giotis, M. Özdemir, T. Bäck, and K. Giannakoglou. 2002. *Metamodel—Assisted Evolution Strategies.* Springer Berlin Heidelberg, 361–370.

[12] S. García and F. Herrera. 2008. An Extension on "Statistical Comparisons of Classifiers over Multiple Data Sets" for all Pairwise Comparisons. *Journal of Machine Learning Research* 9 (2008), 2677–2694.

[13] N. Hansen. 2006. The CMA Evolution Strategy: A Comparing Review. In *Towards a New Evolutionary Computation.* Number 192 in Studies in Fuzziness and Soft Computing. Springer Berlin Heidelberg, 75–102.

[14] N. Hansen. 2009. 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).* ACM, New York, NY, USA, 2389–2396.

[15] N. Hansen, A. Auger, S. Finck, and R. Ros. 2012. *Real-Parameter Black-Box Optimization Benchmarking 2012: Experimental Setup.* Technical Report. INRIA.

[16] N. Hansen, S. Finck, R. Ros, and A. Auger. 2009. *Real-Parameter Black-Box Optimization Benchmarking 2009: Noiseless Functions Definitions.* Technical Report RR-6829. INRIA. Updated February 2010.

[17] N. Hansen and A. Ostermeier. 1996. Adapting Arbitrary Normal Mutation Distributions in Evolution Strategies: The Covariance Matrix Adaptation. In *Proceedings of IEEE International Conference on Evolutionary Computation, 1996.* IEEE, 312–317.

[18] R. Herbrich, T. Graepel, and K. Obermayer. 1999. Support Vector Learning for Ordinal Regression. In *1999 Ninth International Conference on Artificial Neural Networks ICANN 99.*, Vol. 1. 97–102.

[19] Y. Jin, M. Olhofer, and B. Sendhoff. 2001. Managing approximate models in evolutionary aerodynamic design optimization. In *Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat. No.01TH8546)*, Vol. 1. 592–599.

[20] D. R. Jones, M. Schonlau, and W. J. Welch. 1998. Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global Optimization* 13, 4 (1998), 455–492.

[21] S. Kern, N. Hansen, and P. Koumoutsakos. 2006. Local Meta-models for Optimization Using Evolution Strategies. In *Parallel Problem Solving from Nature - PPSN IX (Lecture Notes in Computer Science)*, Vol. 4193. Springer Berlin Heidelberg, 939–948.

[22] J.W. Kruisselbrink, M.T.M. Emmerich, A.H. Deutz, and T. Bäck. 2010. A robust optimization approach using Kriging metamodels for robustness approximation in the CMA-ES. In *2010 IEEE Congress on Evolutionary Computation (CEC).* 1–8.

[23] I. Loshchilov, M. Schoenauer, and M. Sebag. 2012. Self-adaptive surrogate-assisted covariance matrix adaptation evolution strategy. In *Proceedings of the 14th GECCO (GECCO '12).* ACM, New York, NY, USA, 321–328.

[24] I. Loshchilov, M. Schoenauer, and M. Sebag. 2013. BI-population CMA-ES Algorithms with Surrogate Models and Line Searches. In *Genetic and Evolutionary Computation Conference (GECCO Companion).* ACM Press, 1177–1184.

[25] I. Loshchilov, M. Schoenauer, and M. Sebag. 2013. Intensive Surrogate Model Exploitation in Self-adaptive Surrogate-assisted CMA-ES (saACM-ES). In *Genetic and Evolutionary Computation Conference (GECCO).* ACM Press, 439–446.

[26] J. Lu, B. Li, and Y. Jin. 2013. An Evolution Strategy Assisted by an Ensemble of Local Gaussian Process Models. In *Proceedings of the 15th Annual Conference on Genetic and Evolutionary Computation (GECCO '13).* ACM, New York, NY, USA, 447–454.

[27] H. Mohammadi, R.L. Riche, and E. Touboul. 2015. Making EGO and CMA-ES Complementary for Global Optimization. In *Learning and Intelligent Optimization (Lecture Notes in Computer Science)*, C. Dhaenens, L. Jourdan, and M.-E. Marmion

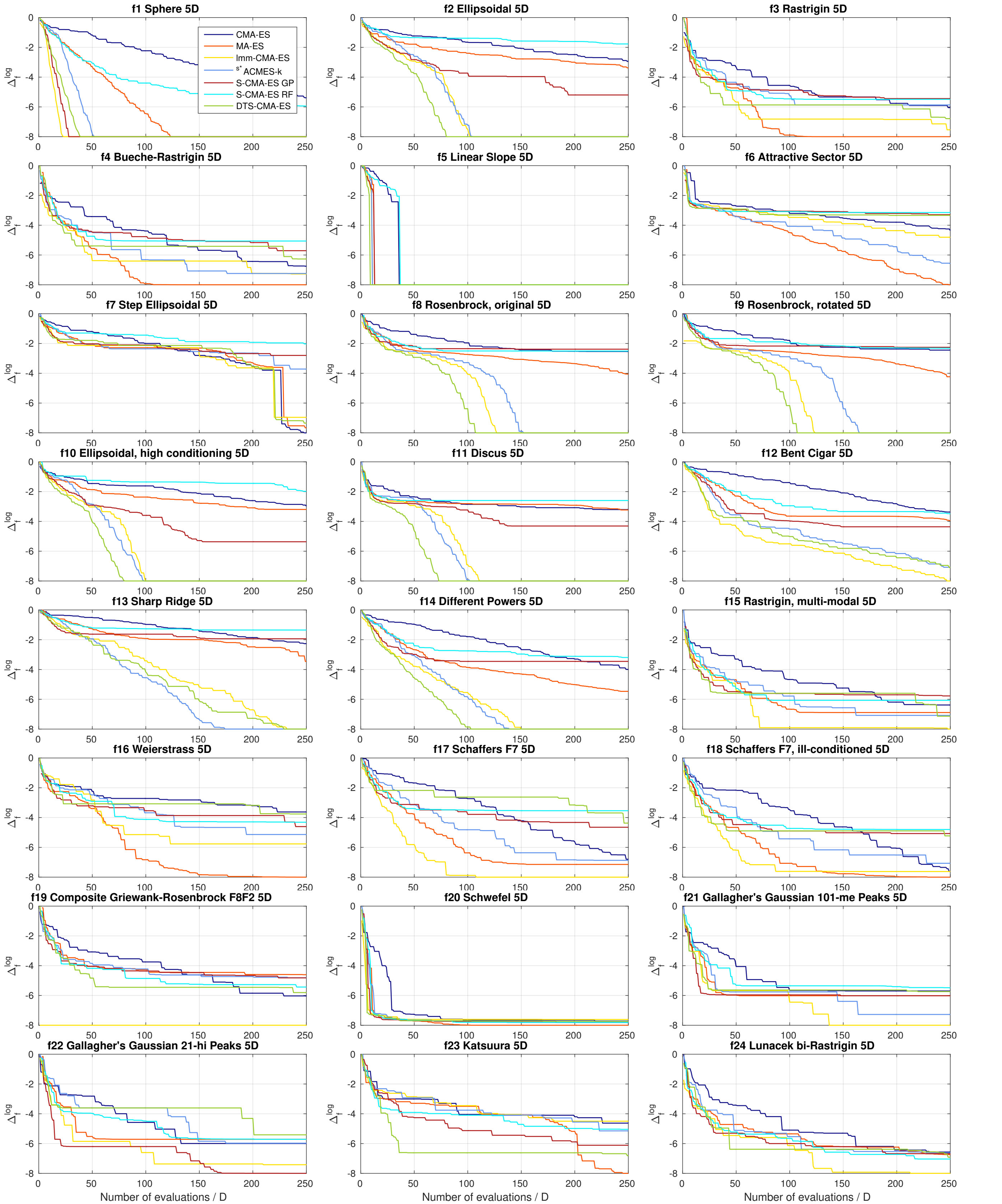

Figure 1: Algorithm comparison on 24 BBOB noiseless functions in $5D$

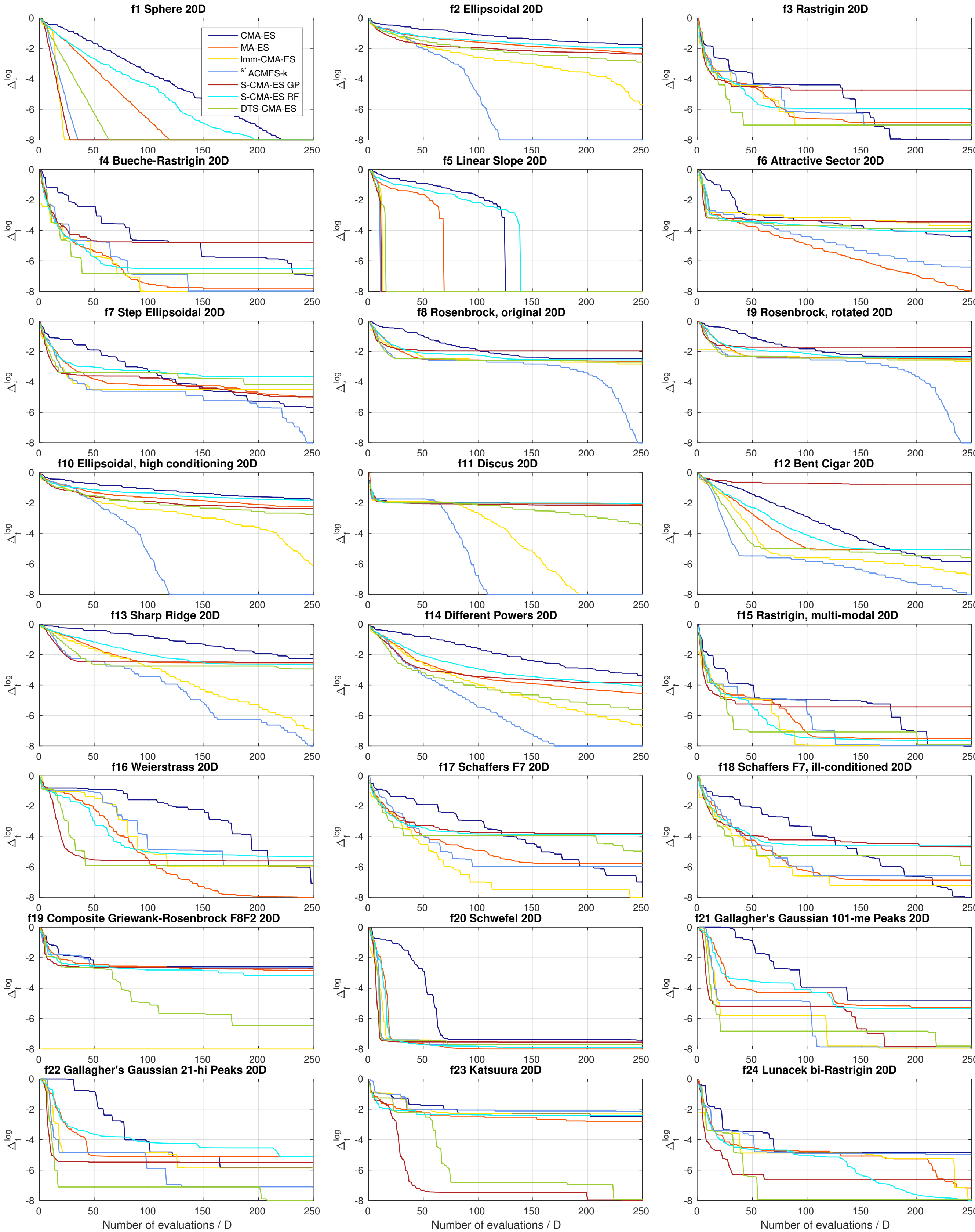

Figure 2: Algorithm comparison on 24 BBOB noiseless functions in *20D*

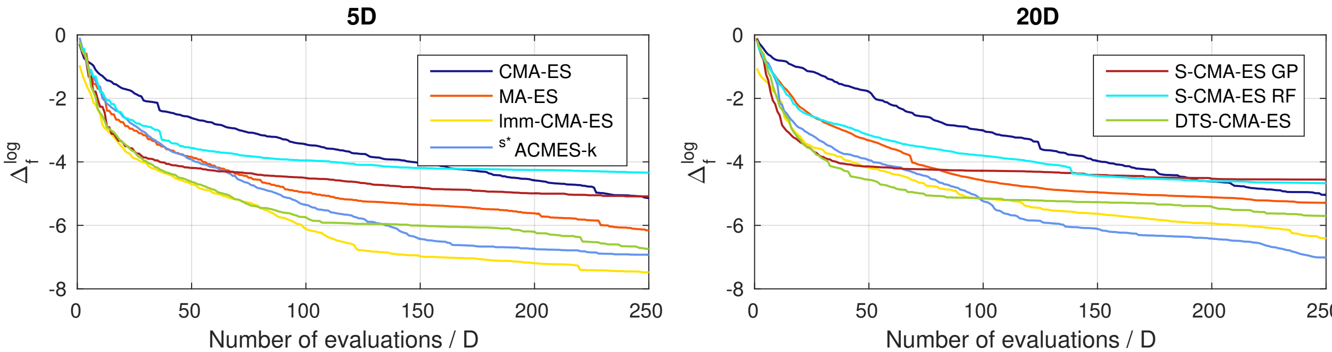

Figure 3: Algorithm comparison using averaged $\Delta_f^{\log}$ values in $5D$ and $20D$

5*D*	CMA-ES		MA-ES		lmm-CMA-ES		^{s*}ACM-ES-k		S-CMA-ES GP		S-CMA-ES RF		DTS-CMA-ES	
$^{\#FEs}/_{\#FE_T}$	⅓	1	⅓	1	⅓	1	⅓	1	⅓	1	⅓	1	⅓	1
CMA-ES	—	—	1	3	1	4	1	5	1	8	4	14	2	7
MA-ES	23*	21*	—	—	8	8	15	10	12	15	20*	21*	12	11
lmm-CMA-ES	23*	20*	16	16	—	—	17	16	17	21*	22*	22*	12	13
^{s*}ACM-ES-k	23*	19*	9	14	7	8	—	—	11	17	15	21*	10	9
S-CMA-ES GP	23*	16	12	9	7	3	13	7	—	—	18	18	8	6
S-CMA-ES RF	20	10	4	3	2	2	9	3	6	6	—	—	6	4
DTS-CMA-ES	22*	17*	12	13	12	11	14	15	16	18	18*	20*	—	—
20*D*	CMA-ES		MA-ES		lmm-CMA-ES		^{s*}ACM-ES-k		S-CMA-ES GP		S-CMA-ES RF		DTS-CMA-ES	
$^{\#FEs}/_{\#FE_T}$	⅓	1	⅓	1	⅓	1	⅓	1	⅓	1	⅓	1	⅓	1
CMA-ES	—	—	3	7	3	7	3	6	1	12	2	11	1	9
MA-ES	21*	17	—	—	11	5	10	7	11	16	18	18	8	7
lmm-CMA-ES	21*	17*	13	19	—	—	13	10	13	21*	17	20*	11	19
^{s*}ACM-ES-k	21*	18*	14	17	11	14	—	—	13	20*	18	20*	15	18
S-CMA-ES GP	23*	12	13	8	11	3	11	4	—	—	13	12	7	5
S-CMA-ES RF	22*	13	6	6	7	4	6	4	11	12	—	—	4	3
DTS-CMA-ES	23*	15	16	17	13	5	9	6	17	19	20	21	—	—

Table 2: A pairwise comparison of the algorithms in $5D$ and $20D$ on the BBOB noiseless functions for different evaluation budgets. The number of wins of i-th algorithm against j-th algorithm over all benchmark functions is given in i-th row and j-th column. The asterisk marks the row algorithm achieving a significantly lower value of the objective function than the column algorithm (on medians over 15 instances taken from all 24 functions) according to the Friedman post-hoc test with the Bergmann-Hommel correction at the family-wise significance level $\alpha = 0.05$.

(Eds.), Vol. 8994. Springer International Publishing, 287–292.

[28] R.H. Myers and D.C. Montgomery. 1995. *Response Surface Methodology: Process and Product in Optimization Using Designed Experiments* (1st ed.). John Wiley & Sons, Inc., New York, NY, USA.

[29] Y.S. Ong, P.B. Nair, and A.J. Keane. 2003. Evolutionary Optimization of Computationally Expensive Problems via Surrogate Modeling. *AIAA Journal* 41, 4 (2003), 687–696.

[30] Z. Pitra, L. Bajer, and M. Holeňa. 2016. *Doubly Trained Evolution Control for the Surrogate CMA-ES.* Springer International Publishing, Cham, 59–68.

[31] Z. Pitra, L. Bajer, and M. Holeňa. 2015. Comparing SVM, Gaussian Process and Random Forest Surrogate Models for the CMA-ES. In *ITAT 2015: Information Technologies - Applications and Theory.* CreateSpace Independent Publishing Platform, North Charleston, USA, 186–193.

[32] Z. Pitra, L. Bajer, J. Repický, and M. Holeňa. 2017. Comparison of Ordinal and Metric Gaussian Process Regression as Surrogate Models for CMA Evolution Strategy. In *Proceedings of GECCO '17 Companion, Berlin, Germany, July 15–19, 2017.* ACM, 8.

[33] M. J. D. Powell. 2006. The NEWUOA software for unconstrained optimization without derivatives. In *Large-Scale Nonlinear Optimization.* Number 83 in Nonconvex Optimization and Its Applications. Springer US, 255–297.

[34] C.E. Rasmussen and C.K.I. Williams. 2006. *Gaussian Processes for Machine Learning.* MIT Press.

[35] T.P. Runarsson. 2004. *Constrained Evolutionary Optimization by Approximate Ranking and Surrogate Models.* Springer Berlin Heidelberg, 401–410.

[36] C. Sun, Y. Jin, R. Cheng, J. Ding, and J. Zeng. 2017. Surrogate-assisted Cooperative Swarm Optimization of High-dimensional Expensive Problems. *IEEE Transactions on Evolutionary Computation* (2017).

[37] S. Swarzberg, G. Seront, and H. Bersini. 1994. S.T.E.P.: The Easiest Way to Optimize a Function. In *Proceedings of the First IEEE Conference on Evolutionary Computation. IEEE World Congress on Computational Intelligence*, Vol. 1. 519–524.

[38] H. Ulmer, F. Streichert, and A. Zell. 2003. Evolution strategies assisted by Gaussian processes with improved preselection criterion. In *The 2003 Congress on Evolutionary Computation, 2003. CEC '03*, Vol. 1. 692–699 Vol.1.