# Assessing the suitability of surrogate models in evolutionary optimization*

Martin Holeňa and Radim Demut

Institute of Computer Science, Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 2, 18207 Prague

**Abstract.** *The paper deals with the application of evolutionary algorithms to black-box optimization, frequently encountered in biology, chemistry and engineering. In those areas, however, the evaluation of the black-box fitness is often costly and time-consuming. Such a situation is usually tackled by evaluating the original fitness only sometimes, and evaluating its appropriate response-surface model otherwise, called surrogate model of the fitness. Several kinds of models have been successful in surrogate modelling, and a variety of models of each kind can be obtained through parametrization. Therefore, real-world applications of surrogate modelling entail the problem of assessing the suitability of different models for the optimization task being solved. The present paper attempts to systematically investigate this problem. It surveys available methods to assess model suitability and reports the incorporation of several such methods in our recently proposed approach to surrogate modelling based on radial basis function networks. In addition to the commonly used global suitability of a model, it pays much attention also to its local suitability for a given input. Finally, it shows some results of testing several of the surveyed methods in two real-world applications.*

## 1 Introduction

An important application area of *evolutionary optimization algorithms* [8, 31] is *black-bock optimization*, i.e., optimization of an objective function (in evolutionary terms called fitness) that cannot be described explicitly, but is known only from its evaluations in a finite number of points in the input space. Frequently, the fitness is evaluated in some empirical way, through measurements or testing. This is typical for applications in biology, chemistry, or materials science [1]. In those domains, however, the fact that evolutionary algorithms rely solely on fitness evaluations can be quite disadvantageous because the evaluation of empirical functions encountered there is usually *time-consuming* and *costly*. For example in the evolutionary optimization of catalytic materials [1, 12], where a fitness describes the suitability of the material for a particular chemical reaction, its evaluation in one generation of the evolutionary algorithm needs several days to several weeks of time and costs several to many thousands of euros.

The usual way of dealing with a time-consuming and costly evaluation of an objective function is to evaluate such a function only sometimes, and evaluate its suitable response-surface model otherwise [19, 25]. In the context of evolutionary optimization, such a model is commonly called *surrogate model* of the fitness, and the approach is called *surrogate modelling* [10, 29, 32, 37] (occasionally also outside that context [3, 23]). Because fitness is typically assumed to be highly nonlinear, nonlinear models are used as surrogate models. So far most frequently encountered have been *Gaussian processes* [7, 28, 37] (inspired by their success in response surface modelling [20, 21, 23, 35]), *radial basis function (RBF) networks* [2, 37] and other kinds of *feedforward neural networks* [16, 18].

Due to the applicability of different kinds of models to surrogate modelling, as well as due to the possibility to construct a variety of models of each kind through an appropriate parametrization, a large number of various surrogate models can always be employed. Therefore, real-world applications of surrogate modelling entail the problem of *assessing the suitability of different models* for the optimization task being solved. Unfortunately, no systematic attention seems to have been paid to that problem so far in the area of surrogate modelling, the research in this area being focused on the integration of surrogate models with evolutionary optimization algorithms, their adaptation to the optimization tasks, and on increasing the accuracy of the constructed models [11, 14, 17, 26]. The present paper is an attempt to change the situation. We survey available methods to assess model suitability, concentrating in particular on local suitability of the model for a given input. Moreover, we give some results of testing several of the surveyed methods on two real-world applications of surrogate modelling.

In the following section, the principles of surrogate modelling are recalled and their usefulness for evolutionary optimization is documented. The key section of the paper is Section 3, in which the most important methods for assessing model suitability are explained and results of their testing are presented.

* The research reported in this paper has been supported by the Czech Science Foundation (GA ČR) grant P202/11/1368.

## 2 Surrogate modelling in evolutionary optimization

In evolutionary optimization, surrogate modelling is an approach in which the evaluation of the *original black-box fitness* is restricted to points considered to be most important in the search for its global maximum, and its appropriate *response-surface model* is evaluated otherwise. Important for searching the global maximum of a fitness function are on the one hand the *highest values* found so far, on the other hand the *diversity* of current population. Therefore, the selection of points in which the original fitness is evaluated is always based on some combination of those two criteria.

The different ways of interaction between the system input/output interface, the generated evolutionary algorithm (EA) and the surrogate model can basically be assigned to one of the following two strategies:

A. The *individual-based strategy* consists in choosing between the evaluation of the original fitness and the evaluation of its surrogate model individual-wise, for example, in the following steps:
   - **(i)** An initial set $\mathcal{E}$ of individuals is collected in which the original fitness $\eta$ was evaluated (e.g., individuals forming several first generations of the EA).
   - **(ii)** The model is trained using pairs $\{(x, \eta(x)) : x \in \mathcal{E}\}$.
   - **(iii)** The EA is run with the fitness $\eta$ replaced by the model for one generation with a population $\mathcal{Q}$ of size $qP$, where $P$ is the desired population size for the optimization of $\eta$, and $q$ is a prescribed ratio (e.g., $q = 10$ or $q = 100$).
   - **(iv)** A subset $\mathcal{P} \subset \mathcal{Q}$ of size $P$ is selected so as to contain those individuals from $\mathcal{Q}$ that are most important according to the considered criteria for the progress of optimization.
   - **(v)** For $x \in \mathcal{P}$, the original fitness is evaluated.
   - **(vi)** The set $\mathcal{E}$ is replaced by $\mathcal{E} \cup \mathcal{P}$ and the algorithm returns to (ii).

B. The *generation-based strategy* consists in choosing between both kinds of evaluation generation-wise, for example, in the following steps:
   - **(i)** An initial set $\mathcal{E}$ of individuals in which the original fitness $\eta$ was evaluated is collected like in the individual-based strategy.
   - **(ii)** The model is trained using pairs $\{(x, \eta(x)) : x \in \mathcal{E}\}$.
   - **(iii)** The EA is run with the fitness $\eta$ replaced by the model for a number $g_m$ of generations, interactively obtained from the user, with populations $\mathcal{P}_1, \ldots, \mathcal{P}_{g_m}$ of size $P$.
   - **(iv)** The EA is run with the original fitness $\eta$ for a prescribed number $g_e$ of generations with populations $\mathcal{P}_{g_m+1}, \ldots, \mathcal{P}_{g_m+g_e}$ (frequently, $g_e = 1$).
   - **(v)** The set $\mathcal{E}$ is replaced by $\mathcal{E} \cup \mathcal{P}_{g_m+1} \cup \cdots \cup \mathcal{P}_{g_m+g_e}$ and the algorithm returns to (ii).

The fact that surrogate modeling is employed in the context of costly or time-consuming objective functions effectively excludes the possibility to use those functions for tuning surrogate modeling methods, and for comparing different models and different ways of their combining with evolutionary optimization. To get around this difficulty, artificial *benchmark functions* can be used, computed analytically but expected to behave in evolutionary optimization similarly to the original fitness. As an example, Fig. 1 shows the application of surrogate modelling to a benchmark function proposed in [34] for the application area of optimization of catalytic materials (cf. [1]). The benchmark function was optimized using the system GENACAT [13, 15], one of several evolutionary optimization systems developed specifically for that application area. The evolutionary algorithm employed by GENACAT is a genetic algorithm (GA) taking into account the composition and properties of catalytic materials. As surrogate model, a RBF-network trained with data from all previous generations was used, combined with the GA according to the individual-based strategy. The results shown in Fig. 1 clearly document that surrogate modelling substantially accelerates the search for the maximum of a fitness function.

## 3 Assessing the suitability of different models

Instead of a single surrogate model, a whole *set of models* $\mathcal{F}$ an be used. Then it is necessary to decide how suitable each of them is to be evaluated instead of the original fitness $\eta$. Typically, the suitability of a model $F \in \mathcal{F}$ is assumed to be indirectly proportional to some error $\varepsilon(F)$, defined on $\mathcal{F}$ and calculated using a given sequence of data not used for the construction of $F$. Consequently, the most suitable surrogate model is the one fulfilling

$$\hat{F} = \arg\min_{F \in \mathcal{F}} \varepsilon(F). \tag{1}$$

There are various ways how $\varepsilon(F)$ takes into account the evaluation $\eta(x)$ by the original fitness and the evaluation $F(x)$ by the model for given inputs $x$, e.g., mean absolute error, mean squared error, root mean square error, relative entropy, Kullback-Leibler divergence, .... There are also two basic ways how to assure that the given sequence of data was not used for the

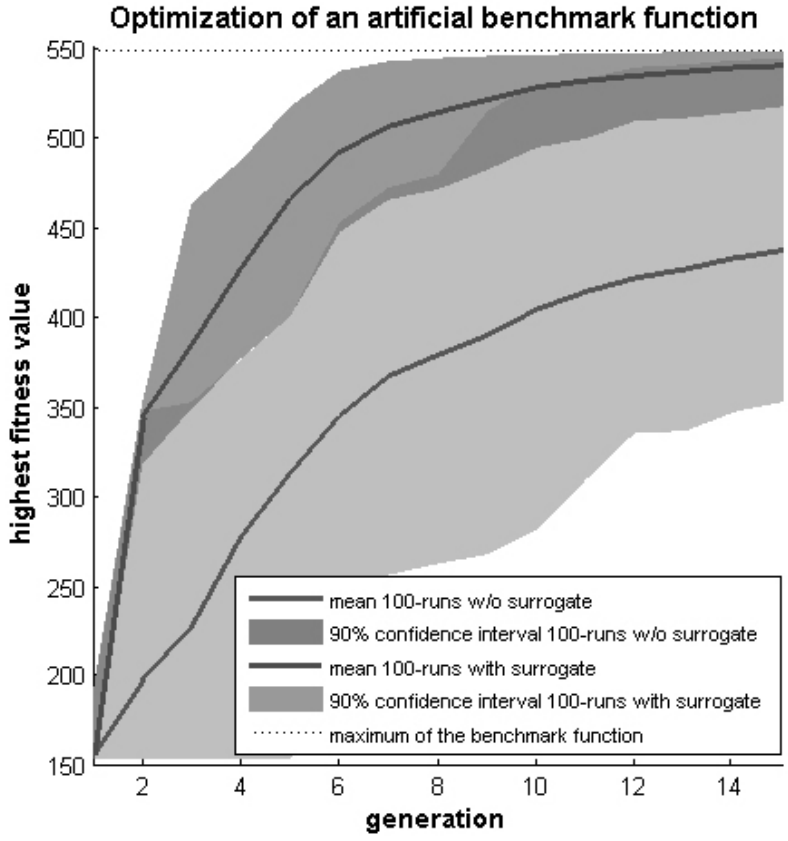


**Fig. 1.** Comparison of the highest values of the benchmark fitness function from [34] found by the evolutionary optimization system GENACAT [13, 15] without surrogate modeling and with an RBF network used as surrogate model, according to the individual-based strategy.

construction of $F$: *single split* and *cross-validation*, the latter having the important advantage that all available data are used both for model construction and for the estimation of model error.

Let us exemplify the calculation of $\varepsilon(F)$ by recalling the definition of the *root mean squared error* on an input data sequence $D$:

$$\mathrm{RMSE}(F) = \mathrm{RMSE}_D(F) = \tag{2}$$
$$= \sqrt{\frac{1}{|D|}\sum_{x\in D}(F(x)-\eta(x))^2},$$

where $|D|$ denotes the cardinality of $D$. Thus the overall root mean squared error of $F$ based on a $k$-fold crossvalidation with folds $D_1, \ldots, D_k$ is:

$$\mathrm{RMSE}(F) = \frac{1}{k}\sum_{i=1}^{k}\mathrm{RMSE}_{D_i}(F) = \tag{3}$$
$$= \frac{1}{k}\sum_{i=1}^{k}\sqrt{\frac{1}{|D_i|}\sum_{x\in D_i}(F(x)-\eta(x))^2},$$

Observe that the most suitable model $\hat{F}$ in (1) depends on the considered set of surrogate models $\mathcal{F}$, but does not depend on the inputs in which it has to be evaluated. Therefore, it can be called *globally most suitable* with respect to the given sequence of data. Its obvious advantage is that it needs to be found only once, and then it can be used for all evaluations, as long as the set $\mathcal{F}$ does not change.

Global suitability of surrogate models based on cross-validation was tested in more than a dozen evolutionary optimization tasks. Here, we show results of testing it in a task where $\mathcal{F}$ was a set of multilayer perceptrons (MLPs) with two hidden layers and different architectures. They were restricted to have $n_I = 14$ input neurons, $n_o = 3$ output neurons, and the numbers of hidden neurons $n_{H1}$ in the first and $n_{H2}$ in the second layer fulfilling the heuristic *pyramidal condition*: the number of neurons in a subsequent layer must not exceed the number of neurons in a previous layer. Consequently,

$$14 \geq n_{H1} \geq n_{H2} \geq 3, \tag{4}$$

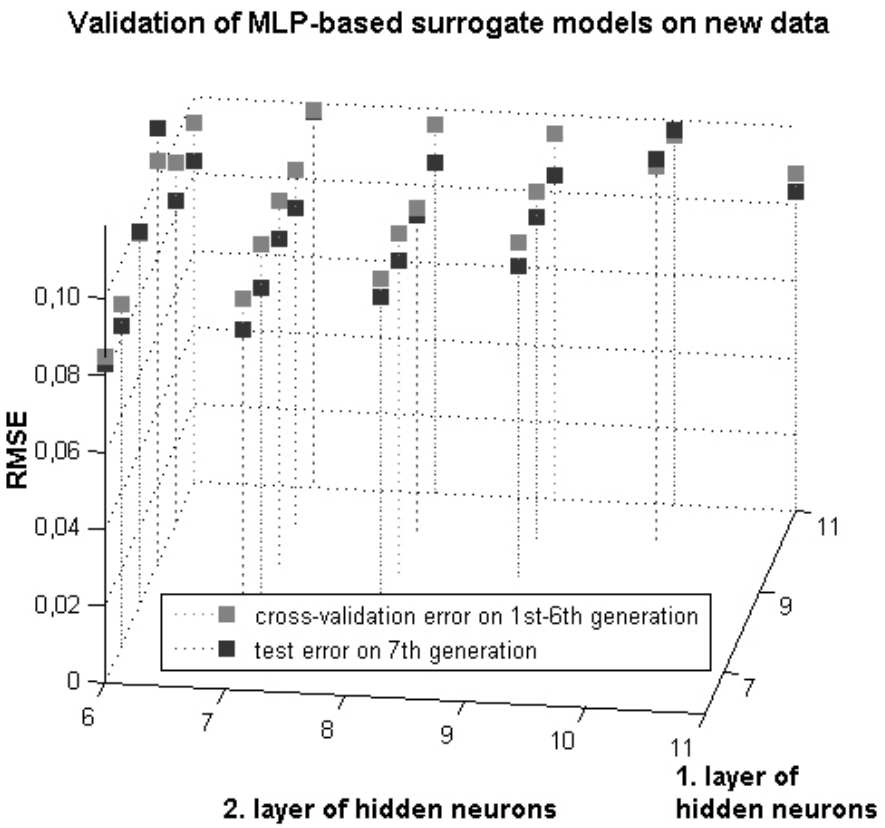


**Fig. 2.** Comparison of the RMSE of 21 surrogate models on test data, i.e., data from the $7^{\text{th}}$ generation of a genetic optimization, with the RMSE-estimate obtained for those models by means of leave-one-out cross-validation on data from the $1^{\text{st}}$–$6^{\text{th}}$ generation.

which yields 78 different MLP architectures. They were tested as follows:

1. The employed GA was run for 6 generations using the original fitness.
2. For each of the considered 78 architectures, one surrogate model was trained using all the available data from the $1^{\text{st}}$–$6^{\text{th}}$ generation.
3. The RMSE of each surrogate model on the data from the $1^{\text{st}}$–$6^{\text{th}}$ generation was estimated using leave-one-out cross-validation, according to (3).
4. The $7^{\text{th}}$ generation $G_7$ of the genetic algorithm was produced.
5. The models obtained in step 2 were used to predict the fitness of $x \in G_7$.
6. For $x \in G_7$, also the original fitness was evaluated.
7. From the results of steps 5–6, the RMSE of each surrogate model on the data from the $7^{\text{th}}$ generation was calculated according to (2), with $D = G_7$.
8. For each surrogate model, the RMSE calculated in step 7 was compared to the RMSE estimate from step 2.

Figure 2 visualizes the results of comparisons in step 8 for a subset of the considered surrogate models, namely for the 21 MLP architectures that, in addition to (4), fulfil $6 \leq n_{H1}, n_{H2} \leq 11$. The visualized results indicate that the rank of models according to the RMSE-based suitability estimated by means of leave-one-out cross-validation on the data from the $1^{\text{st}}$–$6^{\text{th}}$ generation correlates with their rank according to the RMSE on the data from the $7^{\text{th}}$ generation. We also quantified the extent of that correlation, using:

**(i)** Kendall's rank correlation coefficient $\tau$ between the ranks of models according to the suitability based on RMSE and estimated using leave-one-out cross-validation, and according to the RMSE on test data from the $7^{\text{th}}$ generation,

**(ii)** achieved significance level $p_\tau$ of the test of rank independence based on the correlation coefficient $\tau$ obtained in (i).

The results were

$$\tau = 0.77 \text{ and } p_\tau = 1.7 \cdot 10^{-8}, \tag{5}$$

which clearly confirm a strong correlation between the rank of the model suitability and the rank of model RMSE on test data.

### 3.1 Local suitability

Needless to say, the fact that a globally most suitable model achieves the least value of an error $\varepsilon$ calculated using a given sequence of data does not at all mean that it yields the most suitable prediction for every $x$ in which fitness $\eta$ can be evaluated. Therefore, we introduce the model $\hat{F}_x$ *locally most suitable for* $x$ as

$$\hat{F}_x = \arg\min_{F\in\mathcal{F}} \lambda(F,x). \tag{6}$$

Like $\varepsilon$ in (1), $\lambda$ in (6) denotes some error. Differently to $\varepsilon$, however, $\lambda$ is defined on the cartesian product $\mathcal{F}\times\mathcal{X}$, where $\mathcal{X}$ denotes the set of points in which $\eta$ can be evaluated.

Recall that a surrogate model is evaluated in points $x\in\mathcal{X}$ in which $\eta$ has not been evaluated yet. Hence, the calculation of the error $\lambda(F,x)$ must not depend on the value of $\eta(x)$. Though in the context of surrogate modelling, using such errors has not been reported yet, a number of error measures exist that could be used to this end, most importantly:

- widths of *confidence intervals* [33];
- transductive confidence [9, 27, 36];
- estimation of prediction error relying on *density estimation* [4];
- *sensitivity analysis* [5];
- several heuristics based on the *nearest neighbours* of the point of evaluation [4, 30];
- heuristic based on the variance of *bagged models* [6];

We are currently extending the surrogate model presented in [2], which is based on RBF-networks, with three of the above error measures:

**(i)** *Width of prediction intervals* for $x\in\mathcal{X}$, i.e., of confidence intervals for $\hat{F}_x(x)$ in (6) based on the linearization of the surrogate models and on the assumption of independent normally distributed residuals. First, each $F\in\mathcal{F}$ is replaced with its Taylor expansion of the 1st order, which is a linear regression model $F^{\mathrm{LIN}}$ with $d+1$ parameters, where $d$ is the dimensionality of points in $\mathcal{X}$. Hence, $\mathcal{F}$ is replaced with

$$\mathcal{F}^{\mathrm{LIN}} = \{F^{\mathrm{LIN}} : F\in\mathcal{F}\}. \tag{7}$$

Then for each considered $x\in\mathcal{X}$, the element $\hat{F}_x^{\mathrm{MLE}}$ of $\mathcal{F}^{\mathrm{LIN}}$ corresponding to the maximum-likelihood estimate of the $d+1$ parameters is found using a given training sequence of input-output pairs $(x_1,y_1),\ldots,(x_p,y_p)$. That allows to calculate for each $F\in\mathcal{F}$ the error in (6) as

$$\lambda_{(i)}(F,x) = \left| \sum_{j=1}^{p} (y_j - \hat{F}_x^{\mathrm{MLE}}(x_j))^2 - \right.$$
$$\left. - \left(1 + \frac{F_{1-\alpha}[1,p-d]}{p-d}\right) \sum_{j=1}^{p} (y_j - F(x_j))^2 \right|, \tag{8}$$

where $F_{1-\alpha}[1,p-d]$ denotes the $1-\alpha$ quantile of the Fisher-Snedecor distribution with the degrees of freedom 1 and $p-d$.

**(ii)** *Difference between the predicted value and the nearest-neighbours average* is for $k$ nearest neighbours $x_{n_1},\ldots,x_{n_k}$ calculated according to

$$\lambda_{(ii)}(F,x) = \left| \frac{\sum_{j=1}^{k} y_{n_j}}{k} - F(x) \right|. \tag{9}$$

**(iii)** *Bagged variance* requires to have some set $\mathcal{B}$ of basic models and to find, in $\mathcal{B}$, a given number $m$ of models bagged with respect to $(x_1,y_1),\ldots,(x_p,y_p)$, i.e., globally most suitable with respect to bootstrap samples from $(x_1,y_1),\ldots,(x_p,y_p)$. Recall from (1) that to find such a bagged model, an error $\varepsilon(B)$ is needed, calculated using the respective bootstrap sample. In our implementation, we always use RMSE to this end. In its calculation according to (3), hence, $D_1,\ldots,D_k$ are folds of the bootstrap sample. Using the bagged models, the final set of considered surrogate models is defined as

$$\mathcal{F} = \left\{ F : F = \sum_{j=1}^{m} B_j^F \ \&\ B_1^F,\ldots,B_m^F \in \mathcal{B} \right\}, \tag{10}$$

and the bagged variance is calculated according to

$$\lambda_{(iii)}(F,x) = \frac{1}{m}\sum_{j=1}^{m} \left( B_j^F(x) - \frac{1}{m}\sum_{j=1}^{m} B_j^F(x) \right)^2. \tag{11}$$

Because the extension of the surrogate model from [2] with those three error measures has been implemented very recently, it is now in the course of testing in a second evolutionary optimization task. Here, a part of the results from the first task will be shown, concerning the optimization of catalytic materials for high-temperature synthesis of HCN [24]. In that task, $\mathcal{F}$ was a set of five RBF networks, each with a different number of hidden neurons in the range 1–5. Those five networks were tested in a similar way as was employed in the above MLP-case. In particular:

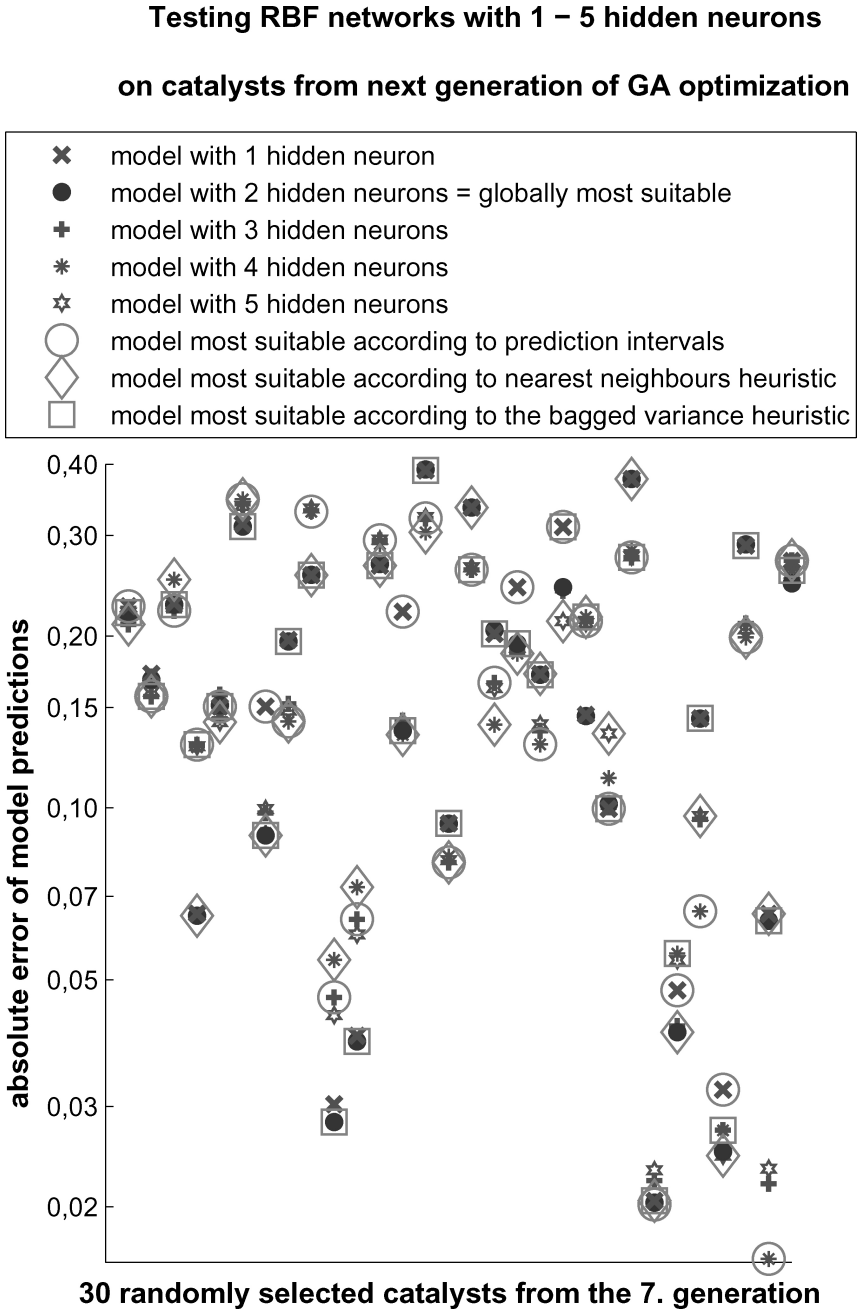


**Fig. 3.** Juxtaposition of the errors of predictions by RBF networks with 1–5 hidden neurons for 30 catalysts randomly selected from the $7^{\text{th}}$ generation of a genetic optimization of catalytic materials for high-temperature synthesis of HCN [24], with the choices of the networks locally most suitable according to $\lambda_{(i)}, \lambda_{(ii)}$ and $\lambda_{(iii)}$, and with the globally most suitable model.

1. The employed GA was again run for 6 generations using the original fitness.
2. For each of the 5 possible numbers of hidden neurons, one surrogate model was trained using all the available data from the $1^{\text{st}}$–$6^{\text{th}}$ generation.
3. The RMSE of each surrogate model on the data from the $1^{\text{st}}$–$6^{\text{th}}$ generation was estimated using 10-fold cross-validation, according to (3).
4. Based on the result of step 3, the globally most suitable model was determined.
5. The $7^{\text{th}}$ generation $G_7$ of the genetic algorithm was produced.
6. The models obtained in step 2 were used to predict the fitness of $x \in G_7$.
7. For $F \in \mathcal{F}$ and $x \in G_7$, the errors $\lambda_{(i)}(F,x)$, $\lambda_{(ii)}(F,x), \lambda_{(iii)}(F,x)$ were calculated.

8. The surrogate models locally most suitable for $x \in G_7$ according to $\lambda_{(i)}, \lambda_{(ii)}$ and $\lambda_{(iii)}$ were determined.
9. For $x \in G_7$, the original fitness was evaluated.
10. For $x \in G_7$, the absolute errors of the considered surrogate models were calculated from the results of steps 6 and 9.

Figure 3 visualizes the locally most suitable models determined in step 8 for a subset of 30 randomly selected catalytic materials from the $7^{\text{th}}$ generation. The fact that for those catalysts the absolute errors of all five RBF networks were calculated in step 10 allows to juxtapose the choices of the locally most suitable models and the globally most suitable model (model with 2 hidden neurons) with the achieved absolute errors. The juxtaposition shows that the model with the lowest absolute error was nearly always assessed as the locally most suitable by some of the three implemented error measures. Unfortunately, no one of them could be relied on in a majority of all cases.

## 4 Conclusion

This paper is, to our knowledge, a first attempt to systematically investigate available methods for assessing the suitability of surrogate models in evolutionary optimization. In addition to the commonly used global suitability of a model, it paid much attention also to its local suitability for a given input. We have incorporated three methods for assessing local suitability into our recently proposed approach to surrogate modelling based on RBF networks. The paper not only surveyed available methods for assessing suitability, but also described their testing and presented some of the testing results.

The presented results clearly confirm the usefulness of methods for assessing the suitability of surrogate models: there is a strong correlation between the ranks of models according to the RMSE-based global suitability estimated by means of leave-one-out cross-validation on data from the $1^{\text{st}}$–$6^{\text{th}}$ generation and according to the RMSE on the data from the $7^{\text{th}}$ generation, and for nearly every unseen input, the model with the lowest absolute error is indeed assessed as the locally most suitable by some of the implemented methods. Unfortunately, none of the methods is able to correctly assess the locally most suitable model for a majority of the $7^{\text{th}}$ generation, which shows that further research in this area is needed. We want to take active part in such research, pursuing the following three directions:

**(i)** Implement and test further methods for assessing local suitability, listed above in Subsection 3.1. We consider particularly interesting the transductive confidence machine [9, 27, 36] because it is a novel method and has solid theoretical fundamentals.

**(ii)** Investigate whether the success of some of the tested methods for assessing local suitability depends on the kind of dataset (in terms of the number of nominal attributes, number of continuous attributes, etc.) on which the surrogate models have been trained, or on their descriptive statistics. To this end, we want to make use of the GAME system [22], which collects a large amount of meta-data about the kind of the processed dataset and about its descriptive statistics.

**(iii)** Modify and combine the tested methods, using the results obtained in (ii), to increase their success in assessing the local suitability of surrogate models.

## References


1. M. Baerns and M. Holeňa: *Combinatorial development of solid catalytic materials. Design of high-throughput experiments, data analysis, data mining.* Imperial College Press, London, 2009.
2. L. Bajer and M. Holeňa: *Surrogate model for continuous and discrete genetic optimization based on RBF networks.* In C. Fyfe, P. Tino, C. Garcia-Osorio, and H. Yin, (eds), Intelligent Data Engineering and Automated Learning. Lecture Notes in Computer Science 6283, pp. 251–258. Springer Verlag, Berlin, 2010.
3. A.J. Booker, J. Dennis, P.D. Frank, D.B. Serafini, Torczon V., and M. Trosset: *A rigorous framework for optimization by surrogates.* Structural and Multidisciplinary Optimization, 17, 1999, 1–13.
4. Z. Bosnić and I. Kononenko: *Comparison of approaches for estimating reliability of individual regression predictions.* Data & Knowledge Engineering, 67, 2008, 504–516.
5. Z. Bosnić and I. Kononenko: *Estimation of individual prediction reliability using the local sensitivity analysis.* Applied Intelligence, 29, 2008, 187–203.
6. L. Breiman: *Bagging predictors.* Machine Learning, 24, 1996, 123–140.
7. D. Büche, N.N. Schraudolph, and P. Koumoutsakos: *Accelerating evolutionary algorithms with gaussian process fitness function models.* IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 35, 2005, 183–194.
8. C.A.C. Coello, G.B. Lamont, and D.A. van Veldhuizen: *Evolutionary algorithms for solving multi-objective problems, 2nd Edition.* Springer Verlag, Berlin, 2007.
9. A. Gammerman, V. Vovk, and V. Vapnik: *Learning by transduction.* In Uncertainty in Artificial Intelligence, pp.148–155. Morgan Kaufmann Publishers, San Francisco, 1998.
10. D. Gorissen, I. Couckuyt, P. Demeester, T. Dhaene, and K. Crombecq: *A surrogate modeling and adaptive sampling toolbox for computer based design.* Journal of Machine Learning Research, 11, 2010, 2051–2055.

11. D. Gorissen, T. Dhaene, and F. DeTurck: *Evolutionary model type selection for global surrogate modeling.* Journal of Machine Learning Research, 10, 2009, 2039–2078.
12. M. Holeňa and M. Baerns: *Computer-aided strategies for catalyst development.* In G. Ertl, H. Knözinger, F. Schüth, and J. Weitkamp, (eds), Handbook of Heterogeneous Catalysis, pp. 66–81. Wiley-VCH, Weinheim, 2008.
13. M. Holeňa, T. Cukic, U. Rodemerck, and D. Linke: *Optimization of catalysts using specific, description based genetic algorithms.* Journal of Chemical Information and Modeling, 48, 2008, 274–282.
14. M. Holeňa, D. Linke, and U. Rodemerck: *Evolutionary optimization of catalysts assisted by neural-network learning.* In Simulated Evolution and Learning. Lecture Notes in Computer Science 6457, pp.220–229. Springer Verlag, Berlin, 2010.
15. M. Holeňa, D. Linke, and U. Rodemerck: *Generator approach to evolutionary optimization of catalysts and its integration with surrogate modeling.* Catalysis Today, 159, 2011, 84–95.
16. M. Holeňa, D. Linke, and N. Steinfeldt: *Boosted neural networks in evolutionary computation.* In Neural Information Processing. Lecture Notes in Computer Science 5864, pp. 131–140. Springer Verlag, Berlin, 2009.
17. Y. Jin: *A comprehensive survery of fitness approximation in evolutionary computation.* Soft Computing, 9, 2005, 3–12.
18. Y. Jin, M. Hüsken, M. Olhofer, and B. Sendhoff: *Neural networks for fitness approximation in evolutionary optimization.* In Y. Jin, (ed.), Knowledge Incorporation in Evolutionary Computation, pp. 281–306. Springer Verlag, Berlin, 2005.
19. D.R. Jones: *A taxonomy of global optimization methods based on response surfaces.* Journal of Global Optimization, 21, 2001, 345–383.
20. D.R. Jones, M. Schonlau, and W.J. Welch: *Efficient global optimization of expensive black-box functions.* Journal of Global Optimization, 13, 1998, 455–492.
21. J.P.C. Kleijnen, W. van Baers, and I. van Nieuwenhuyse: *Constrained optimization in expensive simulation: Novel approach.* European Journal of Operational Research, 202, 2010, 164–174.
22. P. Kordík, J. Koutník, J. Drchal, O. Kovářík, M. Čepek, and M. Šnorek: *Meta-learning approach to neural network optimization.* Neural Networks, 23, 2010, 568–582.
23. S.J. Leary, A. Bhaskar, and A.J. Keane: *A derivative based surrogate model for approximating and optimizing the output of an expensive computer simulation.* Journal of Global Optimization, 30, 2004, 39–58.
24. S. Möhmel, N. Steinfeldt, S. Endgelschalt, M. Holeňa, S. Kolf, U. Dingerdissen, D. Wolf, R. Weber, and M. Bewersdorf: *New catalytic materials for the high-temperature synthesis of hydrocyanic acid from methane and ammonia by high-throughput approach.* Applied Catalysis A: General, 334, 2008, 73–83.
25. R.H. Myers, D.C. Montgomery, and C.M. Anderson-Cook: *Response surface methodology: proces and product optimization using designed experiment.* John Wiley and Sons, Hoboken, 2009.
26. Y.S. Ong, P.B. Nair, A.J. Keane, and K.W. Wong: *Surrogate-assisted evolutionary optimization frameworks for high-fidelity engineering design problems.* In Y. Jin, (ed.), Knowledge Incorporation in Evolutionary Computation, pp. 307–331. Springer Verlag, Berlin, 2005.
27. H. Papadopoulos, K. Poredrou, V. Vovk, and A. Gammerman. *Inductive confidence machies for regression.* In Proceedings of the 13th European Conference on Machie Learning, pp. 345–356, 2002.
28. A. Ratle: *Kriging as a surrogate fitness landscape in evolutionary optimization.* Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 15, 2001, 37–49.
29. T. Ray and W. Smith: *A surrogate assisted parallel multiobjective evolutionary algorithm for robust engineering design.* Engineering Optimization, 38, 2006, 997–1011.
30. S. Schaal and C.G. Atkeson: *Assessing the quality of learned local models.* In Advances in Neural Information Processing Systems 6, pp. 160–167. Morgan Kaufmann Publishers, San Mateo, 1994.
31. R. Schaefer: *Foundation of global genetic optimization.* Springer Verlag, Berlin, 2007.
32. H. Ulmer, F. Streichert, and A. Zell: *Model assisted evolution strategies.* In Y. Jin, (ed.), Knowledge Incorporation in Evolutionary Computation, pp. 333–355. Springer Verlag, Berlin, 2005.
33. E. Uusipaikka: *Confidence intervals in generalized regression models.* CRC Press, Boca Raton, 2009.
34. S. Valero, E. Argente, V. Botti, J.M. Serra, P. Serna, M. Moliner, and A. Corma: *DoE framework for catalyst development based on soft computing techniques.* Computers and Chemical Engineering, 33, 2009, 225–238.
35. J. Villemonteix, E. Vazquez, and E. Walter: *An informational approach to the global optimization of expensive-to-evaluate functions.* Journal of Global Optimization, 44, 2009, 509–534.
36. V. Vovk, A. Gammerman, and G. Shafer: *Algorithmic learning in a random world.* Springer Verlag, Berlin, 2005.
37. Z.Z. Zhou, Y.S. Ong, P.B. Nair, A.J. Keane, and K.Y. Lum: *Combining global and local surrogate models to accellerate evolutionary optimization.* IEEE Transactions on Systems, Man and Cybernetics. Part C: Applications and Reviews, 37, 2007, 66–76.