Author: Echtenbruck, M.M.
Title: Optimally weighted ensembles of surrogate models for sequential parameter optimization
Issue Date: 2020-07-02

Appendices

Appendix A

Supplementary Results

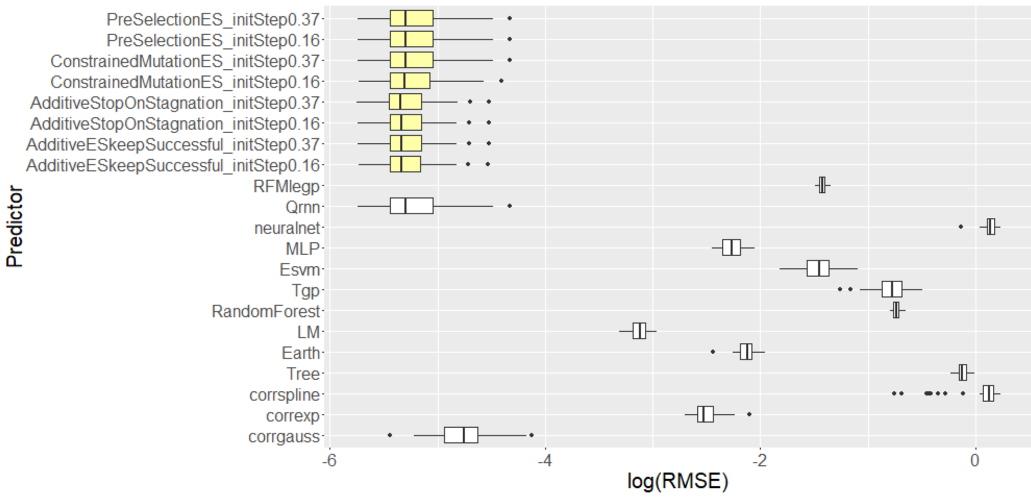
A large number of experiments was carried out for this thesis, and an even larger number of result plots was generated for the evaluation of these experiments. For the sake of readability, only the most relevant plots were shown in the main part of this thesis. In the following, supplementary plots are provided that also may be of interest since they provide further insights into the functionality and performance of the proposed ensemble method.

A.1 N-ary Ensembles on Higher Dimensional Physical Functions

In Section 4.5 experiments were carried out to compare the ensemble building method using different adaptations of the ES to each other and all base models. The results presented were condensed to the relevant information, each time showing only the results of the best performing base model, to preserve the readability. For the sake of completeness and to prove that no important information was omitted, in the following, the results for all base models are presented.

Figure A.1 shows the results for the experiments on the otl-circuit function (cf. Figure A.1a) and on the piston function (cf. Figure A.1b). On these functions, the ensemble adaptations and the best base model show comparable results and perform significantly better than the remaining base models, whose performances are heavily varying.

(a) Result on the otl-circuit function.



(b) Result on the piston function.

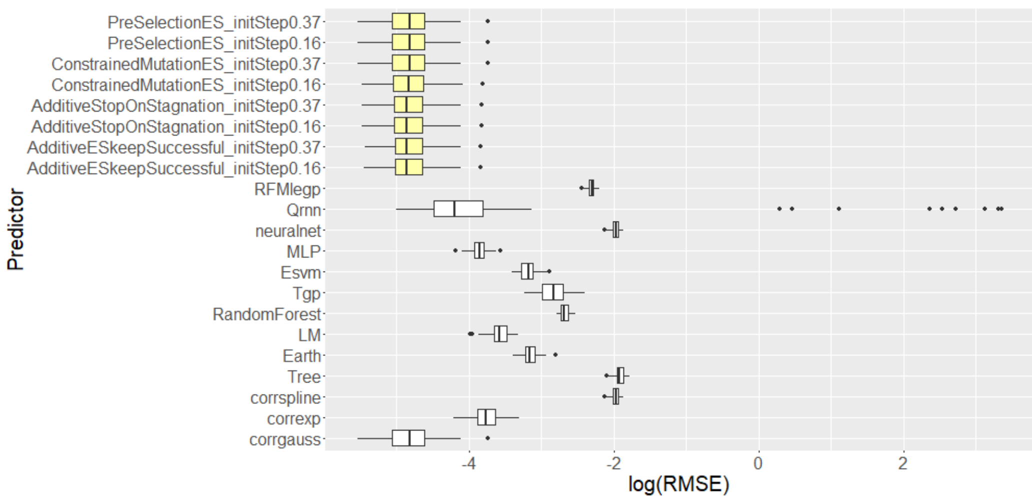
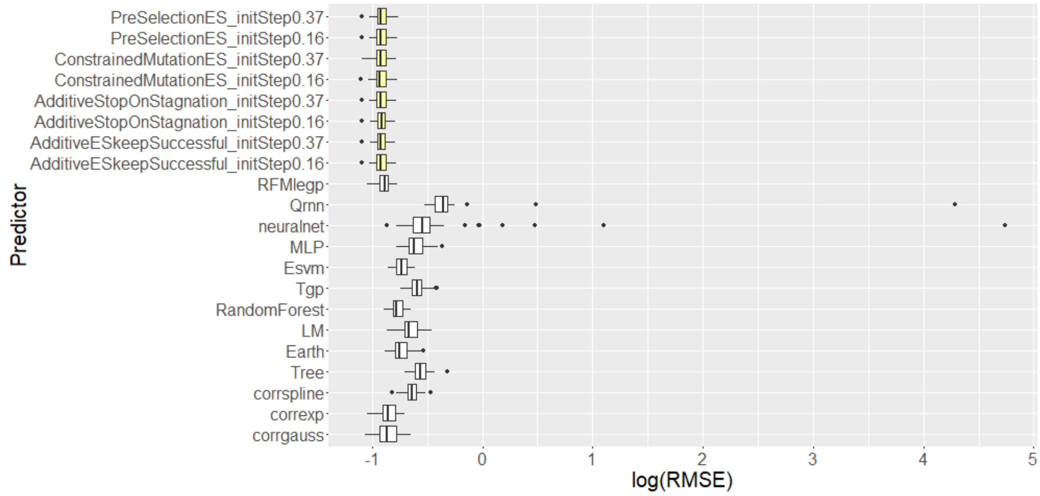


Figure A.1: The plot shows the results of the comparison of the performances, in terms of RMSE, of the different adaptations of the ensemble building method and all base models on the otl-circuit function and the piston function. Ensemble results are colored yellow, the base model result is shown in white. The ensemble adaptations can compete with the best base model and perform clearly better than all remaining base models.

(a) Result on the robot function. The ensemble adaptations can compete with the best base model. Correxp and Corrgauss also show comparable performances, only slightly weaker than the best base model and the ensemble adaptations. All remaining base models perform significantly worse.



(b) Result on the wing weight function. The ensemble adaptations belong to the better performing models.

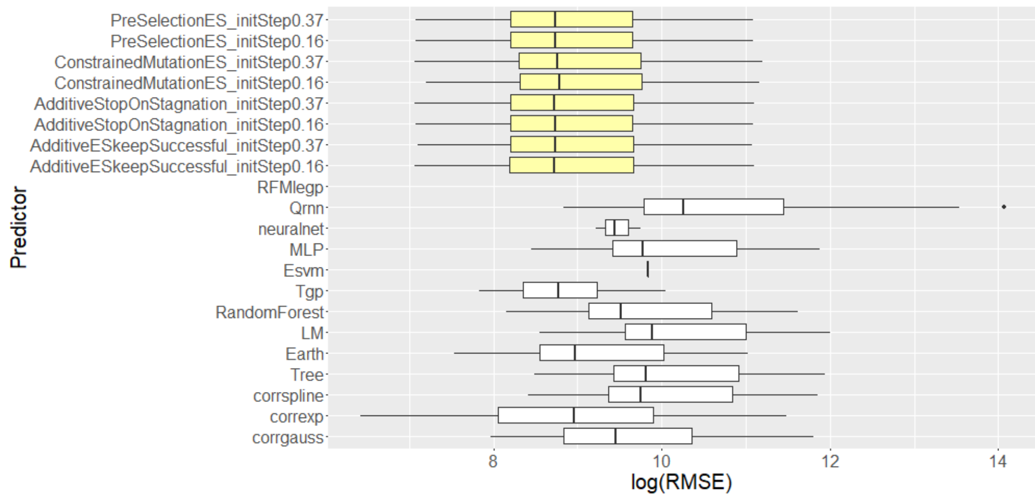


Figure A.2: The plot shows the results of the comparison of the performances, in terms of RMSE, of the different adaptations of the ensemble building method and all base models on the robot function and the wing weight function. Ensemble results are colored yellow, the base model result is shown in white.

Remarkable is that despite the large variance in the performances of the base models, the ensembles show a steady performance with a standard deviation comparable to or even slightly better than the one of the best base model.

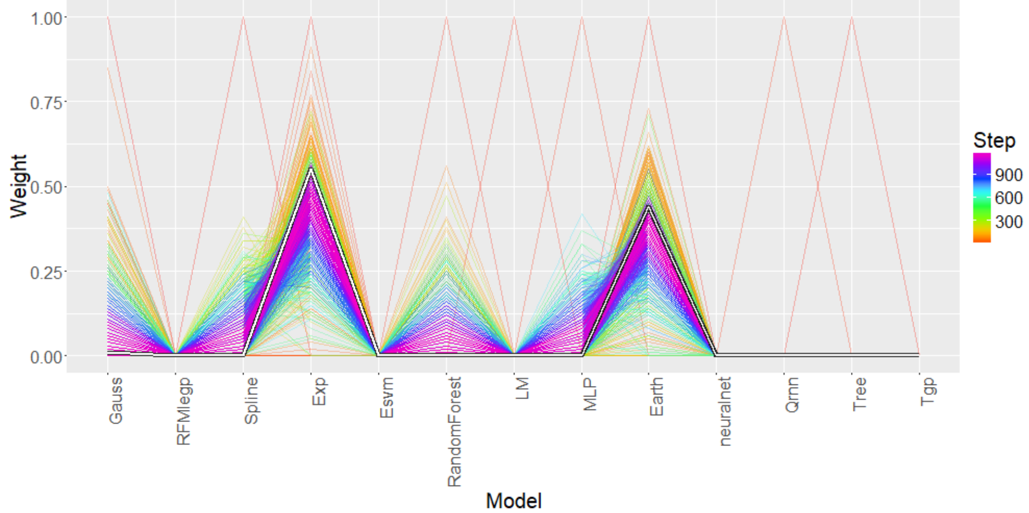
Figure A.2 shows the results for the experiments on the robot function (cf. Figure A.2a) and on the wing weight function (cf. Figure A.2b).
On the robot function, the ensembles and the best base model again show comparable performance. The base models ranked second and third, corrgauss and correxp, already perform slightly worse and all remaining base models perform significantly worse than the ensemble adaptations. Like on the otl-circuit function and the piston function this variance in the performances of the base models does not influence the good performance of the ensemble.
On the wing weight function, the results are more close; only one base model performs significantly worse than all ensemble adaptations. Still, the ensembles are among the best performing models.

In the following, a closer look is taken at the behavior of the different ensemble approaches. Exemplary plots are shown for the optimization of the weights to find the best model on the otl-circuit function (cf. Figures A.3 and A.4). The plots document the development of the weights for each step of the search. Each line represents a single individual considered during the search, and its color marks the related search step when the individual was found.

In all of these plots can be seen, that the base models RFMlegp, Esvm, neuralnet and Tgp did not receive any weight throughout the whole search. This is owed to the fact that these models failed during the preceding cross-evaluation in at least one, but rather most or even all of the evaluations. To obtain a reliable ensemble, models that failed in at least one of the evaluation steps are excluded from the search for the best weights.

Figure A.3 shows the development of the weights during the optimization using the additive approaches. The characteristics of these approaches can be easily read from the plots. The additive approach, which stops after the search stagnates (cf. Figure A.3a), ends the search after the addition of MLP. Previously added base models at least gained small weight during the search, but no improvement could be made with the addition of MLP. As a result, the search stopped.
The additive approach that does not stop on stagnation considered all base models. The course of the optimization can be read from the plot (cf. Figure A.3b). The three base models Gauss, Exp and Earth that were part of the search space from the start on, show nearly the complete range of colors.

(a) Sequential addition with stop on stagnation. After the search stagnates with the addition of MLP the search is stopped. Therefore, the base models LM, Qrnn and tree that are ranked lower than MLP are not considered in this search.



(b) Sequential addition without stop on stagnation. All base models are considered.

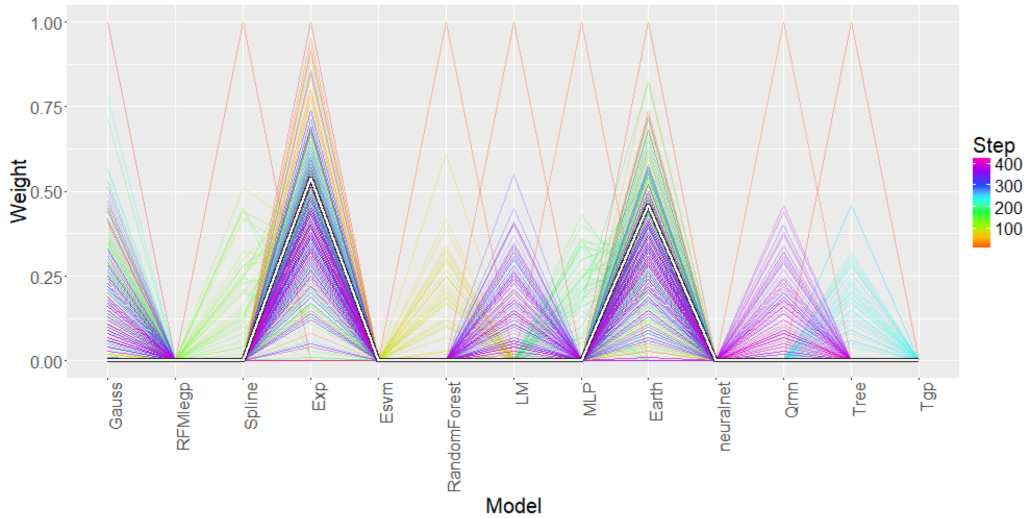
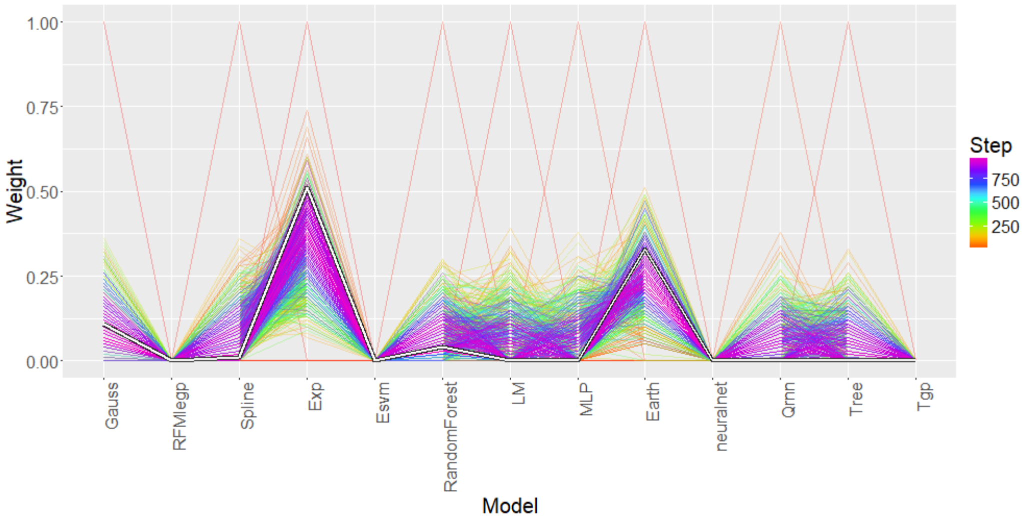


Figure A.3: The plots document the development of the weights during the optimization for the otl-circuit function using the additive approaches. Each line represents a single individual considered during the search, and its color marks the related search step when the individual was found, the white line marks the best weights combination. Though, through the course of the optimization both approaches consider different individuals, in the end, they agree on similar weight combinations.

(a) Restriction of the mutation.



(b) Preselection of models. During the automatized preselection of the base models, the search space is restricted to the three best-ranked base models.

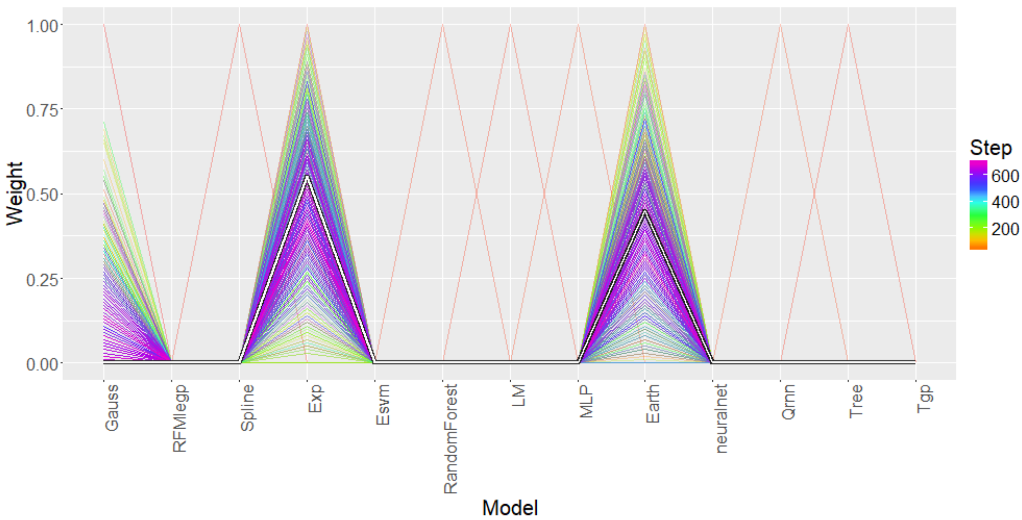


Figure A.4: The plots document the development of the weights during the optimization for the otl-circuit function using the approaches that restrict the search space and the mutation respectively. Each line represents a single individual considered during the search, and its color marks the related search step when the individual was found, the white line marks the best weights combination.

The base models that were added later in the course of the optimization show only a limited range of colors.
In the end, both approaches agree on similar best weights for the best ensemble combination.

Figure A.4 shows the development of the weights during the optimization using the approaches that restrict the considered base models or the mutation respectively. Again, the characteristics of these approaches can be easily read from the plots.
The approach that restricts the mutation searches the complete search space throughout the course of the optimization. No structure in the search can be read from the plot; the course of the search seems random. The final search result visibly differs from the weight combinations preferred by the additive approaches although most weight is given to the same base models.
The approach that initially selects a subset of the available base models for optimization using a rule based on the comparison to the mean predictor as described in Section 4.4.4. Applying this rule, the search is restricted to the base models Gauss, Exp, and Earth. The search on these base models again leads to similar weights as already preferred by the additive approaches.

A.2 The Performance of Dynamical Adapted CCM in SPO

In Section 5.2.5 the CCMs using $\tau = 1, \lambda = 10$ and $\tau = 20, \lambda = 20$ respectively were compared to two strong ensemble competitors and all base models. The results presented were condensed to the relevant information to preserve the readability. For the sake of completeness, in the following, the complete results for all base models are given and discussed.

Table A.1 presents these results. Given are the mean and standard deviation for the results of the optimization processes, best results are marked bold. To allow for a comparison of the models over the different functions, the mean results are function-wise ranked, and the average ranking, as well as the final rank for each model, is given in the two last rows.
Some of the entries show a red 'N/A'. These entries refer to experimental setups where the optimization process using the stated model did not succeed in at least one repetition. Therefore, the related rankings in the two last rows are given in brackets.

	$\tau = 1, \lambda = 10$	correxp	corrgauss	corrspline	Earth	Lm
ackley2D	2.4957 (±1.931)	7.6620 (±1.465)	2.1552 (±3.055)	3.2729 (±3.189)	5.8966 (±1.853)	3.1309 (±1.741)
ackley4D	6.8503 (±3.179)	11.588 (±1.275)	5.3801 (±3.576)	7.8778 (±4.558)	8.1570 (±1.870)	**5.2055** (±1.074)
GLG4D	22.430 (±12.00)	33.021 (±5.766)	**21.101** (±13.17)	26.344 (±10.49)	29.465 (±10.19)	32.479 (±9.483)
GLG8D	29.616 (±9.420)	57.321 (±7.660)	**28.702** (±11.39)	42.634 (±20.82)	60.226 (±4.488)	56.429 (±6.917)
otl-circuit	2.6055 (±0.002)	2.6542 (±0.034)	2.6553 (±0.070)	2.7007 (±0.070)	**2.6039** (±0.000)	N/A
piston	**0.1667** (±0.003)	0.1708 (±0.002)	0.1709 (±0.005)	0.1750 (±0.011)	0.1670 (±0.001)	0.1761 (±0.011)
robot	0.0132 (±0.020)	0.0552 (±0.027)	0.0309 (±0.015)	0.0756 (±0.044)	0.0253 (±0.022)	0.0169 (±0.036)
rosenbrock4D	**2.4138** (±1.522)	288.78 (±300.4)	3.5536 (±2.490)	60.989 (±138.8)	727.23 (±610.5)	70.933 (±63.79)
rosenbrock8D	4127.1 (±2989)	3188.6 (±1474)	6819.3 (±2799)	18887 (±12949)	23688 (±12099)	**696.83** (±582.0)
wingweight	182.50 (±13.11)	174.83 (±14.12)	178.91 (±14.44)	181.05 (±6.350)	175.94 (±6.323)	185.10 (±12.82)
AVG RANK	3.6	7.2	4.2	6.95	7.05	(5.7)
RANK	2	11	3	7	10	(5)

	$\tau = 20, \lambda = 20$	MLP	Neuralnet	RandomForest	Tgp	tree
ackley2D	1.9344 (±2.041)	**0.3425** (±0.694)	7.8315 (±2.263)	2.5598 (±1.198)	1.1200 (±0.522)	5.6125 (±1.443)
ackley4D	6.0215 (±3.242)	9.7943 (±3.098)	15.330 (±2.523)	8.9207 (±1.946)	5.8991 (±1.028)	15.540 (±0.936)
GLG4D	23.576 (±13.02)	25.799 (±10.63)	33.676 (±9.031)	30.801 (±7.160)	21.690 (±12.28)	32.181 (±8.249)
GLG8D	30.335 (±7.166)	48.353 (±7.637)	56.281 (±1.467)	53.407 (±6.857)	59.229 (±6.998)	58.156 (±6.744)
otl-circuit	2.6103 (±0.012)	**2.6039** (±0.000)	2.8980 (±0.191)	2.7849 (±0.116)	2.7580 (±0.045)	3.0001 (±0.092)
piston	0.1704 (±0.008)	0.1669 (±0.001)	0.2107 (±0.014)	0.1706 (±0.003)	0.1770 (±0.002)	0.2027 (±0.016)
robot	0.0181 (±0.018)	**0.0000** (±0.000)	**0.000** (±0.000)	0.0239 (±0.033)	0.06681 (±0.032)	0.0876 (±0.056)
rosenbrock4D	4.1977 (±2.088)	815.76 (±560.3)	N/A	253.76 (±212.2)	306.48 (±222.6)	732.86 (±689.9)
rosenbrock8D	3224.2 (±2041)	17994 (±12828)	N/A	4705.0 (±2657)	6633.4 (±3569)	23605 (±15191)
wingweight	**173.0** (±10.35)	181.99 (±11.02)	N/A	184.59 (±12.15)	182.94 (±9.371)	181.05 (±12.88)
AVG RANK	**3.4**	5.2	(6.65)	7	7	10
RANK	**1**	4	(6)	8.5	8.5	12

Table A.1: The table displays the results for the main experiment setup. Given are mean optimization result with standard deviation for the CCMs in comparison to all base models.

The most important result here is that the two CCMs are ranked first and second, unlike before, where the CCMs were ranked second and third only. This is owed to the fact, that in the previous representation for each function only the best performing base model was considered, and this best base model, though changing for each function, ranked as one in the overall ranking.
Noteworthy are also the results for the rosenbrock functions. The optimization results obtained from the different base models and the two ensembles range between 2.4 and 816 for rosenbrock4D and between 697 and 23688 for rosenbrock 8D. Though the variance in the performances of the different base models is large, the ensembles showed good performances on both functions; the CCM using $\tau = 1, \lambda = 10$ even performed best on the rosenbrock4D function.

Bibliography

[1] Zaefferer, M., Breiderhoff, B., Naujoks, B., Friese, M., Stork, J., Fischbach, A., Flasch, O., Bartz-Beielstein, T.: Tuning multi-objective optimization algorithms for cyclone dust separators. In: Proceedings of the 2014 Annual Conference on Genetic and Evolutionary Computation. GECCO '14, New York, NY, USA, ACM (2014) 1223–1230

[2] Löffler, F.: Staubabscheiden. (Aus der Lehrbuchreihe Chemieingenieurwesen/Verfahrenstechnik). Georg Thieme Verlag Stuttgart, New York (1989)

[3] Slack, M., Prasad, R., Bakker, A., Boysan, F.: Advances in cyclone modelling using unstructured grids. Chemical Engineering Research and Design **78** (2000) 1098 – 1104 Separation Processes.

[4] Elsayed, K., Lacor, C.: Optimization of the cyclone separator geometry for minimum pressure drop using mathematical models and CFD simulations. Chemical Engineering Science **65** (2010) 6048 – 6058

[5] Wasilewski, M., Duda, J.: Application of Computational Fluid Dynamics to optimization of cyclone dust separators operated in the cement industry. Volume Czasopismo Chemik 67/2013. (2013) 985–994

[6] Iman Pishbin, S., Moghiman, M.: Optimization of cyclone separators using genetic algorithm. International Review of Chemical Engineering (I.RE.CH.E.) **2** (2010)

[7] Singh, P., Couckuyt, I., Elsayed, K., Deschrijver, D., Dhaene, T.: Multi-objective geometry optimization of a gas cyclone using triple-fidelity co-kriging surrogate models. Journal of Optimization Theory and Applications **175** (2017) 172–193

[8] Allmendinger, R., Emmerich, M.T., Hakanen, J., Jin, Y., Rigoni, E.: Surrogate-assisted multicriteria optimization: Complexities, prospective so-

lutions, and business case. Journal of Multi-Criteria Decision Analysis **24** (2017) 5–24

[9] A. Arias-Montaño and C. A. Coello Coello and E. Mezura-Montes: Multi-objective airfoil shape optimization using a multiple-surrogate approach. In: 2012 IEEE Congress on Evolutionary Computation. (2012) 1–8

[10] Skinner, S., Zare-Behtash, H.: State-of-the-art in aerodynamic shape optimisation methods. Applied Soft Computing **62** (2018) 933 – 962

[11] Muyl, F., Dumas, L., Herbert, V.: Hybrid method for aerodynamic shape optimization in automotive industry. Computers & Fluids **33** (2004) 849 – 858 Applied Mathematics for Industrial Flow Problems.

[12] Jones, D.R.: Optimization in the Automotive Industry. In: Optimization and Industry: New Frontiers. Springer US, Boston, MA (2003) 39–58

[13] Jameson, A., Fatica, M.: Using computational fluid dynamics for aerodynamics. (2003) White paper, National Research Council Workshop on the Future of Supercomputing, Santa Fe, NM, 24-25 September 2003.

[14] Bartz-Beielstein, T., Lasarczyk, C., Preuss, M.: Sequential Parameter Optimization. In McKay, B., et al., eds.: Proceedings 2005 Congress on Evolutionary Computation (CEC'05), Edinburgh, Scotland, Piscataway NJ, IEEE Press (2005) 773–780

[15] Hutter, F., Hoos, H.H., Leyton-Brown, K.: Sequential model-based optimization for general algorithm configuration. In: International Conference on Learning and Intelligent Optimization, Springer (2011) 507–523

[16] Žilinskas, A., Mockus, J.: On one bayesian method of search of the minimum. Avtomatika i Vychislitel'naya Teknika **4** (1972) 42–44

[17] Jones, D.R., Schonlau, M., Welch, W.J.: Efficient global optimization of expensive black-box functions. Journal of Global optimization **13** (1998) 455–492

[18] Kenneth P. Burnham, D.R.A.: Model Selection and Multimodel Inference - A Practical Information-Theoretic Approach. Springer-Verlag New York (2002)

[19] Bartz-Beielstein, T., Zaefferer, M.: Model-based Methods for Continuous and Discrete Global Optimization. Applied Soft Computing **55** (2017) 154–167

[20] Li, R., Emmerich, M.T.M., Eggermont, J., Bovenkamp, E.G.P., Bäck, T., Dijkstra, J., Reiber, J.H.C.: Metamodel-assisted mixed integer evolution strategies and their application to intravascular ultrasound image analysis. In: 2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence). (2008) 2764–2771

[21] Bartz-Beielstein, T., Stork, J., Zaefferer, M., Rebolledo, M., Lasarczyk, C., Ziegenhirt, J., Konen, W., Flasch, O., Koch, P., Friese, M., Gentile, L., Rehbach, F.: Sequential Parameter Optimization Toolbox (Package "SPOT"). (2018) R package available at `https://CRAN.R-project.org/package=SPOT`, version 1.1.0.

[22] James, G., Witten, D., Hastie, T., Tibshirani, R.: An Introduction to Statistical Learning: With Applications in R. Springer Publishing Company, Incorporated (2014)

[23] Rencher, A.C., Schaalje, G.B.: Linear Models in Statistics. 2nd edn. Wiley-Interscience. John Wiley & Sons, Hoboken, New Jersey (2008)

[24] Lenth, R.: Response-Surface Methods in R, Using rsm. Journal of Statistical Software **32** (2009) 1–17

[25] Breiman, L., Friedman, J., Stone, C.J., Olshen, R.: Classification and Regression Trees. The Wadsworth and Brooks-Cole statistics-probability series. Taylor & Francis (1984)

[26] Breiman, L.: Random forests. Machine learning **45** (2001) 5–32

[27] Therneau, T.M., Atkinson, B., Ripley, B.: rpart: Recursive Partitioning. (2011)

[28] Liaw, A., Wiener, M.: Classification and regression by randomforest. R News **2** (2002) 18–22

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

[30] Zell, A.: Simulation neuronaler Netze. Oldenbourg (2000)

[31] Anderson, J.A., Rosenfeld, E.: Neurocomputing: Foundations of Research. MIT Press, Cambridge, MA, USA (1988)

[32] Ramcón y Cajal, S.: Histology of the Nervous System of Man and Vertebrates. Volume 1. Oxford University Press, 200 Madison Avenue, New York, New York 10016 (1995)

[33] Kandel, E.R., Schwartz, J.H., Jessell, T.M.: Principles of Neural Science. 3rd edn. Elsevier, New York (1991)

[34] Fritsch, S., Guenther, F., Suling, M., Mueller, S.: Training of Neural Networks (Package "neuralnet"). (2016) R package available at `https://CRAN.R-project.org/package=neuralnet`, version 1.32.

[35] Cannon, A.J.: Multi-Layer Perceptron Neural Network with Optional Monotonicity Constraints (Package "monmlp"). (2017) R package available at `https://CRAN.R-project.org/package=monmlp`, version 1.1.3.

[36] Cannon, A.J.: Quantile regression neural networks: implementation in R and application to precipitation downscaling. Computers & Geosciences **37** (2011) 1277–1284

[37] Boser, B.E., Guyon, I.M., Vapnik, V.N.: A training algorithm for optimal margin classifiers. In: Proceedings of the Fifth Annual Workshop on Computational Learning Theory. COLT '92, New York, NY, USA, ACM (1992) 144–152

[38] Vapnik, V.: Statistical learning theory. Wiley (1998)

[39] Smola, A.J., Schölkopf, B.: A tutorial on support vector regression. Statistics and Computing **14** (2004) 199–222

[40] Forrester, A., Sobester, A., Keane, A.: Engineering Design Via Surrogate Modelling: A Practical Guide. John Wiley & Sons (2008)

[41] Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C., Lin, C.: Misc Functions of the Department of Statistics, Probability Theory Group (Package "e1071"). (2018) Available at `https://CRAN.R-project.org/package=e1071`, version 1.6.7.

[42] Chang, C.C., Lin, C.J.: LIBSVM: A Library for Support Vector Machines. ACM Intelligent Systems and Technology **2** (2011) 27:1–27:27

[43] Friedman, J.H., Roosen, C.B.: An introduction to multivariate adaptive regression splines. Statistical Methods in Medical Research **4** (1995) 197–217

[44] Friedman, J.H.: Multivariate Adaptive Regression Splines. The Annals of Statistics **19** (1991) 1–67

[45] Friedman, J.H.: Fast MARS. Technical Report 110, Stanford University Department of Statistics (1993)

[46] S. Milborrow. Derived from mda:mars by T. Hastie and R. Tibshirani.: earth: Multivariate Adaptive Regression Splines (Package "Earth"). (2011) R package available at `http://CRAN.R-project.org/package=earth`, version 4.4.4.

[47] Rasmussen, C.E., Williams, C.K.I.: Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning). The MIT Press (2005)

[48] Krige, D.: A statistical approach to some basic mine valuation problems on the witwatersrand. Journal of the Southern African Institute of Mining and Metallurgy **52** (1951) 119–139

[49] Matheron, G.: Principles of geostatistics. Economic Geology **58** (1963) 1246–1266

[50] Sacks, J., Welch, W.J., Mitchell, T.J., Wynn, H.P.: Design and analysis of computer experiments. Statistical science (1989) 409–423

[51] Hald, A.: On the history of maximum likelihood in relation to inverse probability and least squares. Statist. Sci. **14** (1999) 214–222

[52] Søren N. Lophaven, Hans Bruun Nielsen, J.S.: Dace - a matlab kriging toolbox. Technical report, Technical University of Denmark (2002)

[53] Gramacy, R.B.: tgp: An R package for bayesian nonstationary, semiparametric nonlinear regression and design by treed gaussian process models. Journal of Statistical Software **19** (2007) 1–46

[54] Dancik, G.M., Dorman, K.S.: mlegp: statistical analysis for computer models of biological systems using r. Bioinformatics **24** (2008) 1966

[55] Gorissen, D., Couckuyt, I., Demeester, P., Dhaene, T., Crombecq, K.: A surrogate modeling and adaptive sampling toolbox for computer based design. Journal of Machine Learning Research **11** (2010) 2051–2055

[56] Manuel López-Ibáñez and Jérémie Dubois-Lacoste and Leslie Pérez Cáceres and Mauro Birattari and Thomas Stützle: The irace package: Iterated racing for automatic algorithm configuration. Operations Research Perspectives **3** (2016) 43 – 58

[57] Emmerich, M., Giotis, A., Özdemir, M., Bäck, T., Giannakoglou, K.: Metamodel—assisted evolution strategies. In Guervós, J.J.M., Adamidis, P., Beyer, H.G., Schwefel, H.P., Fernández-Villacañas, J.L., eds.: Parallel Problem Solving from Nature — PPSN VII, Berlin, Heidelberg, Springer Berlin Heidelberg (2002) 361–370

[58] Mockus, J., Tiesis, V., Zilinskas, A.: The application of bayesian methods for seeking the extremum. Towards Global Optimization **2** (1978) 117–129

[59] Bartz-Beielstein, T.: Spot: An r package for automatic and interactive tuning of optimization algorithms by sequential parameter optimization. (2010)

[60] Stoean, C., Preuss, M., Stoean, R., Dumitrescu, D.: Multimodal optimization by means of a topological species conservation algorithm. Trans. Evol. Comp **14** (2010) 842–864

[61] Wong, K.C., Leung, K.S., Wong, M.H.: Protein structure prediction on a lattice model via multimodal optimization techniques. In: Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation. GECCO '10, New York, NY, USA, ACM (2010) 155–162

[62] Qing, L., Gang, W., Zaiyue, Y., Qiuping, W.: Crowding clustering genetic algorithm for multimodal function optimization. Appl. Soft Comput. **8** (2008) 88–95

[63] Gallagher, M., Yuan, B.: A general-purpose tunable landscape generator. IEEE Trans. Evolutionary Computation **10** (2006) 590–603

[64] Bartz-Beielstein, T.: How to Create Generalizable Results. In Kacprzyk, J., Pedrycz, W., eds.: Springer Handbook of Computational Intelligence. Springer Berlin Heidelberg, Berlin, Heidelberg (2015) 1127–1142

[65] Ackley, D.H.: An Empirical Study of Bit Vector Function Optimization. Genetic Algorithms and Simulated Annealing (1987) 170–215

[66] Rosenbrock, H.H.: An automatic method for finding the greatest or least value of a function. The Computer Journal **3** (1960) 175–184

[67] Bäck, T., Schwefel, H.P.: An overview of evolutionary algorithms for parameter optimization. Evol. Comput. **1** (1993) 1–23

[68] Surjanoviv, S., Bingham, D.: Virtual library of simulation experiments. [ONLINE] (2016) Available at: http://www.sfu.ca/ ssurjano/. [Accessed 30 November 2016].

[69] Ben-Ari, E.N., Steinberg, D.M.: Modeling Data from Computer Experiments: An Empirical Comparison of Kriging with MARS and Projection Pursuit Regression. Quality Engineering **19** (2007) 327–338

[70] Kenett, R., Zacks, S.: Modern Industrial Statistics: Design and Control of Quality and Reliability. Volume 41. Cengage Learning (1998)

[71] An, J., Owen, A.: Quasi-regression. Journal of Complexity **17** (2001) 588 – 607

[72] Raymer, D.P.: Aircraft Design: A Conceptual Approach and Rds-student, Software for Aircraft Design, Sizing, and Performance Set (AIAA Education). AIAA (American Institute of Aeronautics & Ast (2006)

[73] Friese, M., Bartz-Beielstein, T., Emmerich, M.: Building ensembles of surrogates by optimal convex combination. In Papa, G., Mernik, M., eds.: Bioinspired Optimization Methods and their Applications. (2016) 131–143

[74] Friese, M., Bartz-Beielstein, T., Bäck, T., Naujoks, B., Emmerich, M.: Weighted ensembles in model-based global optimization. AIP Conference Proceedings **2070** (2019) 020003

[75] Friese, M., Bartz-Beielstein, T., Bäck, T., Emmerich, M.: Optimally Weighted Ensembles of Surrogate Models for Sequential Parameter Optimization. Journal of Global Optimization (submitted to) (2019)

[76] Friese, M., Zaefferer, M., Bartz-Beielstein, T., Flasch, O., Koch, P., Konen, W., Naujoks, B.: Ensemble-Based Optimization and Tuning Algorithms. In Hoffmann, F., Hüllermeier, E., eds.: Proceedings 21. Workshop Computational Intelligence, Universitätsverlag Karlsruhe (2011) 119–134

[77] Gittins, J.C.: Bandit Processes and Dynamic Allocation Indices. Journal of the Royal Statistical Society. Series B (Methodological) **41** (1979) 148–177

[78] Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction (Adaptive Computation and Machine Learning). MIT Press (1998)

[79] Hawking, S.: On the Shoulders of Giants: The Great Works of Physics and Astronomy. Penguin (2003)

[80] Sober, E.: The principle of parsimony. British Journal for the Philosophy of Science **32** (1981)

[81] Schapire, R.E.: The strength of weak learnability. Machine Learning **5** (1990) 197–227

[82] Kittler, J., Hatef, M., Duin, R.P.W., Matas, J.: On combining classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence **20** (1998) 226–239

[83] Symonds, M.R.E., Moussalli, A.: A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. Behavioral Ecology and Sociobiology **65** (2011) 13–21

[84] Akaike, H.: 15. In: Information Theory and an Extension of the Maximum Likelihood Principle. Springer New York, New York, NY (1998) 199–213

[85] Kittler, J., Hojjatoleslami, S., Windeatt, T.: Weighting factors in multiple expert fusion. In: BMVC. (1997)

[86] Breiman, L.: Stacked regressions. Machine learning **24** (1996) 49–64

[87] Polikar, R.: Ensemble based systems in decision making. Circuits and Systems Magazine, IEEE **6** (2006) 21–45

[88] Schein, A.I., Popescul, A., Ungar, L.H., Pennock, D.M.: Methods and metrics for cold-start recommendations. In: Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '02, New York, NY, USA, ACM (2002) 253–260

[89] Hess, S.: Sequentielle modellbasierte Optimierung mit Ensembles durch einen Reinforcement-Learning-Ansatz. Master's thesis, Technische Universität Dortmund (2012)

[90] Ghosh, J., Samanta, T.: Model selection–an overview. Current Science (2001) 1135–1144

[91] Baxt, W.G.: Improving the accuracy of an artificial neural network using multiple differently trained networks. Neural Computation **4** (1992) 772–780

[92] Jakob, W., Gorges-Schleuter, M., Sieber, I., Süss, W., Eggert, H.: Solving a highly multimodal design optimization problem using the extended genetic algorithm gleam. In: 6th Internat.Conf.Computer Aided Optimum Design of Structures (OPTI 99), Orlando, Fla., March 16-18, 1999. (1999) 41.03.01; LK 01.

[93] Claeskens, G., Hjort, N.L.: Model Selection and Model Averaging. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press (2008)

[94] Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences **55** (1997) 119 – 139

[95] Bishop, C.M.: Neural Networks for Pattern Recognition. Oxford University Press, Inc., New York, NY, USA (1995)

[96] van Stein, B., Wang, H., Kowalczyk, W., Emmerich, M., Bäck, T.: Cluster-based kriging approximation algorithms for complexity reduction (2017)

[97] van Stein, B., Wang, H., Kowalczyk, W., Bäck, T., Emmerich, M.: Optimally weighted cluster kriging for big data regression. In: International Symposium on Intelligent Data Analysis, Springer (2015) 310–321

[98] Zerpa, L.E., Queipo, N.V., Pintos, S., Salager, J.L.: An optimization methodology of alkaline surfactant-polymer-flooding processes using field scale numerical simulation and multiple surrogates. Journal of Petroleum Science and Engineering **47** (2005) 197 – 208

[99] Perrone, M., Cooper, L.: When networks disagree: Ensemble methods for hybrid neural networks. Neural networks for speech and image processing (1993)

[100] Goel, T., Haftka, R.T., Shyy, W., Queipo, N.V.: Ensemble of surrogates. Structural and Multidisciplinary Optimization **33** (2007) 199–216

[101] Acar, E., Rais-Rohani, M.: Ensemble of metamodels with optimized weight factors. Structural and Multidisciplinary Optimization **37** (2009) 279–294

[102] Torgo, L.: Functional models for regression tree leaves. In: Proceedings of the Fourteenth International Conference on Machine Learning. ICML '97, San Francisco, CA, USA, Morgan Kaufmann Publishers Inc. (1997) 385–393

[103] Dasarathy, B.V., Sheela, B.V.: A composite classifier system design: Concepts and methodology. Proceedings of the IEEE **67** (1979) 708–713

[104] van Stein, B., Wang, H., Kowalczyk, W., Bäck, T., Emmerich, M.: Optimally weighted cluster kriging for big data regression. In Fromont, E., De Bie, T., van Leeuwen, M., eds.: Advances in Intelligent Data Analysis XIV, Cham, Springer International Publishing (2015) 310–321

[105] Ginsbourger, D., Helbert, C., Carrao, L.: Discrete mixtures of kernels for kriging-based optimization. Qual. Reliab. Eng. Int. **24** (2008) 681–691 Special Issue: The European Network for Business and Industrial Statistics (ENBIS).

[106] Friese, M., Bartz-Beielstein, T., Vladislavleva, K., Flasch, O., Mersmann,

O., Naujoks, B., Stork, J., Zaefferer, M.: Ensemble-based model selection for smart metering data (abstract). Technical report, CIplus (2012)

[107] Flasch, O., Friese, M., Zaefferer, M., Bartz-Beielstein, T., Branke, J.: Learning model-ensemble policies with genetic programming. Technical Report 3/2015, TH Köln (2015)

[108] Beyer, H.G., Schwefel, H.P.: Evolution Strategies: A Comprehensive Introduction. Natural Computing **1** (2002) 3–52

[109] Schwefel, H.P.: Evolution and optimum seeking. Sixth-generation computer technology series (1995)

[110] Bäck, T., Foussette, C., Krause, P.: Contemporary evolution strategies. Springer (2013)

[111] Kennedy, J. In: Particle Swarm Optimization. Springer US, Boston, MA (2010) 760–766

[112] Eggensperger, K., Feurer, M., Hutter, F., Bergstra, J., Snoek, J., Hoos, H., Leyton-Brown, K.: Towards an empirical foundation for assessing bayesian optimization of hyperparameters. In: NIPS workshop on Bayesian Optimization in Theory and Practice. (2013)

[113] Mostaghim, S., Pfeiffer, F., Schmeck, H.: Self-organized invasive parallel optimization. In: Proceedings of the 3rd Workshop on Biologically Inspired Algorithms for Distributed Systems. BADS '11, New York, NY, USA, ACM (2011) 49–56

[114] Chugh, T., Sindhya, K., Hakanen, J., Miettinen, K.: A survey on handling computationally expensive multiobjective optimization problems with evolutionary algorithms. Soft Computing (2017)

[115] Hussein, R., Roy, P.C., Deb, K.: Adaptive Switching Strategy for Metamodeling Based Multi-objective Optimization: Part I,Generative Frameworks. Technical Report COIN Report No. 2019001, Michigan State University (2019)

[116] Hussein, R., Roy, P.C., Deb, K.: Adaptive Switching Strategy for Metamodeling Based Multi-objective Optimization: Part II, Simultaneous and Combined Frameworks. Technical Report COIN Report No. 2019002, Michigan State University (2019)

[117] Chugh, T., Jin, Y., Miettinen, K., Hakanen, J., Sindhya, K.: A surrogate-assisted reference vector guided evolutionary algorithm for computationally

expensive many-objective optimization. IEEE Transactions on Evolutionary Computation **22** (2018) 129–142

[118] Stork, J., Friese, M., Zaefferer, M., Bartz-Beielstein, T., Fischbach, A., Breiderhoff, B., Naujoks, B., Tušar, T.: Open Issues in Surrogate-Assisted Optimization. In: High-Performance Simulation-Based Optimization. Springer (2020) 225 – 244

Glossary

Abbreviations

AIC	Akaike's Information Criterion	37
CCM	Convex Combinations of Models	52
CFD	Computational Fluid Dynamics	3
DOE	Design of Experiment	25
ES	Evolutionary Strategy	64
GMSE	Generalized Mean Squared Error	45
MARS	Multivariate Adaptive Regression Splines	19
MSE	Mean Squared Error	45
RMSE	Root Mean Squared Error	45
SBO	Surrogate Based Optimization	6
SPO	Sequential Parameter Optimization	4
SPOT	Sequential Parameter Optimization Toolbox	4
SVM	Support Vector Machine	17
wRMSE	Weighted Root Mean Squared Error	90

Variables and Identifiers

$\alpha,\ (1-\alpha)$	The Weights Used for the Binary Model Definition	52

α, β, γ	The Weights for the Convex Combination of Models	61
$\hat{y}$	The Predicted Function Value	9
$\mathbf{av}$	A vector of flags, indicating active and suspended models	93
σ	The Step Width of the (1+1)-ES	64
σ_{init}	The Initial Step Width of the (1+1)-ES	64
σ_{min}	Minimum step size of the (1+1)-ES	99
τ	Ensemble Building Interval	88
C	Candidate set for evaluation on the objective function in each sequential step, , with respect to exploration or exploitation respectively (cf. Chapter 2.2)	27
D	Observed Data	25
d	The Dimension	61
M	The Surrogate Model	26
M^*	The Fitted Surrogate Model	26
P	The population of the (1+1)-ES	93
y	The Observed Function Value	9
ρ	Density of the Immediate Neighborhood of a Point	90
$n_{\mathbf{eval}}$	Function Evaluations per Sequential Step	26