Investigation of the Potential of Using Surrogate Models in the Gear Design Process

Jens Brimmers, Marius Willecke, Christoph Löpenhaus and Christian Brecher

This paper was first presented at the International VDI Conference on Gears 2019, 3rd International Conference on High-Performance Plastic Gears 2019, 3rd International Conference on Gear Production 2019, Garching/Munich (VDI-Berichte 2355, 2019, VDI Verlag GmbH, Page 727–740).

State of the Art

Surrogate models, also known as response surface models or metamodels, are approximation models, which are based on mathematical functions (Ref.1). In engineering, surrogate models are used to correlate the input and output variables of experiments and simulations (Refs. 2–10). This is especially true for very time-consuming, costly or high number of experiments/simulations. In this case, the surrogate model can be evaluated much faster in comparison to the experiment or complex simulation. This is most important for design space exploration or optimization where a high number of experiments of simulations is necessary (Ref. 5). In order to reduce the time effort, the extensive simulation is only performed for a reduced number of parameter sets. These initial parameter sets are defined by means of methods of design of experiment (DOE), e.g., full-factorial sampling or latin hypercube sampling (Ref. 11). For computational problems a latin hypercube sampling or the Monte-Carlo approach (random sampling) is often used to identify the initial parameter sets. Once the initial parameter sets are identified, the simulation is performed at these given points. The results of the simulation are used to fit a surrogate model to the given input variables in order to approximate the system behavior of the engineering system. Possible approximation types for surrogate models are shown in Figure 1. The most common modeling types are models based on radial basis functions (RBF), kriging models, also known as Gaussian process models, and models based on multivariate adaptive regression splines (MARS).

RBFs are functions whose value only depends on the Euclidian distance from the origin (Ref. 12). An approximation model consists of a number of different radial basis functions, which are weighted accordingly. The weights of each basis functions are tuned in order to improve the quality of the approximation for the given number of data points. In the example in Figure 1 the function $f(x) = 1 + \sin(x^2)$ was evaluated at six test data points and approximated by the usage of an RBF surrogate model consisting of Gaussian basis functions, as a type of RBF. The approximation follows the trend of the sine function but is not able to predict any of the test data points in high accordance.

Kriging or Gaussian process models originate from geosciences and are usually used to predict the location of certain commodities like oil or gold for which only a finite number of boreholes exists (Ref. 13). The Gaussian process consists of two parts, one global and one local part. The global part can be

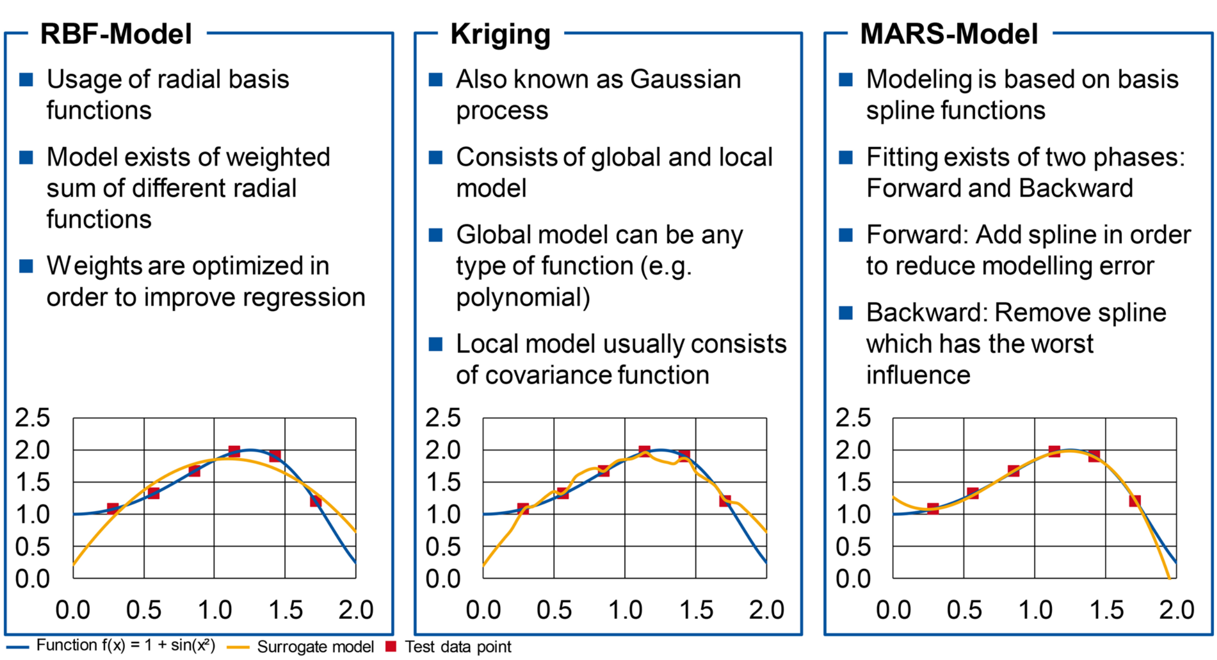

Figure 1 Overview of different surrogate modeling types.

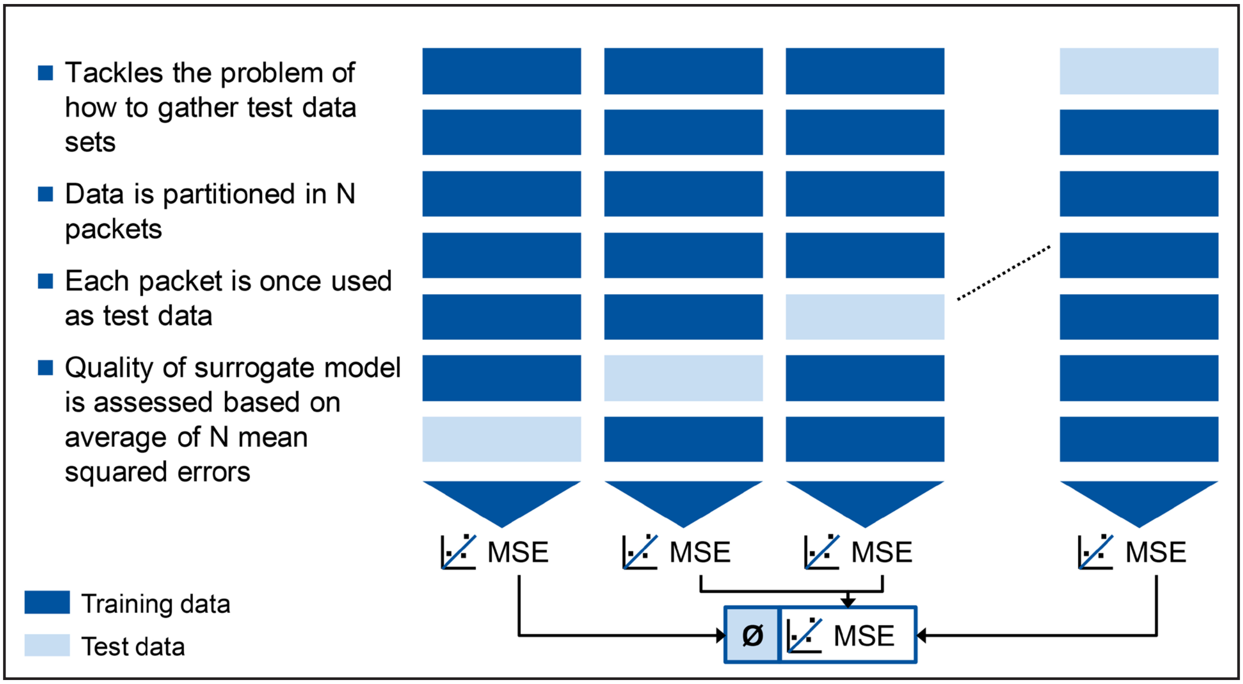

Figure 2 Cross-validation as a method to assess the quality of the surrogate model.

any type of function (e.g., exponential or polynomial), whereas the local part consists of a stochastical process, usually a covariance function. For the example of the sine function (Fig. 1), the Gaussian process model (quadratic global model) follows the trend of the sine function and matches each given test data point (Refs. 14–15).

The third type are multivariate adaptive regression splines (MARS) (Ref. 16). A MARS model consists of different basis splines, e.g., linear function, etc. The fitting process consists of two parts, the first part being the forward-phase, where new basis splines are added in order to improve the quality of the approximation and therefore reduce the residuum; the second phase is the backward-phase, where the number of basis splines is tried to be reduced. The basis spline with the worst influence on the approximation is thus removed. For the example of the sine function (Fig. 1), the MARS model (up to degree of 3) is able to approximate each given test data point and the trend within the boundaries of the test data points with high accordance. Only for the area of extrapolation, higher residuum can be seen.

In order to evaluate the quality of a surrogate model, the error between prediction and given data point (residuum) is calculated. The mean squared error (MSE) is a possible method to calculate the residuum and is widely used in literature for regression problems (Ref. 5). For regression problems, it is necessary to have separate training and test data points in order to evaluate the quality of the prediction for points not being part of the training data. The surrogate model is trained on the training data set and evaluated at the test data points. The mean squared error as a measure of quality of the prediction and the test data is calculated. The lower the error, the better the model is. For engineering problems, where the usage of surrogate models is interesting, it is very hard to come by test data sets, as it takes additional time and money to generate these data points. In order to still be able to differentiate between training and test data sets, the cross-validation method is used (Fig. 2, (Ref. 5).

The given data is partitioned in N small data packets. Each packet is used as test data set once whereas the remaining data sets are used to train the model. Thus, the surrogate model setup is conducted N times. The quality of the surrogate model is the average mean squared error of each test data set.

Objective and Approach

The state of the art shows that surrogate models offer a great potential in reducing the necessary number of long experiments/simulations and have been applied to various fields of engineering problems, most notably to geoscience problems. Until now, there has been no application/comparative study of different surrogate modeling techniques within the field of gear design. Especially for the micro geometry design of gears, a high number of parameter sets have to be investigated. This leads to a great number of variants, which need to be simulated.

The aim of the report is to investigate the potential of using surrogate models within the gear design process. The report focuses on the comparison of different surrogate modeling techniques/types and their suitability for the gear design process. In addition, the surrogate models are used to optimize the micro geometry in respect to the defined design objective.

The approach for using surrogate models within the gear design process is shown (Fig. 3.).

The process is demonstrated based on a flow chart and starts with the definition of the design objective.

There are several gear design objectives (durability, efficiency, NVH, weight, cost), which usually are in conflict with each other (Ref. 17). Thus, a weighted objective function is used to describe the design objective, which considers different objectives. Next, the parameter space is defined. For each micro geometry parameter, lower and upper boundaries are set. Following, DOE methods (e.g., latin hypercube, random sampling) are used to do a design space exploration for a defined number of sets and get initial data points with which to train the surrogate model. These data points are then simulated with the help of the FE-based tooth contact analysis ZAKO3D (TCA) in order to characterize the operational behavior of the gears (Ref. 18). Based on the results, the surrogate model is trained and tested based on the method of cross-validation. If the cross-validation error is low enough, the process is proceeded; otherwise, the hyperparameter (parameter for the model generation) of the surrogate model is tuned in order to reduce the cross-validation error further. Next, the decision whether to use the surrogate model as a regression formula or for an optimization of the objective function is to be taken. If the model should be used as a regression formula, the process is at its end and the surrogate model gives a good approximation formula which can

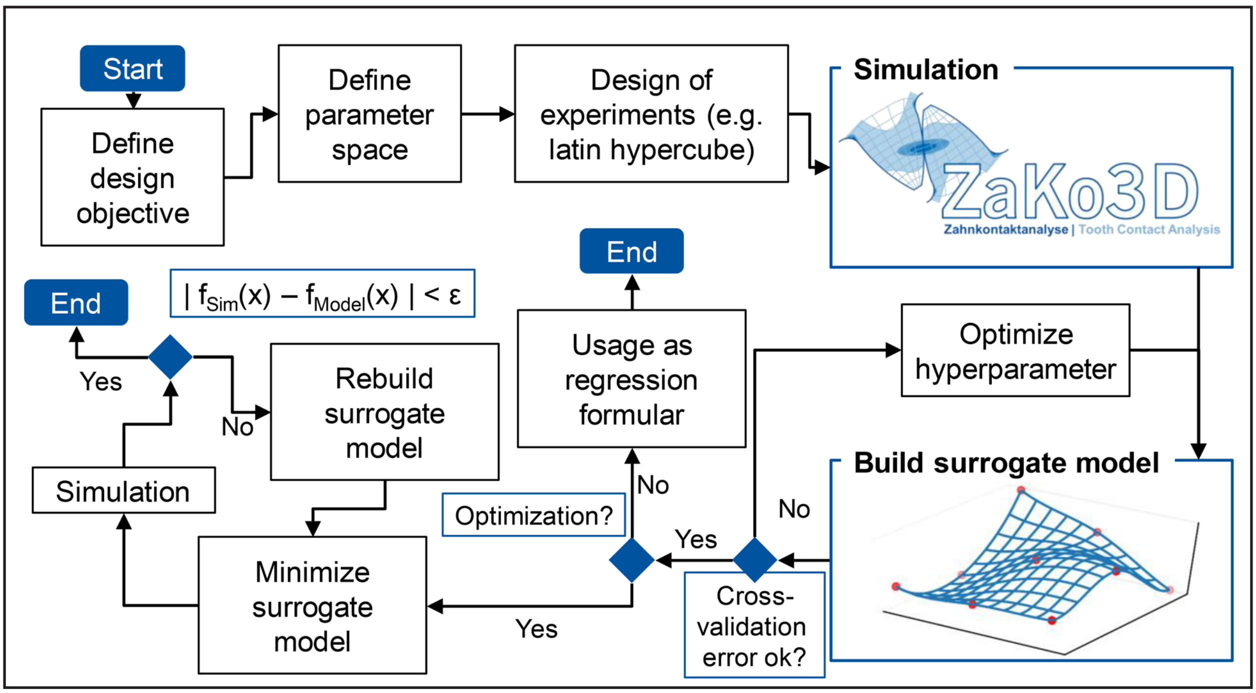

Figure 3 **Overview on how surrogate models can be used for the gear design process.**

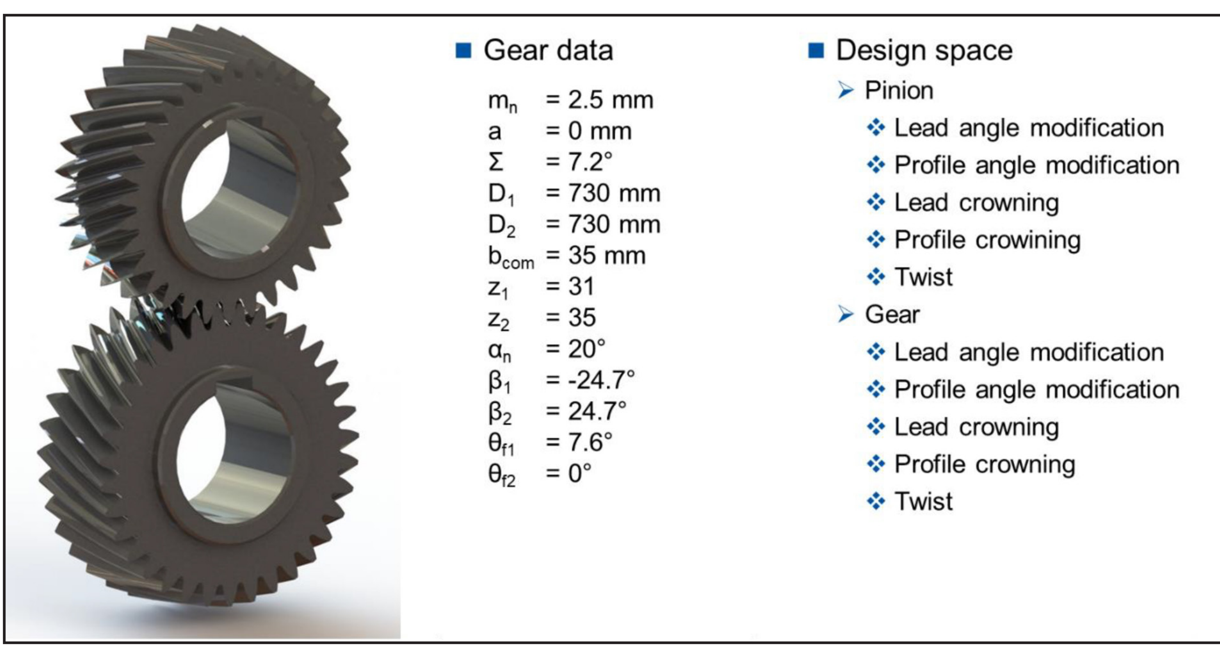

Figure 4 **Gear data of test gear set and design space.**

be easily evaluated in contrast to the simulation method. If the surrogate model should be used for an optimization, it is then used to minimize the approximated objective value with optimization techniques (e.g., gradient-based optimization). The optimal parameter set is recalculated in the simulation (here TCA). Next, the error between the surrogate model and the simulation is checked for the optimal parameter set. If the error is below a certain threshold, the optimization is ended. Otherwise, the surrogate model is retrained adding the optimal parameter set and the iterative process starts again.

Definition of Test Gear Set and Design Space

The test gear set which is used for this report is depicted in Figure 4. In order to investigate the potential of surrogate models, a beveloid gear set from an automobile application is chosen. The beveloid gear set has a normal module of $m_n = 2.5$ mm, number of teeth of $z_1/z_2 = 31/35$, a normal pressure angle of $\alpha_n = 20°$ and a crossing angle of $\Sigma = 7.2°$. The pinion is a beveloid gear with a root cone angle of $\theta_{f1} = 7.6°$, while the gear is a helical gear.

The design space for the investigation consists of ten different micro geometry parameters. For each, pinion and gear, the lead and profile angle modification, the lead and profile crowning as well as the twist are modified. The chosen upper and lower boundary for each parameter can be seen in Table 1.

The design objective for this report is the tooth root stress σ_{F2} acc. to van Mises at the gear for a torque of $T_2 = 100$ Nm. The occurring tooth root stresses are normalized to a scale of [0,1] in order to reduce the effect of the magnitude.

Comparison of Surrogate Modeling Techniques

Surrogate modeling techniques offer a great potential in approximation of complex simulations models. In order to set up a surrogate model, a defined initial set of points for the chosen design space has to be calculated. Design of experiment methods help in order to sample the points and have an even spread in the design space. Thus, the potential of surrogate models in the gear design process is also influenced by the DOE methods chosen to define the initial set of points. In order to investigate the influence of DOE methods, four different DOE methods were chosen. Out of the four methods, three methods are based on a latin hypercube while the last method is a random sampling within the boundaries of the design space (Refs. 11, 19). The sampling was done using the *Python* library *PYDOE2* (Ref. 20). Each of the three latin hypercube methods uses a different criterion for sampling the points. The criterion "center" centers the points within the sampling interval. The criterion "correlation" minimizes the correlation between each of the sampling points. The criterion "maximin" maximizes the minimum distance between each of the sampling points. The four methods were used to define samples for N = 300, 600 and 900 sample points within the abovementioned design space.

The comparison of the four DOE methods was done using a kriging model with a constant global polynomial (Fig. 5). The methods were compared on the value of the root mean squared

Table 1 Boundaries for the design space

Parameter	Lower boundary / µm	Upper Boundary / µm
Lead angle modification Pinion	-60	80
Profile angle modification Pinion	-42.5	0
Lead crowning Pinion	-50	0
Profile crowning Pinion	0	20
Twist Pinion	-30	30
Lead angle modification Gear	-80	80
Profile angle modification Gear	-20	20
Lead crowning Gear	-50	50
Profile crowning Gear	0	20
Twist Gear	-30	30

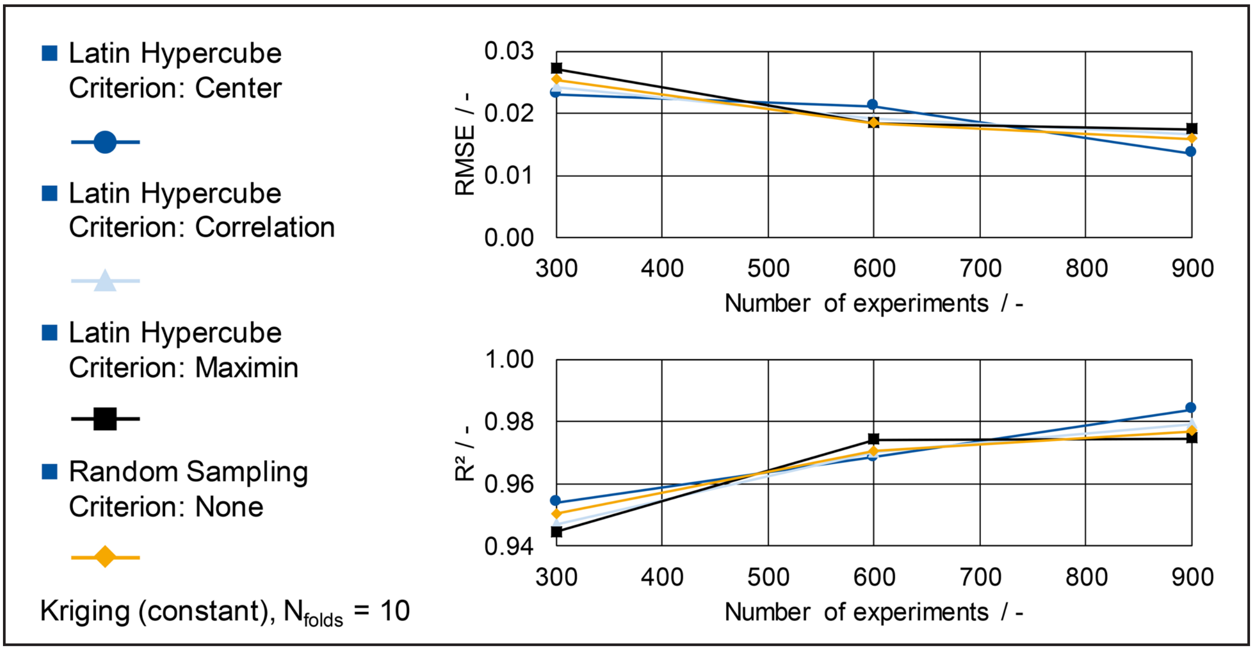

Figure 5 Comparison of DOE sampling methods.

error (RMSE) and the coefficient of determination R^2. Each of the two values is the mean of a cross-validation using $N_{folds} = 10$.

The comparison of the four DOE methods shows no defined influence of the sampling method on the quality of the approximation using the kriging model. For each of the methods a coefficient of determination $R^2 > 0.94$ can be achieved. Thus, a sampling using a latin hypercube with the center criterion is used onwards because of the highest mean coefficient of determination.

The figure also shows the influence of the number of experiments (sampling points). With rising numbers of experiments the quality of the approximation is increasing, which is to be expected. In order to investigate whether the coefficient of determination converges to a certain value with increasing number of experiments, a higher number of experiments is added.

The surrogate modeling techniques mentioned in the state of the art are compared to each other in terms of approximation quality (Fig. 6). The RMSE and R^2 are used as quality criterion for the comparison and are the mean values of a cross-validation using $N_{folds} = 10$. The kriging model uses a linear polynomial as a global function and uses a Gaussian correlation as a local function. The model is set up using the *SMT Toolbox* written in *Python* ((Ref. 1). The RBF model uses no global polynomial and is set up using the *SMT Toolbox*. The MARS model uses basic splines up to the degree of three and a maximum number

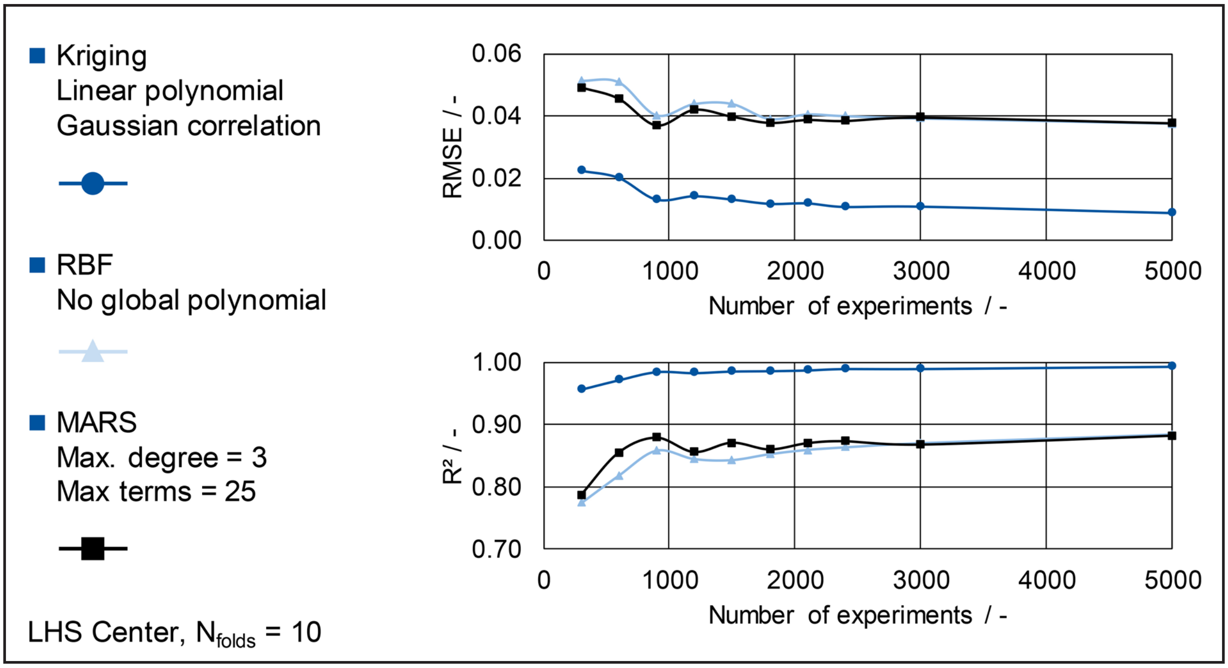

Figure 6 **Comparison of different surrogate modeling techniques.**

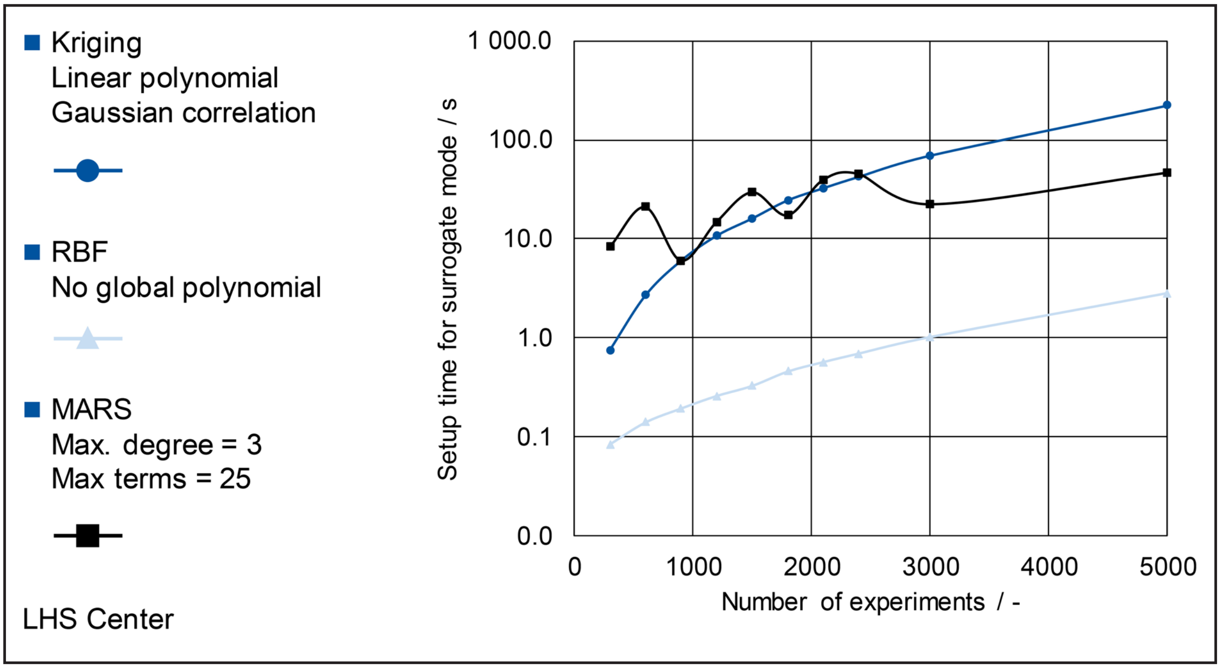

Figure 7 **Comparison of setup time for the different surrogate modeling techniques**

of terms of 25. The model is set up using the *Python* library *PY-EARTH* (Ref. 21).

For each of the three modeling techniques we can see an increase in quality with an increasing number of experiments. Comparing the three techniques, the kriging model has the best overall performance in terms of approximation quality. The RMSE is below 0.023 for each of the selected number of experiments and is steadily decreasing to a RMSE = 0.0093 at N = 5,000 experiments. The RBF and MARS model in contrast show a magnitude which is more than double in comparison to the kriging model. The difference between the RBF and the MARS model is minor, while the MARS model shows a slightly better approximation.

The same observation can be made for the coefficient of determination. The kriging model shows the best approximation with $R^2 > 0.95$ for each of the samples — even at a low number of N = 300 experiments. For N = 5,000 experiments $R^2 = 0.993$ can be achieved. The coefficient of determination for the RBF and the MARS model stays below a $R^2 < 0.9$ for each sample.

The results of the kriging model show that the marginal approximation quality (increase in quality with increase in number of experiments) is almost constant after N = 1,500 experiments. For the given example gear set, a number of N = 1,500 experiments should be enough to build a very good surrogate model.

Based on the results shown, we can conclude that the kriging method is the most suitable surrogate modeling technique for the gear design process of the three chosen techniques. Even for a low number of experiments, the coefficient of determination already shows very good results.

Although the approximation quality is the most important factor when it comes to surrogate modeling, the duration for setting up the model also plays a role in the design process. Therefore, Figure 7 shows the comparison of the setup time for the three used models in a logarithmic scale.

The lowest setup time is achieved by the RBF model, which is faster by a magnitude of 10 in comparison to the kriging and MARS model. For a low number of experiments the kriging model is faster than the MARS model. For a high number of experiments N > 2,500, the setup time for the kriging model is greater than the setup time for the MARS model. The setup time for the MARS model is not heavily influenced by the sample of experiments in comparison to the other two models.

Although the kriging model shows the highest setup time (especially for high number of experiments), the magnitude of the time to setup the surrogate model is still below or in the range of the simulation time for one sample point. Thus, we can conclude that the kriging model offers the highest potential for the usage in the gear design process.

Summary and Outlook

The gear design process is relying heavily on the use of simulation methods for predicting the operational behavior of gears. The cross influences of different geometry parameters (being macro or micro geometry) and positional parameters are oftentimes non-linear. In addition, there usually is a wide design space, which needs to be tested/explored to find the most suitable parameter setting. Thus, a great number of variants needs to be simulated in a tooth contact analysis. The high number of variants causes a high effort in calculation time and resources. In order to reduce the necessary number of experiments and therefore reduce the design time for gears, this paper focusses on the potential of using surrogate modeling techniques in the gear design process.

The investigation is conducted for a beveloid sample gear set with a design space consisting of ten different micro geometry parameters. The design space is explored using four different sampling techniques from the design of experiment (DOE). The results show that the chosen sampling techniques only have a very minor effect on the quality of the approximation. Because of the highest mean coefficient of determination, a centered latin hypercube was chosen for the further investigation. The comparison of the surrogate modeling techniques shows that a kriging model offers the highest approximation quality; this is even true for a low number of experiments. It can also be observed (which was to be expected), that an increasing number of experiments leads to an increase in the approximation quality. For the particular test gear set, a number of N = 1,500 experiments already offers offer a very good approximation.

Future research should focus on the application of the surrogate models in the optimization process. In addition, other modeling techniques, like convolutional neural networks or K-nearest neighbor approaches, should be investigated.

Acknowledgement

The authors gratefully acknowledge the financial support of the German Research Foundation (DFG) – Project Number 362666660. In addition, the authors gratefully acknowledge the support of the RWTH Aachen University by providing computing resources under project rwth0343.

References

1. Bouhlel, M.A., J.T. Hwang, N. Bartoli, R. Lafage, J. Morlier and JRRA Martins. (2019) A *Python* surrogate modeling framework with derivatives. *Advances in Engineering Software:102662. doi:10.1016/j.advengsoft.2019.03.005.*
2. Wang, G.G. and S. Shan. (2007) "Review of Metamodeling Techniques in Support of Engineering Design Optimization." *J. Mech. Des. 129(4):370, doi:10.1115/1.2429697.*
3. Wang, J., W. Shen, Z. Wang, M. Yao and X. Zeng. (2014) "Multi-objective optimization of drive gears for power split device using surrogate models," *J Mech Sci Technol 28(6):2205– 2214. doi:10.1007/s12206-014-0509-4.*
4. Dasari, S.K., A. Cheddad and P. Andersson. (2019) "Random Forest Surrogate Models to Support Design Space Exploration in Aerospace Use-Case," J. *MacIntyre*, I. Maglogiannis, L. Iliadis and E. Pimenidis. (Hrsg) *Artificial Intelligence Applications and Innovations. Springer International Publishing, Cham, S 532–544.*
5. Forrester, A.I.J., A. Sóbester and A.J. Keane. (2008) *Engineering design via surrogate modelling. A practical guide.* Wiley, Chichester.
6. Forrester, A.I.J. and A.J. Keane. (2009) "Recent advances in surrogate-based optimization," *Prog. Aerosp. Sci. 45(1-3):50–79. doi:10.1016/j.paerosci.*2008.11.001.

7. Koziel, S and J.W. Bandler. (2012) "Accurate modeling of microwave devices using kriging- corrected space mapping surrogates." *Int. J. Numer. Model. 25(1):1–14. doi:10.1002/jnm.803.*
8. Mack, Y., T. Goel, W. Shyy and R. Haftka. (2007) *Surrogate model-based optimization framework. A case study in aerospace design Evolutionary computation in dynamic and uncertain environments*, Springer, S 323–342.
9. Queipo, N.V., R.T. Haftka, W. Shyy, T. Goel, R. Vaidyanathan and P.K. Tucker. (2005) "Surrogate- based analysis and optimization." *Prog. Aerosp. Sci. 41(1):1–28.*
10. Razavi, S., B.A. Tolson and D.H. Burn. (2012) "Review of surrogate modeling in water resources," Water Resour. Res. 48(7):559. doi:10.1029/2011WR011527.
11. Siebertz, K. van, D.T. Bebber and T. Hochkirchen. (2010) *Statistische Versuchsplanung, Design of Experiments (DOE),* 1. Aufl. VDI-Buch. Springer, Heidelberg.
12. Lowe, D. (1988) Multi-variable functional interpolation and adaptive networks. Complex Systems 2:321–355.
13. Daya, Sagar B.S., Q. Cheng and F. Agterberg. (Hrsg) (2018) *Handbook of Mathematical Geosciences: Fifty Years of IAMG*, Springer International Publishing, Cham.
14. Simpson, T.W., T.M. Mauery, J.J. Korte and F. Mistree. (2001) "Kriging Models for Global Approximation in Simulation-Based Multidisciplinary Design Optimization," *AIAA Journal 39(12):2233–2241. doi:10.2514/2.1234.*
15. Amine Bouhlel M, N. Bartoli N, R.G. Regis, A. Otsmane and J. Morlier. (2018) "Efficient global optimization for high-dimensional constrained problems by using the Kriging models combined with the partial least squares method," *Eng. Optim. 50(12):2038–2053. doi:10.1080/0305215X.2017.1419344.*
16. Friedman, J.H. (1991) "Multivariate Adaptive Regression Splines." *Ann. Statist. 19(1):1–67. doi:10.1214/aos/1176347963.*
17. Klocke, F. and C. Brecher. (2017) Zahnrad- und Getriebetechnik. Auslegung - Herstellung - Untersuchung - Simulation, 1. Aufl. Carl Hanser, München.
18. Hemmelmann, J.E. (2007) "Simulation des lastfreien und belasteten Zahneingriffs zur Analyse der Drehübertragung von Zahnradgetrieben," Diss., RWTH Aachen University.
19. Simpson, T.W., J.D. Poplinski, P.N. Koch and J.K. Allen. (2001) "Metamodels for Computer-based Engineering Design. Survey and recommendations." *Eng. Comput. 17(2):129–150. doi:10.1007/PL00007198.*
20. Sjögren, R., D. Svensson and A.D. Lee. (2018) *pyDOE2.*
21. Rudy, J. (2013) *Py-earth.*

Dr.-Ing. Jens Brimmers is the head of the gear department at the Laboratory for Machine Tools and Production Engineering (WZL) of RWTH Aachen University since June 2019. He gratuated from RWTH Aachen University with master's degrees in mechanical engineering and business administration. His Ph.D. thesis focussed on beveloid gears and topological tooth flank modifications.

Marius Willeke, M. Sc. studied Mechanical Engineering at RWTH Aachen University 2019, graduating with distinction as Master of Science. Since 2019, he has served as a research assistant at WZL of RWTH Aachen University. His primary interests include end of line testing development of tooth contact simulation software and simulation of acoustic behavior of gearboxes gear measurement.

Dr. -Ing. Dipl. -Wirt. -Ing. Christoph Löpenhaus is since 2021 working with Flender GmbH as head of production for the part manufacturing plants in Bocholt and Voerde, Germany. Flender is a worldwide leading provider of drive systems especially for wind turbine and industry applications. From 2019-2021 he was Business Development Manager Geared Bearings at Cerobear GmbH, a leading manufacturer of hybrid bearings especially in the aerospace, space, and industry segment. From 2014–2019 he held the position as Chief Engineer of the Department of Gear Technology of WZL, RWTH Aachen with research focus on gear manufacturing, design, and testing. He previously worked with WZL as Team Leader and Research Assistant. His educational background is in the field of Industrial Engineering with a diploma in 2009 and a Ph.D. in mechanical engineering in the field of gear technology in 2015. For his scientific achievements he was awarded the Springorum Commemorative Coin in 2010 and the Borchers Badge in 2016.

Prof. Dr.-Ing. Christian Brecher has since January 2004 been Ordinary Professor for Machine Tools at the Laboratory for Machine Tools and Production Engineering (WZL) of the RWTH Aachen, as well as Director of the Department for Production Machines at the Fraunhofer Institute for Production Technology IPT. Upon finishing his academic studies in mechanical engineering, Brecher started his professional career first as a research assistant and later as team leader in the department for machine investigation and evaluation at the WZL. From 1999 to April 2001, he was responsible for the department of machine tools in his capacity as a Senior Engineer. After a short spell as a consultant in the aviation industry, Professor Brecher was appointed in August 2001 as the Director for Development at the DS Technologie Werkzeugmaschinenbau GmbH, Mönchengladbach, where he was responsible for construction and development until December 2003. Brecher has received numerous honors and awards, including the Springorum Commemorative Coin; the Borchers Medal of the RWTH Aachen; the Scholarship Award of the Association of German Tool Manufacturers (Verein Deutscher Werkzeugmaschinenfabriken VDW); and the Otto Kienzle Memorial Coin of the Scientific Society for Production Technology (Wissenschaftliche Gesellschaft für Produktionstechnik WGP).