33RD CONGRESS
OF THE INTERNATIONAL COUNCIL
OF THE AERONAUTICAL SCIENCES
STOCKHOLM, SWEDEN, 4-9 SEPTEMBER, 2022

ICAS 2022 SWEDEN

Enhancement of Flexible Strategy for Augmenting Design Points for Computer Experiments and Application to a MDO of a Fan-Blisk

A. Tamkan[1], P. Flassig[1], M. Meyer[2]

[1]Technical University of Applied Sciences Brandenburg, Department of Engineering
[2]Rolls-Royce Deutschland Ltd & Co KG

Abstract

For every design exploration problem, the question of how many design points should be used to explore design space sufficiently is remained to be answered. Few points might result in insufficient approximation and lack of understanding of the design problem. In contrast, too many points might be costly, both in terms of time and resources especially if the model is multi-dimensional and complex. After execution of all designs of a defined explorative study, it is often the case that models fitted to the data are not sufficiently accurate which leads to the need for additional design points or a new Design of experiments (DoE). A usual approach that users follow is initializing a new DoE with increased number of design points. This however could lead to the complete waste of the first investigation. Therefore, it is necessary to add design points wherever and whenever is necessary in order to help users to explore the design space efficiently. In this regard, this study proposes a highly efficient and universal method to augment an existing DoE which is successfully applied to a multidisciplinary design optimization (MDO) problem.

1. Introduction

DoE is a systematic method to plan and evaluate experiments for complex models by variation of input factors in order to obtain knowledge about the model and its output factors. Exploring the design space or getting as much as information possible regarding the model behavior using a defined minimum number of design points or, respectively, to obtain a required model understanding by examining a minimum number of experiments is desirable. If a global explorative study or an approximation model of a costly to evaluate transfer function is desired, the first step is often to generate a space-filling DoE. The number of design points depends on the design space dimension, the complexity of the transfer function, and the time or computational budget. From an industrial point of view, one strives to minimize the number of runs for a sufficient level of accuracy in fitting an approximation model. Unfortunately, this number is not known a priori for most cases, and heuristics, combined with empirical knowledge are used. After execution of all designs, it is often the case that models fitted to the data are not sufficiently accurate, leading to the need for additional design points to be evaluated. This can be achieved by either defining a new DoE or adding additional points. Of course, when the former is pursued, this new DoE will be completely new which may be in that case unacceptably wasteful.

2. Methodology

Various researches have been conducted to improve the capability and efficiency of the existing Latin Hypercube Design (LHD) which is one of the most popular DoE techniques. It has been stated that the LHD might be enhanced by reducing the correlation among the inputs of the design matrix for a multitude of dimensions [1]. Another research suggests that pairwise correlation and inter-site distances should be considered simultaneously while constructing an LHD [2]. With regards to the augmentation of design plans, numerous methods have been published. In one study, it has been shown that various exploration-based sequential design strategies are more efficient compared to the popular and proven space-filling optimal Latin Hypercube (LH) approaches [3]. In [4] a method is proposed where an LHD with specific probability density functions (pdf) and correlation structure can be extended. The number of points to be augmented needs to be twice the number of initial sampling points. This could be a downside when fewer numbers are required for a required level of system understanding. A highly efficient and universal method to augment an existing DoE with additional points while maintaining space-filling properties (uniform and orthogonal distributed samples) has been presented in [5]. This method is capable of dealing with input DoE plans which are plain Monte Carlo (MC), Latin Hypercube (LH), or optimal LH plans, as given in Figure 1.

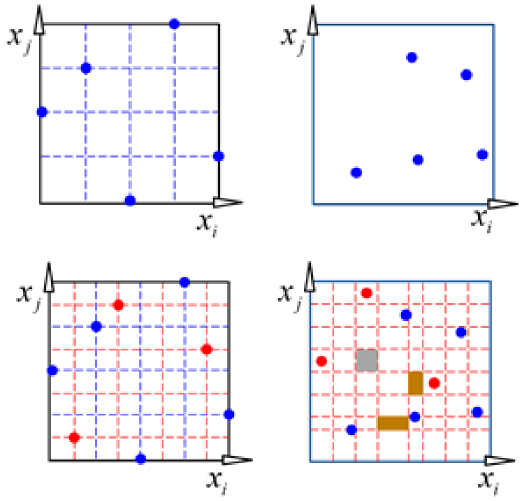

Figure 1 – Basic idea of 1 level augmentation with 3 additional points on a 2D design space with an initial sample size of 5 for a proper Latin Hypercube Sampling (LHS) input (left), and MC input (right).

Additionally, the space-filling properties of the final design plan are nearly independent of the augmentation sequence, i.e. from the number of levels and batch size. Basically, the final sample plan size of an augmentation sequence is defined by

$$N = N_0 + \sum_{i=1}^{l} m_i \quad (1)$$

with the initial number of samples, the number of levels and batch size are given as N_0, l, m_i, respectively. For demonstration Figure 2 shows empirical results of random augmentation sequences that are repeated 100 times, i.e. with a random number of levels and random batch sizes, for a 50 dimensional design space, and an initial sample size of $N_0 = 64$ and a maximum sample size of $N = 150$. As space-filling criterion the maximum absolute pairwise correlation coefficient given as

$$\rho_{map} = \max_{i=1(1)d, j<i} \left|\rho_{ij}^{P}\right| \quad (2)$$

where the indices i, j and the the design space dimension d define all relevant pairs of design parameter combinations [5]. The lower green line in Figure 2 represents the reference quality that has been obtained by the creation of direct optimal LHS with varying sample sizes in the range between 80 and 150. The black points show exemplarily how the space-filling property of the 2nd random augmentation (from the 100) evolves from each augmentation level, i.e. the first level augments the initial design plan with $N_0 = 64$ by $m_1 = 19$ points. This leads to a sample size of 83 and a maximum absolute pairwise correlation coefficient $\rho_{map} = 0.38$. With the second augmentation with $m_2 = 50$, ρ_{map} is reduced to 0.31.

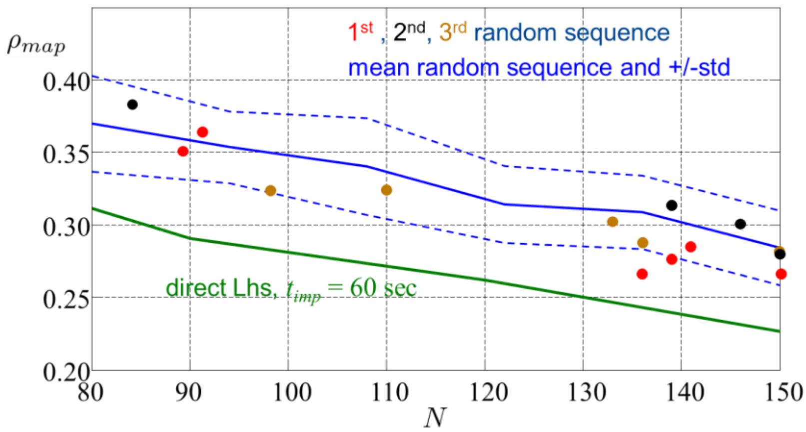

Figure 2 – Empirical results of random augmentation sequences according to [5], repeated 100 times to demonstrate universality.

It is noted that methods such as descriptive sampling (e.g. SOBOL sequences [6]) can be used to define a DoE with a very high number of designs where each additional design bases on the previous ones, making it possible to increase the number of design successively. However, these methods struggle with space filing properties (even for large numbers of design points) for design space dimensions greater than ten [5].

For MDO design problems and emulation techniques for rapid design space exploitation, it is often required that design points need to be condensed in certain areas of the design space because emulated understanding is not sufficient and prediction accuracy is too low for these areas. One could simply augment the whole design space until sufficient understanding has been gained. This approach leads to more design evaluation than necessary, of course. Therefore, this paper focuses on enhancing the method presented in [5] by adding the capability to not only allow design plans with evenly spread design points but also non-uniform distributed points defined by arbitrary probability density functions.

To motivate basic idea of the developed method the *Branin* test function

$$f(x_1, x_2) = \left(x_2 - \frac{51}{40\pi^2}\,{x_1}^2 + \frac{5}{\pi}\,x_1 - 6\right)^2 + 10\left(1 - \frac{1}{8\pi}\right)\cos(x_1) + 10 \qquad (3)$$

is used. As can be seen in Figure 3 an initial optimal LHS with $N_0 = 15$ points has been created and the *Branin* test function has been evaluated for these points. Based on the results a radial basis function (RBF) based approximation model $\hat{f}(x_1, x_2)$ according to [7] has been trained. A qualitative comparison of the prediction capability of the RBF-based model with the exact function can be seen in Figure 3. Since both models are not costly to be evaluated the absolute relative error between *Branin* and RBF-based approximation in percent

$$\varepsilon \;=\; 100\%\,\left|(f - \hat{f})/f\right| \qquad (4)$$

can be calculated at hundreds of points across the two-dimensional design space. As a result, three areas with very high differences can be identified, cf. right plot in Figure 3. Nevertheless, this is only possible for a pre-defined or known design space.

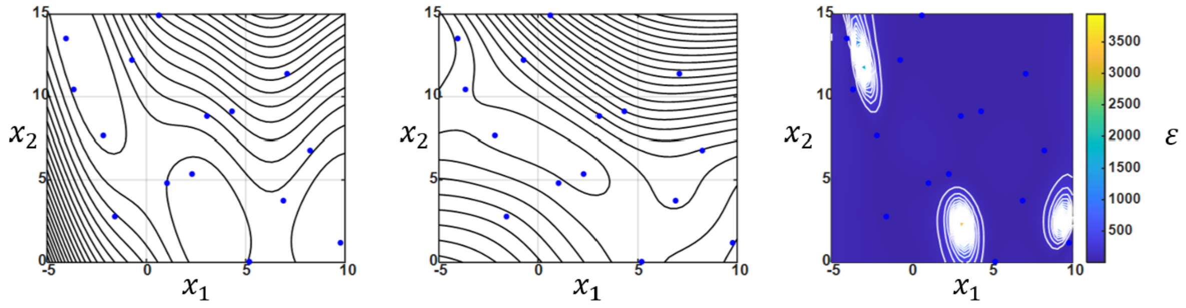

Figure 3 – Approximation of *Branin* test function (left) based on an optimal LHD with $N_0 = 15$ points using RBF (middle) and resulting absolute relative errors in percent (right).

For a costly to evaluate transfer function prediction capability can be estimated by use of *leave-one-out* cross validation (LOOCV). In principle, the procedure of LOOCV is to remove each sample once from the training data set to train RBF models for each reduced set of samples and to predict the exact function values at the removed data point to be able to compare the differences between the predicted and known value. The advantage of LOOCV is to use the existing data only without a need for analytical definition. For the test model, the LOOCV results in terms of the difference between RBF approximation and *Branin* test function can be indicated with the *Pearson correlation coefficient* ρ^P obtained from the scatter plot which is shown in Figure 4c. For the initial RBF model, the correlation coefficient between *Branin* function values and approximation values is $\rho^P = 0.84$. Also, the relative error ε according to Equation 4 is shown in Figure 4b as a contour plot. Comparing Figure 4b and Figure 3 right, one can see that the LOOCV results match the existing differences between the *Branin* test function and the RBF approximation. The two curves that are shown in Figure 4 are the error pdfs that are derived by projecting the LOOCV errors to the individual axes and scaling to ensure the basic requirements of a pdf are illustrated in Figure 4a and Figure 4d, respectively. As shown in Figure 4d the density of the $f_\varepsilon(x_1)$ has maxima for lower x_1 and towards upper bound which is also can be identified from the three peaks of Figure 4b.

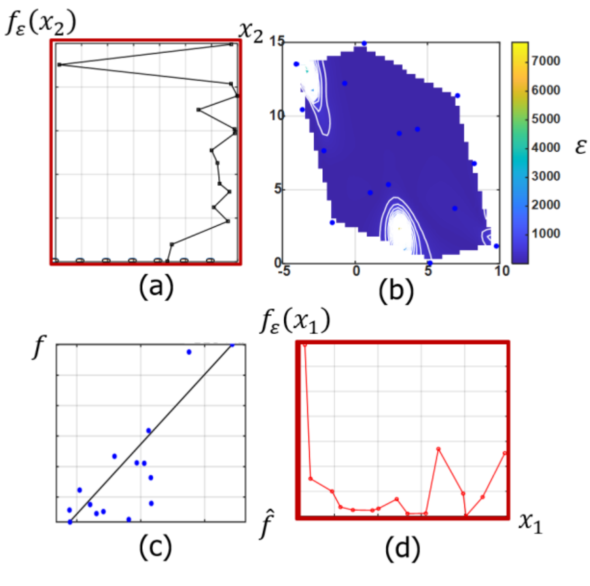

Figure 4 – LOOCV prediction results in form of a contour graph for ε for the RBF-based approximation of the *Branin* test function shown in Figure 3 (b), Pearson correlation ρ^P (c), and projections of the errors for x_1 and x_2 in form of pdfs $f_\varepsilon(x_1)$ (d) and $f_\varepsilon(x_2)$ (a).

Although $\rho^P = 0.84$ can be acceptable for some use cases, sometimes a better approximation model might be needed. To further increase predicting capability, the two augmentation models (i) standard and (ii) a newly developed enhanced pdf-based, each with a batch size of $m_1 = 15$ are applied. Figure 5, shows the augmented points coming from the two methods and the LOOCV results from the two RBF models using $N_1 = 30$ points.

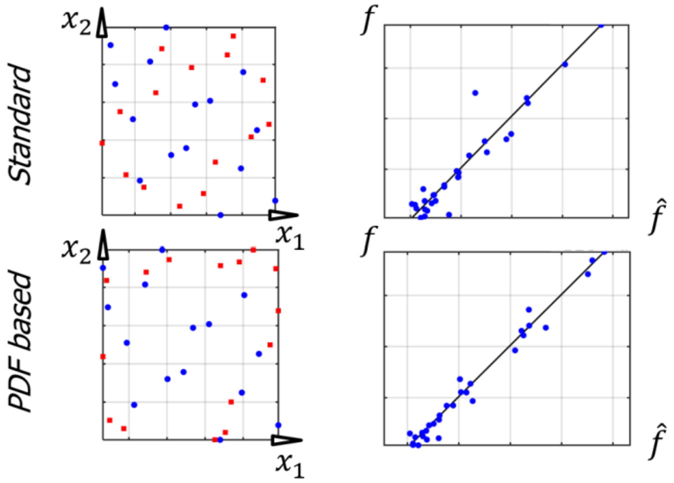

Figure 5 – Standard augmentation (top) and enhanced pdf-based augmentation (bottom) using results, i.e. pdfs $f_\varepsilon(x_1)$ and $f_\varepsilon(x_2)$, shown in Figure 4 for the 15 initial points (blue circles) and augmented points (red squares) with their respective correlation coefficients.

Comparing the standard augmentation with Figure 3 and Figure 4, it can be seen that a lot of augmented points are located where predicting quality was already acceptable. To overcome this inefficiency, the additional points are not added as evenly distributed but according to error pdfs for each of the two design variables. Compared to the standard augmentation the proposed approach augments design points where the error is high. As illustrated in Figure 4, the error for parameters x_1 and x_2 are yielded high for lower and higher ends, and thus the pdf-based augmentation shifts the points to the areas where prediction capability is low. This also results in better emulator prediction capability with the increased ρ^P and consecutively a better approximation model as shown in Table 1.

	Initial model	Standard augmentation model	Enhanced PDF based augmentation model
Number of sample points	15	30	30
Pearson correlation coefficient	0.84	0.95	0.99

Table 1 – Comparison of LOOCV results in terms of Pearson correlation coefficient for the *Branin* test function

3. Application

The design of an aero-engine fan blade is a demanding task due to the large design space, attempting to meet contradicting objectives, stringent certification requirements, insufficient analytical design tools, coupling non-linear dynamics, lack of experimentation, required high-precision simulation methods, and time-consuming process chain. The need to operate under a wide range of flight conditions during the flight, demanding not only good design and off-design performance characteristics but also large margins of aerodynamic stability to avoid stalls and surges makes the design of the fan blade further challenging. As a consequence, state-of-the-art simulation methods are compulsory to adequately solve such multi-faceted problems involving various disciplines. Multidisciplinary design optimization indeed proposes a solution, providing the necessary sophistication using advanced algorithms and numerical techniques.

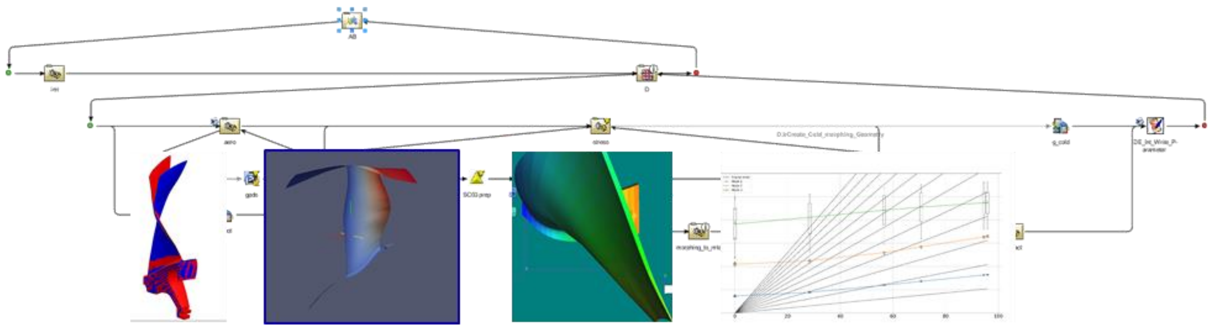

Figure 6: MDO *Isight* workflow for a fan-blisk with geometry creation for different running conditions, 3D CFD, stress and postprocessing.

To demonstrate industrial applicability, the standard and enhanced augmentation methods are applied to a Multidisciplinary Design Optimization (MDO) workflow for a Fan-Blisk. Figure 6 shows the current design process, that has been automated by use of the process and automation tool *Isight*, where fully parametrized blade creation, 3D CFD, and stress assignments can be performed.

A standard exemplary design task is to optimize the aerofoil subjected to constraints from various disciplines. For proof of concept, two simple single-objective optimization problems have been formulated. The first one focuses on structural aspect and given as

$$\min_{\boldsymbol{x}^l \leq \boldsymbol{x} \leq \boldsymbol{x}^u} f_1(\boldsymbol{x}) \quad s.t. \;\; \boldsymbol{h}(\boldsymbol{x}) \leq \boldsymbol{0} \in R^r \,, \quad \boldsymbol{x} \in R^6 \,, \tag{5}$$

where $\boldsymbol{x}$ corresponds to six-dimensional design parameter vector including two maximum thickness values at hub- and mid-span, two axial-shift values at mid- and tip-span, and two theta-shift values at mid- and tip-span ranging in between lower $\boldsymbol{x}^l$ and upper $\boldsymbol{x}^u$ bounds. The balancing criterion $f_1(\boldsymbol{x})$ is defined by different stress values on the aerofoil. The feasible design space is defined by l number of different frequency and high cycle fatigue (HCF) constraints $\boldsymbol{h}(\boldsymbol{x}) \in R^r$.

The initial design space for the first example is initially populated using 40 points whereas only $N_0 = 32$ blades successfully pass through the whole process. Subsequently, it is then augmented using 20 points by using standard augmentation and enhanced augmentation. From the 20 points, however, $m_1 = 18$ designs, and $m_1 = 16$ designs have converged for respective methods. For the demonstration purpose, a hard to emulate HCF constraint is selected out of the $l + 1$ trained RBF models. The LOOCV results for the initial model, and for the both augmentations have been illustrated in Figure 7 and tabulated in Table 2. Since the correlation coefficient of the enhanced augmentation method has been improved by a small fraction, the LOOCV mean error

$$\mu_\varepsilon = \frac{100\%}{N} \sum_{i=1}^{N} \frac{f_i - \hat{f}_i}{f_i} \tag{6}$$

and the standard deviation

$$\sigma_\varepsilon = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(100\% \cdot \frac{f_i-\hat{f}_i}{f_i} - \mu_\varepsilon\right)^2} \quad (7)$$

have been also introduced and tabulated in Table 2. As it can be seen from Table 2, the errors of the enhanced method are lower indicating that the RBF using the enhanced augmented data has better prediction capability compared to the standard augmentation.

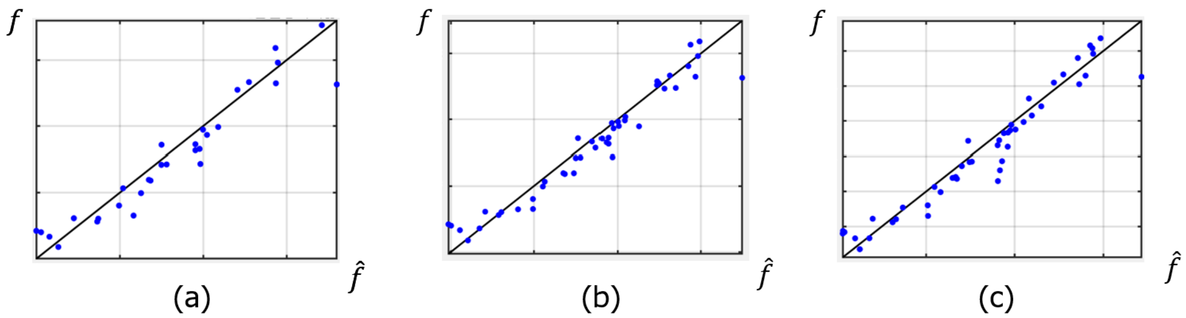

Figure 7: Comparison of LOOCV results and correlations for initial model (a), standard augmentation (b) and enhanced augmentation (c).

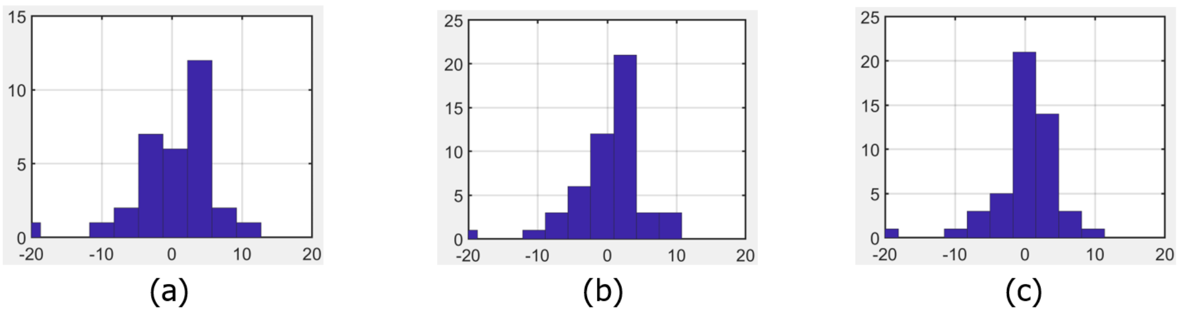

Figure 8: Comparison of LOOCV relative errors for initial model (a), standard augmentation (b) and enhanced augmentation (c).

	Initial model	Standard augmentation model	Enhanced PDF based augmentation model
Number of sample points	32	50	48
Pearson correlation coefficient	0.89	0.92	0.924
Mean error	0.28%	0.20%	0.17%
Standard deviation error	6.20%	5.03%	4.92%

Table 2 – Comparison of LOOCV results in terms of Pearson correlation coefficient, the mean error and standard deviation error according to Equation 6 and 7 for the first use case

The second fan blisk problem focuses on the aerodynamic performance and reads as

$$\min_{x^l \leq x \leq x^u} f_2(\boldsymbol{x}) \quad s.t. \quad \boldsymbol{h}(\boldsymbol{x}) \leq \boldsymbol{0} \in R^l, \quad \boldsymbol{x} \in R^{12}, \tag{8}$$

where $\boldsymbol{x}$ corresponds to twelve-dimensional design parameter vector including two blade inlet angle values at hub- and tip-span, two blade outlet angle values at hub- and tip-span, two chord values at hub- and tip-span, two maximum thickness values at hub- and tip-span, two axial-shift values at mid- and tip-span, and two theta-shift values at mid- and tip-span ranging in between lower $\boldsymbol{x}^l$ and upper $\boldsymbol{x}^u$ bounds. The aerodynamic efficiency $f_2(\boldsymbol{x})$ is obtained from 3D computational fluid dynamics (CFD) and inequality constraints analogous to the Equation 5.

In contrast to the previous example, this time two reference LH experiments using 50 and 100 have been executed to investigate the potential differences between sequential and a single batch DoE. However, $N_0 = 44$ and $N_0 = 85$ number of design points have been completed successfully. The results from the LOOCV have been illustrated in Figure 9a and 9b, respectively.

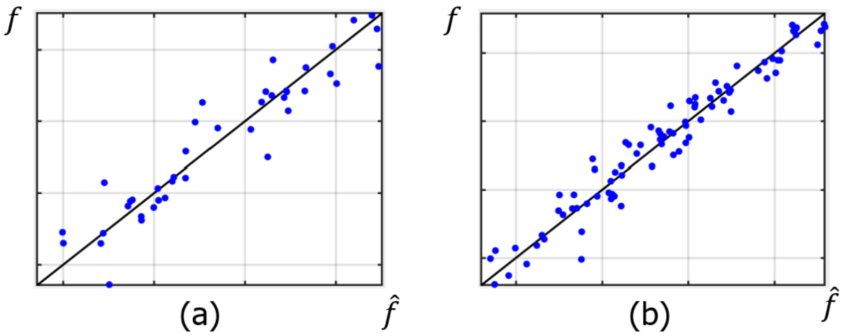

Figure 9: Comparison of LOOCV results and correlations for first reference model (a), and second reference model (b)

Single batch LH	First model	Second model
Number of sample points	44	85
Pearson correlation coefficient	0.95	0.97

Table 3 – Comparison of LOOCV results in terms of Pearson correlation coefficient and number of sample points for the first and second reference single batch DoE

Similar to the first case, the initial design space is initially populated using 50 points whereas only $N_0 = 44$ blades successfully pass through the whole process. The first augmentation has been performed using 20 points by using standard augmentation and enhanced augmentation. From the 20 points, however, $m_1 = 19$ designs have converged for both methods. After that, the second augmentation has been performed with 20 points, however, $m_1 = 19$ designs, and $m_1 = 18$ designs have converged for standard and enhanced methods. For the demonstration purpose, the aerodynamic efficiency is picked out of the $l + 1$ trained RBF models. The LOOCV results for the standard and enhanced methods have been illustrated in Figure 10, Figure 11, and tabulated in Table 4 and 5, respectively.

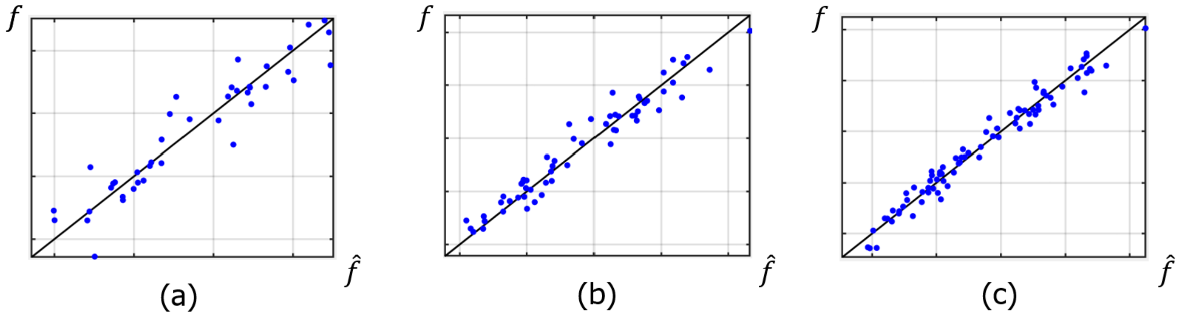

Figure 10: Comparison of LOOCV results and correlations for initial model (a), first augmentation (b) and second augmentation (c) for the standard method.

Standard augmentation	Initial model	First augmentation	Second augmentation
Number of sample points	44	63	82
Pearson correlation coefficient	0.95	0.97	0.98

Table 4 – Comparison of LOOCV results in terms of Pearson correlation coefficient and number of sample points for the standard augmentation case

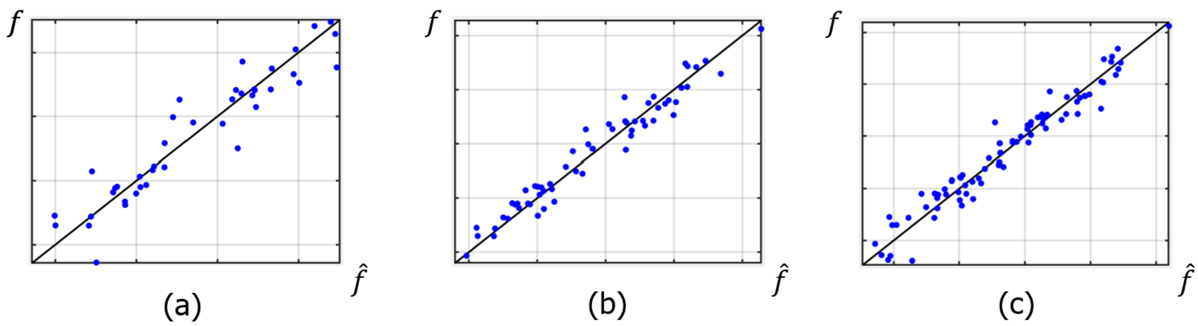

Figure 11: Comparison of LOOCV results and correlations for initial model (a), first augmentation (b) and second augmentation (c) for the enhanced method.

Enhanced augmentation	Initial model	First augmentation	Second augmentation
Number of sample points	44	63	81
Pearson correlation coefficient	0.95	0.98	0.98

Table 5 – Comparison of LOOCV results in terms of Pearson correlation coefficient and number of sample points for the enhanced augmentation case

The first observation from the second problem is that residual that comes from the second reference DoE executed with a single batch $N_0 = 85$ is higher compare to the first enhanced augmentation with $N_0 = 63$. Even the first standard augmentation with $N_0 = 63$ is yielding similar results despite having less number of design points being used for the model. From Table 4 and 5, the results show that the models improve with the first augmentation but do not improve much with the second augmentation. The difference that should be mentioned is that the first augmentation of the enhanced method has the same capability with the second augmentation of the normal method. This creates a numerical advantage and clearly saves time and since the second augmentation will not be needed if the enhanced method is used.

4. Conclusions

In this paper, a newly developed enhanced pdf-based augmentation method has been introduced. The method has been developed by using the error pdfs that are obtained by projecting the leave-one-out-cross-validation results to individual design parameter. The efficiency of the method first has been tested and proved using the *Branin* test function on two-dimensional design space at first, to apply later on for an industrial application. The industrial application is an MDO of a Fan-Blisk workflow where the design task is the single objective optimization of an aerofoil considering the constraints from multiple disciplines. The problem at hand is divided into two subsets to demonstrate different approaches that one can possibly take for every design exploration. In this regard, the differences between the single batch approach, standard augmentation, and the developed enhanced pdf-based augmentation have been examined. It is shown that the new enhanced method is superior to the both the other two methods both in two design dimensions $d = 6$ and $d = 12$

5. Acknowledgments

This work has been carried out in collaboration with Rolls-Royce Deutschland. Rolls-Royce Deutschland’s permission to publish this work is greatly acknowledged. The authors also would like to thank Franck Ngalamou and Clara Henkel for their contributions to the presented work.

6. Contact Author Email Address

mailto: flassigp@th-brandenburg.de, tamkan@th-brandenburg.de, marcus.meyer@rolls-royce.com,

7. Copyright Statement

References

[1] Hernandez S, Lucas T, Carlyle M. Constructing Nearly Orthogonal Latin Hypercubes for Any Nonsaturated Run-Variable Combination. *ACM Transactions on Modeling and Computer Simulation 22*, 20:1-20:17, 2012.

[2] Joseph R, Hung Y. Orthogonal-Maximin Latin Hypercube Designs. *Statistica Sinica* 18, pp 171-186, 2008.

[3] Crombecq K, Learmans T, Dhaene T. Efficient Space-filling and Non-Collapsing Sequential Design Strategies for Simulation-based Modelling. *European Journal of Operational Research,* 214, 2011.

[4] Hirsch C, Wunsch D. *Uncertainty Management for Robust Industrial Design in Aeronautics, Notes on Numerical Fluid Mechanics and Multidisciplinary Design* 140. Chapter: Latin Hypercube Sampling-Based Monte Carlo Simulation: Extension of the Sample Size and Correlation Control, Springer, 2019

[5] Flassig P, Bates A. A Flexible Strategy for Augmenting Design Points for Computer Experiments. *7th Dresdner Probabilistik Workshop,* Dresden, 2014

[6] Sobol I. On the Distribution of Points in a Cube and the Approximate Evaluation of Integrals. *USSR Computational Mathematics and Mathematical Physics,* Vol 7, Issue 4, pp 86-112, 1967

[7] Schaback R. Comparison of Radial Basis Function Interpolants. *Multivariate Approximation: from CAGD to Wavelets,* World Scientific, Singapore, pp 293-305, 1993