RESEARCH PAPER

Predictive modelling to support sensitivity analysis for robust design in aerospace engineering

Siva Krishna Dasari[1] · Abbas Cheddad[1] · Petter Andersson[2]

Abstract

The design of aircraft engines involves computationally expensive engineering simulations. One way to solve this problem is the use of response surface models to approximate the high-fidelity time-consuming simulations while reducing computational time. For a robust design, sensitivity analysis based on these models allows for the efficient study of uncertain variables' effect on system performance. The aim of this study is to support sensitivity analysis for a robust design in aerospace engineering. For this, an approach is presented in which random forests (RF) and multivariate adaptive regression splines (MARS) are explored to handle linear and non-linear response types for response surface modelling. Quantitative experiments are conducted to evaluate the predictive performance of these methods with Turbine Rear Structure (a component of aircraft) case study datasets for response surface modelling. Furthermore, to test these models' applicability to perform sensitivity analysis, experiments are conducted using mathematical test problems (linear and non-linear functions) and their results are presented. From the experimental investigations, it appears that RF fits better on non-linear functions compared with MARS, whereas MARS fits well on linear functions.

1 Introduction

The performance of a product may vary to a great extent depending on variations in design parameters, work environment, etc. Given these uncertainties, designers are highly interested in evaluating how to design a product while minimizing variations in its performance—this is referred to as robust design. For instance, to estimate how long time that product lasts under different environmental conditions. One of the tools for robust design is sensitivity analysis which is the study of the independent variables' influence on a dependent variable in the design space (Ma et al. 2015). This kind of analysis helps design engineers to enhance their knowledge about design variables and their interactions, and about the unknown underlying function that maps these design variables to an output variable. However, it is not feasible to study all possible variations in design variables because this analysis would involve enormous computational cost for running engineering simulations. Furthermore, the time is limited for design engineers to make decisions about design variables and design concepts. One way to solve this problem is to use response surface models—they are also known as surrogate models and metamodels—for sensitivity analysis (Ceylan et al. 2014; Chen et al. 2006; Ma et al. 2015; Sathyanarayanamurthy and Chinnam 2009; Taflanidis et al. 2011). The sensitivity analysis–based response surface models allow for efficient studies of how uncertainties in input variables affect system performance.

✉ Siva Krishna Dasari
siva.krishna.dasari@bth.se

Abbas Cheddad
abbas.cheddad@bth.se

Petter Andersson
petter.andersson@gknaerospace.com

[1] Department of Computer Science, Blekinge Institute of Technology, Karlskrona, 37141, Sweden

[2] Engineering Method Development, GKN Aerospace Engine Systems Sweden, Dept. 9635 - TL3, Trollhättan, SE-461 81, Sweden

The construction of response surface models—to mimic the complex behaviour of underlying simulation models—requires a dataset of inputs and known outputs. These known outputs are produced from simulations. Since simulations are expensive to conduct, the datasets are small in real-world context. From the literature studies, we have learned that popular methods such as polynomial models, Kriging, Radial basis functions and multivariate adaptive regression were studied to construct surrogate models for design problems, and their advantages and disadvantages have been highlighted in Ceylan et al. (2014); Huang et al. (2011); Sathyanarayanamurthy and Chinnam (2009); Shan and Wang (2009); and Wang and Shan (2007). However, as complexity of design problems increases, the challenges such as curse-of-dimensionality and dealing with highly non-linear design spaces remain the same. Furthermore, we have learned that tree models, random forests, can handle challenges which are high-dimensionality, small sample sizes and non-linearity in surrogate modelling. Also, in our previous published paper, we have compared random forests with support vector machines (SVM), linear regression and M5P methods (Dasari et al. 2015). In these comparisons, we observed that RF is capturing non-linearities much better compared with SVM, linear regression and M5P methods.

Although the RF method has been shown to be an accurate and a successful tool in other application tasks, the applicability of this method has been rarely investigated in aerospace design engineering for surrogate modelling, and we cite herein two of the studies found in Graening et al. (2008) and Dasari et al. (2015). Therefore, this paper opted RF to bridge this gap by investigating its performance to handle linear and non-linear datasets for response surface modelling. Furthermore, it has been indicated that extracting some knowledge from models can guide the design and optimization process (Graening et al. 2008). We believe that tree models can provide decision rules that can help design engineers for a better understanding of design parameters and space for further analysis. Also, we believe that rules extraction is an advantage of using tree models over other methods such as Kriging, support vector machines and artificial neural networks. However, the rules' extraction is out of the scope of this paper. Nevertheless, we demonstrated how to extract decision rules from RF models for the studied design tasks in our previous paper (Dasari et al. 2019).

In this study, we use response surface models to support sensitivity analysis for robust design of Turbine Rear Structure (TRS). In contrast to some previous studies (Sathyanarayanamurthy and Chinnam 2009; Taflanidis et al. 2011), we perform sensitivity analysis not only to analyze how sensitive the output variable is to variations of input variables but also to analyze the sensitivity spans of the output variable (more details are provided in Section 5.3). This paper provides contribution in the following aspect.

RF, which is rarely studied in response surface modelling, is selected in this paper for response surface modelling and sensitivity analysis. We selected MARS due to its popularity in response surface modelling as a benchmark method in order to determine the performance of RF. Furthermore, RF has been investigated on mathematical test functions that have been used in comparative studies literature for response surface modelling. From this investigations, we shared insights about RF to handle the linear and non-linear response types in order to support robust design in comparison with MARS.

We present aim and scope in Section 2, background in Section 3, related work in Section 4 and our approach to perform sensitivity analysis for robust design in Section 5. Section 6 states our experimental design. We present results and analysis in Section 7, discussions in Section 8 and conclusions in Section 9.

2 Aim and scope

This paper aims to support sensitivity analysis for robust design. For this, the paper focusses on sharing insights on how different response surface methods can contribute to model different types of characteristics of TRS data. For the investigations, we have used two datasets which describe two design studies of TRS where the response types exhibit linear and non-linear behaviour. The design and analysis that are needed for designing the TRS component are out of the scope of the paper. Hence, we do not focus on discussing details on how the calculations are done for design variables and objectives, for instance, welding life or FBO load cases. Instead, the purpose here is to investigate how should different response types of TRS analysis results be handled when building response surface models. For example, FBO analysis often provides non-linear data results that are not easily captured with low-degree polynomial order methods. For the construction of response surface models, we opted RF to explore its performance to handle linear and non-linear response types. In order to determine the performance of RF, we use the baseline method as MARS since it has been investigated widely for response surface modelling. Furthermore, mathematical test functions were studied for generalization of the selected methods to handle linear and non-linear response types and to perform sensitivity analysis to determine the performance of RF when comparing with MARS.

3 Background

In this section, we present response surface modelling briefly.

In response surface modelling, the aim is to determine a continuous function $\hat{f}$ (model) of a set of design variables $\mathbf{x} = x_1, x_2, ..., x_n$ from a limited amount of available data D. The available data D represents the exact evaluations of the function f, and in general cannot carry sufficient information to uniquely identify f. Thus, response surface modelling deals with two problems which are constructing a model $\hat{f}$ from the available data D and evaluating the error ε of the model (Mack et al. 2007).

The prediction of the simulation based model output using response surface modelling approach is formulated as follows:

$$f(\mathbf{x}) = \hat{f}(\mathbf{x}) + \varepsilon(\mathbf{x}) \tag{1}$$

where $\hat{f}(\mathbf{x})$ is the predicted output and $\varepsilon(\mathbf{x})$ is the error in the prediction. The construction of the response surface model involves several steps (shown in Fig. 1) (Queipo et al. 2005):

(1) Design of experiments: A set of design variables ($\mathbf{x} = (x_1, x_2, ..., x_d)^T$) and their values are selected.

(2) Numerical simulations: Let f be the black-box function (simulations) and evaluate f on design points $y_i = f(\mathbf{x}_i)$ where $\mathbf{x}_i \in \mathbb{R}^d$ and $y_i \in \mathbb{R}$.

(3) Construction of response surface model: Consider the data $D = \{(\mathbf{x}_1, y_1), ..., (\mathbf{x}_n, y_n)\}$; given the data, a continuous function $\hat{f}$ is determined to evaluate new design point $\hat{y} = f(\hat{x_i})$.

(4) Model validation: Assessing the predictive performance of $\hat{f}$ from the available data D.

In the machine learning terminology, response surface models can be referred to as supervised regression models or prediction models (Min et al. 2017). In engineering design, these prediction models can be used in different phases, for instance, sensitivity analysis and design optimization. This study focuses on response surface models for sensitivity analysis.

Sensitivity analysis allows the study of the behaviour of different design variables (inputs) influence on design objectives (output). Also, it can be used to identify least important design variables, variable prioritization and parameter interactions (Queipo et al. 2005). Several techniques, for instance, response surface models, Fourier amplitude sensitivity test, variance-based sensitivity analysis and Sobol technique, have been used for sensitivity analysis (Ma et al. 2015). In our study, we use response surface models to perform sensitivity analysis.

4 Related work

Several techniques have been used to construct response surface models for sensitivity analysis. Sathyanarayanamurthy and Chinnam (2009) constructed response surface models by support vector machine, radial basis function and Kriging for sensitivity analysis to determine the influence of input variables on the outcome variable using two test problems. The experimental results of this study indicate that Kriging is a suitable method for response surface modelling in the context of probabilistic engineering design— probabilistic engineering design deals with uncertainties in engineering analysis and design. Although Kriging has been shown as an accurate method, it is difficult to obtain and use because of its global optimization process that is used to identify the maximum likelihood estimators (Wang and
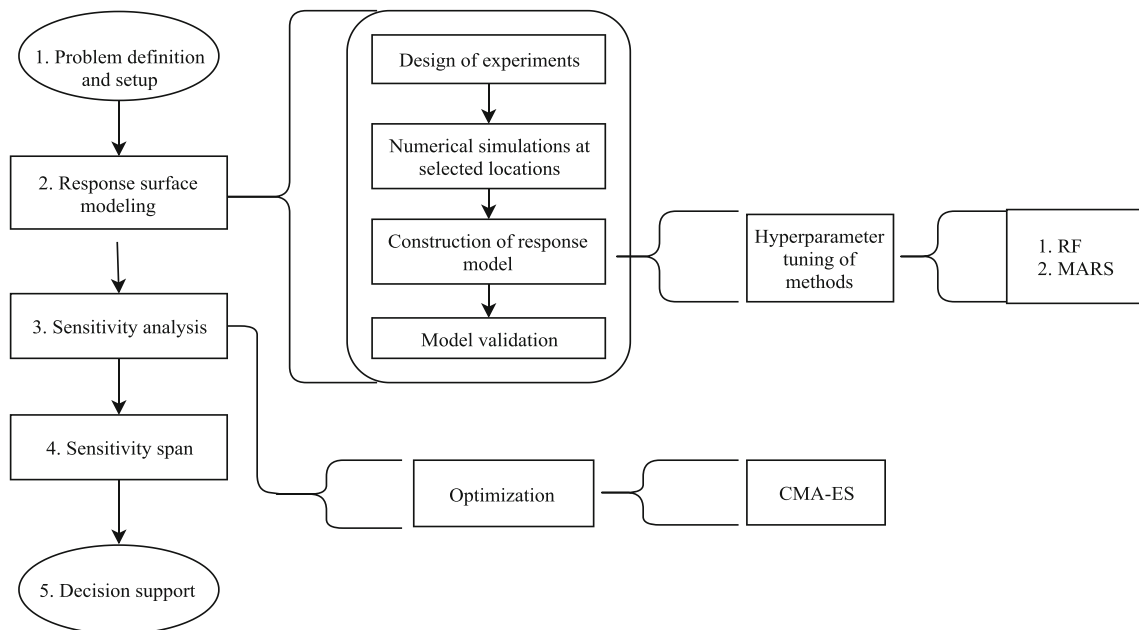

Fig. 1 An approach to perform sensitivity analysis using response surface modelling

Shan 2007). Furthermore, the investigations were limited to two-dimensional problems.

In another study, Taflanidis et al. (2011) constructed response surface model with moving least squares method to perform sensitivity analysis to determine the influence of input variables on the outcome variable for the robust design of offshore energy conversion devices . Recently, Ceylan et al. (2014) constructed response surface models by multivariate linear regressions and artificial neural networks (ANN) to perform sensitivity analysis for robust design of concrete pavement design, and the authors concluded that ANN is a robust and accurate method to capture the nonlinearities between variables. However, for our datasets, ANN did not perform well in our initial investigations; hence, we exclude the ANN method from this paper. We believe the reason might be due to the size of our datasets.

In contrast to these studies (Ceylan et al. (2014); Sathyanarayanamurthy and Chinnam (2009); Taflanidis et al. (2011)), we perform sensitivity analysis not only to analyze how sensitive the output variable is to variations of input variables but also to analyze the sensitivity spans of the output variable (more details are provided in Section 5.3). For the response surface models construction, we consider using RF and MARS as a baseline. Since both RF and MARS can handle non-linearities and high-dimensionality of data (Biau and Scornet 2016; Friedman 1991), we selected these methods to investigate their applicability to sensitivity analysis. Furthermore, we believe the hyperparameter tuning for these methods is easier than for Kriging and ANN.

5 An approach to perform sensitivity analysis using response surface modelling

In this section, we demonstrate our approach as shown in Fig. 1 to support sensitivity analysis for robust design of TRS through response surface modelling. First, we introduce our domain case study followed by mathematical test functions. Second, we describe briefly how response surface models are constructed for these functions with RF and MARS. Last, we present sensitivity analysis based on response surface models for our case study.

5.1 Problem definition and setup

5.1.1 A case study—studying thermal variations on TRS

We present our case study details where our goal is to support sensitivity analysis for robust design of TRS through response surface models (RSM).

TRS is located at the rear part of the aircraft engine that provides a support structure for low-pressure shaft and redirects exhaust flow from low-pressure turbine to the exit nozzle (Forslund et al. 2011). This is a challenging multidisciplinary problem that involves complex manufacturing solutions with high temperatures. Therefore, each aspect of this component needs to be studied to fulfill the design requirements and constraints.

We have 118 alternative design concepts that were used to investigate the design space of the TRS of an aircraft engine. Each design concept represents 27 design variables which includes five design variables of vane, three design variables of turbine hub, 15 different groups of thicknesses and four thermal variations T_1, T_2, T_3 and T_4. Simulations were conducted to understand functional behaviour and to predict possible failure modes in design concepts. The task in the TRS design is to understand how sensitive a design objective (for instance, welding life, stiffness and weight of the turbine vane) is to variations of four thermal variations T_1, T_2, T_3 and T_4. These four thermal variations are from the four zones which are nacelle, gas path, low-pressure turbine disc cavity and inner hub and plug as shown in Fig. 2. The right side of Fig. 2 shows the TRS with variations in thermal zones. We want to investigate what is the lowest and the highest predicted values of design objectives when we vary the four thermal variations. For this purpose, we use optimization using covariance matrix adaptation evolution strategy method. The highest and the lowest values of T_1, T_2, T_3 and T_4 will be used as the upper and the lower boundary for optimization. The gap between the highest and the lowest predicted value of a design objective is a number (span) that illustrates how sensitive a particular design concept is to variation of the temperatures T_1, T_2, T_3 and T_4. For instance, let us denote the minimum life of welding as Min_{life} and the maximum life of welding as Max_{life}, then the sensitivity span is the difference between Max_{life} and Min_{life} under T_1, T_2, T_3 and T_4 for welding life. The span values of welding life help to analyze the life of welding under different thermal variations for each design concept and analyze their impact on strength, aero performance and cost. The term cost is related to producibility assessment for business, for instance, the process time and costs. In welding manufacturability of TRS, the cost estimation is done by choosing price per weld length to model the costs. It is measured based on selected welding method, joint filler material, welding length, number of welds and average weld speed (Kveselys 2017). The cost assessment is out of the scope of this paper and the cost of TRS is studied separately in Heikkinen et al. (2016); Bertoni et al. (2018); and Kveselys (2017). The intention behind mentioning the cost is to merely give an overall idea that each design concept could be analyzed in different aspects, for instance, mechanical and aerodynamics performance to be balanced with producibility aspects.
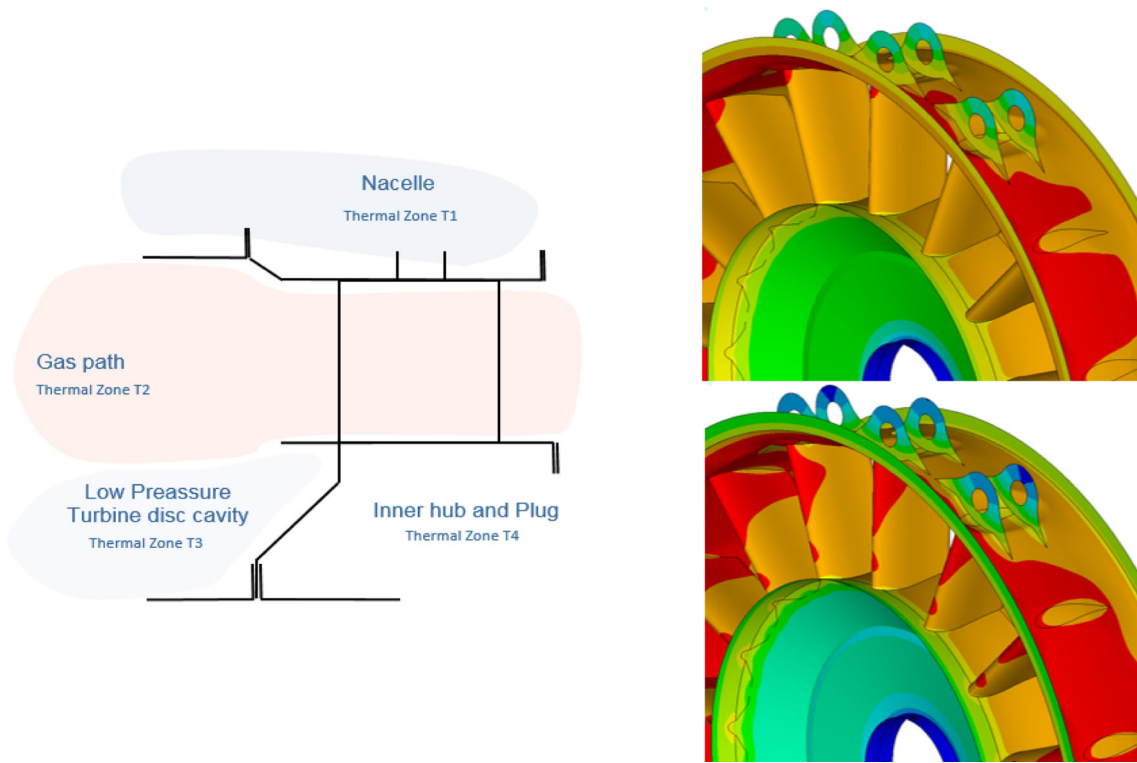
Fig. 2 Thermal variations in four thermal zones (left side), and their studied influence on TRS (right side)

The task is to get the optimal minimum and maximum predictions of design objectives of the TRS datasets to calculate sensitivity span (Fig. 3, explained in Section 5.3) using sensitivity analysis based on RSMs for

decision-making to help design engineers in their assessment criteria. For this, we conduct two experiments. In experiment 1, we use our TRS case study datasets to construct response surface models with RF and MARS (more
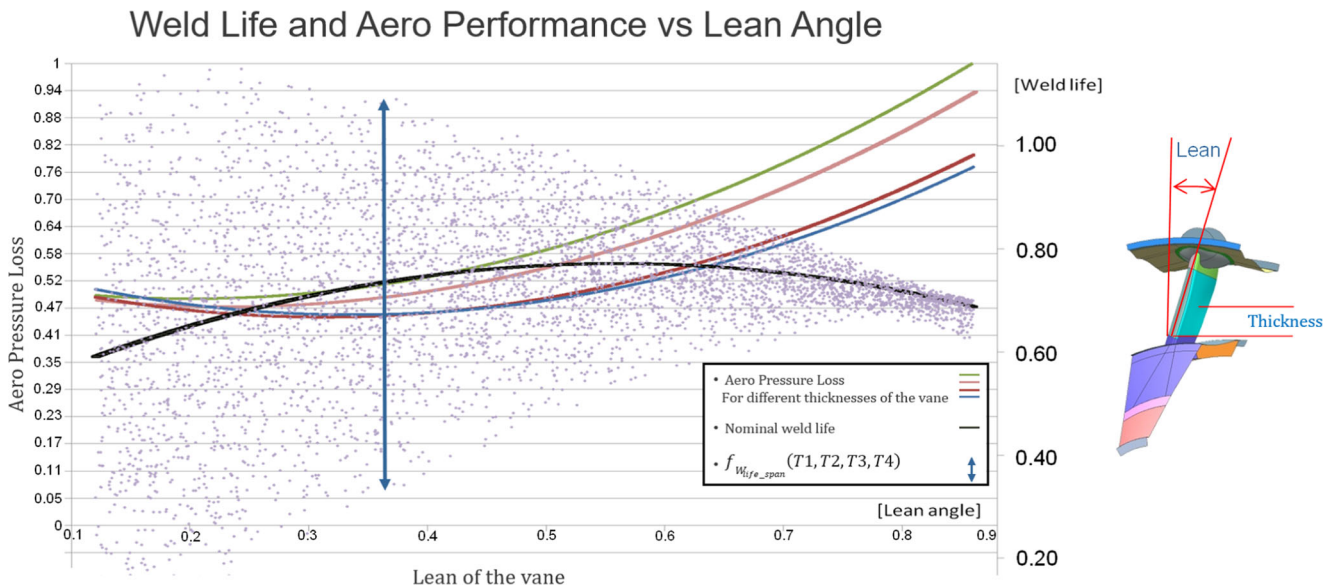

Fig. 3 Sensitivity analysis: varying four thermal variations for welding life, and analyzing its results with aero pressure loss for four thicknesses of vane and nominal welding life with respect to lean of the vane

details about the datasets are provided in Section 6.1). The next step is to perform sensitivity analysis to study thermal variations. However, we do not have the ground truth in our case study to evaluate the results from RF and MARS for sensitivity analysis to determine which estimation of sensitivity span is the closest to the ground truth.

Hence, we demonstrated the sensitivity analysis to determine the performance of the selected methods in experiment 2 with mathematical test problems. For this, we use 8 non-linear mathematical functions (termed herein, NL1, NL2, NL3... NL8), which are used to study the behaviour of different response surface modelling methods (Hock and Schittkowski 1980; Jin et al. 2001; Zhao and Xue 2010; Song et al. 2018). Furthermore, we have both linear and non-linear response types our domain case. Therefore, we selected two linear mathematical functions (L9 and L10) from Hock and Schittkowski (1980). In the following subsection, we present these 10 mathematical problems.

5.1.2 Mathematical test functions

The eight non-linear mathematical functions and the two linear mathematical functions (L9 and L10) are shown in Table 1.

5.2 Response surface modelling using RF and MARS

We used RF and MARS to construct response surface models (step 2 in Fig. 1). The details of these methods can be found in Breiman (2001); Friedman (1991); and Liaw and Wiener (2002).

Root mean squared error (RMSE) is used to measure the predictive performance of RF and MARS. Previous studies show that hyperparameter tuning for methods could improve their performance (D'iaz-Uriarte and De Andres 2006; Duroux and Scornet 2018; Probst et al. 1804). Thus, we conducted hyperparameter tuning for both RF and MARS using the following procedure (shown also in Fig. 4).

1. Set configuration values for the selected hyperparameters (for instance, K is a parameter for RF). More details are explained in Section 6.
2. Generate possible configurations by varying the hyperparameters' values from step 1.
3. Construct models for the problem of interest using the selected methods (RF and MARS).
4. Repeat step 3 for each possible hyperparameter configuration value.
5. Evaluate the model performance for each hyperparameter configuration value.
6. Select the model with the least RMSE.

We repeat the above procedure for every mathematical function.

Table 1 Mathematical test functions

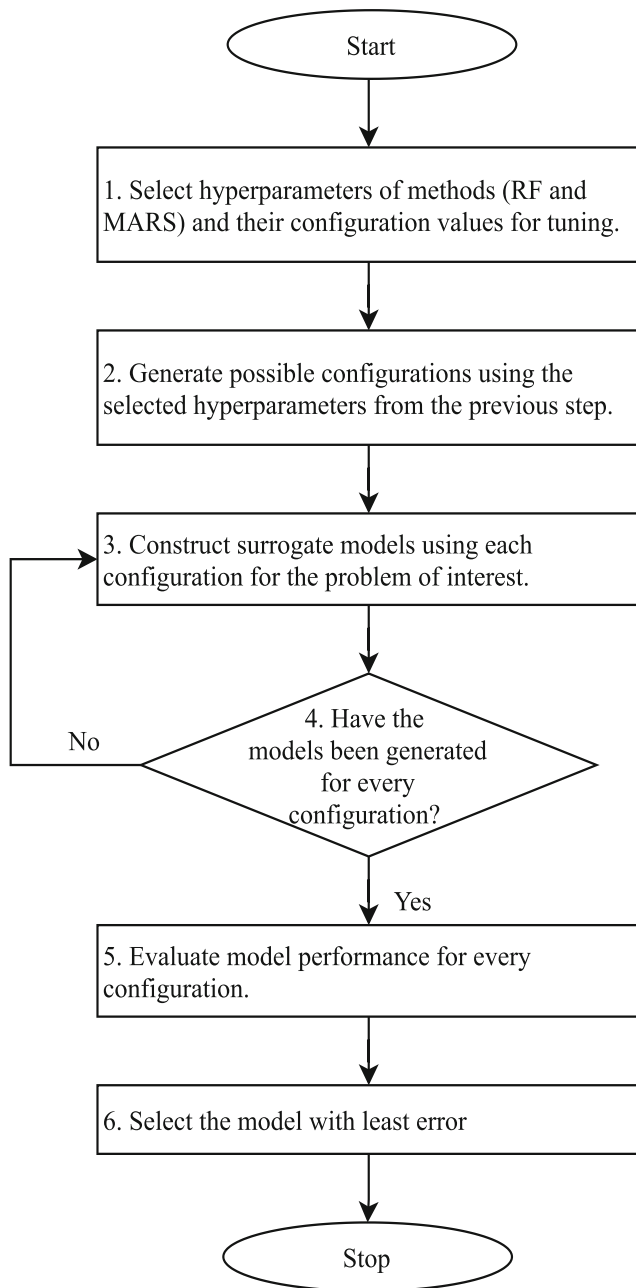| P.no : mathematical functions	Bounds
$NL1 : f(x) = (x_1 - 10)^2 + 5(x_2 - 12)^2 + x_3^4 + 3(x_4 - 11)^2 + 10x_5^6 + 7x_6^2 + x_7^4 - 4x_6 x_7 - 10x_6 - 8x_7$	$[0, 1]^D$
$NL2 : f(x) = x_1^2 + x_2^2 + (x_1 x_2) - 14x_1 - 16x_2 + (x_3 - 10)^2 + 4(x_4 - 5)^2 + (x_5 - 3)^2 + 2(x_6 - 1)^2 + 5x_7^2 + 7(x_8 - 11)^2 + 2(x_9 - 10)^2$ $+ (x_{10} - 7)^2 + 45$	$[0, 1]^D$
$NL3 : f(x) = sin(\pi x / 12) cos(\pi y / 16)$	$[-10, 10; -20, 20]$
$NL4 : f(x) = sin(x + y) + (x - y)^2 - 1.5x + 2.5y + 1$	$[-1.5, 4; -3, 3]$
$NL5 : f(x) = 0.5x_1^2 + x_2^2 - x_1 x_2 - 7x_1 - 7x_2$	$[-1, 1]^D$
$NL6 : f(x) = [30 + (x_1 sin(x_1)][4 + exp(-x_2^2)]$	$[-2, 2]^D$
$NL7 : f(x) = (6x_1 - 2)^2 sin[2(6x_1 - 2)]$	$[0, 1]^D$
$NL8 : f(x) = \frac{sin(10\pi x_1)}{2x_1} + (x_1 - 1)^4$	$[0.5, 2.5]^D$
$L9 : f(x) = 4.3x_1 + 31.8x_2 + 63.3x_3 + 15.8x_4 + 68.5x_5 + 4.7x_6$	$[0, 1]^D$
$L10 : f(x) = 24.55x_1 + 26.75x_2 + 39x_3 + 40.50x_4$	$[0, 1]^D$

Start

1. Select hyperparameters of methods (RF and MARS) and their configuration values for tuning.

2. Generate possible configurations using the selected hyperparameters from the previous step.

3. Construct surrogate models using each configuration for the problem of interest.

4. Have the models been generated for every configuration?

No

Yes

5. Evaluate model performance for every configuration.

6. Select the model with least error

Stop

Fig. 4 Hyperparameter tuning as presented in Section 5.2

5.3 Sensitivity analysis based on response surface models

In this section, we present the task of sensitivity analysis for one design objective, *weld life*, and the same procedure is used for every design objective.

Consider the available data D as a set of design concepts $D = \{c_1, c_2..., c_n\}$. Each design concept (an instance in machine learning terminology) represents a set of design variables ($\mathbf{x} = (x_1, x_2, ..., x_d)^T$) and a design objective (y) where $\mathbf{x} \in \mathbb{R}^d$ and $y \in \mathbb{R}$. Given these design variables

and design objective, a continuous function $\hat{f}$ is determined using RF and MARS to predict the outcome $\hat{y}$ for new design concepts where $\hat{y} \in \mathbb{R}$. Later, this function $\hat{f}$ is used for sensitivity analysis.

In Fig. 3, the *x*-axis represents the *lean of the vane* which is an independent variable and the *y*-axis represents the *weld life*. The welding life is one of the outputs of TRS design study and we focus on modelling the non-linearities of this output in order to get accurate predictions during sensitivity analysis. Hence, the purpose is to investigate how the selected response surface methods handle these non-linearities; thus, we do not include any details on how welding life is calculated.

To perform sensitivity analysis, for each value of *lean of the vane* (let us call it x_1), we vary the value of uncertain variables T_1, T_2, T_3 and T_4 (let us call them x_2, x_3, x_4 and x_5), while keeping the remaining x_n, $n \in \{6, ..., d\}$ fixed, to analyze the *weld life* (let us call it $\hat{y}$). The function $\hat{f}$ is used to predict the outcome $\hat{y}$ for $\mathbf{x}$. Instead of only predicting the outcome $\hat{y}$, we predict $\hat{y}_{\min}$ and $\hat{y}_{\max}$ given variation in $\mathbf{x}$. These predictions are shown in Fig. 3 in which every dot represents a prediction (simulated design) based on $\hat{f}$ (surrogate model determined by RF and MARS). We call the difference between $\hat{y}_{\min}$ and $\hat{y}_{\max}$ the sensitivity span as shown in Fig. 3 (the vertical double arrow line). On the right hand side of Fig. 3, the lean and thickness of vane are shown. The axes are normalized due to the confidentiality of data.

$$f_{\text{Wlife-span}}(T_1, T_2, T_3, T_4) = \hat{y}_{\max} - \hat{y}_{\min} \qquad (2)$$

Together with tuned welding results, aero pressure loss for different thickness of the vane and nominal weld without tuning thermal variations also added to Fig. 3. This type of analysis help design engineers for life assessment of the product and its impact on performance. For instance, one observation from this analysis is that the span is decreasing when the angle of lean is increasing. In this case, the objective is to maximize the sensitivity span of design concepts that satisfies the design constraints and requirements. What we seek is not only selecting the design concept that has the highest span but also balancing with aero pressure loss for different thickness groups, and other design requirements. The idea is to get a better understanding of design parameters and their effects on system performance, and to make informed decisions about them. Furthermore, it could help (1) to identify interesting regions in design space which can be further explored and (2) to identify which design variables cause higher or lower sensitivity span.

Optimization is performed in order to tune as many variations of the uncertain variables as needed to search for the minimum and maximum predictions for a selected output. For this purpose, we use the covariance matrix

adaptation evolution strategy (CMAES) as an optimization method—which is a powerful evolutionary algorithm for global optimization (Hansen and Ostermeier 2001). The CMAES method is a non-linear optimization technique and is based on evolutionary strategy to find the optimum solution. In this method, the candidate solutions are samples from multivariate normal distribution. Furthermore, the convergence of this method is based on covariance matrix which is used to update the rule for this algorithm to determine the best way to proceed to the next iterations. Here, the covariance matrix represents the pairwise dependencies between the variables. We used CMAES optimization technique to find the optimal predictions. However, the focus is not on optimization algorithms instead we used a non-linear optimization technique to find the optimal predictions when performing sensitivity analysis using response surface models. Initially, we have used another method called the multi-level single linkage (MLSL) which is a stochastic optimization method and compared it with CMAES. Then, we observed that both methods can be used for our purpose and we choose CMAES as it is computationally faster than MLSL. Furthermore, the CMAES method has been used for model based optimization using RF response surface models in Preuss et al. (2012). Hence, we believe this method can be used for our purpose to find the optimal predictions.

6 Experimental design

The aim of the experiments is to determine which method (RF or MARS) performs better to support sensitivity analysis in engineering design. For this, we conducted two experiments. In experiment 1, we evaluate the performance of RF and MARS for response modelling using our case study datasets. In experiment 2, we evaluate the results from sensitivity analysis to determine which estimation of sensitivity span is the closest to the ground truth using 10 mathematical functions (explained in Section 5.1.2). This section describes our experimental design which includes datasets description, parameter and configuration selection, evaluation procedure and performance measures.

6.1 Dataset descriptions

Datasets for experiment 1

In experiment 1, two datasets were collected from a major aerospace sub-system manufacturer to investigate how RF and MARS can contribute to model characteristics of TRS data, specifically, the linear and non-linear response types. These datasets are two design studies of TRS where
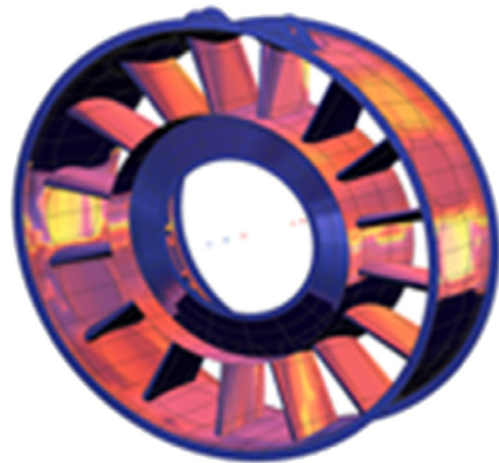

Fig. 5 TRS case

the design space was investigated by design engineers. In the following, we explain these datasets in detail.

D1 is a dataset with 118 alternative design concepts that were used to investigate the design space of the TRS. Each design concept contains 27 design variables which includes:

– Temperatures: There are 4 temperature zones accounted for in the component. The temperatures in these zones are then varied to enable a sensitivity analysis. These four temperature zones are shown in Fig. 2.
– Thickness parameters: The TRS is divided into 15 groups/zones where the thickness is adjusted simultaneously. It is important to choose these zones carefully to avoid the "weakest point of a chain" situation. Figure 5 shows the TRS case.
– Vane-related parameters: The following are the parameter related to the vane which are also shown in Fig. 6.

  1. Axial distance of the coord at the vane hub intersection
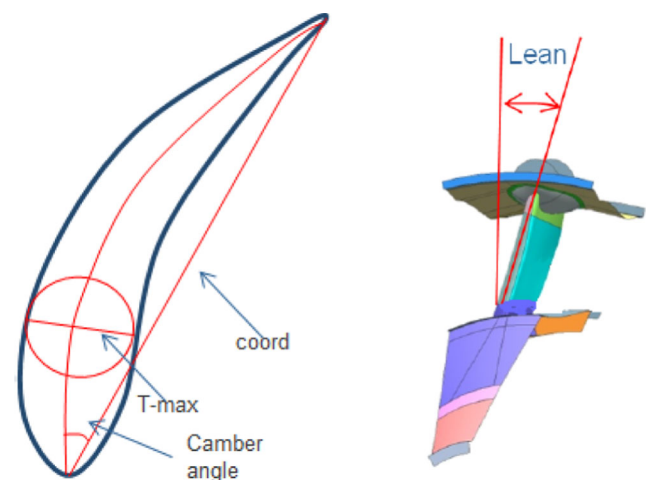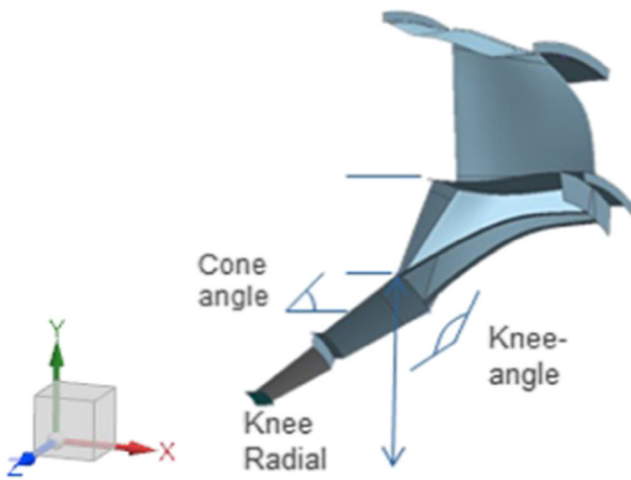

Fig. 6 Parameters related to the vane

Fig. 7 Parameters related to the hub in dataset D1

   2. Axial distance of the coord at the vane shroud intersection
   3. Leading Edge Camber angle
   4. Max thickness of the vane at the vane hub intersection
   5. Lean (the leaning of the vane).

- Hub-related parameters: Figure 7 illustrates the hub parameters which are the hub cone angle, the hub knee radial position and the hub cone angle.

D2 is another study result of TRS with 49 alternative design concepts where we have 14 thickness groups of TRS, 4 temperatures, struts where the number of vanes were varied as shown in an example in Fig. 8 and the lean of the vane is shown in Fig. 6 (right hand side).

Furthermore, hub rear stiffness is added together with the hub cone angle and the hub knee radial position as shown in Fig. 9. In both Figs. 7 and 9, the *x*-axis is the engine axis FLA (Forward Looking Aft), and the knee radial distance—which is the distance from the knee point to the engine axis—lies on the YZ plane (equally around the circumferential geometry).

We have 12 output variables for dataset D1 and 7 output variables for dataset D2. The design objectives (outputs) of D1 and D2 are related to aero performance (swirl: a measurement for how much the air rotates when it exits the engine and pressure loss: a measurement for aero performance), mechanical functions (buckling: fan blade out load case and over turning moment) and welding life of turbine vane design. We have two independent models using RF and MARS for each output variable. Thus, we have a total of 19 single output datasets for experiments. We named them as D1-1, D1-2 ... D2-19. The details (type, dataset, description, response type) of these datasets are shown in Table 2. Our domain expert provided the response type for these datasets.

Datasets for experiment 2

Different sample sizes are reported for the selected test problems in Jin et al. (2001) and Zhao and Xue (2010). However, we choose 410 sample set to train models since that is the highest dataset size we have in our application. To evaluate the performance of these models, we selected 1000 samples as similar to the study of Jin et al. (2001). We used these constructed response surface models to perform sensitivity analysis using 100 samples.

6.2 Hyperparameter configurations

For RF, we selected F (the number of random features) and K (the number of trees) hyperparameters to tune. The motivation is a previous study states that increasing the size of the tree (K) can decrease the forest error rate, and decreasing the number of random features (F) can reduce the strength of a tree (increases the error rate) (Oshiro et al. 2012; Dasari et al. 2019). Also, a threshold range is provided for K which is 1 to 128 to increase accuracy for classification tasks (Oshiro et al. 2012). We also selected similar range of K configuration values between 10 and 130 with a step size of 10. We select F configuration values from the minimum to the maximum number of features from datasets with a step size of 1. According to the R documentation for RF that we use, F is *mtry* and K is *ntree*. For MARS, we selected degree d for parameter interactions. The default value for this parameter is 1. The suitable value for d is in the range $2 \leq d \leq 4$ (Friedman 1991). Thus, we selected the d values from 1 to 4 with a step size of 1.

6.3 Evaluation procedure

We used cross-validation to maximize training set size and to avoid testing on training data (Kohavi et al. 1995). The procedure of cross-validation is as follows: a dataset is divided into k sub-samples. A single sub-sample is chosen as testing data and the remaining $k - 1$ sub-samples are used as training data. The procedure is repeated k times, in which each of the k sub-samples is used exactly once as testing data. Finally, all the results are averaged and a single estimation is provided. In our experiments, we choose $k = 10$ for all the datasets. We used R software packages[1,2,3] to conduct the experiments.

6.4 Performance measures

Predictive accuracy: we used RMSE to measure the predictive performance of RF and MARS.

---

[1]Optimization: https://cran.r-project.org/web/packages/cmaes
[2]MARS: https://cran.r-project.org/web/packages/earth/earth.pdf
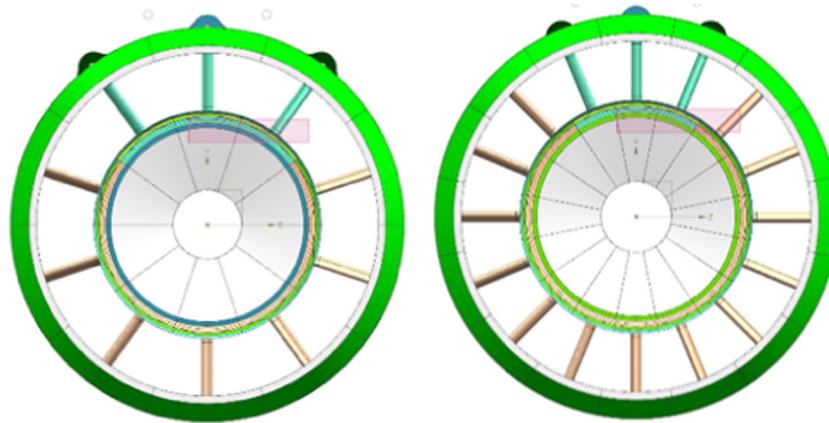[3]RF: https://cran.r-project.org/web/packages/randomForest

Fig. 8 Strut parameter

Sensitivity analysis: we used RMSE to measure the performance of RF and MARS for sensitivity analysis. The RMSE is calculated using the estimated sensitivity span—which is calculated from the optimal minimum and maximum predictions from the optimization using RF and MARS—and the ground-truth sensitivity span. The ground-truth sensitivity is also a span from the optimal minimum and maximum values from the optimization using mathematical test functions (shown in Section 5.1.2).

6.5 Experiment setup

We used the following procedure to perform sensitivity analysis in experiment 2. For experiment 1, the first and second steps are used to construct response models using our domain datasets. The below experiment setup is also shown in Fig. 10.

1. Construct response surface models with RF and MARS using hyperparameters tuning for each function (shown in Section 5.1.2).
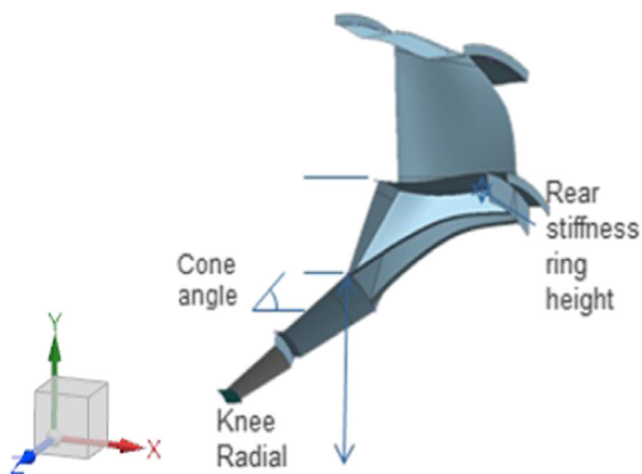2. Evaluate the predictive performance of response surface model (RF and MARS) using RMSE.

Fig. 9 Parameters related to the hub in dataset D2

3. Use the constructed response surface model as an input function for the optimization
4. Use the highest and the lowest values of the uncertain variables that exist in the data set as the upper and the lower boundaries for the optimization study (explained in Section 5.3). We set the maximum number of function evaluations as 1000 in the optimization.
5. Run the optimization for a concept (an instance) by varying the selected uncertain variables and fixing the rest of variables in the dataset.
6. Get minimum and maximum prediction for an output, and calculate the difference as a sensitivity span.
7. Repeat steps 3 to 6 for every concept in the dataset.

7 Results and analysis

7.1 Experiment 1

In this section, we present the case study results of RSM obtained from the experiment using RF and MARS. Table 2 shows the normalized RMSE of the two methods. We ranked the performance of each method by 1 or 2 depending on the RMSE results (rank 1 for low RMSE and 2 for high). The last two rows of Table 2 show the average rank of each method for linear and non-linear response types. The results show that RF yields the lowest error for 12 datasets, and MARS yields the lowest error for 7 datasets. The average rank of the linear response datasets shows that MARS (1.4) performs better than RF (1.6). The average rank of non-linear response datasets shows that RF (1.28) performs better than MARS (1.71).

7.2 Experiment 2

Since sensitivity analysis is performed using response surface models, we first evaluated the performance of response surface models. Later, we evaluated the results
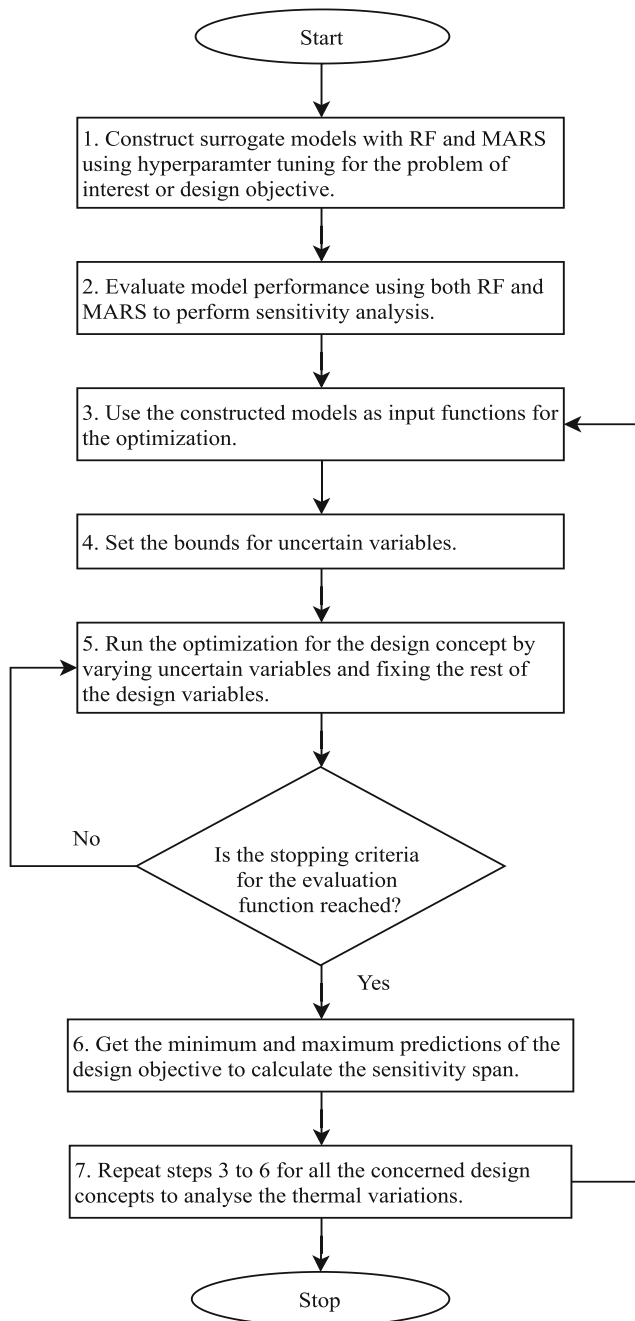
Fig. 10 Experiment setup as presented in Section 6.5

from sensitivity analysis to determine which estimation (either RF or MARS) of sensitivity span—which is calculated from optimal predictions of RF and MARS response surface models—gets closest to the ground truth.

Predictive accuracy of response surface models Table 3 shows the normalized RMSE. As shown in Table 3, for the non-linear functions, RF has a higher predictive accuracy compared with MARS on average. For linear problems (L9 and L10), MARS has higher predictive accuracy compared with RF.

Sensitivity analysis Table 4 shows the results from sensitivity analysis (normalized RMSE). The RMSE is calculated from the sensitivity span by RF and MARS, and the ground-truth sensitivity span. We started analyzing results with a hypothesis that the response surface model with a higher predictive accuracy (from Table 3) will lead to a lower RMSE which is calculated from sensitivity span results. As it is shown in Table 4, RF has lower RMSE for non-linear problems NL3, NL4, NL6, NL7 and NL8, whereas MARS has lower RMSE for NL1, NL2 and NL5. Regarding the linear problems L9 and L10, the results show that MARS yields a lower RMSE, and it also has a higher prediction accuracy (Table 3). Thus, the result from sensitivity analysis supports our hypothesis that the model with higher prediction accuracy (Table 3) will lead to lower RMSE of sensitivity span (Table 4) except the case of NL5. We believe that it is due to low order of non-linearity in NL5.

7.3 Analysis

Figure 11 shows plots for RF and MARS and analytical models for NL 3, NL4, NL7 and NL8, and we can observe that RF has captured the actual function better than MARS. We observed that all these four test problems have periodic functions (cosine or sine) as shown in Section 5.1.2. One of these four problems, NL3, was studied in a study by Jin et al. (2001), and the authors showed grid plots for radial basis function, Kriging, MARS, polynomial regression (PR) and analytical function. If we compare our plots with their study plots, we can observe that RF has captured the function behaviour similar to Kriging and radial basis function methods, and better than PR and MARS. Also from Fig. 11, we observed that the optimal minimum and maximum that RF has produced is the closest to the analytical functions. Consequently, RF also yielded the span closest to the span of analytical functions.

On the other hand, MARS captures the 2D function behaviour for NL7 and NL8 as rigid segments, however, not as smooth as RF. The reason is that MARS has hinge functions with knots to fit non-linear models. These hinge functions are formed from piece-wise linear functions. This might be the reason that the curve looks linear from one hinge function to another and together they form a non-linear function. For linear problems L9 and L10, MARS has shown the best performance. This is expected since MARS is an extension of linear models and in the way it fits the data. We also observed from MARS model (equation) that MARS has similar behaviour as the actual function (L9 and L10). Furthermore, MARS also yields the best performance for NL1, and we observed the equation from MARS model that it captured the behaviour of NL1 well. NL1 is a polynomial function with degree 6; hence, we selected another problem, NL2, with a quadratic polynomial

Table 2 RSM performance comparison on 19 datasets: Normalized RMSE (rank)

Type	Dataset	Description	ResponseType	RF	MARS
Aero performance	D1-1	Swirl	Linear	0.9367 (1)	1.1571 (2)
	D1-2		Linear	1.1387 (2)	0.6814 (1)
	D1-3	Pressure loss	Linear	0.9540 (1)	1.1086 (2)
	D1-4		Non-linear	0.9317 (2)	0.5814 (1)
Mechanical	D1-5	Buckling - fan blade out load case	Non-linear	0.8867 (1)	1.0163 (2)
	D1-6		Non-linear	0.9972 (2)	0.9539 (1)
	D2-7		Non-linear	0.9065 (1)	0.9207 (2)
	D2-8		Non-linear	0.9222 (1)	1.0241 (2)
	D1-9	Over turning moment	Non-linear	1.1425 (2)	0.5971 (1)
	D2-10		Linear	1.0535 (2)	0.8012 (1)
	D2-11		Linear	1.0516 (2)	0.7116 (1)
Life	D1-12	Welding life	Non-linear	1.0809 (2)	0.8942 (1)
	D1-13		Non-linear	0.9849 (1)	1.0933 (2)
	D1-14		Non-linear	0.9366 (1)	1.1025 (2)
	D1-15		Non-linear	1.0165 (1)	1.0315 (2)
	D1-16		Non-linear	0.9758 (1)	1.0671 (2)
	D2-17		Non-linear	0.7125 (1)	0.8935 (2)
	D2-18		Non-linear	0.9441 (1)	1.0835 (2)
	D2-19		Non-linear	0.9746 (1)	0.9923 (2)
			lAvg.rank (L)	1.6	*1.4*
			lAvg.rank (NL)	*1.28*	1.71

Italic entries highlight the method with the lowest RMSE average

degree to analyze if MARS retains its performance. The results of NL2 show MARS performed best for NL2. It seems that MARS is better at fitting polynomial functions; however, we only studied two polynomial functions.

Sensitivity analysis For this, we have plotted the optimal min predictions for a linear and a non-linear response types which are the swirl and the welding life, respectively. These predictions were obtained when performing sensitivity

Table 3 RSM performance comparison for 10 test problems

P.no	RF	MARS
NL1	1.8022 (2)	*0.1978 (1)*
NL2	1.5674 (2)	*0.4326 (1)*
NL3	*0.1643 (1)*	1.8357 (2)
NL4	*0.1849 (1)*	1.8151 (2)
NL5	*0.8622 (1)*	1.1378 (2)
NL6	*0.4477 (1)*	1.5523 (2)
NL7	*0.1830 (1)*	1.8170 (2)
NL8	*0.1260 (1)*	1.8740 (2)
L9	1.4754 (2)	*0.5246 (1)*
L10	2.0000 (2)	*4.03E-14 (1)*
Avg.rank(NL)	*1.28*	1.75
Avg.rank(L)	2	*1*

Italic entries highlight the method with the lowest RMSE average

Table 4 Sensitivity analysis—RMSE calculated from the sensitivity span by RF and MARS, and the ground-truth sensitivity span derived from mathematical test functions

P.no	RF	MARS
NL1	1.8542 (2)	*0.1458 (1)*
NL2	1.4758 (2)	*0.5242 (1)*
NL3	*0.1255 (1)*	1.8745 (2)
NL4	0.4893 (1)	1.5107 (2)
NL5	*1.9803 (2)*	*0.0197 (1)*
NL6	*0.3052 (1)*	1.6948 (2)
NL7	*0.1143 (1)*	1.8857 (2)
NL8	*0.1759 (1)*	1.8241 (2)
L9	2.0000 (2)	*8.10E-11 (1)*
L10	2.0000 (2)	*3.20E-10 (1)*
Avg.rank(NL)	*1.37*	1.62
Avg.rank(L)	2	*1*

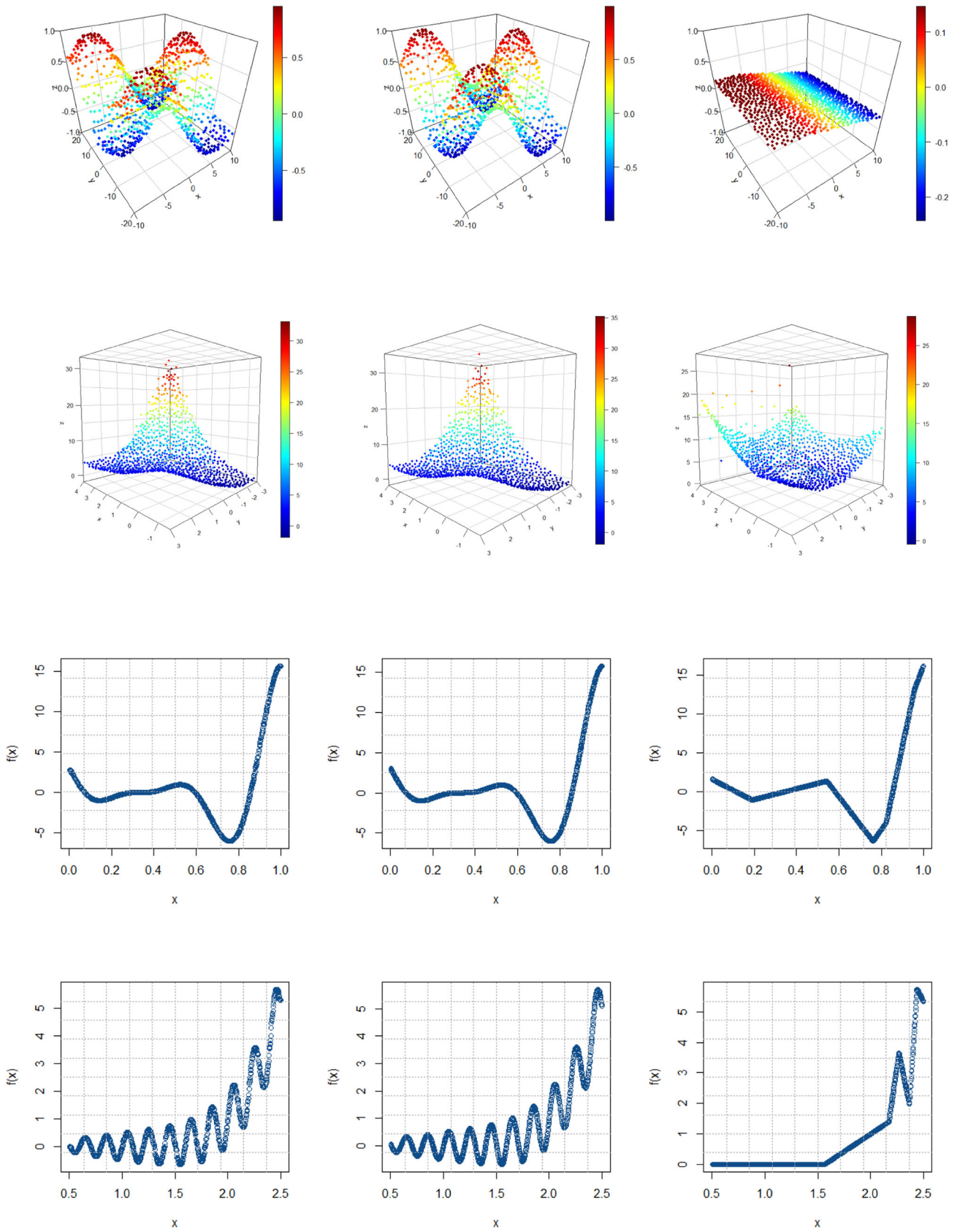Italic entries highlight the method with the lowest RMSE average

Fig. 11 Predictions are plotted in 3D and 2D for NL3, NL4, NL7 and NL8 in rows, respectively. First column: RF, middle column: analytical function (ground truth) and last column: MARS

analysis based on response surface models. The purpose of these plots is to analyze the optimal predictions of RF in comparison with MARS. Figure 12 shows plots for welding life (D1-14) and swirl (D1-2) as shown in Table 2. The RF optimal prediction was denoted as **1** in the *x*-axis and MARS was denoted as **2** in the *x*-axis in both plots in Fig. 12. The *y*-axis shows the normalized predictions for the welding life and the swirl. By examining Fig. 12 (left hand side), we can see that RF provides better optimal prediction for welding life (non-linear output) compared with MARS. Also for this output, RF response model has lower error as shown in Table 2, whereas MARS has lower error for swirl (linear type output); consequently, MARS provides better optimal prediction for the swirl output as shown in Fig. 12 (right hand side).

Statistical analysis We also performed statistical analysis to see the performance differences between RF and MARS for response surface modelling using RMSE results of Tables 2 and 3. For this purpose, we used the Mann-Whitney-Wilcoxon signed-rank test (Sheskin 2003). The Wilcoxon signed-rank test is a non-parametric statistical test that can be used to measure performance differences between two independent groups (Sheskin 2003). The statistical hypothesis which we set is:

H_o: *RF and MARS methods perform equally well with respect to predictive performance.*

H_a: *There is a significant difference between the performance of the methods.*

The statistical test produces a *p* value of 0.7819. The *p* value is greater than the 0.05 significance level. We therefore cannot reject the null hypothesis; hence, both RF and MARS perform equally well with respect to predictive performance.
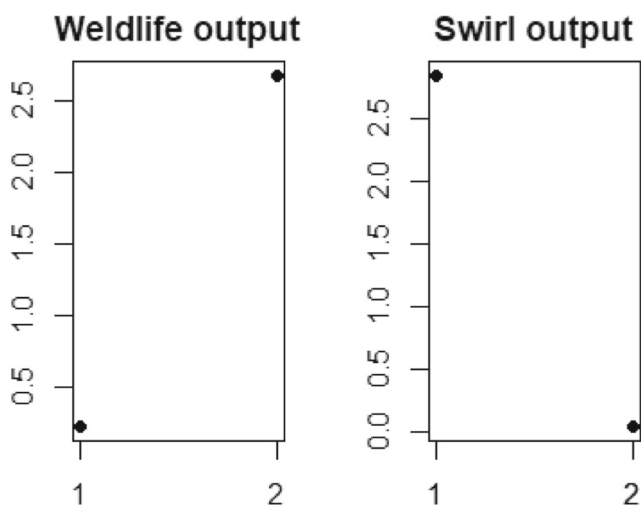

Fig. 12 SA: optimal predictions from RF (indicates 1 in *x*-axis) and MARS (indicates 2 in *x*-axis)

8 Discussions

In this section, we present our discussions around the results of RF and MARS. One must consider that our observations about these methods are based on the 10 test functions and our case study datasets that we studied. Since we studied linear and non-linear response types, the evaluation results of this study could help to choose either RF or MARS for those type of problems.

Experiment 1 focussed on response modelling using RF and MARS with TRS datasets. The results indicate that RF performed well on non-linear response type of problems and MARS performed well on linear response type on average. Experiment 2 focussed also on response modelling using RF and MARS but with 10 test mathematical functions (both linear and non-linear). From the results of experiment 2, we have similar observations as in experiment 1, that is RF performs well on non-linear problems and MARS has a better performance on linear problems. Hence, experiment 2 supports rxperiment 1's results of response modelling. Therefore, these observations about RF and MARS can be generalized to some extent. Furthermore, experiment 2 focussed on sensitivity analysis and it shows that RF performs well on non-linear problems and MARS performs well on linear problems.

Regarding MARS: Since we know the response types (four mathematical functions) beforehand and since we know that MARS is an extension of linear models, we expected that MARS performs well for the linear responses. As we expected, the results indicate that MARS has a higher prediction accuracy for the linear responses on average. Although MARS also models the non-linearities between variables, the result indicates that it could not outperformed RF. However, it does perform well for NL2 and it has low non-linearity where the polynomial degree is 2 and also 6 for NL1.

Regarding RF: As stated in the literature that RF is capable of handling non-linearities, it does show better performance compared with MARS. Especially in fitting smooth functions as shown in Fig. 11, RF shows the best fit. Furthermore, we compared our plots for NL3 to the study by Jin et al.; it seems RF fits the function similar to Kriging and radial basis function. Thus, RF can be suggested as an alternative method to construct response surface models in engineering design.

Hyperparameter tuning: As previous studies show that hyperparameter tuning could improve predictive accuracy, we performed such experiments. We observed that the performance of RF and MARS is improved by tuning their hyperparameters. Furthermore, we observed that the optimal hyperparameter configurations are different for each problem that we studied. Hence, we can recommend our hyperparameter configuration settings to tune for both RF and MARS (shown in Section 6.2). For a large dataset or

a large number of trees, RF can take quite a bit of memory (more execution time) (Liaw and Wiener 2002).

Other advantages of both methods: Both methods can give variable importance which explains the predictor variable influence on a dependent variable. This variable importance could be used for screening. Screening identifies relevant input variables and removes less important variables in the problems of interest. It reduces the dimensionality of the problems which contributes to saving computational cost (Shan and Wang 2010). Furthermore, RF not only provides variable importance but also provides human understandable decision rules using the *InTrees* framework (Deng 2014). The *InTrees* framework extracts information from the trees in the form of rules, and prunes redundant rules and leaves the non-redundant rules in the forest. Those rules may provide some information to engineers regarding the prediction procedure or relationship between input and output variables. Both the screening and the rule extraction are out of scope of this study.

9 Conclusions and future work

We explored the use of response surface models to support sensitivity analysis for robust design of Turbine Rear Structure (TRS). For this, we investigated the applicability of random forests (RF) and multivariate adaptive regression splines (MARS) for TRS case study and for 10 test mathematical functions (both linear and non-linear). We presented our approach to support sensitivity analysis, and we conducted experiments to determine which method estimation (either RF or MARS) of sensitivity span is the closest to the ground truth. We used root mean squared error (RMSE) to evaluate the performance of RF and MARS. From the experimental analysis, we observed that RF is better at capturing the non-linearities of mathematical test functions compared with MARS, whereas MARS performs better on linear functions. Furthermore, parameter tuning is recommended for both RF and MARS to improve the predictive accuracy. In future work, we will focus on screening to identify irrelevant design variables in order to reduce the dimensionality of the design space. Furthermore, we plan to compare RF with another widely studied method, Kriging, for response surface modelling.

10 Replication of results

This paper conducted two experiments, whose details are presented in this section to facilitate the replication of our results. The system specification that we used is a 64-bit Windows 7 Operating System with 2.7 GHz Intel Core i7-4600U CPU and 8 GB RAM.

Software: the R software (version 3.5.3) has been used for the experiments in this paper. The R software and all used R packages are freely available online.

Datasets: In experiment 1, the use-case datasets have been collected from our industry partner. The design parameters and design objectives of the use-case datasets were presented in Section 5.1.1. However, due to the confidentiality of the industry, the datasets are not included. This is one of the reasons why we used the mathematical test functions in experiment 2 to allow researchers to replicate and compare the results. These mathematical test functions and the input bounds of parameters were presented in Table 1. For the sample generation, Latin Hypercube Sampling R package[4] with the type of design "maximin" was used for both the training and the testing datasets, comprising 410 and 1000 samples, respectively. We set the random seed for the training datasets to 18 and for the testing data to 17. For the sensitivity analysis in experiment 2, 100 more samples were generated using random seed 2021.

Surrogate model generation: for both methods RF and MARS, the R packages have been provided in Section 6.3. Also, the experimental setup for the surrogate model generation was presented in 7 steps in Section 5.2 and the hyperparameter settings for these methods were presented in Section 6.2. For the sensitivity analysis, the used optimization R package was provided in Section 6.3. For the normalization of the results in experiment 2, we used the clusterSim R package (type n9).[5] For plotting results in Fig. 11, we used plot3D package.[6] For the statistical significance test, we used the wilcox.test.[7] For all used R packages, only specified hyperparameters in this paper were tuned and the rest of settings were set to default values.

Compliance with ethical standards

---

[4]https://cran.r-project.org/web/packages/DoE.wrapper/index.html

[5]https://cran.r-project.org/web/packages/clusterSim/clusterSim.pdf

[6]https://cran.r-project.org/web/packages/plot3D/index.html

[7]https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html

References

Bertoni A, Dasari SK, Hallstedt S, Petter A (2018) Model-based decision support for value and sustainability assessment: applying machine learning in aerospace product development. In: 15th international design conference, dubrovnik, the design society, vol 6, pp 2585–2596

Biau G, Scornet E (2016) A random forest guided tour. Test 25(2):197–227

Breiman L (2001) Random forests. Mach Learn 45(1):5–32

Ceylan H, Kim S, Gopalakrishnan K, Schwartz CW, Li R (2014) Sensitivity analysis frameworks for mechanistic-empirical pavement design of continuously reinforced concrete pavements. Constr Build Mater 73:498–508

Chen W, Jin R, Sudjianto A (2006) Analytical global sensitivity analysis and uncertainty propagation for robust design. J Qual Technol 38(4):333–348

Dasari SK, Lavesson N, Andersson P, Persson M (2015) Tree-based response surface analysis. In: Machine learning, optimization, and big data. Springer, pp 118–129

Dasari SK, Cheddad A, Andersson P (2019) Random forest surrogate models to support design space exploration in aerospace use-case. In: IFIP international conference on artificial intelligence applications and innovations. Springer, pp 532–544

Deng H (2014) Interpreting tree ensembles with intrees. arXiv:14085456

D'iaz-Uriarte R, De Andres SA (2006) Gene selection and classification of microarray data using random forest. BMC Bioinform 7(1):1

Duroux R, Scornet E (2018) Impact of subsampling and tree depth on random forests. ESAIM: Probability and Statistics 22:96–128

Forslund A, Soéderberg R, Loéoéf J, Galvez AV (2011) Virtual robustness evaluation of turbine structure assemblies using 3d scanner data. In: ASME 2011 international mechanical engineering congress and exposition, american society of mechanical engineers, pp 157–165

Friedman JH (1991) Multivariate adaptive regression splines. The Annals of Statistics, pp 1–67

Graening L, Menzel S, Hasenjäger M, Bihrer T, Olhofer M, Sendhoff B (2008) Knowledge extraction from aerodynamic design data and its application to 3d turbine blade geometries. Journal of Mathematical Modelling and Algorithms 7(4):329

Hansen N, Ostermeier A (2001) Completely derandomized self-adaptation in evolution strategies. Evolutionary Computation 9(2):159–195

Heikkinen T, Stolt R, Elgh F, Andersson P (2016) Automated producibility assessment enabling set-based concurrent engineering. In: Proceedings of the 23rd ISPE Inc. International Conference on Transdisciplinary Engineering, Parana, Curitiba, October 3–7, 2016. IOS Press, pp 947–956

Hock W, Schittkowski K (1980) Test examples for nonlinear programming codes. J Optim Theory Appl 30(1):127–129

Huang Z, Wang C, Chen J, Tian H (2011) Optimal design of aeroengine turbine disc based on kriging surrogate models. Computers & Structures 89(1-2):27–37

Jin R, Chen W, Simpson TW (2001) Comparative studies of metamodelling techniques under multiple modelling criteria. Structural and Multidisciplinary Optimization 23(1):1–13

Kohavi R et al (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Ijcai, vol 14, pp 1137–1145

Kveselys D (2017) Weld producibility assessment system: evaluation of producibility assessment tools using set-based approach in multi-disciplinary aerospace design

Liaw A, Wiener M (2002) Classification and regression by randomforest. R news 2(3):18–22

Ma T, Zhang W, Zhang Y, Tang T (2015) Multi-parameter sensitivity analysis and application research in the robust optimization design for complex nonlinear system. Chinese Journal of Mechanical Engineering 28(1):55–62

Mack Y, Goel T, Shyy W, Haftka R (2007) Surrogate model-based optimization framework: a case study in aerospace design. In: Evolutionary computation in dynamic and uncertain environments. Springer, pp 323–342

Min ATW, Sagarna R, Gupta A, Ong YS, Goh CK (2017) Knowledge transfer through machine learning in aircraft design. IEEE Comput Intell Mag 12(4):48–60

Oshiro TM, Perez PS, Baranauskas JA (2012) How many trees in a random forest? In: Machine learning and data mining in pattern recognition. Springer, pp 154–168

Preuss M, Wagner T, Ginsbourger D (2012) High-dimensional model-based optimization based on noisy evaluations of computer games. In: Learning and intelligent optimization. Springer, pp 145–159

Probst P, Wright MN, Boulesteix AL (1804) Hyperparameters and tuning strategies for random forest. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, p e1301

Queipo NV, Haftka RT, Shyy W, Goel T, Vaidyanathan R, Tucker PK (2005) Surrogate-based analysis and optimization. Progress in Aerospace Sciences 41(1):1–28

Sathyanarayanamurthy H, Chinnam RB (2009) Metamodels for variable importance decomposition with applications to probabilistic engineering design. Computers & Industrial Engineering 57(3):996–1007

Shan S, Wang GG (2009) Survey of modeling and optimization strategies to solve-dimensional design problems with computationally-expensive black-box functions. Struct Multidiscip Optim 41(2):219–241. https://doi.org/10.1007/s00158-009-0420-2

Shan S, Wang GG (2010) Survey of modeling and optimization strategies to solve high-dimensional design problems with computationally-expensive black-box functions. Struct Multidiscip Optim 41(2):219–241

Sheskin DJ (2003) Handbook of parametric and nonparametric statistical procedures. CRC Press, Boca Raton

Song X, Lv L, Li J, Sun W, Zhang J (2018) An advanced and robust ensemble surrogate model: extended adaptive hybrid functions. Journal of Mechanical Design 140(4):041402

Taflanidis A, Loukogeorgaki E, Angelides D (2011) Analysis and design of offshore energy conversion devices under modeling uncertainties. In: OCEANS 2011. IEEE, pp 1–10

Wang GG, Shan S (2007) Review of metamodeling techniques in support of engineering design optimization. Journal of Mechanical Design 129(4):370–380

Zhao D, Xue D (2010) A comparative study of metamodeling methods considering sample quality merits. Struct Multidiscip Optim 42(6):923–938