Efficient Motor Blade Design: A Sensitivity Analysis and Kriging Approach

Ibrahima S WADE

March 2023

Abstract

In this paper, we investigate the efficiency of motor blades in the context of various quantities of interest, such as performance, torque, pressure difference, and acoustics, which depend on input parameters, primarily the blade geometry. Our goal is to enhance the manufacturing process by identifying the most influential input variables on the output, allowing for a focused allocation of resources towards improving these inputs, consequently leading to improved outputs.

Sensitivity analysis is a method of assessing the effect of variations in input parameters on the output of a model or a system. It examines the changes in output resulting from changes in input values, and can help to identify the most important input parameters in terms of their impact on the output. This method is proved to be extremely valuable since performing multiple experiments with all input variables would generally be prohibitively expensive. We also employ Kriging, a powerful statistical interpolation technique, to create surrogate models mimicking the relationship between input variables and outputs.

By combining sensitivity analysis with Kriging, we aim to efficiently identify the most influential input variables and propose targeted improvements for the motor blade manufacturing process, ultimately leading to enhanced efficiency and performance with minimal cost.

1 Introduction

Sensitivity analysis is defined as the study of how uncertainty in the output of a numerical model or system can be attributed to different sources of uncertainty in input variables. Sensitivity analysis results can be used to simplify the model, gain a better understanding of the system, or even verify its accuracy. Sensitivity analysis methods can be divided into two main families: local methods and global methods. When we carry out a sensitivity analysis around a specific point of interest in the input space of the model, the most suitable tools to satisfy our problem are local sensitivity analysis methods, can be grouped into three main types: the One At a Time (OAT) method, the scenario decomposition method, and the spiderweb graph method. Contrary to local methods, global sensitivity analysis methods study the influences of uncertain inputs over the entire input space. Because of this, global methods are considered to be much more reliable tools for sensitivity analysis. Among these, we can mention the Morris method, the sequential bifurcation method, and the Sobol' indices method. In the following, we will mainly focus on the latter, describing its principle and the different methods used to estimate these indices.

An other important aspect of sensitivity analysis is the building of meta-models using Kriging methods as their main tool. Originally conceived as tool for mining engineering, the Kriging has been used, since its mathematical formalization, to compute at low cost data in simulations with known entries and exits. In the ensuing, we will detail how different the results given by a Kriging-based meta-model can be when its parameters differ, and this for same dataset.

The article is organized as follows. Introduction corresponds to Section 1. In Sections 2 and 3, we present a reminder first on Sobol' indices to perform global sensitivity analysis and second on Kriging, mainly its conception and its mathematical formalization. Section 4 is dedicated to the different methods used and the results will be shown in Section 5. A conclusion and some perspectives are given in Section 6.

2 Sobol' indices for SA

2.1 Principle and definition

Sobol'indices are sensitivity indices used to quantify the influence of an input on the output. Named after mathematician Ilya Meïérovitch Sobol, these indices are based on a variance decomposition [4].
Let f be a black box function and $Y \in L^2$ such that $Y = f(X)$. To determine the importance of the influence of an input variable X_i on the variance of the output Y, we need to study how much this variance decreases when we fix the variable X_i to a value $x_i^* : V(X_i = x_i)$. The disadvantage of this indicator is the choice of the value x_i^* of X_i , which is resolved by taking into account the expected value of this quantity for all possible values of $x_i^* : E(V(X_i))$. In this way, the more important the variable X_i is with respect to the variance of Y, the lower this quantity will be.
Considering the formula for the total variance such as [5]:

$$V(Y) = V(E(X_i)) - E(V(X_i))$$

We can equivalently use the quantity: $V(E(X_i))$. This quantity will then be large proportionally to the importance of the variable X_i with respect to the variance of Y. To use a normalized indicator, we define the sensitivity index of Y to X_i:

$$S_i = \frac{V(E(Y|X_i))}{V(Y)} \tag{1}$$

This newly created index corresponds to the first-order sensitivity index by Sobol. Its role is to quantify the sensitivity of the output Y to the input variable X_i by calculating the part of Y's variance due to the variable X_i. Sobol introduced this sensitivity index by decomposing the f model function into a sum of functions of increasing dimensions.

$$Y = f(X_1, .., X_p) = f_0 + \sum_{i=1}^{p} f_i(X_i) + \sum_{1 \le i < j \le p} f_{ij}(X_i, X_j) + ... + f_{1,...,p}(X_1, ..., X_p),$$

where

$$\begin{aligned}
f_0 &= E(Y), \\
f_i(X_i) &= E(Y|X_i) - E(Y), \\
f_{i,j}(X_i, X_j) &= E(Y|X_i, X_j) - E(Y|X_i)E(Y|X_j) + E(Y), \\
f_{i,j,k}(X_i, X_j, X_k) &= E(Y|X_i, X_j, X_k) - E(Y|X_i, X_j) - E(Y|X_i, X_k) - E(Y|X_j, X_k)...
\end{aligned}$$

2.2 Estimation of Sobol' indices

Several procedures for estimating Sobol indices have been proposed and studied, some based on the Monte Carlo or quasi-Monte Carlo experiment model. There are also other estimation procedures based on other experimental designs, such as polynomial chaos, when applying it the model output is represented as an expansion of orthogonal polynomials, with the polynomials chosen according to the probability distribution of the input parameters. These polynomials create a surrogate model that efficiently approximates the original model. However, effective estimation of Sobol' indices can also be achieved through the "Pick Freeze" method and Rank-based estimation, that we will detail in the following sections.

2.2.1 Pick Freeze Method

Estimation by the Pick Freeze method is used for models with a reasonable number of inputs, usually less than a thousand [1]. When the number of inputs is large, this estimation will then require a large number of model evaluations, generally too important to put into practice. This is referred to as a "high-dimensional context". Thus, we will describe here how to set up a multiple Pick Freeze.
Let $F \subset \{1, ..., p\} \quad \forall p \in N^*$ a set of indices, $F^C = \{1, ..., p\} \backslash F$, and f a black box.
We define Y^F such that:

$$Y^F = f(X^F) \quad where \quad X^F = \begin{cases} X_i & if \quad i \in F \\ X_i^{'} & if \quad i \in F^C \end{cases}$$

This method is named after the fact that, to generate Y^F variables, all input parameters whose indices are included in F are frozen.
In a classic Pick Freeze estimation, only one variable is frozen at that time, namely = i.
This allows us to define:

$$S_F = \frac{Cov(Y; Y^F)}{V(Y)}$$

S_F allows us to admit the natural estimator:

$$\hat{S}_F = \frac{\frac{1}{N}\sum Y_k Y_k^F - (\frac{1}{N}\sum \frac{Y_k + Y_k^F}{2})^2}{\frac{1}{N}\sum \frac{(Y_k)^2 + (Y_k^F)^2}{2} - (\frac{1}{N}\sum \frac{Y_k + Y_k^F}{2})^2}.$$

This methodology is based on a particular design of experiment that is not available for our case. That is why we are interested in an estimator based on a N-sample only like the rank-based estimation.

2.2.2 Rank-based estimation

Rank-based estimation of Sobol' indices is a method to estimate sensitivity analysis parameters without requiring any assumption on the probability distribution of inputs. In this section we detail its methodology.

Let us take up the notations used a little earlier in (1) .
Here, it is assumed that the inputs X_i for $i = 1; ...; p$ are scalars and we want to estimate the Sobol index S^1 with respect to X_1:

$$S^1 = \frac{Var(E(Y|X_1))}{Var(Y)}$$

To do this, we consider an N-sample of input/output (X_1^i, Y_1) pair given by:

$$(X_1^1, Y_1), (X_1^2, Y_2), ..., (X_1^N, Y_N)$$

These pairs are then rearranged such that:

$$X_1^{(1)} < ... < X_1^{(N)}$$

We then introduce [5]:

$$S^1_{N,Rank} = \frac{\frac{1}{N}\sum_{i=1}^{N-1} Y_{(i)} Y_{(i+1)} - (\frac{1}{N}\sum_{i=1}^{N} Y_i)^2}{\frac{1}{N}\sum_{i=1}^{N} Y_i^2 - (\frac{1}{N}\sum_{i=1}^{N} Y_i)^2}$$

Thus, we have the following theorem [5]:

Theorem (Gamboa, Gremaud, Klein, Lagnoux, 2021)
One has $S^1_{N,Rank} \xrightarrow[N\to\infty]{\text{a.s.}} S^1$. (2) If the X_i 's are uniformly distributed and under some mild assumptions on f, then

$$\sqrt{N}\left(S^1_{N,\text{ Rank }} - S^1\right) \xrightarrow[N\to\infty]{\mathcal{L}} \mathcal{N}_1\left(0, \sigma_R^2\right).$$

However, this method only works for one-dimensional problems. One of the main focuses of this article, as you will see in Section 4 will be finding ways to estimate higher order indices using a new generation of estimators.

3 Kriging

To return to our study, we are intending to conduct a sensitivity analysis on an important set of data with multiple entries and exits. However, we have to compose with two major constraints. The first one is the lack of information concerning the existence of a possible link between the corresponding entries and exits. The second rests in the cost of the black box which does not allow to generate data as we please.
To answer these constraints, we must build a meta-model allowing us to generate as much data as we wish. A meta-model corresponds to a modelling of a studied phenomenon which has the advantage of being inexpensive opposite to the black box, but as close as possible to it to mimic it. Also, it will be necessary to conduct a sensitivity analysis on the errors we will encounter compared with the chosen parameters. Thus, a Kriging-based meta-model represents the best solution to solve our problem at a low cost.

Named after its creator the mining engineer D.G. Krige, the Kriging is a method of spatial interpolation basing itself on a series of statistical empirical methods whose goal is the determination of ores' spatial distribution based on a set of drillings [3]. However, the method's formalization was released by Georges Matheron using the correlations between the drilling to estimate their spatial distribution.
The spatial interpolation corresponds to an estimation problem of a function $f(x)$, where $x = (y; z)$, in a point x_0 of the plan based on known values of f at a certain number m of surrounding points x_i :

$$f(x_0) = \sum_{i=1}^{m} W_i \times f(x_i)$$

The problem consists to determinate the ponderation, namely the W_i, for each of the surrounding points. Concerning the choice of the weight, the Kriging chooses them based on the degree of the similarity between the values of f. In other words, the choice is made from the covariance between the points, in depending on the distance between these ones.
The Kriging consists to compute the W_i of the equation above with the help of the values of the function $\gamma(h)$ corresponding to the m chosen points.

There are three types of univariate Kriging :
- simple Kriging with a stationary variable with a known mean
- ordinary Kriging with a stationary variable of unknown mean
- universal Kriging with a non-stationary variable which contains a tendency

In a more mathematical context, the Kriging works with the following principle [6].
The method's main postulate is that f is supposed to be a realization of a process $Y : D \subset \mathbb{R}^d \to \mathbb{R}$ with a mean-function $k : (x; y) \in \mathbb{R}^d \times \mathbb{R}^d \to k(x; y) \in \mathbb{R}$.

The choice of the weights is done in such a way that the prediction made is unbiased and with minimal variance error.
In the context of the implementation of a simple Kriging, the process Y is assumed to be centered and stationary of order 2 with a known covariance function k.
Thanks to this, we are able to compute the vector $r_N(x_0)$ and the squared-matrix R_N with an $N \times N$ size of the observation point's covariances such as :

$$r_N(x_0) = (Cov[Y(x_0); Y(x_1)]; ...; Cov[Y(x_0); Y(x_N)])^T = [k(x_0; x_1); ...; k(x_0; x_N)]^T$$

$$R_N = \begin{pmatrix} Cov[Y(x_1); Y(x_1)] & \dots & Cov[Y(x_1); Y(x_N)] \\ \vdots & \ddots & \vdots \\ Cov[Y(x_N); Y(x_1)] & \dots & Cov[Y(x_N); Y(x_N)] \end{pmatrix} = \begin{pmatrix} k(x_1; x_1) & \dots & k(x_1; x_N) \\ \vdots & \ddots & \vdots \\ k(x_N; x_1) & \dots & k(x_N; x_N) \end{pmatrix}$$

Thus, the aleatory prediction of the Kriging is written as :

$$\hat{Y}(x_0) = \sum_{i=1}^{N} \lambda_i^*(x_0) Y(x_i)$$

where the optimal vector is found by minimizing the quadratic error :

$$\lambda^*(x_0) = [\lambda_1^*(x_0); ...; \lambda_N^*(x_0)]^T = R_N^{-1} r_N(x_0)$$

One of the main upsides of Kriging is the fact that its variance is explicitly known :

$$\sigma_N^2(x_0) = E[[Y(x_0) - \hat{Y}(x_0)]^2] = k(x_0; x_0) - r_N^T(x_0) R_N^{-1} r_N(x_0)$$

This knowledge allows us to build confidence intervals.

4 Methods

4.1 Kriging-based meta-models

In the following section, our goal is to describe the logic and tools used during the construction of our multiple Kriging-based meta-models.

Firstly, the most important parameter to define throughout the elaboration of a Kriging-based meta-model is the covariance kernel, named C here, because of the consequences it provokes on the whole obtained meta-model, especially if the trend is known or supposed constant.
With this information in mind, C must be chosen inside a set of defined-positive kernels in order for it to be allowable [7]. However, verifying the considered positive definition of given C as truthful is not an easy task and performing a non-parametric estimation seems unrealistic. Usually, and in this application too, when a Kriging is performed, we choose beforehand a family-parameter of known defined positive covariance kernels.
One of the usual restrictions during their use for a meta-model, is to consider the kernels as solely depending of the increment $u - v$, called "stationary kernels". Another profitable way to obtain eligible covariance kernels in high dimensions consists in the selection of the tensor products of the eligible kernels $1 - d$. Such kernels are called "separable" and are the most commonly used in the literacy of computing experiments.

Concerning the tools used during the meta-models' creation process, we used exclusively the statistical-language R, threw its interface RStudio, and the package the following packages : "dplyr", "DiceKriging" and "jmcm".
To come back to the covariance kernels' choice, the ones currently used in the package "DiceKrging" are based on the following model, up to the multiplicative constant $\sigma^2 > 0$:

$$c(h) := C(u; v) = \sigma^2 \prod_{j=1}^{d} g(h_j; \theta_j)$$

where $h = (h_1; \ldots; h_d) := u - v$ and g is a covariance kernel of dimension 1.

Although it makes sense in certain scenarios, the package does not allow the mix by dimensions of different covariance kernels. The parameters θ_j are selected to be physically intelligible in the same unit as the corresponding variables. Those parameters are called "scales of characteristic lengths" (Rasmussen and Williams, 2006).

In our case, the defined positive covariance kernels are the following for g, with $\theta > 0$ and $0 < p \leq 2$:

Matèrn 5/2	$g(h) = (1 + \frac{\sqrt{5}\|h\|}{\theta} + \frac{5h^2}{3\theta^2}) \times exp(-\frac{\sqrt{5}\|h\|}{\theta})$
Matèrn 3/2	$g(h) = (1 + \frac{\sqrt{3}\|h\|}{\theta}) \times exp(-\frac{\sqrt{3}\|h\|}{\theta})$
Gaussian	$g(h) = exp(-\frac{h^2}{2\theta^2})$
Exponential	$g(h) = exp(-\frac{\|h\|}{\theta})$
Power-Exponential	$g(h) = exp(-(\frac{\|h\|}{\theta})^p)$

Globally, the package "DiceKriging" represents, in our case, the most efficient tool to perform the computations in a Kriging-based meta-model thanks to its accessibility and versatility. Thus, for the construction of our meta-models, we used its function "km" to define the parameters of the Kriging we wanted to perform. Concerning the function "km", the scope of Kriging it can performed is pretty large thanks to the multiple parameters the users can modify, such as the trend, the nugget or the noise effect.

Without doing an entire review of our code, we will briefly describe the data set we worked on and the main parameters we chose to implement in our meta-models.
The data set that was given to us contains 600 observations of motor blades' performances split between a total of 18 variables, with the first 15 representing the inputs and the last 3 the outputs.
With this starting configuration, we decided to build a total of 15 Kriging-based meta-models to simulate our data. More precisely, for each output variable we simulated a Kriging with all the covariance kernels listed above (3 exits $\times$ 5 kernels = 15 meta-models). Also, for each simulation we chose to use a constant trend, which corresponds to the default choice in our package.
Finally, we needed to perform a Cholesky decomposition in order for the output vector to be defined positive. To achieve that, we implemented in all of our meta-models the following parameter "nugget = 1e-8*var(y)", with y the name of the output vector in R.

4.2 Sobol' indices estimation for higher orders

In this section, we aim to estimate the Sobol' coefficients with respect to two or more input variables, thus estimating the second order Sobol indices. Given the set of data points:

$$(X_i^1, X_j^1, Y^1), ..., (X_i^n, X_j^n, Y^n)$$

We can compute the Sobol' indices as follows:

$$S_{\{i,j\}} = \frac{V\left(E\left[Y \mid X_i, X_j\right]\right)}{V(Y)}$$

To calculate the second-degree Sobol' indices, we use the following formula:

$$S_{i,j} = \frac{V\left(E\left[Y \mid X_i, X_j\right]\right) - V\left(E\left[Y \mid X_i\right]\right) - V\left(E\left[Y \mid X_j\right]\right)}{V(Y)}$$

One clear challenge that arises is how to rearrange the data points in the current context, given that we are dealing with vectors in the Euclidean plane. One solution is to utilize the traveling salesman algorithm, defining the rank based on the optimal path obtained from this algorithm.
A clear challenge that arises is how to rearrange the data points in the current context, given that we are dealing with vectors in the Euclidean plane. One solution is to utilize the traveling salesman algorithm, defining the rank based on the optimal path obtained from this algorithm.
The Traveling Salesman Problem (TSP) is a classical combinatorial optimization problem that has been widely studied in the fields of operations research and computer science. The problem can be described as follows:

given a set of cities and the distances between each pair of cities, the objective is to find the shortest possible route that visits each city exactly once and returns to the starting city.

Formally, the TSP can be defined on a complete weighted graph G = (V, E), where V is a set of vertices representing the cities, E is a set of edges representing the connections between cities, and a weight function w : E $\rightarrow$ R assigns a non-negative distance to each edge. The goal is to find a Hamiltonian cycle (a cycle that visits each vertex exactly once) in G with the minimum total weight.

Various algorithms and heuristics have been proposed to solve the TSP, ranging from exact algorithms, such as branch and bound, dynamic programming, and integer linear programming, to approximate algorithms and meta-heuristics, including genetic algorithms, simulated annealing, and ant colony optimization. Despite its simple description, the TSP is an NP-hard problem, which means that finding an optimal solution becomes increasingly difficult as the number of cities grows.
The rationale behind using the traveling salesman problem (TSP) to define the ranks in our estimator lies in the fact that the TSP algorithm can find an optimal path in a higher-dimensional metric space.

As mentioned in this article [2], by solving the TSP in higher dimensions, we can extend Chatterjee's method to estimate any sensitivity index of any order. Specifically, for an index of order k, we need to consider a permutation that solves the TSP in $\mathbb{R}^k$. Moreover, the same n sample can be used to estimate all Sobol indices of any order. This approach allows us to consider problems whose inputs are not real-valued but take their values in any metric space. The use of the TSP to define the ranks in our estimator ensures that we can effectively deal with the multi-dimensional nature of the problem. Although finding the optimal solution to the TSP is computationally challenging, approximation algorithms can be used to obtain near-optimal solutions efficiently. The critical aspect is to propose a permutation without a fixed point such that $\sigma(i)$ is close to i for any generic i. This way, the TSP-based method enables us to estimate the Sobol' indices for higher-dimensional problems effectively and accurately, thus making it an ideal choice for our sensitivity analysis.

Following the rationale for using the traveling salesman algorithm in our sensitivity analysis, we implemented the TSP algorithm in Python to effectively process our dataset. We utilized the solve_tsp library, which provides efficient algorithms for solving the TSP in a variety of contexts. This library allowed us to compute the optimal path for our data points, taking into account the multi-dimensional nature of our problem.

After solving the TSP on the dataset obtained by the Kriging method, we implemented the formulas to calculate the second-order Sobol indices for each pair of input variables. By combining the TSP solution with the Kriging-based surrogate models, we were able to effectively identify the most influential input variables on the motor blade performance. This comprehensive approach, which integrates the traveling salesman algorithm, Kriging, and sensitivity analysis, enabled us to propose targeted improvements for the motor blade manufacturing process. Ultimately, our methodology leads to enhanced efficiency and performance with minimal cost, demonstrating the power and utility of this combined approach.

5 Results

5.1 The data

Having generated the data from the Kriging model, we examine the following input variables: Calage_R1, Calage_R2, Calage_R3, Calage_R4, Calage_R5, Lcorde_R1, Lcorde_R2, Lcorde_R3, Lcorde_R4, and Lcorde_R5. The target variables are Delta_P_Q2_WBP and Torque_Q2_WBP.

To analyze the relationships between the input variables, we computed the correlation coefficients between each pair of variables and visualized the results using a heatmap. The correlation matrix is shown below. As

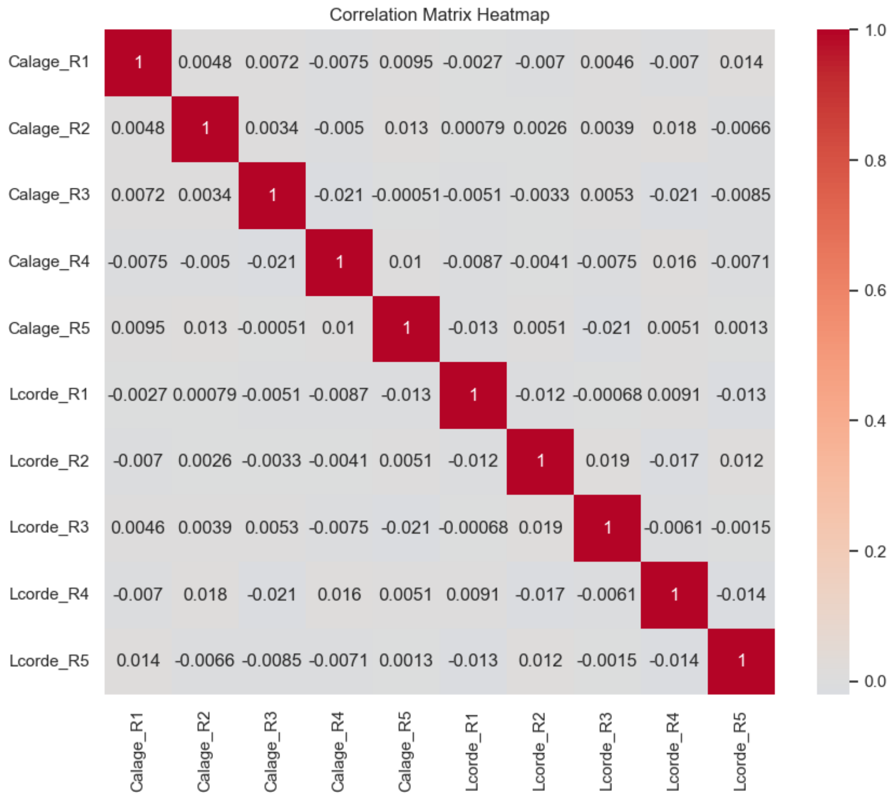

Figure 1: Heatmap of the correlation matrix between input variables

seen in Figure 1, the correlation coefficients between input variables are close to zero, indicating that there are no strong linear relationships between any pairs of input variables. This suggests that the input variables can be considered as independent of one another, which is advantageous when building predictive models, as it reduces the risk of multicolinearity and increases the stability of the model's estimates. In addition to the correlation matrix, we also visualized the relationships between the input variables using scatterplots. Scatterplots provide a more intuitive way to identify patterns and trends in the data. They allow us to see the distribution and relationships between variables clearly, helping us to understand the structure of the data.

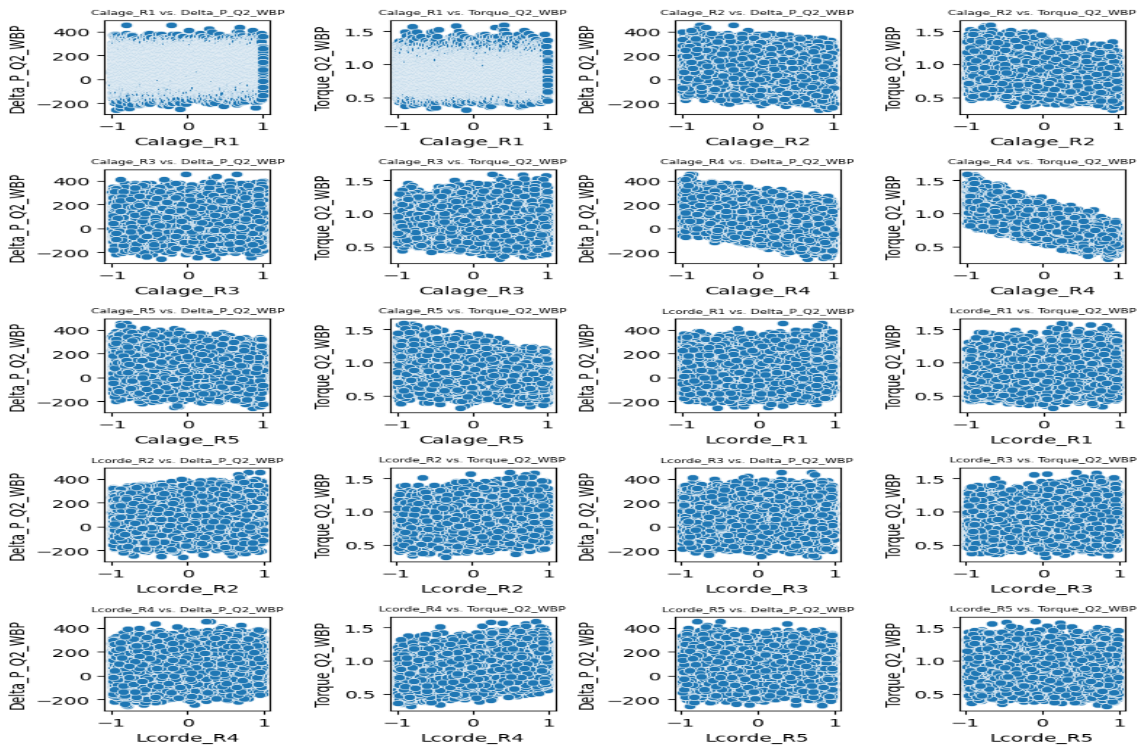

Figure 2: Scatterplots of the relationships between input variables

As can be seen in Figure 2, most of the input variables have weak relationships with each other, which is consistent with the findings from the correlation matrix heatmap. This observation further supports our conclusion that multicollinearity is not a significant concern in our dataset.

To further understand the structure of our dataset, we analyzed the distribution of the output variables. This allows us to identify any potential biases or unusual patterns in the target variables, which could impact the performance of our predictive model.

Figure 3 shows the distributions of the output variables, namely Delta_P_Q2_WBP and Torque_Q2_WBP. The distributions provide insights into the range and variability of the output variables, as well as any potential skewness or unusual patterns that may need to be addressed during the modeling process.

5.2 The Kriging-based meta-models

After the execution of our 15 Kriging-based meta-models, we found the following final values, called F in R, with rounding to the hundredth :

Covariance kernel	First output	Second output	Third output
Matèrn 5/2	$3162,86$	$303,88$	$-919,30$
Matèrn 3/2	$3196,03$	$335,73$	$-907,17$
Gaussian	$3177,80$	$287,91$	$-925,89$
Exponential	$3428,53$	$673,28$	$-750,97$
Power-exponential	$3165,72$	$303,96$	$-887,88$

Before analyzing these numbers we must precise that, since the Kriging process uses some randomness in its computations, some slight differences can be observed in the returned values between multiple executions of a same meta-model.

Concerning the final values that were returned to us, F correspond to the acquisition function which is a criterion

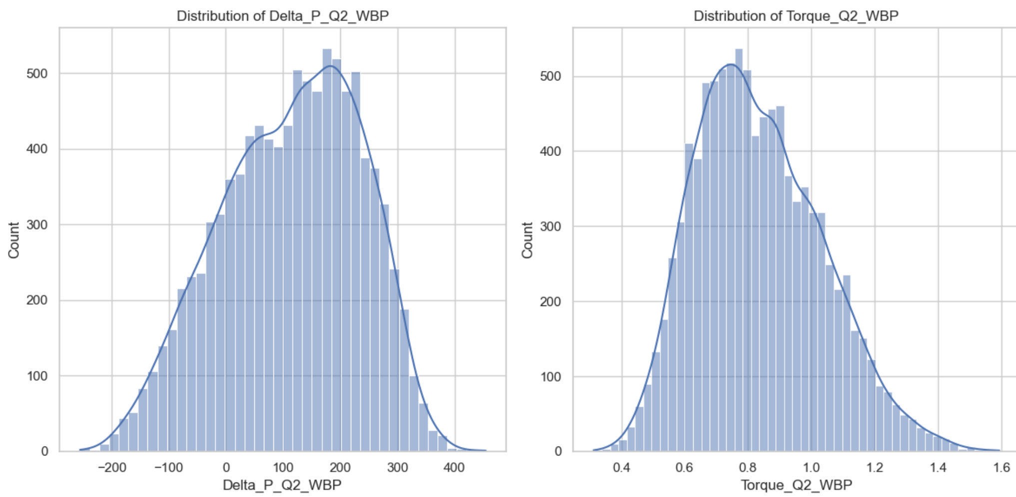

Figure 3: Distributions of the output variables

used to determine the next following point to sample. Basically, F quantifies the potential of a point to improve the optimization process by giving information on the uncertainty or expected improvement in the value of the unknown function at that point. Basically, the higher the acquisition function gets for a specific point, higher are the chances of that point to be selected as the next point to sample in the optimization process.
With this knowledge in mind, we could say that the usage of a Exponential covariance kernel gives the better results because it presents the higher returned values. However, our meta-models need improvements because specially the ones concerning the third output which all stopped after the 101^{st} iteration.
Finally, we were able to plot all the meta-models the leave-one-out cross validation, the standardized residuals and their normal QQ-plot. We give below an example of the plots created. As we previously said, this work is just the beginning of the construction process of satisfying Kriging-based meta-models. The next step will be to predict future values with our already existing meta-models or creating new ones with the inclusion of other characteristics, like trend or noise effect.

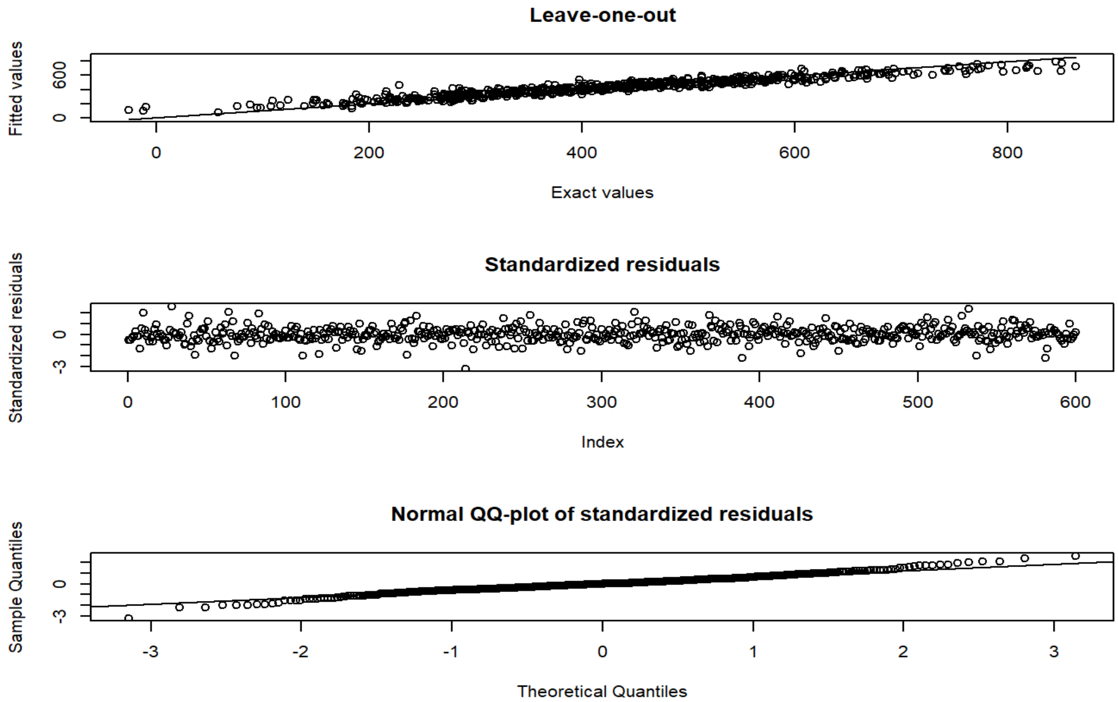

Figure 4: Plot of the Exponential-kernel meta-model for the first output

5.3 Estimated Indices

In this section, we present the estimated indices for the two output variables, namely Delta_P_Q2_WBP and Torque_Q2_WBP. The estimated indices provide valuable insights into the relative importance of the input variables on the output variables. To visualize the results, we have created heatmaps for each output, which can be found in Figure 5.

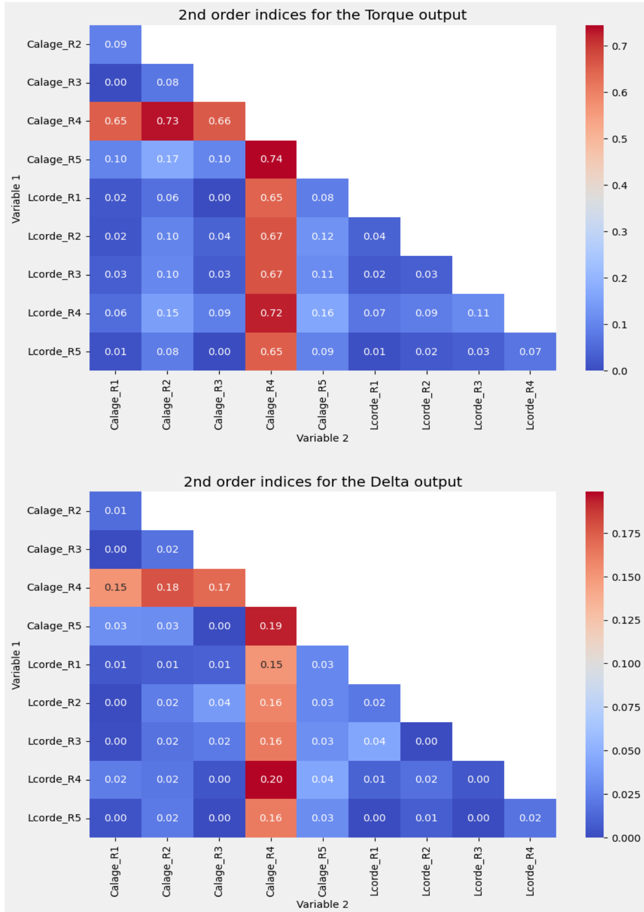

Figure 5: Heatmaps of the estimated indices for Delta_P_Q2_WBP and Torque_Q2_WBP

The heatmaps in Figure 5 display the estimated indices for each input variable, with color gradients representing the magnitude of the indices. A higher value indicates a stronger influence of the input variable on the corresponding output variable. By examining the heatmaps, we can identify the most influential input variables for each output and better understand their relationships.

From the heatmaps, we can observe that Calage_R4 has high values of indices when paired with any other variable. This indicates that Calage_R4 has a strong influence on both output variables. Interestingly, the most influential combination of variables for the second-degree order is Calage_R4 and Calage_R5 for both outputs. This observation is quite intriguing, as it suggests that these two variables have a particularly strong interaction effect on the outputs.

This information can be used to inform decisions about variable selection and to prioritize further investigation of the most influential variables, particularly Calage_R4 and Calage_R5. Additionally, the heatmaps can be used to verify if the model's assumptions are met and if multicollinearity is not a concern, as discussed in the previous sections.

5.4 Practical Implications and Recommendations

These findings have significant real-life implications for the design and optimization of motor blades at Valeo. As Calage_R4 and Calage_R5 have been identified as the most influential variables on both output variables, the company should pay special attention to these parameters when designing and manufacturing their motor blades. The strong interaction effect between Calage_R4 and Calage_R5 suggests that adjusting these parameters in tandem could lead to significant improvements in performance, potentially increasing the efficiency and effectiveness of the motors.

Valeo can utilize these results to prioritize research and development efforts, focusing on the optimization of those two variables in their motor blade designs. By doing so, the company can potentially reduce production costs, improve product quality, and gain a competitive advantage in the market. Furthermore, the insights from this analysis can also be used to refine and streamline the testing and validation process for new motor blade designs, enabling faster innovation and shorter time-to-market for new products.

In conclusion, by leveraging the results of this analysis, Valeo can make more informed decisions in the design and production of their motor blades, ultimately leading to better performance and increased customer satisfaction.

6 Conclusion

In this study, we have demonstrated the effectiveness of a new approach combining Kriging and Sobol' sensitivity analysis for understanding and optimizing the performance of motor blades. This methodology involves the use of the Traveling Salesman Problem (TSP) for rank-based estimation, which has been shown to be a reliable estimator in recent research. Given the constraints of the pick-freeze method in this problem, the TSP-based estimation proved to be particularly beneficial in our case.

Our findings indicate that Calage_R4 has significant influence when coupled with other variables, and the combination of Calage_R4 and Calage_R5 is the most influential for both output variables. By focusing on these key variables, manufacturers can optimize their design and manufacturing processes to enhance the performance of their motor blades.

The combination of Kriging and Sobol' sensitivity analysis has proven to be a valuable tool for gaining insights into the complex relationships between input variables and their effects on performance metrics. By using these techniques, companies can make informed decisions in the design and manufacturing stages to improve product performance and efficiency, ultimately leading to increased customer satisfaction and market competitiveness.

Moreover, the integration of TSP within the process of rank-based estimation allows for more efficient and cost-effective sensitivity analysis. This innovative approach enables manufacturers to achieve better results with a reduced number of experiments, which translates into reduced costs, improved resource allocation, and faster product development cycles.

In conclusion, the innovative approach presented in this study, combining Kriging, Sobol' sensitivity analysis, and TSP-based rank estimation, offers a powerful and cost-effective solution for companies looking to optimize their processes, reduce costs, and improve the overall quality of their products.

References

[1] Y. de Castro and A. Janon. Randomized pick-freeze for sparse sobol indices estimation in high dimension. *ESAIM: Probability and Statistics*, 19:725–745, 2015.

[2] F. Gamboa, P. Gremaud, T. Klein, and A. Lagnoux. Global sensitivity analysis: A novel generation of mighty estimators based on rank statistics. *Bernoulli*, 28(4):2345–2374, 2022.

[3] Y. Gratton. Le krigeage: la méthode optimale d'interpolation spatiale. *Les articles de l'Institut d'Analyse Géographique*, 1(4), 2002.

[4] J. Jacques. Pratique de l'analyse de sensibilité: comment évaluer l'impact des entrées aléatoires sur la sortie d'un modèle mathématique. *Lille: sn*, 2011.

[5] A. Lagnoux. Partie i – from sobol' indices to universal indices. `https://perso.math.univtoulouse.fr/lagnoux/files/2013/12/CIROQUO`$_K L_p artI.pdf$.

[6] A. Lagnoux, T. N. Nguyen, B. Demory, and M. Henner. Kriging and expected improvement applied to an industrial context-prediction of new geometries increasing the efficiency of fans. 2021.

[7] O. Roustant, D. Ginsbourger, and Y. Deville. Dicekriging, diceoptim: Two r packages for the analysis of computer experiments by kriging-based metamodeling and optimization, 2012.