OPTIMIZATION OF MULTI-FIDELITY DATA USING CO-KRIGING FOR HIGH DIMENSIONAL PROBLEMS

K. Elsayed [*†] and C. Lacor [‡]

ABSTRACT

This paper deals with an efficient and multi-fidelity design strategy for high dimensional industrial problems. The most significant factors have been determined based on the Muschelknautz method of modeling (MM) using the screening approach. For cyclone separator, only four (from seven) geometrical parameters are significant. An optimized sampling plan based on random Latin hypercube (LHS) has been used to fit Co-Kriging based on CFD data and an analytical model for estimation of pressure drop. Co-Kriging exhibits better accuracy than ordinary Kriging and blind Kriging if only the high fidelity data is used. For global optimization, the Co-Kriging surrogate in conjunction with genetic algorithms (GA) is used. CFD simulations based on the Reynolds stress turbulence model are also used in this study. A new set of geometrical ratios (design) has been obtained (optimized) to achieve minimum pressure drop. A comparison of numerical simulation of the new design and the Stairmand design confirms the superior performance of the new design compared to the Stairmand design.

KEY WORDS

Surrogate Models, Kriging, Co-Kriging, Blind Kriging, Surrogate based optimization

[*] Postdoc researcher, Department of Mechanical Engineering, Vrije Universiteit Brussel, Pleinlaan 2 -1050 Brussels- **Belgium**, Email: khairy.elsayed@vub.ac.be.

[†] Assistant Professor, Helwan University, Faculty of Engineering - Mattaria, Department of Mechanical Power Engineering, Masaken El-Helmia, 11718 Cairo, **Egypt**, Email: kelsayed75@gmail.com.

[‡] Professor, Department of Mechanical Engineering, Vrije Universiteit Brussel, Pleinlaan 2 -1050 Brussels- **Belgium**, Email: chris.lacor@vub.ac.be

.

INTRODUCTION

Engineering design problems is concerned with the making of decisions based on analysis, which directly impact the product or service being designed. To accomplish this, engineers may need months of experimental or computational investigations. One approach of gaining this desirable increased insight into the problems being studied is using surrogate (or meta) models. Such models seek to provide cheap answers in terms of computing power. The meta-models can also be used to bridge between various levels of sophistication afforded by varying fidelity physics based simulation codes, or between predictions and experiments. Their role is to aid understanding and decision taking by wringing every last drop of information from the analysis and data sources available to the design team and making it available in a useful and powerful way [1].

Surrogate models as the name means commonly used to replace the expensive simulations or experiments to predict the effect of certain input parameters on the output variables (curve fitting) or to estimate the objective function value for optimization process. Ordinary Kriging, radial basis function approximation are typical examples of this branch of usage.

Recently, the meta-models have been used as tool to calibrate the less accurate (simplified) codes. Such multi-fidelity or multi-level approaches can also deal with experimental data and CFD simulations data or for CFD simulation on different grid levels, or fully developed flow results and developing flow results. Co-Kriging is a typical example of surrogate models which can handle multi-fidelity data. To filter the noise of the experimental data or the computational noise stems from the used schemes [1], meta-models can help the engineer as well. A typical example is the Regression Kriging. The surrogate models have been used as a data-mining tool several decades ago, when the polynomial regression was the cheap tool for data-mining.

Constructing The Surrogate Model

The application of met-model instead of the experimental or computational approach is not an easy task. The following difficulties (rocks) should be overcome (passed). The following summary is inspired by the book of Forrester et al.[1].

Many parameters affecting this selection; number of independent variables, the expected input-output relationship (linear, second order, etc.), the possible interaction between independent variables. These factors need some prior knowledge of the problem and the surrogate modeling.

Generally speaking, not all the independent variables are significant, some are negligible. To efficiently decide that, a screening study is needed, one common approach is via first order polynomial regression after creating a design of experiment, then from the analysis of variance, one could decide the most significant parameters. A better and more accurate (but expensive) is by using Morris algorithm [2], where a surrogate model is fitted with limited number of points and then the distribution of the elementary effects is plotted (mean against standard deviation) [1]. Moreover, the dimensionless number is also helpful in reducing the number of design parameters.

The accuracy of the surrogate model depends on the sampling plan. Sometimes, we use data collected by someone else or collected from different source. In this case, there is no choice to design the experiment but still a room of accuracy enhancement via sampling updating if one can perform extra experiments or simulations. The full factorial, central composite and Box-Behnken design of experiments [3] are widely used sampling plan to fit a multi-dimensional polynomial (including interactions) [4]. This sampling plan is known as space-filling. It requires uniform spread of points, consequently it leaves a gap in the projections [1]. Latin hypercube sampling (LHS) (first introduced by McKay [5]) is a better alternative to the stratified sampling plan. In LHS, the design space is splitted into equal sized hypercubes (bins) and placing points in the bins (one in each), making sure that from each occupied bin we could exit the design space along any direction parallel with any of the axes without encountering any other occupied bins [1]. In order to evaluate the uniformity (space-fillingness) of a LHS plan, the maximin metric introduced by Johnson et al. [6] could be used. The result of this optimization process will be a better sampling plan, called Optimal LHS [1]. For multi-fidelity problems a new issue appears, how to select the sampling points for high-fidelity simulations among the low-fidelity sampling plan. The exchange algorithm [1] is a good strategy using the optimal LHS obtained from the Morris-Mitchell criterion [7]. A rule of thumb for the number of points which should be used in the sampling plan is $n \geq 10\,d$, where d is the number of design variables [8,9]. More sampling points will result in a more accurate surrogate. This may need an iterative procedure to achieve a saturation behavior of the constructed surrogate model (no considerable change in r^2 or RMSE). For expensive simulations with up to seven design variables, one may start with the sampling number corresponding to Box-Behnken design (For $d = 3 - 7, n = 15,27,46,54,62$ points respectively) [3].

If one passed all the above-mentioned difficulties, he has to start deal with a lot of mathematics to write a code for constructing the surrogate model.Two other issues should be taken into consideration. (1) Normalize all independent variables into a unit cube (i.e., the design space $D = [0,1]^d$, where d is the number of design variables), or $D = [-1,1]^d$. This step safeguards against scaling issues [1]. The second issue, is avoiding overfitting. Overfitting mean the constructed model is so flexible to fit even the noise. To avoid overfitting, only part of the data can be used to construct the meta-model (training data), and the other part will be used for testing. If the collected data has initially noise, the regression Kriging (not interpolation) can handle it effectively [10].

It is time now to test the model. For interpolation meta-models, this test is performed via validation. If we have large data set (say 25 d), $1/4$ of the data will be kept for testing, and the remaining part will be used to construct the meta-model. Then the accuracy of the model can be tested using the root mean squared error (RMSE). RMSE$= \sqrt{\sum_{i=0}^{n_t} (y^{(i)} - \hat{y}^{(i)})^2 / n_t}$, where y is the given value, $\hat{y}$ is the corresponding predicted value and n_t is the number of tested points. Alternative metric is the correlation coefficient (r^2), $r^2 = \left(cov(\mathbf{y},\hat{\mathbf{y}}) / \sqrt{var(\mathbf{y})\ var(\hat{\mathbf{y}})}\right)^2$. If the accuracy of the given data is known via the standard deviation σ^2, it would be better to have RMSE close to σ^2, otherwise the normalized RMSE (RMSE/(range of tested value)) of 10% is acceptable as a rule of thumb [1]. Moreover, $r^2 > 0.8$ is an acceptable value. The above-mentioned approach can be used to test the effect of number of training points

on the constructed surrogate model. If no valuable change happen in the correlation coefficient, the meta-model is called globally saturated, whereas a better performance will be obtained locally at the added points. In many engineering problems, the data set is limited and the simulation/experiment is expensive, so subdividing the data into 1:3 is not possible. In this case, the cross validation is the suitable approach to judge the meta-model accuracy.

Cross-validation is a technique for assessing how the results of a statistical analysis will generalize to an independent data set [11,12]. It is mainly used in settings where the goal is prediction, and one wants to estimate how accurately a predictive model will perform in practice. Cross-validation is important in guarding against testing hypotheses suggested by the data (called "Type III errors":Solving the wrong problem), especially where further samples are hazardous, costly or impossible to collect. There are two widely used types; k-fold cross validation and one-leave out cross validation. In the k-fold cross-validation, the original sample is randomly partitioned into k subsamples. Of the k subsamples, a single subsample is retained as the validation data for testing the model, and the remaining $k-1$ subsamples are used as training data [13]. The cross-validation process is then repeated k times (the folds), with each of the k subsamples used exactly once as the validation data. The k results from the folds then can be averaged (or otherwise combined) to produce a single estimation. The advantage of this method over repeated random sub-sampling is that all observations are used for both training and validation, and each observation is used for validation exactly once [12]. 10-fold cross-validation is commonly used [14]. As the name suggests, leave-one-out cross-validation (LOOCV) involves using a single observation from the original sample as the validation data, and the remaining observations as the training data. This is repeated such that each observation in the sample is used once as the validation data. This is the same as a k-fold cross-validation with k being equal to the number of observations in the original sample. Leave-one-out cross-validation is computationally expensive because it requires many repetitions of training [12], but for limited data it is preferred than the k-fold cross validation.

The constructed surrogate can be enhanced by adding extra points. If the target is to obtain a more accurate optimal solution, more points can be added iteratively close to the optimum obtained from the previously constructed surrogate model, the process continue until no valuable change is observed in the optimal solution. This approach called local exploitation. Another way is via adding points at the location of highest uncertainty in the prediction, which is available in Gaussian process based surrogate e.g., Kriging. The authors think this approach is only valid for convex problem (no-local optimum). Moreover, a stochastic optimization technique is essential e.g., genetic algorithms. The more preferred approach in adding extra point is called balanced exploitation and exploration, one good example is *the expected improvement infill strategy* [1]. In this strategy, more points will be added at the location where the expected improvement is highest. The iterative procedure stops when the expected improvements falls below certain threshold.

Objectives

The main targets of this study are:
- Apply the screening to identify the most significant factors using the elementary effect approach.

- Investigation the potential of Co-Kriging in high dimensional industrial problems.
- Compare Co-Kriging against ordinary Kriging, gradient-enhanced Kriging and blind Kriging.
- Apply the fitted surrogate model for single-objective optimization using genetic algorithms.
- Perform CFD simulations to compare the basic design and the optimal one.

THEORY

A surrogate model is an engineering method used when an outcome of interest cannot be easily directly measured, so a model of the outcome is used instead [15]. Most engineering design problems require experiments and/or simulations to evaluate design objective and constraint functions as function of design variables. For example, in order to find the optimal airfoil shape for an aircraft wing, an engineer simulates the air flow around the wing for different shape variables (length, curvature, etc.). For many real world problems, however, a single simulation can take many minutes, hours, or even days to complete. As a result, routine tasks such as design optimization, design space exploration and sensitivity analysis become impossible since they require thousands or even millions of simulation evaluations. *One way of alleviating this burden is by constructing approximation models, known as surrogate models* (also known as, response surface models or metamodels) that mimic the behavior of the simulation model as closely as possible while being computationally cheap(er) to evaluate.

Kriging (KG)

Originally, Kriging was developed - by the South African mining engineer Daniel Krige - for interpolation in geostatistical or spatial sampling. Later on, Kriging was applied to the input/output (I/O) data of deterministic simulation models; see [16,17]. Kriging may enable adequate approximation of the simulation I/O function, even when the simulation experiment covers a big input area; i.e., the experiment is global, not local. Kriging assumes that the function being studied is a realization of a Gaussian stochastic process $Y(x) = \mu + Z(x)$ where x is a point in a d-dimensional search space, μ is its mean, and $Z(x)$ is a zero-mean, stationary, Gaussian stochastic process [18]. Kriging is categorized as ordinary Kriging if μ is a constant. This is the most popular incarnation of Kriging in the engineering sciences [9]. There is also simple Kriging where $\mu = 0$.

In universal Kriging, $\mu = \sum_{i=0}^{m} \mu_i v_i(\mathbf{x})$ where the v_i are some known functions and the are μ_i unknown parameters. Usually μ_i takes the form of a low-order polynomial regression. The idea is that μ_i captures known trends in the data and basis functions added to this will fine-tune the model, thus giving better accuracy than ordinary Kriging where a constant μ_i is used. However, we do not usually have a priori knowledge of the trends in the data and specifying them may introduce inaccuracies. Hence the popularity of ordinary Kriging. More recently, blind Kriging has been applied in optimization problems [19]. In blind Kriging, v_i are identified through some data -analytic procedures. Hopefully, if the underlying trends can be identified, the ensuing model will be more accurate than ordinary Kriging [20].

Co-Kriging

Recently, the meta-models have been used as tool to calibrate the less accurate (simplified) codes. Such multi-fidelity or multi-level approaches can also deal with experimental data and CFD simulations data or for CFD simulation on different grid levels, or fully developed flow results and developing flow results [1]. Co-Kriging is a typical example of surrogate models which can handle multi-fidelity data. Co-kriging can be considered as a natural extension to ordinary Kriging.

This way of adding a surrogate based on plentiful cheap data to a hopefully more simple surrogate based on sparse expensive data could employ any surrogate modelling method, but Kriging's multi- output variant, Co-Kriging, with its parametric Gaussian basis functions is particularly attractive. More recently Kennedy and O'Hagan [21] used Co-Kriging to correlate finite element analyses based on a coarse mesh with those based on a fine mesh. They used a formulation which assumes the complex process is equal to the simple process, multiplied by some ratio, plus a difference process. The ratio and difference process are determined using a set of collocated data. In engineering design this method has been cast in a global optimization context and used to optimize a transonic wing design by combining an empirical drag prediction code with computational fluid dynamics (CFD) [22].

Efficient methods for global aerodynamic optimization using computational fluid dynamics simulations should aim to reduce both the time taken to evaluate design concepts and the number of evaluations needed for optimization. Forrester [23] investigate methods for improving such efficiency through the use of partially converged computational fluid dynamics results. These allow surrogate models to be built in a fraction of the time required for models based on converged results. The proposed optimization methodologies increase the speed of convergence to a global optimum while the computer resources expended in areas of poor designs are reduced. A strategy which combines a global approximation built using partially converged simulations with expected improvement updates of converged simulations is shown to outperform a traditional surrogate-based optimization.

Screening And Sampling Plans

To build global models of unknown landscapes, a sampling plan with a uniform, but not regular, spread of points across the design space makes intuitive sense. We also wish to use a sample of points whose projections onto each variable axis are uniform, the logic being that it is wasteful to sample a variable more than once at the same value. The Latin hypercube [5] sampling technique is more favorable [24-29] especially using the Morris and Mitchell's optimal Latin hypercube approach [1,7].

There are three levels of sampling plan (also known as design of experiment (DoE)). The first level is for the screening step. i.e. selection of the most significant factors. This step is only needed if the most significant factors are not known a priori, or to reduce the number of independent variables. The second level is for constructing the surrogate model for the most significant factors. The third level is to augment the constructed surrogate. This procedure is known as infill criteria.

Initial Sampling

The term `Screening Design' refers to an experimental plan that is intended to find the few significant factors from a list of many potential ones [30] i.e., to identify significant main effects, rather than interaction effects, the latter being assumed an order of magnitude less important. Even when the experimental goal is to eventually fit a response surface model of second order, the first experiment should be a screening design when there are many factors to consider. Phoa et al. [31] stated that, It is risky to neglect the effect of interaction between factors during the screening design. Ignoring the interaction can result in wrong statistical inferences, including biased estimates, missing out on important factors and detection of spurious factors.

To determine the most significant factors, the authors suggest two approaches. In case of using polynomial regression surrogate, after the design of experiment using say central composite design, Box-Behnken design or similar DoE approach, the analysis of variance and Pareto chart, can assist in deciding the most significant factors [32]. Whereas this approach is more accurate than the widely used one with only first order terms, this approach cannot detect any interaction than assumed in the design of experiment. The more promising approach is the screening algorithm proposed by Morris [2]. The proposed experimental plans are composed of individually randomized one-factor-at-a-time designs, and data analysis is based on the resulting random sample of observed elementary effects (cf. Sec. 2.3.2).

The straightforward way of sampling a design space in a uniform fashion (using rectangular grid of points, e.g., full-factorial design) will satisfy the uniformity criteria but suffer from two major drawbacks [1]. (1) It is only defined for designs of certain sizes, i.e. number of points $n = \Pi_{i=1}^{d} q_i$, where q_i is the number of level for factor i and d is the problem dimensionality. (2) Sets of points will overlap on the projections, and some gaps in the design space may occur. This drawbacks can be overcome using the Latin hypercube sampling (LHS).

As stated before in Sec. 1.1, in LHS, the design space is splitted into equal sized hypercubes (bins) and placing points in the bins (one in each), making sure that from each occupied bin we could exit the design space along any direction parallel with any of the axes without encountering any other occupied bins [1]. The generated LHS satisfy the multi-dimensional stratification (projections on the axes are uniformly spread). There is no guarantee that the generated LHS is a good (fill the available space uniformly) or bad design of experiment.

In order to measure the uniformity of the sampling plan (aka, space-fillingness), the *maxmin* metric introduced by Johnson et al. [6] is widely used. The criterion based on this metric can be explained as follow:
Let $d_1, d_2, \cdots, d_m$ be the list of the **unique** values of distances between all possible pairs of points in a sampling plan **X**, sorted in ascending order. Further, let $J_1, J_2, \cdots, J_m$ be defined such that J_j is the number of pairs of points in **X** separated by the distance d_j. Johnson et al. [6] call **X** a maximin plan among all available plans if it maximizes d_1 and, among plans for which this is true, minimizes J_1. This definition might yield several *maximin* designs. Therefore Forrester et al. [1] recommend to use what is called the more complete a tie-breaker definition of Morris and Mitchell [7]. They call **X** the maximin plan among all available plans if it

maximizes d_1, among plans for which this is true, minimizes J_1, among plans for which this is true, maximizes d_2, among plans for which this is true, minimizes $J_2, \cdots,$ minimizes J_m [1].

The distance metric widely used is the p-norm of the space, Eq. (1)

$$d_p\left(\mathbf{x}^{(i_1)}, \mathbf{x}^{(i_2)}\right) = \left(\sum_{j=1}^{k} |x_j^{i_1} - x_j^{i_2}|^p\right)^{1/p} \tag{1}$$

$p = 2$ yields the Euclidean norm and $p = 1$ is the rectangular distance, which is computationally cheap, and will be used hereafter. Morris and Mitchell [7] defined the following scalar-valued criterion function $\boldsymbol{\Phi}_q$ used to rank competing sampling plans. *The smaller the value of $\boldsymbol{\Phi}_q$, the better the space-filling properties of $\boldsymbol{X}$ will be* [1].

$$\boldsymbol{\Phi}_q(\mathbf{X}) = \left(\sum_{j=1}^{m} J_j d_j^{-q}\right)^{1/q} \tag{2}$$

Morris and Mitchell [7] recommend minimizing $\boldsymbol{\Phi}_q$ for $q = $ 1,2,5,10, 20, 50 and 100 after testing all these qs values, one can choose the best of the resulting plans according to the actual maximin definition. In order to obtain the optimal Latin hypercube sampling, an optimization procedure using genetic algorithms. After many iterations, the sampling plan which will produce the minimum $\boldsymbol{\Phi}_q$. For more details, we refer to Forrester et al. [1] and Morris and Mitchell [7].

Screening

For surrogate based optimization, it is always preferred to minimize the number of design variable. This can be achieved when a sufficient knowledge about the research study is known apriori. The application of dimensionless numbers (e.g. Reynolds number) reduces the number of variables considerably and helps in generalizing the obtained results. In shape design optimization, the design variables are numerous, and no sufficient knowledge is available apriori. To efficiently decide that, a screening study is needed, one common approach is via polynomial regression model after creating a design of experiment, then from the analysis of variance, one could decide the most significant parameters, e.g., [32]. In this study, the Morris algorithm [1,2] will be applied. A brief description is given below.

Elementary effects using Morris algorithm
It is worthful to mention that to avoid scaling issues (the constructed meta-model will be biased by the higher values), each x_i will be assumed scaled in the interval [0,1]. The region of interest D is d-dimensional unit hypercube. If we know the function $y = y(x_i)$, so based on the derivative $\partial_i(\mathbf{x}) := \partial y\, \partial x_i|_{\mathbf{x}}$ the following cases may be obtained. $\partial_i(\mathbf{x})$ may be (a) zero over all values of $\mathbf{x}$, (b) a non zero constant over all values of $\mathbf{x}$, (c) a non-constant function of x_i, or (d) a non-constant function of one or more $x_j, (j \neq i)$. These corresponding to the following effect of x_i on y: (a) negligible, (b) linear, (c) non-linear or (d) involved with interaction with other inputs [2] . The major role of the preliminary experiment is to determine which inputs may have effects which are either (a) negligible (then remove it from the design variable), (b) linear, and (c or d) using limited number of experiments.

Let us restrict our design space D to a d-dimensional, p-level full factorial grid, that is $x_i \in \{0, 1/(p-1), 2/(p-1), \cdots, 1\}$, for $i = 1, \cdots, d$. For a given baseline value $\mathbf{x} \in D$, let $d_i(\mathbf{x})$ denote the elementary effect of x_i, where

$$d_i(\mathbf{x}) = \frac{y(x_1, x_2, \cdots, x_{i-1}, x_i + \Delta, \cdots, x_d) - y(\mathbf{x})}{\Delta} \tag{3}$$

where $\Delta = \xi/(p-1)$, ξ is a step length factor and $\mathbf{x} \in D$ such that its components $x_i \leq 1 - \Delta$.

According to Morris's method, large measure of central tendency indicates a variable with an important influence on the objective function and a large measure of spread indicates a variable involved in interactions and/or in nonlinear terms. Practically, the sample mean and the sample standard deviation are estimated of a set of $d_i(\mathbf{x})$ values calculated in different parts of the design space. The sampling plan should produce a certain number (say, r) elementary effects for each variable, independently drawn with replacement [1]. Forrester et al. [1] described the steps as follows. Let $\mathbf{B}$ denotes a $(d+1) \times d$ sampling matrix of 0s and 1s with the property that for every column $i = 1, 2, \cdots, d$ there are two rows of $\mathbf{B}$ that differ only in their i^{th} entries. Then compute a random orientation of $\mathbf{B}$, denoted by $\mathbf{B}^*$ [1]:

$$\mathbf{B}^* = (\mathbf{1}_{d+1 \times 1}\mathbf{x}^* + (\Delta/2)[(2\mathbf{B} - \mathbf{1}_{d+1 \times d})\mathbf{D}^* + \mathbf{1}_{d+1 \times d}])\mathbf{P}^* \tag{4}$$

where $\mathbf{D}^*$ is a d-dimensional diagonal matrix, where each element on the diagonal is either +1 or -1 with equal probability, $\mathbf{1}$ is a matrix of 1s, $\mathbf{x}^*$ is a randomly chosen point in our discretized, p-level design space and $\mathbf{P}^*$ is a $d \times d$ random permutation matrix in which each column contains one element equal to 1 and all others equal to 0 and there is no two columns have 1s in the same position. To obtain r elementary effects for each variable, the screening plan is built from r random orientations, i.e., $\mathbf{X} = [\mathbf{B}_1^*, \mathbf{B}_2^*, \cdots, \mathbf{B}_r^*]â€™$. The next step is computing the value of f for each row of $\mathbf{X}$ and store these objective function values in the $r((d+1) \times 1)$ column vector $\mathbf{t}$. Taking one random orientation at a time, the adjacent rows of the screening plan and the corresponding function values from $\mathbf{t}$ can be inserted into Eq. (4) to obtain d elementary effects (one for each variable). The final output is a graphical representation of the sample means $\bar{d}_i$ against sample standard deviations S_i. Morris used $r = p = 4$ and $\xi = p/2$ in his example [2]. The same values will be used in the current study. If we do not have any input/output relationship for the objective function (even based on mathematical model), we will need to construct a surrogate model first, possibly with few samples say $5d$.

BENCHMARK FUNCTIONS AND INDUSTRIAL TEST CASE

Benchmark Functions

For demonstration of the screening of high dimensional problems, the authors suggest a second order polynomial in four variables. Moreover, both the Sphere and the Rosenbrock function are tested.

Simple Four-Dimensional Problem

The first example to demonstrate the application of screening is using the following simple analytical function:

$$y = 20\,x_1 - 50\,x_2\,x_3 + x_3^2 + 0.5x_4^2 \qquad (5)$$

Using $d = 4, p = 4, r = 4, \xi = 2$, Fig. 3 (a) indicates the relatively negligible effect of x_4 as it is located close to zero mean and standard deviation. x_1 has the biggest linear effect. x_2 and x_3 have higher order relationships or interactions.

Another helpful and cheap plots, are the Tile plot and four-variables Nested plot presented in Forrester et al. [1]. Whereas, they used it as a data visualization plots, the authors recommend them as a visual inspection of the effect of each variable after constructing the surrogate model (and before any further enhancement). These plots can be assist to take the decision of neglecting some variables for high dimensional problems.

Tile plot represents an array of contour plots. The baseline values (means) and the ranges were used to generate contour plots of the function (or its surrogate) by varying the inputs pairwise and keeping the remaining variables at the baseline value. Figure 1(c)   represents the tile plot for Eq. ((5)) using a range of $[0; 1]^4$. The left bottom contour plot represent the function variation with x_1 and x_4 while holding x_2 and x_3 at the baseline values ($[0]^2$); The tile plot clearly present that: (a) no interaction between x_4 and x_1, x_2, x_3. Moreover, it has minor effect on the function value (vertical lines means no change with the y-axis variable). (b) strong interaction between x_2 and x_3 (nonlinear curves) (c) Both x_1, x_2, x_3 have major effects on the function, and the effect of x_2, x_3 are close (almost the same line slope).

The four-variable nested plot is a kind of tile plot, but here x_3 varies along the horizontal axis of each tile, x_4 along the vertical axes, while the values of x_1 and x_2 can be read off the bottom of each column of tiles and the beginning of each row respectively. Figure 1(d) represent the four-variable nested plot for Eq. (5) using a range of $[0; 1]^4$. The left bottom contour plot represent the function variation with fixed values for $x_1 = 0$ and $x_2 = 0$ whereas x_3 varies on x-axis and x_4 varies in y-axis from 0 to 1. The following conclusions can be drawn from the visual inspection of Fig. 1(d): (a) When $x_1 = 0, x_2 = 0$, the function has a nonlinear behavior with the change in x_3 and x_4. (b) Once x_1, x_2 become non-zero, the effect of x_4 on y becomes minor (the curved lines at the bottom left plot becomes vertical by increasing either x_1 or x_2). The insignificant effect becomes very clear if we change the variable order of plotting as shown in Fig. 1(e). For more than four variable, many nested plots can be used to understand the input-output relationship.

Sphere function
The sphere function for four variables is given as, $y = \sum_{i=1}^{4} x_i^2$. Figure 2 represents the 3D plots, the elementary effect distributions and the Tile plot. The quadratic function behavior is clear from the 3D plots and the Tile plot. The elementary effect distributions exhibit the significant influence of all the four variables.

Rosenbrock function
In this example, a modified version of Rosenbrock function [33] has been used with $d = 4$. In this study, $x_i \in [-1; 1]^4$.

$$y = \frac{1}{206}\left[\sum_{i=1}^{d-1} 100(x_{i+1} - x_i^2)^2 + (1 - x_i)^2 + \sum_{i=1}^{d} 75\, sin(5(1 - x_i)) - 300\right] \quad (6)$$

Figure 3 represents the 3D plots, the elementary effect distributions and the Tile plot. The highly nonlinear relationship function behavior is clear from the 3D plots and the Tile plot. The elementary effect distributions exhibit the significant influence of all the four variables. x_2 is highly nonlinear with possible interaction with the other variables (high standard deviation). x_1 exhibits the lowest possible interaction with the other variables.

Cyclone Separator Pressure Drop

Cyclones are one of the most widely used separators, which rely on centrifugal forces to separate particles from a gas stream. The primary advantages are the economy, simplicity in construction and adaptability to a wide range of operating conditions. Reversed flow cyclones with a tangential inlet are the most common cyclone design as shown in Fig. 4(a). It consists of seven main geometrical parameters: inlet section height a and width b, cylinder height h, cyclone total height H_t, dust exit diameter (cone tip diameter) B_c, gas outlet diameter (also, called the vortex finder diameter) D_x and vortex finder length S. All these parameters always given as ratios of the cyclone body diameter D. It is generally known that these seven dimensions characterize the collection efficiency (cut-off diameter) and the pressure drop of the cyclone separator [13, 32,34, 35].

The Euler number
The pressure drop across cyclone essentially depends on the cyclone dimensions and operating conditions. Generally, it is proportional to the average dynamic pressure at the inlet and is often defined as [13,36]

$$\Delta P = E_u \left(\frac{1}{2}\rho_g V_{in}^2\right) \quad (7)$$

where E_u is Euler number (the dimensionless pressure drop, also called pressure drop coefficient [36]). The Euler number is a complex nonlinear function of the cyclone geometrical dimensions and is not affected by operating conditions in the high Reynolds number ($R_e > 5E4$) [32,37]. The Euler number will be constant for any cyclone configuration regardless of size as long as the dimension ratios remain the same, although pressure drop varies with different operating conditions (due to the effect of ρ_g and V_{in}). Therefore, pressure drop can be established by determining experimentally or theoretically for a particular cyclone design and also be modified by the semi-empirical correlations to take the effect of solid loading [36].

In order to determine the Euler number more accurately, all eight dimensions of the cyclone are selected to establish the surrogate models because they have the effect on the Euler number to different extent [36,38]. Usually, these dimensions can be characterized by the barrel diameter D and expressed as seven dimensionless geometric ratios [13,36]:

$$E_u = f\left(\frac{D_x}{D}, \frac{a}{D}, \frac{b}{D}, \frac{S}{D}, \frac{H_t}{D}, \frac{h}{D}, \frac{B_c}{D}\right) \qquad (8)$$

According to Eq. (8), seven independent dimensionless geometrical variables and one dependent variable (the Euler number of the cyclone) are selected as respectively the input and output parameters in the surrogate model, as presented in Table 1. For simplicity the division of each factor by the barrel diameter D will be dropped hereafter.

Screening using Muschelknautz Method of Modeling (MM)

The authors have performed in a previous article [32] a sensitivity analysis for the effect of geometrical parameters on the dimensionless pressure drop (Euler number) of gas cyclones. In that study, we fitted a second order polynomial regression for a table of experiment generated using Box-Behnken design of experiment. The analysis of variance (ANOV) [32] and both the main effect plot and Pareto chart [32] indicated that the most significant factors are four; namely, the vortex finder diameter D_x, the inlet height a, the inlet width b and the total cyclone height H_t as shown in Fig. 4(a). Figure 4(b) presents 3D plots for the MM model. Note that, the actual values for the seven geometrical parameters are given in Table 2 but these values are scaled in the range of zero to unity for all variables as shown in Fig. 4(b).

The screening method proposed by Morris [2] has been applied to produce Fig. 4(c). It is clear that, both the cone tip diameter B_c, the vortex finder length S and the cyclone barrel height h are insignificant factors. Moreover, D_x has a negative effect, and both a and b have positive effect on the Euler number, which are the same conclusions stated before [32]. The interaction between variables can be deduced from the Tile plot. The interaction between a with both b and D_x is very clear. Moreover, the significant effect of D_x is very clear from the Tile plot (Fig. 4(d)).

OPTIMUM DESIGN USING CO-KRIGING

In this study, the high fidelity data has been obtained from CFD simulations using the Reynolds stress turbulence model [39], and the low fidelity data are calculated using the Sphered and Lapple model [40] ($Eu = 16\ a\ b/(D_x^2)$). The high fidelity data represents the variation of the most significant (four) variables; the vortex finder diameter D_x, the inlet section width b and height a and the total cyclone height, whereas, the low-fidelity model depicts only the effect of the first three significant factors. This can be considered as a sever approximation (i.e., difficult test for the multi-fidelity model). For better results, the MM model is a good alternative [13].

Figure 5 indicates the superior performance of the Co-Kriging model (in terms of the absolute error) over the ordinary Kriging, DACE and the blind Kriging using only the high-fidelity data.

Previous Optimization Studies

In 1951, Stairmand [43] presented the geometrical ratios for high-efficiency cyclones. Until now, these ratios are still in use. Elsayed and Lacor [32] reported the following

shortages in the Stairmand model for pressure drop calculation [44] which has been used to obtain these geometrical ratios: (1) the velocity distribution has been obtained from a moment-of-momentum balance, estimating the pressure drop as entrance and exit losses combined with the loss of static pressure in the swirl, i.e., neglecting the entrance loss by assuming no change of the inlet velocity occurs at the inlet area; (2) assuming a constant friction factor [37]. Due to the wide range of industrial applications of the cyclone separator, it was a matter of study for decades. However, the optimization studies on it is quite limited in literature. Moreover, many of these studies are not coherent studies or focus on a specific problem.

Ravi et al. [45] carried out a multi-objective optimization study on a set of N identical reverse-flow cyclone separators in parallel by using the non-dominated sorting genetic algorithms (NSGA). Two objective functions were used: the maximization of the overall collection efficiency and the minimization of the pressure drop. Non-dominated Pareto optimal solutions were obtained for an industrial problem in which 165 m^3/s of air was treated. In addition, optimal values of several decision variables, such as the number of cyclones and eight geometrical parameters of the cyclone, are obtained. Their study shows that the diameters of the cyclone body and the vortex finder, and the number of cyclones used in parallel, are the important decision variables influencing the optimal solutions. Moreover, their study illustrates the applicability of NSGA in solving multi-objective optimization problems involving gas-solid separations. The main drawbacks are: (1) They used the model of Shepherd and Lapple [40] for predicting the dimensionless pressure drop (Euler number). In Shepherd and Lapple model, the Euler number depends on only three factors ($Eu = 16a\,b/D_x^2$) and they used it to optimize the seven geometrical parameters. (2) The barrel diameter, number of parallel cyclones and the gas velocity have been included into the optimization design space.

Consequently, it is not devoted to the geometrical ratio. (3) They used many side constraints on the geometrical values ($0.4 \le a/D \le S/D$, $0.15 \le b/D \le (1 - D_x/D)/2$ if $0.5 \le D_x/D \le 0.6$) these constrains prevent searching for the global optimization geometrical ratios for the seven geometrical parameters. (4) No table for the non-dominated Pareto front points is presented from which the designer can select a certain geometrical ratios set. Safikhani et al. [46] performed a multi-objective optimization study on cyclone separators. First, they simulated many cyclones to obtain the pressure drop and the cut-off diameter and used artificial neural network approach to obtain the objective function values. Finally, a multi- objective genetic algorithms are used for Pareto based optimization of cyclone separators considering two conflicting objectives. However, the design variables were only four (instead of seven), the barrel height, the cone height, the vortex finder diameter and length. So they ignored the effect of the inlet dimensions, which has been acknowledged by other researchers as significant geometrical parameters for the cyclone flow field and performance (cf. Elsayed and Lacor [47, 32, 13, 35]). Moreover, they did not explain why they selected these particular parameters. Furthermore, they applied four side constraints on the four tested variables, which prevent searching for the global optimization. Safikhani et al. [48] carried out a multi-objective optimization study using the genetic algorithms to obtain the best vortex finder dimension (diameter and length) and shape (convergent and divergent). Four design variables have been investigated; vortex finder diameter, angle, the upper-part and lower-part lengths of the vortex finder. They applied neural networks to obtain a meta-model for the

pressure drop and collection efficiency from CFD dataset. The main shortage of the Safikhani et al. [48] study are: (1) They used dimensional values instead of dimensionless. Moreover, they applied side constraints which prevent the optimization procedure from obtaining global optimization. (2) The selection of only the vortex finder dimensions as the design variables and neglecting the interaction with the other dimensions, specifically the inlet dimensions [35, 13, 49].

Optimization Using Genetic Algorithms

Genetic algorithms (GAs) are one of the optimization methods finding wide application in optimization problems [50]. Genetic algorithms searches stochastically through the real space of the problem by generating a random initial population. GA technique is one kind of evolution-based systems which measures the fitness of each individual in population, and then selects the fittest individuals until reproducing the intermediate population. The genetic operators affect some individuals in this population and produce the next population for the new generation through selection, crossover, and mutation operations. The GA could optimize linear and nonlinear objective functions by exploring the space of the problem.

The genetic algorithms optimization technique has been applied to obtain the geometrical ratios for minimum pressure drop (Euler number). The objective function is the Euler number (using the trained Co-Kriging model). The design variables are four geometrical dimensions, the inlet height a, the inlet width b, the vortex finder diameter D_x and the total cyclone height H_t. These four variables are the most significant factors which affect the cyclone performance [32].

Table 3 presents the settings used to obtain the optimum design for minimum pressure using global optimization Matlab toolbox in Matlab 2012b commercial package. Table 4 gives the optimum values for cyclone geometrical parameters for minimum pressure drop estimated by CFD simulations (cf., Section 4.3). It is clear from Table 4 that the new optimized design is very close to the Stairmand design in many geometrical parameters, whereas the new ratios will result in the minimum pressure drop.

CFD Comparison between the Stairmand and Optimal Design

The Fluent solver has many turbulence models available for simulating turbulent flow. It is generally recognized that only the Reynolds stress model (RSM) and large eddy simulation (LES) can capture the main features of the highly complicated swirling flow in cyclone separators [32, 35, 51- 60]. The Reynolds stress turbulence model has been used in this study to reveal the turbulent flow in the two cyclone separators. For the detailed governing equation the reader can refer to Elsayed and Lacor [35,39]. The geometrical values are given in Table 5 for the two cyclones.

Numerical settings
The air volume flow rate Q_{in}=50 l/min for the two cyclones, air density is 1.0 kg/m^3 and dynamic viscosity 2.11E-5 $Pa\ s$. The turbulent intensity equals 5% and characteristic length equals 0.07 times the inlet width [61]. A velocity inlet boundary condition is applied at inlet, outflow at gas outlet and wall boundary conditions at all other boundaries. The finite volume method has been used to discretize the partial

differential equations of the model using the SIMPLEC (Semi-Implicit Method for Pressure-Linked Equations-Consistent) method for pressure velocity coupling and QUICK scheme to interpolate the variables on the surface of the control volume. The implicit coupled solution algorithm was selected. The unsteady Reynolds stress turbulence model (RSM) was used in this study with a time step of 0.0001s. The interested reader can refer to Elsayed and Lacor [39] for more details about the numerical settings.

Flow field pattern

The dimensionless static pressure distribution presented in Fig. 6 for the two cyclones indicates that the highest dimensionless static pressure for the Stairmand design is more than twice that of the new design at all sections whereas the central value is almost the same for the two cyclones. This indicates that, the new design has a lower dimensionless pressure drop than the Stairmand design. However, these results are obtained at different inlet velocity for the two cyclones (to have the same air flow rate). The same Euler number values would be obtained if the two cyclones work at the same inlet velocity because the Euler number is not a function of flow velocity if the Reynolds number is higher than 2E4 [32].

The tangential velocity profile is composed of two regions. In the inner region, the flow rotates approximately like a solid body (forced vortex), where the tangential velocity increases with radius. After reaching its peak the velocity decreases with radius in the outer part of the profile (free vortex). The tangential velocity distributions for the two cyclones are similar in pattern (Rankine profile). The inner part of the tangential velocity distribution of the two cyclones is very similar. The outer part for the new design is lower than that for the Stairmand cyclone. This implies less swirl in the optimal cyclone where the particles are subjected to smaller centrifugal force and consequently less collection efficiency. This is expected result because the vortex finder diameter has conflicting effects on the Euler number and the collection efficiency [62]. To avoid this pitfall, a multi-objective optimization process is needed.

CONCLUSIONS

The following conclusions can be drawn from analysis of the obtained results:
- Fitting a surrogate model using multiple levels of fidelity can enhance the accuracy of a surrogate model.
- Co-Kriging super perform the ordinary Kriging and blind Kriging.
- Co-Kriging can be applied to industrial test cases.
- A new optimal cyclone design for minimum pressure drop has been obtained.

As a future extension of this study, the following issues still need more investigation.
- Consider multi-objective optimization problem (Euler number and Stocks number).
- Apply Co-Kriging using MM model as a high-fidelity model and Sphered and Lapple model as a low-fidelity model with optimized LHS and exchange algorithm. In addition to use the expected improvement algorithm for sampling updating.
- Apply the infill criteria and develop a new one for multi-objective optimization problems.

- Test more surrogate models for high-dimensional problems such as support vector machine.

REFERENCES

[1]   A. I. J. Forrester, A. Sobester, A. J. Keane, Engineering design via surrogate modelling: A practical guide, J. Wiley, Chichester, West Sussex, England, 2008.

[2]   M. Morris, Factorial sampling plans for preliminary computational experiments, Technometrics 33 (2) (1991) 161-174.

[3]   G. Box, D. Behnken, Some new three level designs for the study of quantitative variables, Technometrics 2 (1960) 455-475.

[4]   R. H. Myers, D. C. Montgomery, C. M. Anderson-Cook, Response Surface Methodology: Process and Product Optimization Using Designed Experiments, 3rd Edition, Wiley series in probability and statistics, Wiley, Hoboken, N.J., 2009.

[5]   M. D. McKay, R. J. Beckman, W. J. Conover, Comparison of three methods for selecting values of input variables in the analysis of output from a computer code, Technometrics 21 (2) (1979) 239-245.

[6]   M. E. Johnson, L. M. Moore, D. Ylvisaker, Minimax and maximin distance designs, Journal of Statistical Planning and Inference 26 (1990) 131-148.

[7]   M. Morris, T. Mitchell, Exploratory designs for computational experiments, Journal of Statistical Planning and Inference 43 (3) (1995) 381-402.

[8]   D. R. Jones, M. Schonlau, W. J. Welch, Efficient global optimization of expensive black-box functions, Journal of Global Optimization 13 (4) (1998) 455-492.

[9]   A. I. J. Forrester, A. J. Keane, Recent advances in surrogate-based optimization, Progress In Aerospace Sciences 45 (1-3) (2009) 50 - 79.

[10]   I. Couckuyt, T. Dhaene, P. Demeester, ooDACE toolbox-A Matlab Kriging toolbox: Getting started v1.3, http://www.sumo.intec.ugent.be/?q=ooDACE (2012).

[11]   R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, 1995, pp. 1137-1143.

[12]   Wikipedia,Cross-validation, http://en.wikipedia.org/wiki/Cross-validation_(statistics)   (2011).

[13]   K. Elsayed, C. Lacor, Modeling, analysis and optimization of aircyclones using artificial neural network, response surface methodology and CFD simulation approaches, Powder Technology 212 (1) (2011) 115-133.

[14]   G. J. McLachlan, K.-A. Do, C. Ambroise, Analyzing microarray gene expression data, Wiley Series in Probability and Statistics, Wiley, 2004.

[15]   Wikipedia, Surrogate model,   http://en.wikipedia.org/wiki/Surrogate_model (2012).

[16]   J. Sacks, W. J. Welch, T. J. Mitchell, H. P. Wynn, Design and analysis of computer experiments, Statistical science 4 (4) (1989) 409-435.

[17]   T. J. Santner, B. J. Williams, W. Notz, The Design and analysis of computer experiments, Springer, New York, 2003.

[18]   J. Kleijnen, W. van Beers, I. van Nieuwenhuyse, Expected improvement in efficient global optimization through bootstrapped kriging, Journal of Global Optimization (2011) 1-15.

[19]   I. Couckuyt, A. Forrester, D. Gorissen, F. D. Turck, T. Dhaene, Blind kriging: Implementation and performance analysis, Advances in Engineering Software 49 (0) (2012) 1 - 13.

[20]   V. R. Joseph, Y. Hung, A. Sudjianto, Blind kriging: a new method for developing metamodels, Journal of Mechanical Design 130 (3) (2008) 31-102.

[21]   M. C. Kennedy, A. O'Hagan, Predicting the output from complex computer code when fast approximations are available, Biometrika 87 (1) (2000) 1-13.

[22]   A. I. Forrester, A. Sobester, A. J. Keane, Multi-fidelity optimization via surrogate modelling, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science 463 (2088) (2007) 3251-3269.

[23]   A. I. J. Forrester, A. J. Keane, N. Bressloff, Design and analysis of noisy computer experiments, AIAA Journal 44 (10) (2006) 2331-2336.

[24]   J. L. Deutsch, C. V. Deutsch, Latin hypercube sampling with multidimensional uniformity, Journal of Statistical Planning and Inference 142 (3) (2012) 763 - 772.

[25]   H. Janssen, Monte-carlo based uncertainty analysis: Sampling efficiency and sampling convergence, Reliability Engineering and System Safety 109 (2013) 123 - 132.

[26]   C. Diaz, M. Victoria, O. M. Querin, P. Mart, Optimum design of semi-rigid connections using metamodels, Journal of Constructional Steel Research 78 (0) (2012) 97 - 106.

[27]   R. Stocki, T. Szolc, P. Tauzowski, J. Knabel, Robust design optimization of the vibrating rotor-shaft system subjected to selected dynamic constraints, Mechanical Systems and Signal Processing 29 (0) (2012) 34 - 44.

[28]   Z. Khatir, H. Thompson, N. Kapur, V. Toropov, J. Paton, Multi-objective computational fluid dynamics (cfd) design optimisation in commercial bread-baking, Applied Thermal Engineering (2012) 1-7 http://dx.doi.org/10.1016/j.applthermaleng.2012.08.011.

[29]   T. Figueiro, J. Tortai, P. Schiavone, Sensitivity analysis for accurate determination of PSF parameters, Microelectronic Engineering 97 (0) (2012) 77 - 80.

[30]   C. Croarkin (Ed.), Screening designs, NIST/SEMATECH e-Handbook of Statistical Methods, 2012, Ch. 5, http://www.itl.nist.gov/div898/handbook/pri/section3/pri3346.htm.

[31]   F. Phoa, W. Wong, H. Xu, The need of considering the interactions in the analysis of screening designs, Journal of Chemometrics 23 (2009) 545-553.

[32]   K. Elsayed, C. Lacor, Optimization of the cyclone separator geometry for minimum pressure drop using mathematical models and CFD simulations, Chemical Engineering Science 65 (22) (2010) 6048-6058.

[33]   A. Sobester, S. Leary, A. Keane, A parallel updating scheme for approximating and optimizing high fidelity computer simulations, Structural and Multidisciplinary Optimization 27 (2004) 371-383.

[34]   R. Xiang, K. W. Lee, Exploratory study on cyclones of modified designs, Particulate Science and Technology 19 (4) (2001) 327-338.

[35]   K. Elsayed, C. Lacor, The effect of cyclone inlet dimensions on the flow pattern and performance, Applied Mathematical Modelling 35 (4) (2011) 1952-1968.

[36]   B. Zhao, Y. Su, Artificial neural network-based modeling of pressure drop coefficient for cyclone separators, chemical engineering research and design 88 (2010) 606-613.

[37]   A. C. Hoffmann, L. E. Stein, Gas cyclones and swirl tubes: Principle, Design and Operation, 2nd Edition, Springer, 2008.

[38]   G. Ramachandran, D. Leith, J. Dirgo, H. Feldman, Cyclone optimization based on a new empirical model for pressure drop, Aerosol Science and Technology 15 (1991) 135-148.

[39]   K. Elsayed, C. Lacor, CFD modeling and multi-objective optimization of cyclone geometry using desirability function, artificial neural networks and genetic algorithms, Applied Mathematical Modelling Accepted Manuscript, http://dx.doi.org/10.1016/j.apm.2012.11.010.

[40]   C. B. Shepherd, C. E. Lapple, Flow pattern and pressure drop in cyclone dust collectors cyclone without intel vane, Industrial & Engineering Chemistry 32 (9) (1940) 1246-1248.

[41]   R. Hooke, T. A. Jeeves, `` direct search'' solution of numerical and statistical problems, J. ACM 8 (2) (1961) 212-229.

[42]   S. Lophaven, H. Nielsen, J. Sondergaard, Aspects of the Matlab Toolbox DACE, Report IMM-REP-2002-13, Informatics and Mathematical Modelling, DTU, http://www.imm.dtu.dk/ hbn/publ/TR0213.ps (2002).

[43]   C. J. Stairmand, The design and performance of cyclone separators, Industrial and Engineering Chemistry 29 (1951) 356-383.

[44]   C.J. Stairmand, Pressure drops in cyclone separators, Industrial and Engineering Chemistry 16 (B) (1949) 409-411.

[45]   G. Ravi, S. K. Gupta, M. B. Ray, Multiobjective optimization of cyclone separators using genetic algorithm, Ind. Eng. Chem. Res. 39 (2000) 4272-4286.

[46]   H. Safikhani, A. Hajiloo, M. Ranjbar, N. Nariman-Zadeh, Modeling and multi-objective optimization of cyclone separators using CFD and genetic algorithms, Computers & Chemical Engineering 35 (6) (2011) 1064-1071.

[47]   K. Elsayed, C. Lacor, Multi-objective optimization of gas cyclone based on CFD simulation, in: ECCOMAS thematic conference, CFD & Optimization, Antalya, Turkey, 2011.

[48]   H. Safikhani, S. Nourbakhsh, N. Nariman-zadeh, Modeling and multi-objective optimization of cyclone vortex finder using CFD and neural networks, in: 2nd International Conference on Engineering Optimization, September 6-9, 2010, Lisbon, Portugal, 2010, pp. 1-9.

[49]   K. Elsayed, C. Lacor, Numerical modeling of the flow field and performance in cyclones of different cone-tip diameters, Computers & Fluids 51 (1) (2011) 48-59.

[50]   S. Rahmani, S. M. Mousavi, M. J. Kamali, Modeling of road-traffic noise with the use of genetic algorithm, Applied Soft Computing 11 (1) (2011) 1008 - 1013.

[51]   M. D. Slack, R. O. Prasad, A. Bakker, F. Boysan, Advances in cyclone modeling using unstructured grids, Trans IChemE. 78 Part A, (2000).

[52]   A. J. Hoekstra, Gas flow field and collection efficiency of cyclone separators, Ph.D. thesis, Technical University Delft (2000).

[53]   F. Kaya, I. Karagoz, Performance analysis of numerical schemes in highly swirling turbulent flows in cyclones, Current Science 94 (10) (2008) 1273 - 1278.

[54] S. Bernardo, M. Mori, A. Peres, R. Dionisio, 3-D computational fluid dynamics for gas and gas-particle flows in a cyclone with different inlet section angles, Powder Technology 162 (3) (2006) 190 - 200.

[55] T. G. Chuah, J. Gimbun, T. S. Choong, A CFD study of the effect of cone dimensions on sampling aerocyclones performance and hydrodynamics, Powder Technology 162 (2006) 126 - 132.

[56] B. Zhao, Y. Su, J. Zhang, Simulation of gas flow pattern and separation efficiency in cyclone with conventional single and spiral double inlet configuration, Chemical Engineering Research and Design 84 (2006) 1158-1165.

[57] R. B. Xiang, K. W. Lee, Numerical study of flow field in cyclones of different height, Chemical Engineering and Processing 44 (2005) 877-883.

[58] J. Gimbun, T. Chuah, T. Choong, Y. Fakhru'l-Razi, Prediction of the effects of cone tip diameter on the cyclone performance, Aerosol Science and Technology 36 (2005) 1056-1065.

[59] J. Gimbun, T. G. Chuah, T. S. Y. Choong, A. Fakhru'l-Razi, A CFD study on the prediction of cyclone collection efficiency, International Journal for Computational Methods in Engineering Science and Mechanics 6 (3) (2005) 161 - 168.

[60] J. Gimbun, T. Chuah, A. Fakhru'l-Razi, T. S. Y. Choong, The influence of temperature and inlet velocity on cyclone pressure drop: a CFD study, Chemical Engineering & Processing 44 (1) (2005) 7-12.

[61] A. J. Hoekstra, J. J. Derksen, H. E. A. Van Den Akker, An experimental and numerical study of turbulent swirling flow in gas cyclones, Chemical Engineering Science 54 (1999) 2055-2065.

[62] K. Elsayed, C. Lacor, The effect of cyclone vortex finder dimensions on the flow pattern and performance using les, Computers & Fluids 71 (2013) 224 - 239.
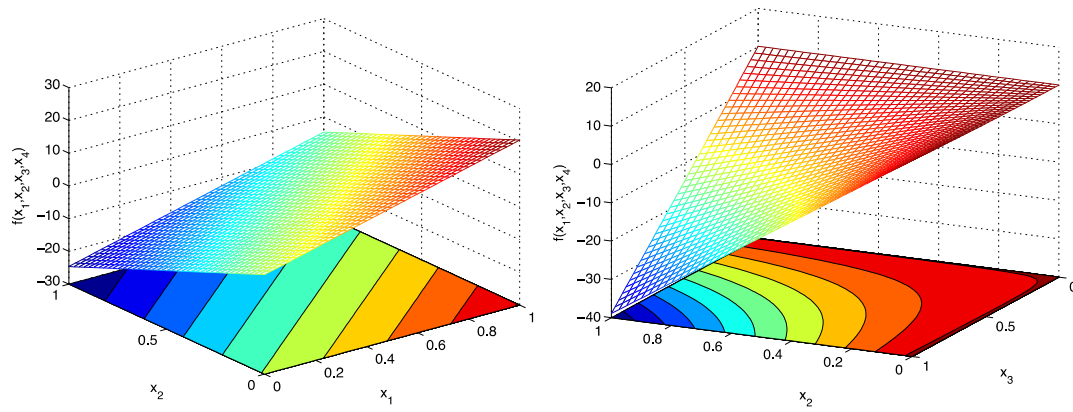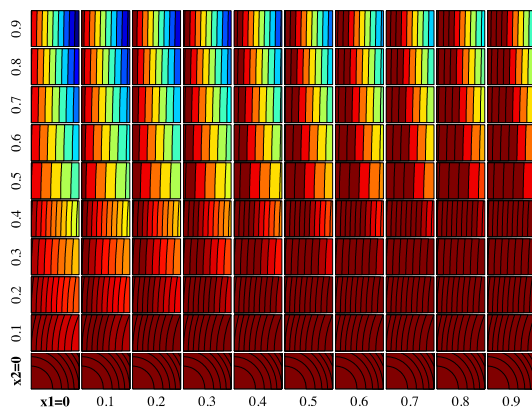
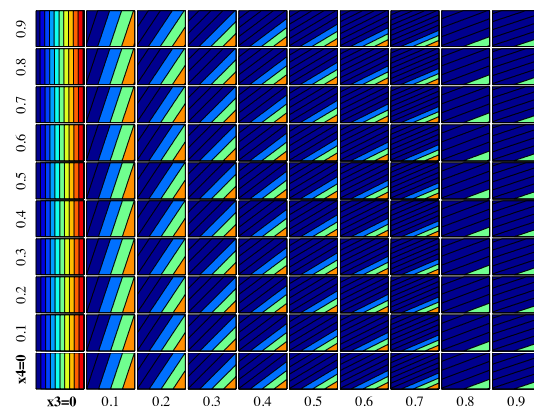(a) 3-D plots of the given function
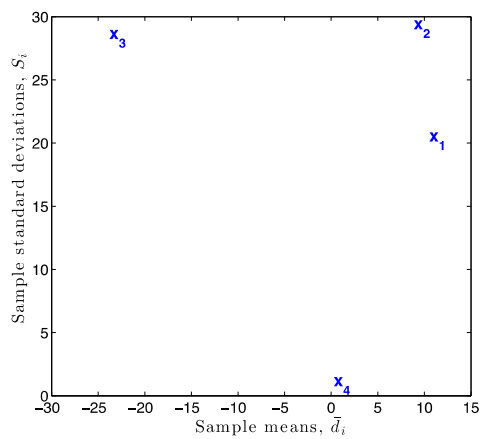
(b) Estimated means and standard deviations of the elementary effect distributions
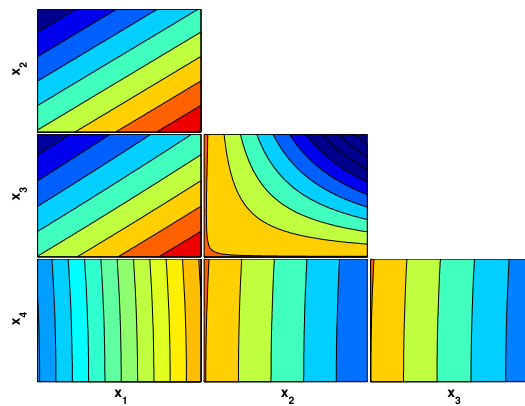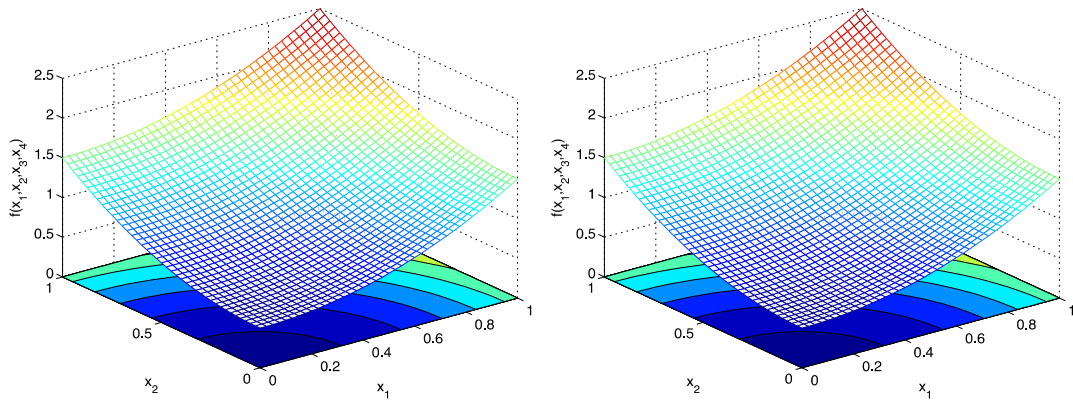

(c) Tile plot
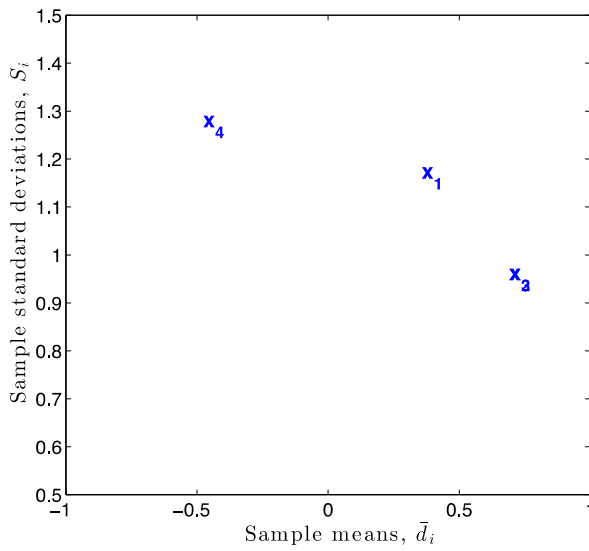

(d) Nested plot (order x1, x2, x3, x4)
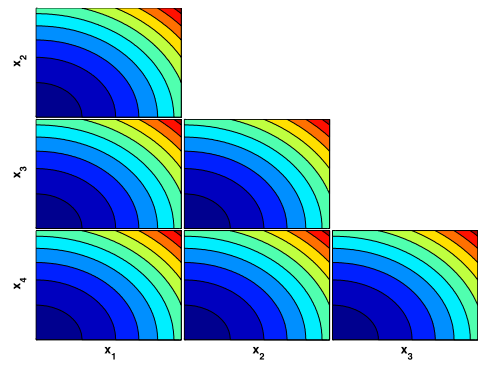

(e) Nested plot (order x3, x4, x1, x2)

Fig. 1. 3D plots, elementary effect distributions of each of the four variables of Eq. (5) and the corresponding tile-plot and the Four-variable nested plot.
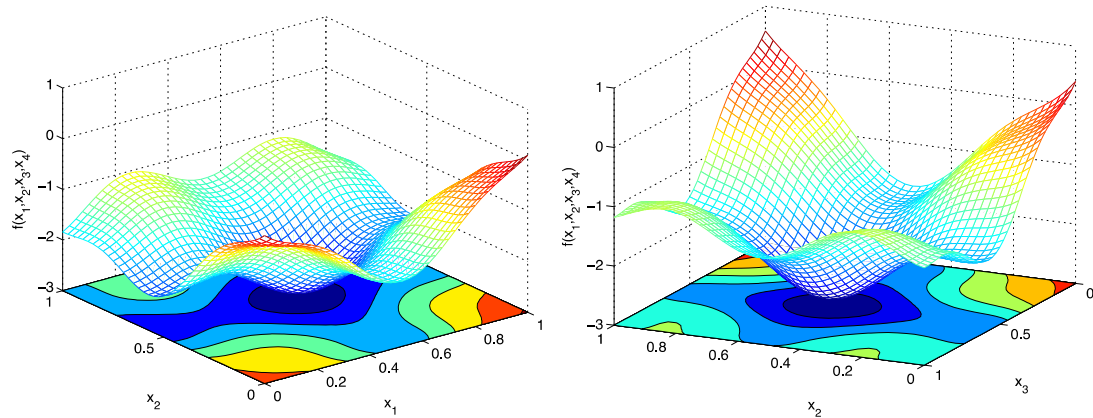
(a) 3-D plots

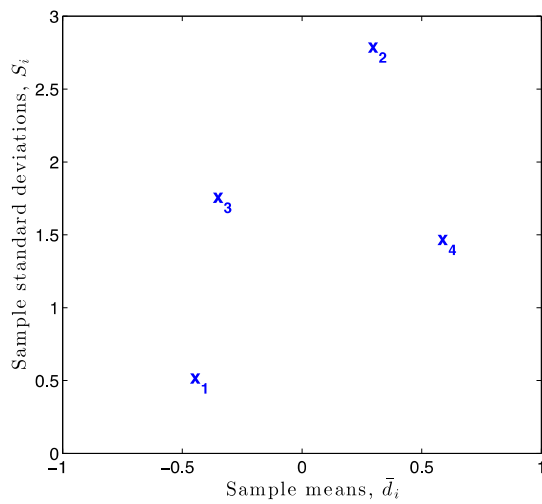(b) Elementary effect distributions

(c) Tile plot

Fig. 2. 3D plots, elementary effect distributions and the corresponding tile-plot for the sphere function. Note: In (b) x_2 and x_3 are coincide

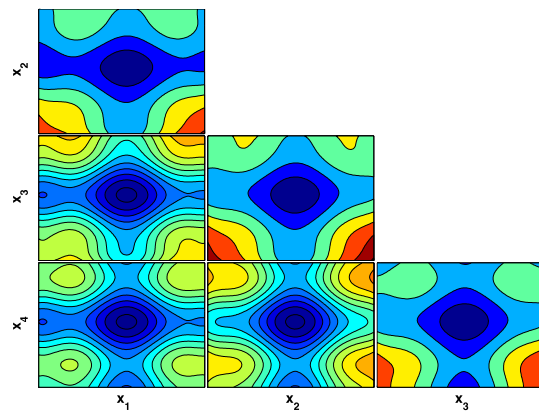**Table 1:** The input and output variables for surrogate model

	Input parameters							Output parameter
Variables	x_1	x_2	x_3	x_4	x_5	x_6	x_7	$\hat{y}$
Specification	D_x	a	b	S	H_t	h	B_c	E_u

(a) 3-D plots of the Rosenbrock function

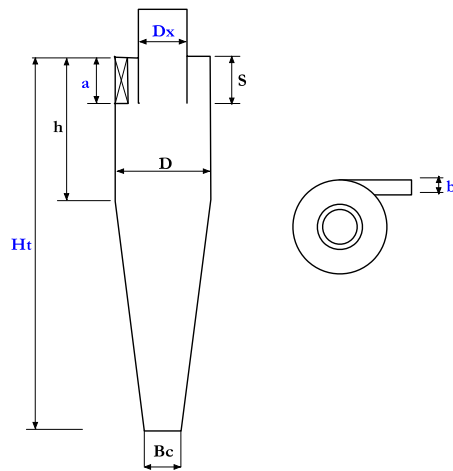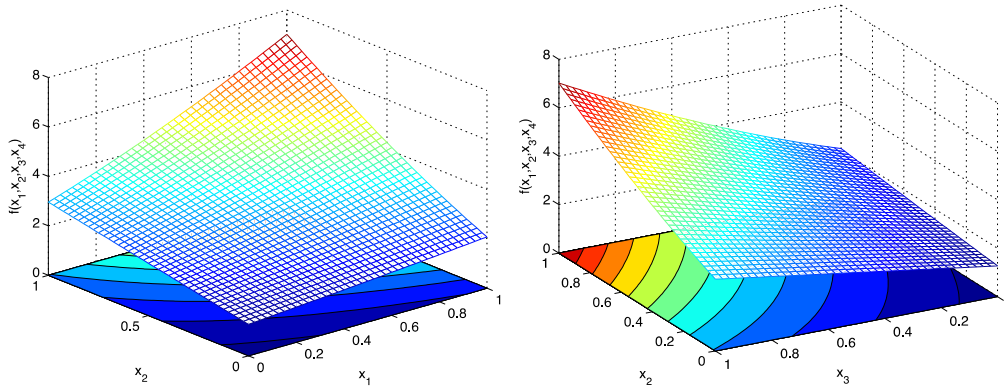(b) Elementary effect distributions

(c) Tile plot

Fig. 3. 3D plots, elementary effect distributions and the corresponding tile-plot

**Table 2:** The values of the independent variables [32]

Variables	Minimum	Maximum
Vortex finder diameter, $D_x/D = x_1$	0.2	0.75
Inlet height, $a/D = x_2$	0.4	0.7
Inlet width, $b/D = x_3$	0.14	0.4
Vortex finder length, $S/D = x_4$	0.4	2.0
Total cyclone height, $H_t/D = x_5$	3.0	7.0
Cylinder height, $h/D = x_6$	1.0	2.0
Cone tip diameter, $B_c/D = x_7$	0.2	0.4

(a) Cyclone geometry (the most significant dimensions in blue)

(b) 3-D plots of the MM model (x1=a, x2=b and x3=Dx)
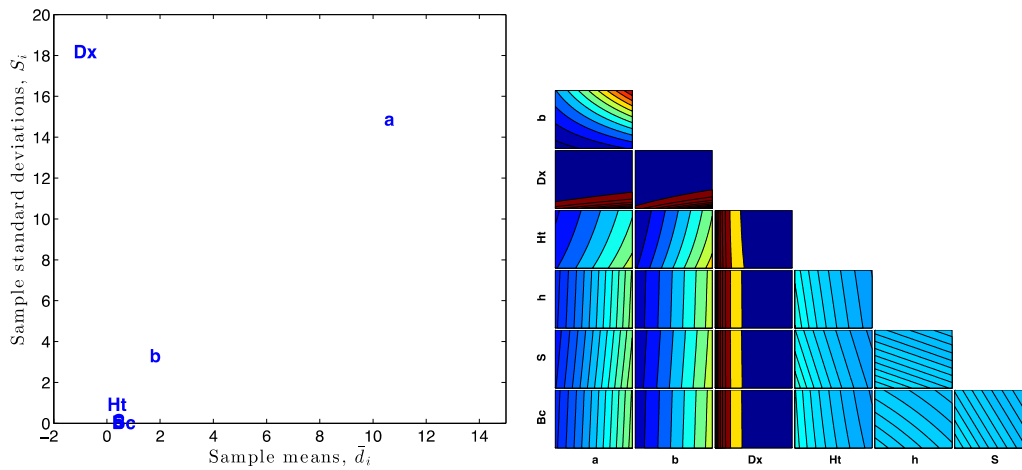
(c) Elementary effect distributions

(d) Tile plot

Fig. 4. Cyclone geometry, 3D plots, elementary effect distributions and the corresponding Tile-plot using MM model.

(a)   Expensive and cheap data

(b)   Comparison between different Krigings model

(c)   Bar plot for the predicted values for randomly selected untrained 15 test cases

(d)   Error line plots for the absolute error for randomly selected untrained 15 test cases
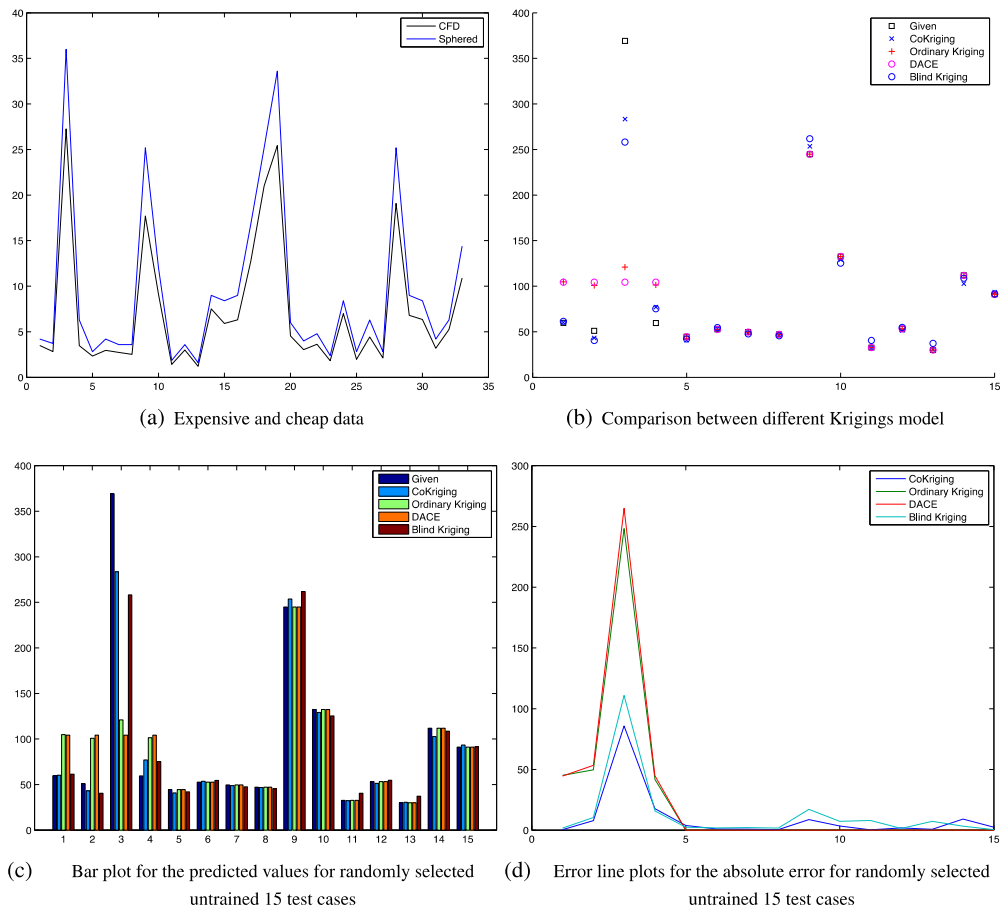
Fig. 5. Comparison between the Co-Kriging model and OK, DACE and BK models for the Euler number. Note that: DACE used a modified version of the Hooke and Jeeves [41] pattern search technique (cf. Section 6 in [42]), whereas the ordinary Kriging used genetic algorithms to obtain the optimum value of the hyperparameter θ.

Table 3. Genetic operators and parameters for single objective optimization

Population type	Double vector
Initial range	[0.25 0.15 0.25 3; 0.5 0.375 0.75 5] for a, b, D_x and H_t respectively
Fitness scaling	Rank
Selection operation:	Tournament (tournament size equals 4)
Elite count	2
Crossover fraction	0.8
Crossover operation	Intermediate crossover with the default value of 1.0
Mutation operation	The constraint dependent default
Maximum number of generations:	800
Population size	200

Table 4. The optimized cyclone separator design for minimum pressure drop using GA[*]

Factor	Low	High	Stairmand design	Optimum design
a	0.25	0.5	0.5	0.499
b	0.15	0.375	0.2	0.150
D_x	0.25	0.75	0.5	0.425
H_t	3.0	5.0	4.0	4.876
Euler number			5.606	1.236

[*] The values for the other three variables are kept identical to the Stairmand design $S = 0.5$, $h = 1.5$ and $B_c = 0.375$.

Table 5. The geometrical parameters for the two designs (D=31E-3 m).

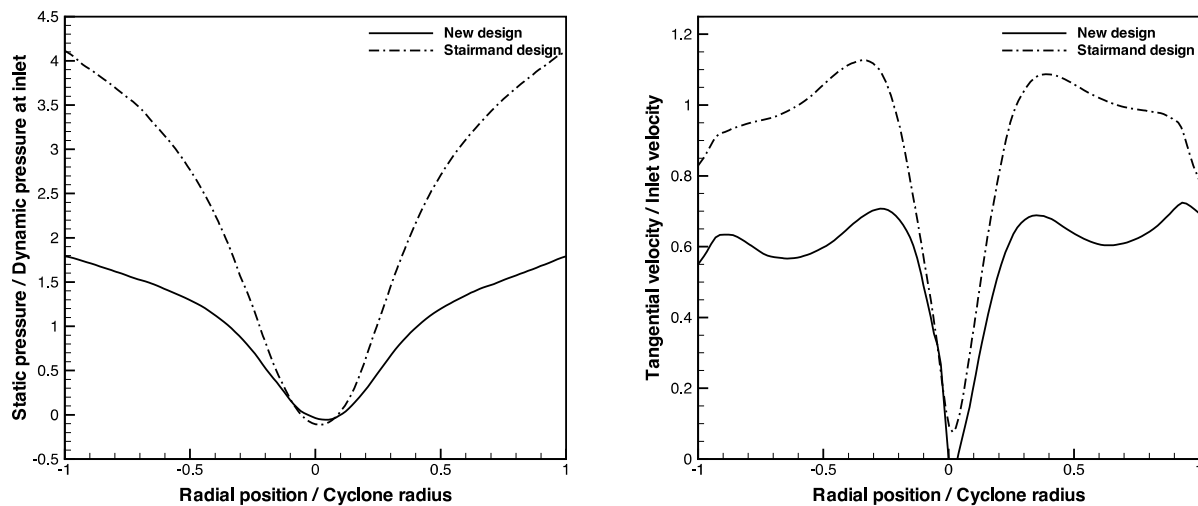| Cyclone	a/D	b/D	D_x/D	H_t/D	h/D	S/D	B_c/D	L_i/D	L_e/D
Stairmand design	0.5	0.2	0.5	4	1.5	0.5	0.375	1.0	0.5
New design	0.499	0.15	0.425	4.876	1.5	0.5	0.375	1.0	0.5

Fig. 6. The radial distribution of time-averaged static pressure, tangenial and axial velocity at a station at $0.75D$ from the inlet section top.