

Innovative Methodologies for Robust Design Optimization with Large Number of Uncertainties Using ModeFRONTIER



Alberto Clarich and Rosario Russo

Abstract This paper describes the methodologies that have been developed by ESTECO during the first phase of UMRIDA European Project, in the field of Uncertainty Management and Robust Design Optimization, and that have been implemented in the software platform modeFRONTIER. In particular, in the first part there are proposed two methodologies, one based on SS-ANOVA regression applied directly to the uncertainties variables and one based on a stepwise regression methodology applied to the Polynomial Chaos terms used for the uncertainty quantification. Aeronautical test cases proposed by UMRIDA consortium are used to verify the validity of the methodologies. In the second part, the state of art methodologies for Robust Design Optimization are compared with a new proposed approach, based on a min-max definition of the objectives, and the application of Polynomial Chaos coefficients for an accurate definition of percentiles (reliability-based robust design optimization). Also in this case an Aeronautical CFD test case is proposed to validate the methodologies.

1 Introduction

As proved by recent studies [1], one of the most efficient methodology to manage accurately the uncertainties is the application of Polynomial Chaos expansion [2]. This methodology however requires a minimum number of samples which increases heavily with the number of uncertainties, and a typical industrial optimization case (for instance at least 10 simultaneous uncertainties) can be hardly treated as a feasible task.

For this reason, we propose in this paper some approaches to handle efficiently industrial problems of this kind, both on the side of Uncertainty Management (UQ) and on the side of Robust Design Optimization (RDO).

A. Clarich (✉) · R. Russo
ESTECO SpA, Area Science Park, Padriciano 99, 34151 Trieste, Italy
e-mail: engineering@esteco.it

For the UQ, the proposed solution is to use methodologies that allow to identify which are the uncertainties having higher statistical effects on the performances of the system, this way allowing to apply the Polynomial Chaos expansion with a smaller number of uncertainties, and therefore a significant smaller number of samples, on a system which is statistically equivalent to the original. As alternative, the uncertain parameters effects analysis can be applied directly to the Polynomial Chaos terms, this way reducing the number of unknown coefficients and therefore the number of needed samples to complete the UQ, without discarding necessarily at all an uncertain variable from the problem (keeping this way an higher accuracy).

For the RDO methodologies, we propose a methodology based on the min-max formulation of objectives, which guarantees the reduction of objectives numbers with respect to a classical RDO approach [3], and therefore the possibility of reducing drastically the number of configurations to be evaluated and of simulations to be performed. In order to guarantee an accurate application of this methodology, we developed an approach which is based on the exploitation of Polynomial Chaos coefficients to evaluate accurately the percentiles of the quantities to be optimized/constrained. This methodology is also called reliability-based design optimization [4], and the solution we propose, based on Polynomial Chaos exploitation, is innovative and very promising in terms of efficiency.

2 UQ of Large Number of Variables: SS-ANOVA and Stepwise Regression for Sparse Collocation

Smoothing Spline ANOVA (SS-ANOVA) [5] models are a family of smoothing methods suitable for both uni-variate and multi-variate modeling/regression problems characterized by noisy data, given the assumption of Gaussian-type responses. In particular, SS-ANOVA is a statistical modeling algorithm based on a function decomposition similar to the classical analysis of variance (ANOVA) decomposition and the associated notions of main effect and interaction. Each term - main effects and interactions—can be used to reveal the percentage contribution of each single uncertain parameter, and of any uncertain variables pair, on the output global variance, since in a statistical model the global variance can be explained (decomposed) into single model terms.

A generic multi-variate regression problem could be stated as a constrained minimization problem, that can be expressed in Lagrangian terms as:

$$\min L(f) + \frac{\lambda}{2} J(f) \quad (1)$$

where $L(f)$ is defined as *minus log likelihood* of the model $f(x)$ given the data, to be minimized to maximize the data fit, and $J(f)$ is defined as a *quadratic roughness functional*, to be subjected by a constraint— $J(f) \leq \rho$ —that can be used to preserve

the overfitting (a large roughness guarantees a smoother model, while smaller values imply rougher functions but better agreement to the data).

We can generally assume that the regression model $f(\mathbf{x})$ can be expressed as a sum of N independent components $f_j(x_j)$, each one function of a single variable x_j . By this approximation, the regression model would take into account only main effects (the effect of each single variable).

A more complete regression model, which has to consider also interaction effects, will include in $f(\mathbf{x})$ also the interaction terms $f_{ij}(x_i, x_j)$. The smoothing parameters, needed to solve the regression problem, can be determined by a proper data-driven procedure, such as the generalized cross validation (GCV), as described in [6]. The number of decomposition terms, which is equal also to the minimum number of needed sampling points, is equal to $N(N - 1)/2$ with N number of variables.

We can therefore apply the definition of internal product projecting the $f(\mathbf{x})$ to any component f_k obtaining the value of its contribution (or probability), by the (normalized) expression:

$$\pi_k = \frac{\langle f_k, f \rangle}{\|f\|^2} \tag{2}$$

Expression 2 is called *contribution index* k and expresses the relative significance of the different terms composing the model, therefore the contribution of each variable main effect or interaction effect.

As alternative to the first methodology for UQ here proposed, we have adopted another approach [7], which consists in applying a regression analysis directly on the Polynomial Chaos expansion (PCE) expression, in other words the PCE will keep only those terms which actually affect the output, discarding the others.

The methodology consists first in ranking the terms using a Least Angle Regression (LAR) technique [8] and then in assessing how many PCE terms should be kept.

The LAR ranking is accomplished by the following procedure (\mathbf{P}_{a_i} represents a generic PCE term).

- Set residual **Res** = **output** – mean (**output**)
- The first selected polynomial term \mathbf{P}_{a_1} is the one with the highest correlation with **Res** namely: \mathbf{P}_{a_1} such that $\text{corr}(\mathbf{P}_{a_1}, \mathbf{Res}) = \max(\text{corr}(\mathbf{P}_{a_i}, \mathbf{Res}))$
- Set $\mathbf{P}_{a_i} = \mathbf{P}_{a_1}$
- For k from 1 to the number of PCE terms to be ranked do:
 - Set **Res** = **Res** – $\lambda \mathbf{P}_{a_i}$ where λ is such that: $\text{corr}(\mathbf{P}_{a_i}, \mathbf{Res}) = \text{corr}(\mathbf{P}_{a_j}, \mathbf{Res})$, the polynomial \mathbf{P}_{a_j} is selected.
 - Solve a least square problem: find c_i and c_j that minimize $(c_i \mathbf{P}_{a_i} + c_j \mathbf{P}_{a_j} - \mathbf{Res})^2$
 - Set $\mathbf{P}_{a_i} = c_i \mathbf{P}_{a_i} + c_j \mathbf{P}_{a_j}$ (new direction)
- Next k

The order of selection of the PCE terms will reflect a ranking based on how much each term affects the output. Once the ranking is done it is necessary to establish a

way to choose how many PCE terms should be kept. The criterion for this is based on the Mean Leave one Out Error (Err_{LOO}).

$$Err_{LOO} = \frac{1}{N} \sum_{i=1}^N \Delta_i^2 \quad (3)$$

where N is the number of samples and

$$\Delta_i = \text{output}(\mathbf{x}_i) - \widehat{M}(\mathbf{x}_i) \quad (4)$$

i.e. the difference between the output corresponding to the i -th sample and the output computed from the PCE obtained excluding from the training samples the i -th.

It is possible to show [6] that Δ_i can be estimated by the following expression:

$$\Delta_i = \frac{\text{output}(\mathbf{x}_i) - M(\mathbf{x}_i)}{1 - h_i} \quad (5)$$

where $M(\mathbf{x}_i)$ is the output evaluated by the PCE computed this time using all the samples and

$$h_i = \text{diag} \left(\mathbf{P} (\mathbf{P}^T \mathbf{P})^{-1} \mathbf{P}^T \right) \quad (6)$$

where $P_{ij} \equiv \mathbf{P}_{a_j}(\mathbf{x}_i)_{i=1\dots N; j=1\dots Nterms}$

Using the previous expressions, given a certain number of PCE terms, it is possible to compute the corresponding Err_{LOO} .

The criterion to select the number of terms consists in monitoring the two quantities:

$$R_{LOO} \equiv 1 - \frac{Err_{LOO}}{\text{var}(\mathbf{output})} \quad (7)$$

$$R_{squared} \equiv 1 - \frac{Err_{squared}}{\text{var}(\mathbf{output})} \quad (8)$$

where $Err_{squared}$ is the squared error sum, i.e. the sum of the squared differences between each sample output and the corresponding output value estimated using the PCE (this time using all the samples); $\text{var}(\mathbf{output})$ is the output variance considering all the samples.

$R_{squared}$ and R_{LOO} are functions of the number of PCE terms: the first will generally increase as the number of terms increases, while the second tends initially to increase as the number of terms increases, but from a certain number of terms on, it starts showing a decreasing trend.

$R_{squared}$ is sensitive on how much the PCE expansion is able to approximate the output, while R_{LOO} is sensitive to overfitting problems: the ideal number of terms should guarantee a good compromise, namely $R_{squared}$ around 0.9, or higher, and R_{LOO} close to its maximum before the decreasing trend due to overfitting.

Table 1 List of uncertainties

Uncertainty	Most likely value (m)	Minimum value (a)	Maximum value (b)	PDF-type
Free stream mach	0.729 (case 6) 0.734 (case 9)	95% m	105% m	Symmetric beta-PDF
Angle of attack	2.31° (case 6) 2.54° (case 9)	98% m	102% m	Symmetric beta-PDF
Thickness-to-chord	Nominal value from geometry	97% Nominal	103% Nominal	Symmetric beta-PDF
Camber line	Nominal value from geometry	Nominal-0.01% chord	Nominal +0.01% chord	Symmetric beta-PDF

Of course any different strategy for ranking the PCE terms and choosing the proper number of terms could be chosen, it is important however to employ a strategy somehow sensitive to both accuracy and overfitting control. In any case the proposed strategy has got the advantage of keeping separated the ranking algorithm from the number of terms choice.

The described approach gives the important benefit of reducing the global number of unknown coefficients for the PCE expansion, and therefore giving the possibility as well of reducing the number of sampling points, needed for the PCE training.

Conversely from the SS-ANOVA methodology, however, the great advantage is that any uncertainty is not necessarily discarded, but its effect might be included in a smaller set of polynomial terms.

3 UQ Test Case Application

To validate the methodologies proposed in previous section, we have applied them to the test case BC-02 of UMRIDA European Project [9].

The test case consists in the UQ quantification of a RAE 2822 airfoil [10], for a specified conditions, and for a total of 13 uncertainties (operational and geometrical). Nominal parameters and uncertainties type and parameters are defined in Table 1.

In particular, the geometrical uncertainties refer to the camber line and the thickness-to-chord ratio of the nominal profile, which have been fitted by a Bezier parametric curve [11], of respectively 7 and 8 control points, uniformly spaced in the abscissas. Since the extreme points are fixed, we consider a total of 5 uncertainties for the (ordinates of) control points of the thickness-to-chord curve (named from *Y1_thickness* to *Y5_thickness*), and 6 for the camber curve (named from *Y1_chord* to *Y6_chord*).

Figure 1 reports the process workflow created for the set-up of this test case in modeFRONTIER software from ESTECO. Each component of the process is defined by dedicated modules (input variables, CAD/CAE interfaces, output variables) interconnected between them, in order to allow the automatic execution of the simulations

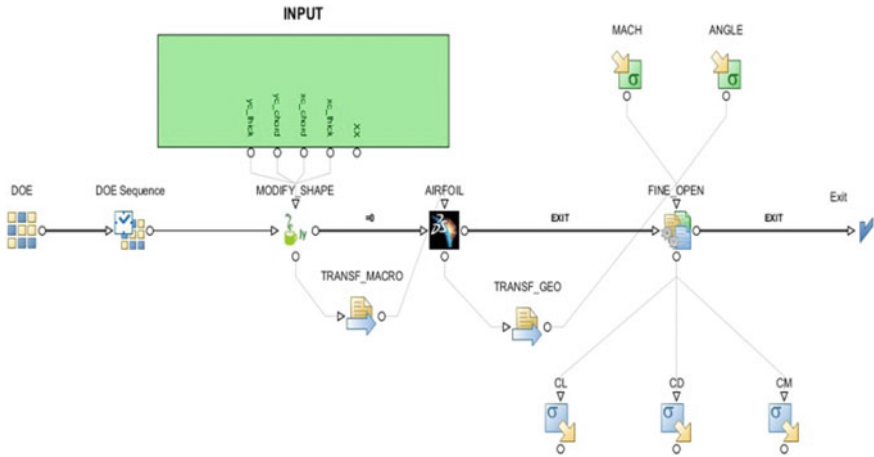


Fig. 1 Workflow for process automation in modeFRONTIER

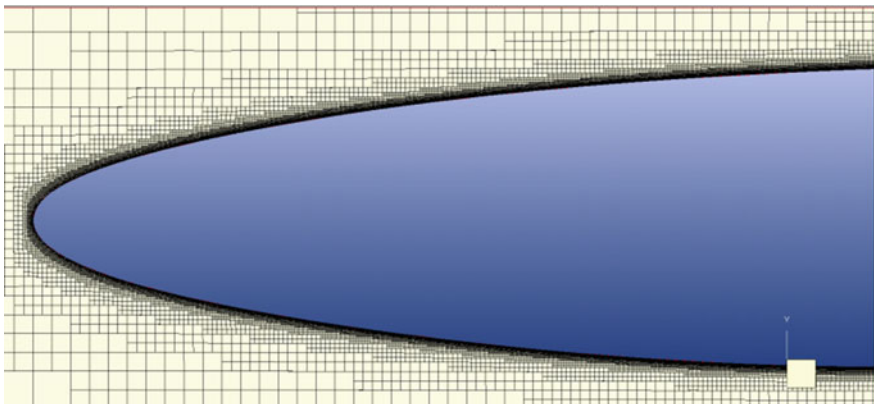


Fig. 2 Mesh overview and detail in FINE/open model

for each design sample which is proposed by the selected algorithm for the UQ. modeFRONTIER software contains as well all the tools needed to complete automatically the UQ of the required parameters.

The mesh provided for this test case has been elaborated by ESTECO in FINE/Open software from NUMECA. The mesh is characterized by an overall number of cells equal to about 1/2 million, which require an average time to complete the simulation of one design sample in about 1 h, using a 2-cpu machine. Around the airfoil the mesh is refined, because it is important to reduce the effect of numerical uncertainties, which are not considered in the problem.

Figure 2 reports a detail of the mesh used.

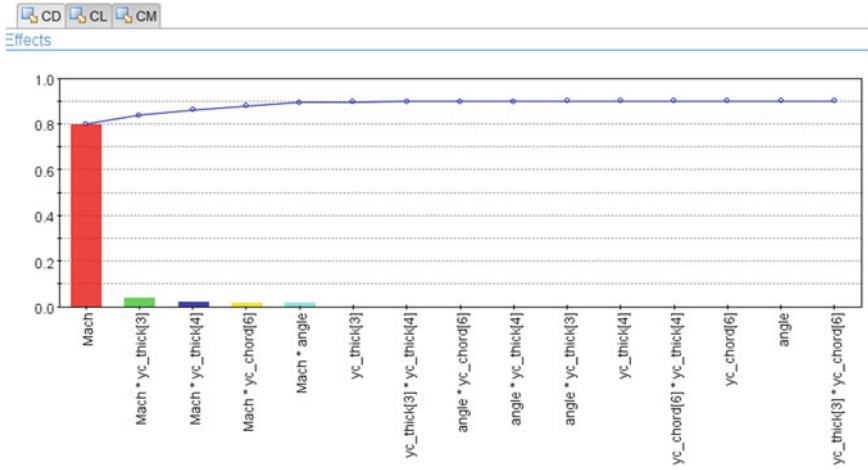


Fig. 3 SS-ANOVA (modeFRONTIER) for each output

Inlet conditions are specified in function of the Reynolds number defined for this test case ($Re = 6.5 \times 10^6$) and of the Mach number specified for this test case (see Table 1).

In the remaining lateral boundaries of the model, an outlet condition is specified, while a symmetry condition is specified on the planes parallel to the flow.

A full turbulent (*Spalart-Allmaras*) model is used; an adaptive meshing procedure is defined, to refine the mesh where the gradient of the pressure is higher.

The first step in the UQ of the test case is the definition of a large DOE (Design of Experiments) using a Latin Hypercube algorithm, considering all the 13 original uncertainties.

For this purpose we have evaluated a series of 105 designs, which is the minimum number of samples to apply a Polynomial Chaos Expansion of order 2 for the UQ, and then repeated the analysis for a larger number (200) of samples.

These samples are also used to apply the SS-ANOVA screening analysis, which indicates the relative effect of each parameter for the selected output (Cd, Cl and Cm). Figure 3 in fact illustrates for one of the outputs (Cd) the relative effect of each uncertain parameter (using the name conventions described above in this section), including in the analysis also the interaction effects.

Considering for each output a cumulative effect of at least 90%, we can conclude that the parameters most important are:

- For CD: *Mach*, *Y3_thickness*, *Y4_thickness*;
- For CL: *Mach*, *angle*, *Y1_chord*, *Y5_chord*;
- For CM: *Mach*, *Y2_chord*.

Table 2 UQ results for test case defined in Table 1

	Cd mean	Cd Sigma	CI mean	CI Sigma	Cm mean	Cm Sigma
200 samples-13 uncertain	1.951E-2	2.12E-3	6.39E-1	1.93E-2	1.035E-1	4.04E-3
100 samples-13 uncertain	1.953E-2	1.96E-3	6.39E-1	1.82E-1	1.035E-1	3.86E-2
	err: 0.1%	err: 7.6%	err: 0%	err: 8.4%	err: 0%	err: 8.6%
45 samp/es-SS-ANOVA (7 uncertain)	1.951E-2	2.09E-3	6.39E-1	2.01 E-2	1.035E-1	4.11E-3
	err: 0%	err: 1.4%	err: 0%	err: 4.1%	err: 0%	err: 1.7%
30 samples-stepwise regression	195E-2	2.14E-3	6.39E-1	1.86E-2	1.035E-1	3.96E-3
	err: 0%	err: 0.6%	err: 0%	err: 3.5%	err: 0%	err: 2.0%

So, globally, the 7 common parameters most important are: *Mach*, *angle*, *Y3_thickness*, *Y4_thickness*, *Y1_chord*, *Y2_chord*, *Y5_chord*.

In other words, we could exclude from the analysis the less significant uncertainties, keeping statistically almost the same information on the UQ of the outputs.

To validate this hypothesis, we have in a second step fixed the 6 not significant uncertainties (to their nominal values), and defined a UQ problem of 7 uncertainties only.

For this reduced problem, a much smaller number of simulations is required and precisely at least 45 samples to apply a Polynomial Chaos expansion of order 2; Table 2 more over will report the results of the UQ analysis.

Applying instead the second methodology (LAR), we have found that only 10 terms are needed to give acceptable errors on a database of 30 samples only, and precisely (the number in the terms notation below refer to the defined order of the variables, which is: 1-*Mach*, 2-*Angle*, 3-*Y1_camber*, 4-*Y2_camber*, 5-*Y3_camber*, 6-*Y4_camber*, 7-*Y5_camber*, 8-*Y6_camber*, 9-*Y1_thickness*, 10-*Y2_thickness*, 11-*Y3_thickness*, 12-*Y4_thickness*, 13-*Y5_thickness*; the apex refers to the exponent of the term, and the _ character refers to an interaction between two terms):

- For CD: 1, 1^2, 11, 2, 1_12, 2_9, 1_3, 10, 1_7, 1_11;
- For CL: 1, 1^2, 2, 1_7, 1_13, 13, 6, 10, _13, 3;
- For CM: 1, 1^2, 2, 13, 1_7, 8, 1_13, 11, 10, 1_4.

Table 2 therefore reports the UQ results for the test case, after the application of Polynomial Chaos expansion (in modeFRONTIER software) of order 2 for the different DOEs analyzed, and in particular: (1) 200 samples with all the uncertainties, (2) 105 samples with all the uncertainties, (3) 45 samples with 7 most important uncertainties (SS-ANOVA application), (4) 30 samples with 10 most important PCE terms.

The results of the test are satisfactory. Applying the first methodology, using 45 samples only it was possible to determine the main momentum averages with a practically absolute accuracy and the standard deviation by an error (computed on the basis of the largest DOE results) between 1 and 4% (higher for lift coefficient and lower for drag).

Applying the second methodology, the results are even slightly improved, since by a lower number of samples, 30, the highest error on standard deviation has been reduced from 4.1 to 3.5%. This second method, in addition, is independent from the significance of the single parameters, that in this particular test case may have given advantage to the first method (having one variable, Mach, predominant in the global variance).

In general, it emerges clearly that a significant reduction of the number of needed samples, reachable by any of the two methodologies, produces an accurate evaluation of the statistical moments, with a contained maximum estimated error.

4 RDO: Classical Versus MINMAX Approach

In order to apply RDO to a problem of industrial relevance, i.e. a problem characterized by a large number of uncertainties and by simulation times which are expansive, it is not just needed to define only an efficient UQ methodology, which can give accurate results with few simulations, but also an efficient optimization approach.

The first approach that we propose in this chapter as the state of the art, is the classical RDO approach [3] based on the definition of a multi-objective optimization problem, consisting generally on the optimization of the mean value of the performances and on the minimization of their standard deviation.

This approach guarantees the definition of a complete Pareto frontier as trade-off of the optimal solutions, in terms of mean performance and in terms of their stability or robustness. This means that at the end of the optimization the designer has the freedom to select the best solutions accordingly to a large variety of possibilities depending on which criteria should be privileged.

The problem of this approach is that a Multi-objective Optimization algorithm is to be chosen, since the definition of a single objective as weighted sum of the different criteria cannot be proposed for the impossibility of knowing a priori the proper weights of the particular optimization problem. Multi-objective Optimization algorithms are in fact generally very robust, but they require a number of simulations generally very much consistent with respect to a single objective optimization case, and for a RDO problem the number of simulations may be not feasible from a practical point of view (this number being multiplied by the sampling size for each design to obtain the overall number of simulations required).

In order to reduce the overall number of simulations for a RDO problem, we propose in this section another approach, described in one of our previous works [4].

The basic idea is to reduce the number of objectives, so that a single-objective algorithm, which requires much less simulations for the convergence, could be applied.

To achieve this purpose the so called *min-max* or *max-min* approach is followed. The idea is to maximize the minimum or worst performance of a distribution function that is to be maximized (for instance the aerodynamic efficiency of a wing), or to minimize the maximum or worst limitation that is to be minimized (for instance the drag coefficient of a wing).

The effect of this approach is the “shift” of the performance distribution in the desired direction, so in a certain sense both the average performance and the stability at the uncertainties are optimized. Considering for instance the drag coefficient distribution of an airfoil: the optimized configuration distribution by this approach will be shifted below the baseline distribution, since we minimize the maximum value of the distribution or the value of its higher tail.

Besides of this one objective, of course other criteria shall be considered (like lift and momentum), but if they can be expressed as constraints, a single-objective algorithm could still be applied.

At this point, before analyzing the possible single-objective algorithms that may be chosen, it is opportune to discuss about the definition of maximum and minimum values of a distribution.

In the case of a Normal distribution of the performance, since it is unlimited, the concept of the extremes may be replaced by a given percentile of the distribution, for instance 95 or 99%. Usually, the reference value is 99.73% because for a Normal distribution it corresponds to the 3Sigma level.

This analysis is also called Sig-Sigma, since six time standard deviation corresponds to the 99.73% of the complete distribution, a value that can be assumed enough representative of the whole distribution.

Now, since by Polynomial Chaos analysis we can compute mean and standard deviation with high accuracy, the computation of the maximum or minimum value with the expression $MEAN \pm 3\sigma$ can be made for each design of the RDO optimization, therefore the objective function can be defined this way.

The limitation of this approach occurs when the performance does not follow a Normal distribution: in this case, the Six-Sigma formulation may not correspond exactly to the correct percentile of the distribution, so from design to design the computation of the objective function could be not accurate. This problem is even more evident for a particular class of RDO problems, the Reliability-based design optimization, where any constraint should be defined accurately on a given percentile of the distribution.

In next chapter we will propose a new methodology, based on the application of Polynomial Chaos on the Reliability-RDO, in order to solve the problem of the accuracy of the min-max approach and make the RDO optimization more efficient. For the moment, as illustration of the state of the art, we follow the Six-Sigma approach for the min-max strategy, compared with the classical two-objectives approach.

In this case we consider a test case derived from the one illustrated in Table 1. For simplicity, we consider only 3 uncertainties for the RAE2822 airfoil, with nominal values equal to 0.734 for free stream Mach number, 2.79° for angle of attack and Reynolds number equal to $6.5E6$. The uncertainties are given by a Normal distribution for Thickness-to-chord profile (a single uncertainty factor which multiplies the thickness profile), Mach number and angle of attack, defined by a standard deviation respectively equal to 0.005, 0.005 and 0.1.

The first optimization strategy applied is the multi-objective approach (3.1), considering the following objectives and constraints:

- Obj.1: Minimize mean value of Cd
- Obj.2: Minimize standard deviation of Cd
- Constraint 1: mean value of $C_m + 3 \cdot \text{standard deviation of } C_m < 0.1305$
- Constraint 2: mean value of $C_l - 3 \cdot \text{standard deviation of } C_l > 0.9$

The last two constraints are imposed by the necessity to guarantee a minimum value of C_l and a maximum value of C_m respectively less and higher than an arbitrary extreme percentile of the baseline distributions, here 99.97%, that can be approximated considering a Normal distribution using the expressions above (Six Sigma rule). A number of 10 sampling points for design was found to be necessary to guarantee an accurate UQ using a Polynomial Chaos expansion of the second order.

The multi-objective approach then consists in the minimization of the mean value and of the standard deviation of the drag coefficient, and we applied a Game Theory algorithm [12] (MOGT in modeFRONTIER) in order to obtain good compromise results by a lower number of simulations than a classical GA algorithm.

Nonetheless, after the evaluation of more than 50 designs (for a total of 500 CFD simulations, that corresponds to about 20 days using a double cpu machine), it was practically impossible to find feasible solutions that improve the original baseline. The optimization approach has been stopped, because the optimization time was considered already excessive for a problem of industrial relevance.

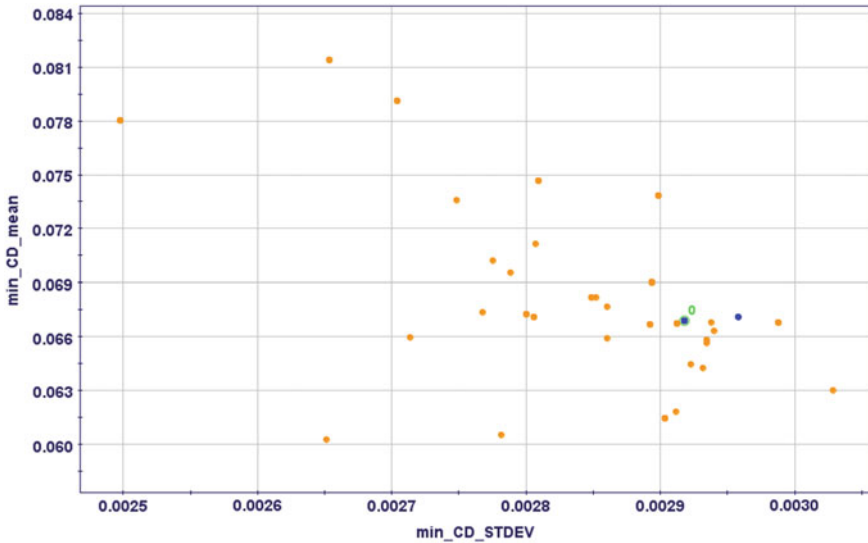


Fig. 4 Optimization results using classical RDO approach

Figure 4 above reports the results obtained following this approach: the two objectives are reported in ordinate (average) and in abscissa (standard deviation), and each point represents a different design proposed during the optimization. The orange color indicates that the design is unfeasible, i.e. that does not respect the constraints, while the blue color indicates that all the constraints are respected.

Beside design 0 (the baseline), only another feasible design has been obtained, without however improving significantly the objectives (standard deviation is higher).

At this point we have then decided to adopt the second methodology, i.e. considering a single objective optimization problem, following the *max-min* approach.

The optimization problem becomes then described as follows:

- Obj.1: Minimize mean value of $Cd + 3 * \text{standard deviation of } Cd$
- Constraint 1: mean value of $Cm + 3 * \text{standard deviation of } Cm < 0.1305$
- Constraint 2: mean value of $Cl - 3 * \text{standard deviation of } Cl > 0.9$

Besides the two constraints on Cl and Cm distributions, mean and standard deviation of Cd have been compacted together into a single objective, which is the minimization of a given high percentile (still 99.97) of the Cd distribution, that can be considered as the “maximum” target value.

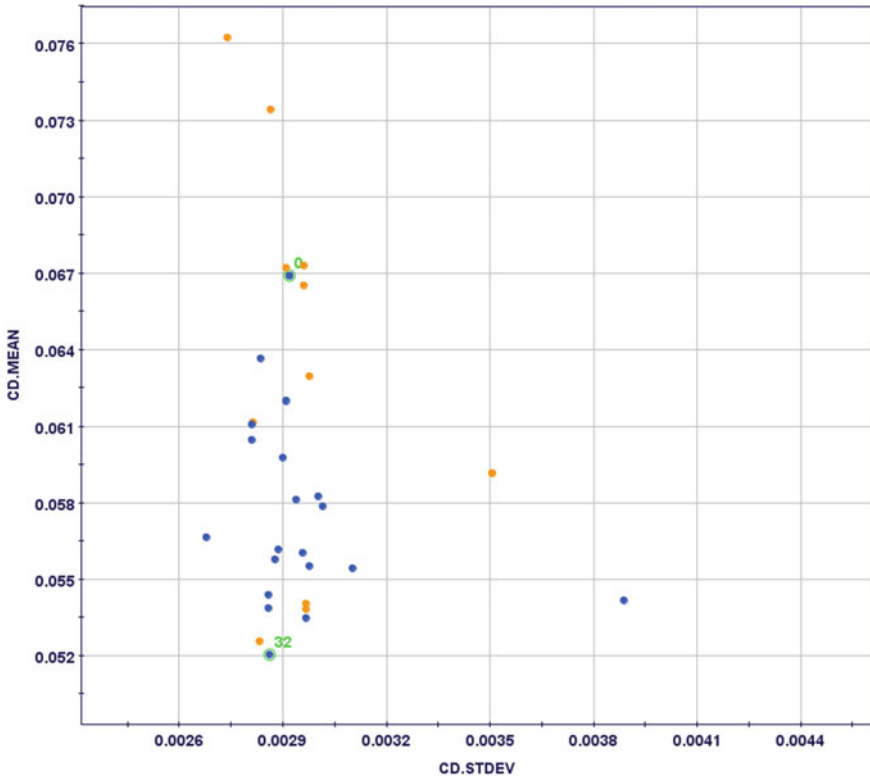


Fig. 5 Optimization results using min-max approach

Table 3 UQ results with different methodologies

Design	Mean Cd	STDV Cd	Mean Cl	STDV Cl	Mean Cm	STDVCm
Baseline	6.693E-2	2.918E-3	9.491E-1	1.575E-2	1.286E-1	6.237E-4
Optimized	5.207E-2	2.860E-3	9.423E-1	1.251E-2	1.127E-1	6.318E-4

To solve efficiently this single-objective optimization, we have applied a *Simplex* algorithm [13], with a global number of simulations not higher than the one considered for the multi-objective case. Figure 5 reports the results obtained, with same parameters (mean and standard deviation) reported in the axis; in green, baseline point and optimized point are highlighted.

The results are in this case much more satisfactory than the previous approach: the percentage of feasible designs is much higher than before, and after just few iterations a possible convergence trend is found, improving in an important way the baseline performance (similar standard deviation, but much lower mean value; see Table 3 for more details).

For the purpose of this comparison, we decided to stop the optimization after less than 40 designs, with a total number of CFD simulations equal to 370, which corresponds to about two weeks of analysis.

Table 2 reports the results obtained following the min-max approach, which are definitively satisfactory.

5 Reliability-Based RDO

We have pointed out in previous section that the main moments of the distribution of the performance of any design can be used to quantify its robustness, i.e. they can be used as criteria for a RDO problem (for instance, one could maximize the mean performance and minimize the standard deviation). Conversely, the min-max approach or more in general a Reliability-based Optimization (RBDO) problem, needs for the optimization criteria the definition of a reliability index or a failure probability. This approach can in fact be used, as noted above, to define accurately a min-max criteria (objective or constraint) also when the output performance is not necessarily of Normal type.

Many methodologies exist in literature to determine the failure probability, such as FORM/SORM [13], which for a RDO optimization could be very expensive from the numerical point of view. For this reason, we propose here a different methodology, based indeed on the Polynomial Chaos polynomial exploitation.

In fact, the evaluation of the performance function in industrial cases can be very demanding, since they often involve expensive CFD or structural numerical simulations. In the approach we propose, these expensive evaluations are required only to determine the coefficients of the PCE (Polynomial Chaos expansion). Once found them, it is possible to express the CDF (cumulative distribution function) of any system response using directly the PCE polynomial, which can be considered as a meta-model of the response, practically free in terms of CPU. Once the CDF is accurately obtained, from the given constraint value we can easily retrieve the corresponding percentage of the distribution, i.e. the *failure probability*.

In this way, a Robust Design Optimization problem can be defined, using as criteria for the optimization the minimization of the *failure probability*: in other words, we search for a new design whose failure probability for the given uncertainties distribution is minimum, either for a new design for which a given percentile (e.g. 99%) of its distribution is minimum. The big advantage of this approach with respect to using FORM/SORM methodologies is the reduced number of sampling points needed to obtain the Polynomial Chaos based meta-model, if compared to the iterations needed to compute the reliability index for each design required by FORM/SORM methodologies.

To validate the efficiency of the methodology proposed in this chapter, we have applied it to the same benchmark case used to describe the State of the Art techniques of RDO in previous section.

Table 4 RDO constraints accordingly to 6a and PCE-RBDO

Constraints	6 sigma definition	PCE-based
Constraint 1: min lift	mean_Cl - 3 * σ Cl = 0.90	0.03% = 0.887
Constraint 2: max drag	mean_Cd + 3 * σ _Cd = 0.0757	99.97% = 0.0747
Constraint 3: max momentum	mean_Cm + 3 * σ _Cm = 0.1305	99.97% = 0.1305

Table 5 Results of *Six-Sigma* based and RBDO method

Design	Cd-max	Cl-min	Cm-max
Baseline	7.47E-2	8.87E-1	1.302E-1
Optimized by 6sigma	*6.06E-2 **(6.33E-2)	*9.04E-1 **(8.94E-2)	*1.145E-1 **(1.129E-1)
Optimized by reliability	6.05E-2	8.95E-1	1.127E-1

In the previous case, we have approximated the needed percentile distributions of the performances (99.97%) by a *Six-Sigma* interval, that is however correct only under the hypothesis, not verifiable a priori, of a Normal distribution of the responses.

Following the new approach proposed, we can instead compute accurately the needed percentile distribution (99.97%) directly from the CDF distribution function defined by the Polynomial Chaos expansion, which is more accurate and valid also if the output distributions are generally different from a Normal distribution.

The definition of the constraints are therefore slightly different from the other approach, and Table 4 above reports the constraints values in the two cases.

In Table 5 we report a comparison of the performances (mean and standard deviation values) of the baseline configuration and of the optimized configuration obtained in each approach (also in this case *SIMPLEX* has been used).

The optimized solution is generally slightly different, considering the performance distributions, following the two approaches, but in both cases the constraints are respected and the selected objective is minimized.

Nevertheless, if in the first approach (*Six-Sigma*) the hypothesis followed is not necessarily correct (the performance distribution of the response does not necessarily follow a Normal distribution) and therefore not necessarily the 99.97% of the distribution really take the values estimated, by the new approach we can estimate with much more accuracy the needed percentile of the distributions, therefore we have can assume with an higher accuracy that the extreme values of the distributions take the values indicated, therefore respecting accurately the constraints.

To prove this assumption, we have re-evaluated the performances of the optimal design found by the *Six-Sigma* approach (* in Table 5), using this time the reliability approach, i.e. extracting the real 99.97%-ile value from the Polynomial Chaos expansion. These corrected values are reported inside brackets in the second row (** in Table 5). As we can note, the performances originally estimated by the *Six-Sigma* approach are in reality worse, and even though the differences are in this case not very

large, the results obtained applying the reliability criteria for the whole optimization (third row of Table 5), i.e. following the new methodology proposed in this section, are better, in particular for what concerns the objective function (drag minimization: $6.05E-2$ instead of $6.33E-2$).

In conclusion, to solve with highest efficiency the *min-max* approach for a RDO problem, a reliability-based approach is needed, and the Polynomial Chaos Expansion approach here described has revealed to be the most efficient approach.

6 Conclusion

In this paper we have illustrated some innovative methodologies for the Robust Design Optimization with large number of uncertainties, which is a typical requirement from the industry.

Two different UQ methodologies have been proposed, one based on SS-ANOVA and one based on a step-wise regression methodology, which can be used to reduce the number of sampling points for an accurate uncertainty quantification (either reducing the number of significant parameters, or reducing the number of Polynomial Chaos terms).

In addition, a methodology for efficient Robust Design Optimization (based on the application of min-max criteria combined with reliability-based optimization formulation and Polynomial Chaos exploitation for percentiles estimation) has been presented.

All the methodologies have been validated by the application of selected test cases in aeronautical field; in the future steps of UMRIDA Project, the proposed methodologies will be applied to an industrial problem of challenging relevance.

Acknowledgements This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. ACP3-GA-2013-605036-UMRIDA.

References

1. Clarich, A., Pediroda, V.: Robust design applications with modeFRONTIER, applying NODESIM-CFD tools. In: NODESIM-CFD Workshop on Quantification of CFD Uncertainties, Bruxelles, 29–30 October 2009
2. Loeven, G., Witteveen, J., Bijl, H.: Probabilistic collocation: an efficient non-intrusive approach for arbitrarily distributed parametric uncertainties. In: 45th AIAA Aerospace Science Meeting and Exhibit, Reno, Nevada (2007)
3. Kalsi, M., Hacker, K., Lewis, K.: A comprehensive robust design approach for decision trade-offs in complex systems design. *J. Mech. Des.* **123**(1), 1–10 (2001)
4. Clarich, A., Marchi, M., Russo, R.: Reliability-based optimization applying polynomial chaos expansion. In: 6th European Conference on Computational Fluid Dynamics (ECFD VI), July 20–25, 2014, Barcelona, Spain

5. Gu, C.: Smoothing Spline ANOVA Models. Springer, New York (2002)
6. Allen, D.M.: The relationship between variable selection and data augmentation and a method for prediction. *Technometrics* **16**, 125–127 (1974)
7. Blatman, G., Sudret, B.: Adaptive sparse polynomial chaos expansion based on Least Angle Regression. In Elsevier Science, June 2010
8. Efron, B., Hastie, T., Johnstone, I., Tibshirani, R.: Least angle regression. *Ann. Stat.* **32**, 407–499 (2004)
9. <https://sites.google.com/a/numeca.be/umrida/home>
10. Various Authors: Experimental data base for computer program assessment. Report of the Fluid Dynamics Panel Working Group, AGARD-AR-138 (1979). ISBN 92-835-1323-1
11. Farin, G.E., Hoschek, J., Kim, M.-S.: Handbook of Computer Aided Geometric Design, pp. 4–6. Elsevier (2002). ISBN 978-0-444-51104-1
12. Clarich, A., Pediroda, V., Periaux, J., Poloni, C.: Comparison between different game theory methodologies in robust design optimisation, Santorini, acts of Eccomas2005
13. Nelder, J.A., Mead, R.: A simplex method for function minimization. *Comput. J.* **7**, 308–313 (1965)