PCA-based sensitivity analysis of response fields using LS-OPT®

Charlotte Keisser[1], Pierre Glay[1], Max Hübner[2], Tobias Graf[2], Katharina Liebold[2], Anirban Basudhar[3], Imtiaz Gandikota[3], Nielen Stander[3]

[1]DYNAmore FRANCE SAS, Versailles, France
[2]DYNAmore GmbH, Stuttgart, Germany
[3]LST LLC, Ansys, Livermore, California, USA

1 Introduction

When performing an optimization, it is important to avoid introducing unnecessary variables that do not impact the design objectives and constraints. Such variables increase the design space size and lead to unnecessary sample evaluations, which can significantly increase the overall computation time or cost. A sensitivity analysis can be performed to quantify the significance of the variables; only the important variables are then used in the sampling and optimization, thus reducing the computational cost.

Linear ANOVA (Analysis of Variance) and Global Sensitivity Analysis (GSA/Sobol) sensitivity measures, which have been available in LS-OPT® for several years, allow the users to analyze the sensitivity of scalar responses to parameters. However, sensitivity analysis of vector entities such as time histories was not available earlier.

Since LS-OPT 7.0, a new sensitivity measure based on PCA (Principal Component Analysis) is available to analyze the importance of parameters for non-scalar entities. This sensitivity measure is available for both time histories and for multi-point responses (spatial data), which were also added in LS-OPT 7.0. This measure will be extended to multi-histories (time varying spatial data) in the future.

PCA is a method similar to Proper Orthogonal Decomposition (POD), which is commonly used in exploratory data analysis or for model prediction. It performs a dimensionality reduction of possibly correlated data sets by projecting each data point onto only a few principal components (PCs) forming a linearly uncorrelated orthonormal basis while preserving the data's variation. The PCs are sorted in decreasing order according to their variance, so that the first PC accounts for the largest variability in the data set. The new sensitivity measure provides the contribution of the parameters to each PC as well as the contribution of each PC to the results. A cumulative history/multi-point responses sensitivity measure is also available, allowing the analysis of the overall model.

A component test of a B-pillar is used to highlight this new feature. In crash scenarios, the nature of impact as well as the critical spatial points to consider on the structure may not be known a priori. Therefore, it may be useful to consider multi-point responses/histories, as demonstrated through this work.

2 Sensitivity Analysis with LS-OPT

A result can depend on many variables, and it could be tempting to play with all of them when performing an optimization. Unfortunately, the computational effort of an optimization increases drastically as the number of variables increases.

In most cases, only a few variables are significant. In such situation, a sensitivity analysis is very useful to determine the significance of design variables. This helps in the understanding of the simulation model's behavior and in selecting the most significant design variables.

Before LS-OPT 7.0, two sensitivity measures were implemented in LS-OPT: Linear ANOVA and GSA/Sobol (Section 2.1 and 2.2). Both are evaluated on responses (scalar results) using the metamodel. A sensitivity analysis method based on PCA was added in version 7.0 for fields.

2.1 Linear ANOVA

The Analysis of Variance (ANOVA) is a linear sensitivity measure used to rank the design variables for screening purposes. The procedure only requires a single iteration using a polynomial regression (even though the results are produced after every iteration of a normal optimization procedure).

ANOVA is a regression-based sensitivity measure with

$$b_j = \frac{\partial f}{\partial x_j} \cdot \Delta x_j \text{ , } j = 1, \ldots, N,$$

Where f is the linear approximation, Δx_j the size of the design space of variable x_j, and N the number of variables, Fig.1:.

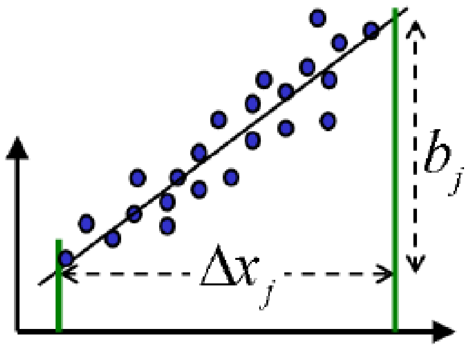

Fig.1: Definition of ANOVA value b_j

ANOVA depicts positive or negative influence and the significance of a variable j regarding to the response is represented by the b_j value. The higher $|b_j|$ is, the most significant the variable is to the response.
As ANOVA is computed based on linear approximations, it may not be accurate. So, in order to guide the user, a confidence interval is available to show the trustworthiness of the computed value. In Fig.2:, $100(1-\alpha)\%$ represents the level of confidence that b_j will be in the computed interval. LS-OPT shows a 95% confidence interval by default, meaning that we are 95% sure that the ANOVA value is inside that interval.

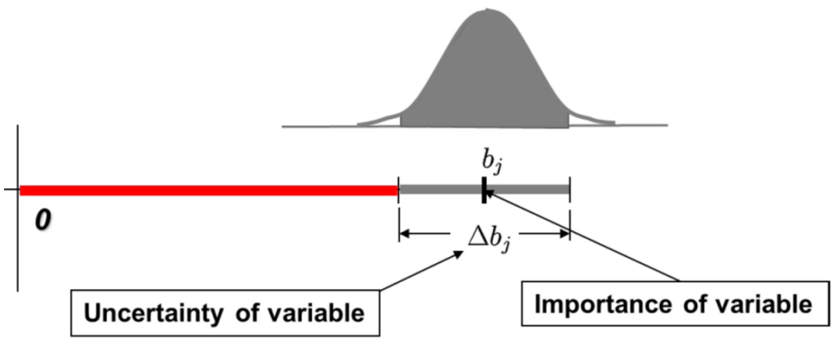

Fig.2: ANOVA value with $100(1-\alpha)\%$ confidence interval

2.2 GSA/Sobol

While the ANOVA is a very popular method to observe the contribution of different regression terms, Global Sensitivity Analysis (Sobol's method) is a non-linear sensitivity measure widely used to study the importance of different variables for higher order models. In this method, a function is decomposed into sub-functions of different variables such that the mean of each sub-function is zero and that each variable combination appears only once:

$$f(x_i, \cdots, x_n) = f_0 + \sum_{i=1}^{n} f_i(x_i) + \sum_{i=1}^{n}\sum_{j=1}^{n} f_{ij}(x_i, x_j) + f_{1,\cdots,,n}(x_1, \cdots x_n).$$

Then, each sub-function variance represents the variance of the function with respect to that variable combination:

$$V_{i,-,j} = \int_0^1 \cdots \int_0^1 f_{i,\cdots,j}^2(x_i, \cdots, x_n) dx_i \cdots dx_j.$$

GSA/Sobol has the particularity to show the absolute influence value and these values are normalized so they can be summed up to determine the influence of a variable on multiple responses.

2.3 Principal Component Analysis

Principal Component Analysis (PCA) is a multivariate statistical technique, meaning that it can be used for multi-points data like curves (histories) or spatial data (multi-point responses). The method uses an orthogonal transformation to convert a large set of possibly correlated observations into a small set of uncorrelated values called Principal Components. This technique is very useful to reduce the size of the data set without losing information.
The Principal Components are sorted in decreasing order according to the variance, so that the first Principal Component accounts for the largest variability in the data set. The variability in the data set of the k-th Principal Component is represented by the inertia λ_k which is the eigenvalue of the correlation matrix.
$H_k = XV_k$ is the k-th Principal Component, with $X^T X = V \Lambda V^T$ the eigenvalue decomposition of the correlation matrix of data set X.
Then to analyze the effect of design variables on the data set, a method similar to ANOVA is applied to the Principal Components. The importance of a variable i to the k-th Principal Component is represented but a sensitivity index:

$$SI_{i,k} = \|S_i H_k\|^2 / \lambda_k \in [0,1]$$

where S_i is the orthogonal projection of the variable values.
Unlike Linear ANOVA and GSA/Sobol, the advantage of PCA is to allow the study of the variables' effect on multi-points data. So instead of looking at the maximal or minimal or final value of a result (e.g. the force), the user could have a sensitivity analysis based on the entire result data set varying in space or time.

2.4 DOE task

A Design of Experiment (DOE) study is used to explore the design space or to calculate sensitivities. The DOE task performs a global approximation of the results using at least $2 * (n + 1)$ recommended samples, n being the number of variables. The greater the non-linearity of the response is, the more points are needed to represent the non-linearity. The number of simulation points is a compromise between accuracy and computational effort. The samples and results can be reused for the subsequent optimization after screening out the unimportant variables, which are treated as constants.

3 B-Pillar model

An LS-DYNA B-pillar model has been used in this work to demonstrate the sensitivity analysis and variable reduction features. The model was extracted from a full-scale model of a Toyota-Yaris developed by George Washington University National Crash Analysis Center [/www.nhtsa.gov/crash-simulation-vehicle-models]. About 15 to 30 cm of the adjacent dwellers were added to the B-pillar and rails were also extracted to represent an adequate support for the B-pillar. The rails and dwellers were constrained using *BOUNDARY_SPC node sets as can be seen in Fig.3:.

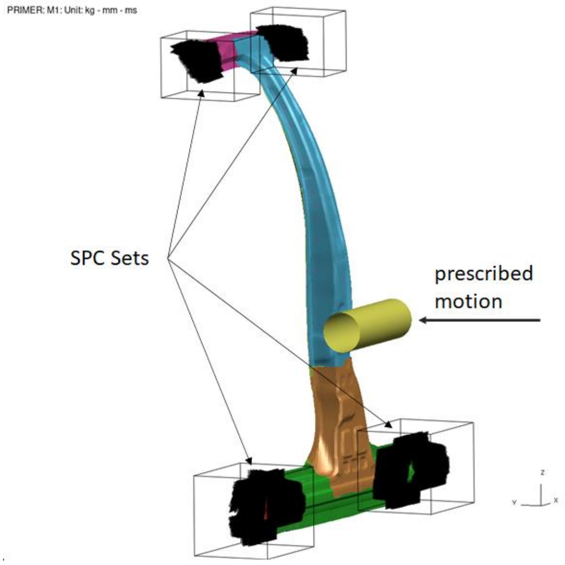

Fig.3: B-pillar with SPC node sets.

The original mesh of the Toyota Yaris was redefined in order to create a mesh size of about 3.5mm edge length. The mesh consists of quadrilateral elements using ELFORM 16, fully integrated shell elements. The refinement leads to a model with about 111.000 shell elements and a little under 200 solid elements as connection elements.
MAT_24 is used to describe the material behavior of the modelled parts. The mesh size of the B-pillar used for the simulations is shown in Fig.4:.

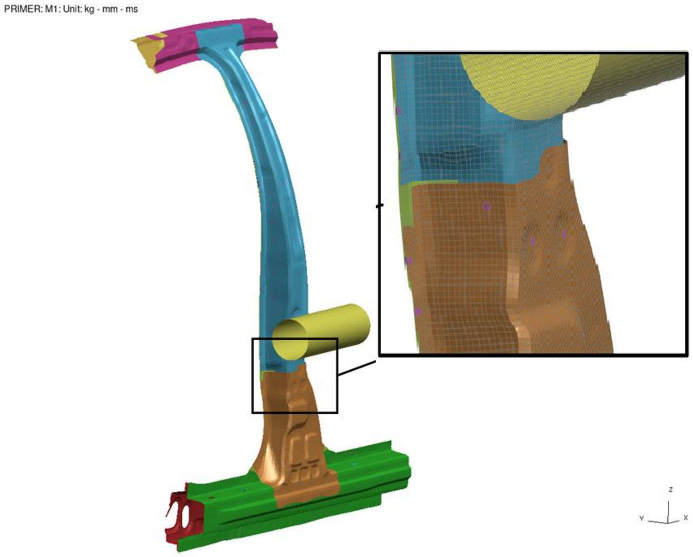

Fig.4: B-pillar mesh size

During the simulation, the B-pillar is experiencing an impact at one third of its height by a cylindrical impactor of 10 cm diameter. The impactor is modelled as rigid and impacts the B-pillar with an initial velocity of 150mm/ms with the other degrees of freedom being blocked. The direction of the impactor striking the B-pillar is shown in Fig.5:.

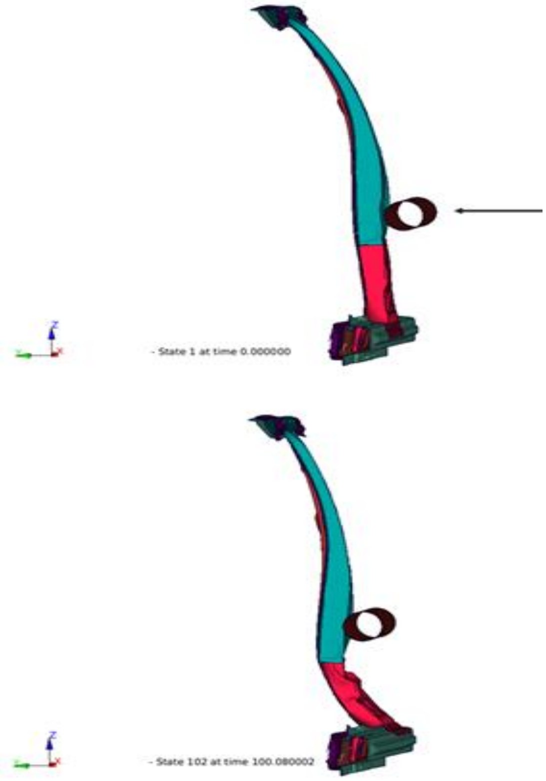

Fig.5: Impactor striking the B-pillar

4 Results: B-pillar Sensitivity Analysis and Optimization

4.1 Design Formulation and Sensitivity Analysis

In order to highlight the Principal Component Analysis sensitivity measure, the B-Pillar model described in Section 3 was used in a DOE study. This section presents the design formulation for the eventual optimization problem and the quantities of interest for sensitivity analysis, as well as the results of the analyses using a DOE task in LS-OPT. The design formulation is as follows.

$$\begin{cases} min \quad Ydisp_{bottom}(final\ time) \\ s.t. \quad Ydisp_{ratio} \leq 0.33,\ where\ Ydisp_{ratio} = \frac{Ydisp_{top}(final\ time)}{Ydisp_{bottom}(final\ time)} \end{cases}$$

For this model, we are interested in 7 design variables:
- ***sfa*** a scaling factor for the compressive yield stress of the aluminum material (the basic value of the yield stress being 400 MPa, so the variation will be between 380 MPa and 420 MPa);
- 6 thickness variables ***t462_t***, ***t_462_b***, ***t484***, ***t491***, ***t493***, ***t494***.

Design Variable	Initial Value	Minimum	Maximum
Sfa	1	0.95	1.05
t462_t	0.9855	0.9	1.8
t462_b	0.9855	0.9	1.8
t484	1.19	0.8	1.7
t491	2.56	2.1	3.1
t493	2.12	1.6	2.6
t494	2.12	1.7	2.2

Table 1: Design Space for the DOE task.

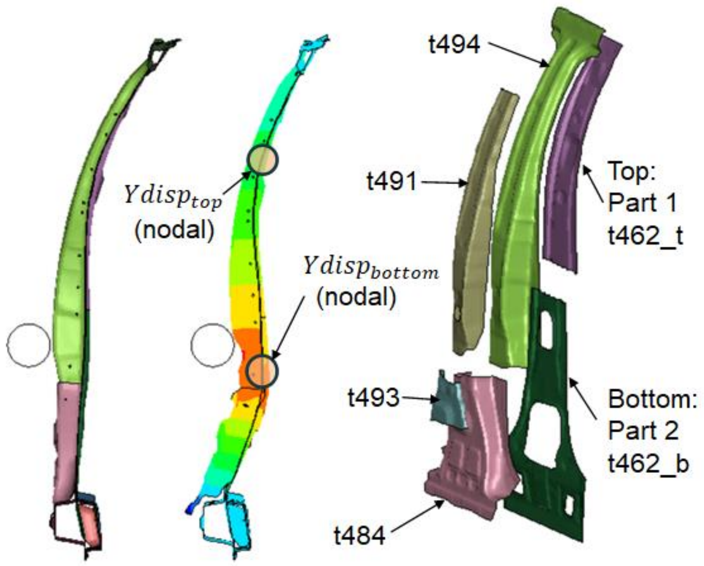

Fig.6: Part IDs of the B-Pillar

Based on the design formulation, the entities of interest for sensitivity analysis are shown below. Though in reality maximum displacement of the b-pillar due to impact is critical for occupant safety, since the crash test data consists of final displacement values, auto manufactures rely on meeting the final displacement requirements. Therefore, final Y displacement values were extracted as design response and multi-response (spatial data) from LS-DYNA analyses of the DOE study. Moreover, the final displacement values are in general noisier than the maximum displacement due to variations in b-pillar spring back or recovery after the impact.

- Responses:
  - The Y displacement of the bottom node at the final time;
  - The Y displacement ratio between the top node and the bottom node at the final time;
- Histories:
  - The Y displacement of the bottom node (the full curve);
  - The Y displacement difference between the bottom node and the top node (the full curve);
- Multi-responses:
  - The Y displacement of all the nodes of the B-Pillar bottom part at the final time;
  - The Y displacement of all the nodes of the B-Pillar top part at the final time.

Responses	Histories	Multi-responses
$Ydisp_{bottom}(final\ time)$	$Ydisp_{bottom}$	$Ydisp_{bottom\_part}$
$Ydisp_{ratio} = \frac{Ydisp_{top}(final\ time)}{Ydisp_{bottom}(final\ time)}$	$Ydisp_{diff} = (Ydisp_{bottom} - Ydisp_{top})$	$Ydisp_{top\_part}$

Table 2: List of observed results

Linear ANOVA and GSA/Sobol were used for response sensitivities and PCA was used for histories and multi-responses. The metamodels were constructed using Feedforward Neural Network with 55 space-filling samples.

4.1.1 Response results

First, having a look at the ANOVA results for the 2 responses (final Y displacement nodal bottom and Y displacement ratio), the variable ***t484*** has a very large influence on both responses and ***t462_t*** also is important for both responses. These 2 variables should stay for the optimization afterwards. Variable ***t494*** appears at the 4th important variable for both results. Variables ***t462_b*** and ***t491*** appears in the 4th important variables for at least one of the responses. Just looking at these results is not easy to decide whether we should keep or not these variables for the optimization.
Regarding the scaling factor ***sfa***, it seems to not be important for any of the responses and should be excluded for the optimization. It is interesting to note that GSA ignores the importance of ***t462_b***, although this relates to the part of b-pillar with highest deformation.

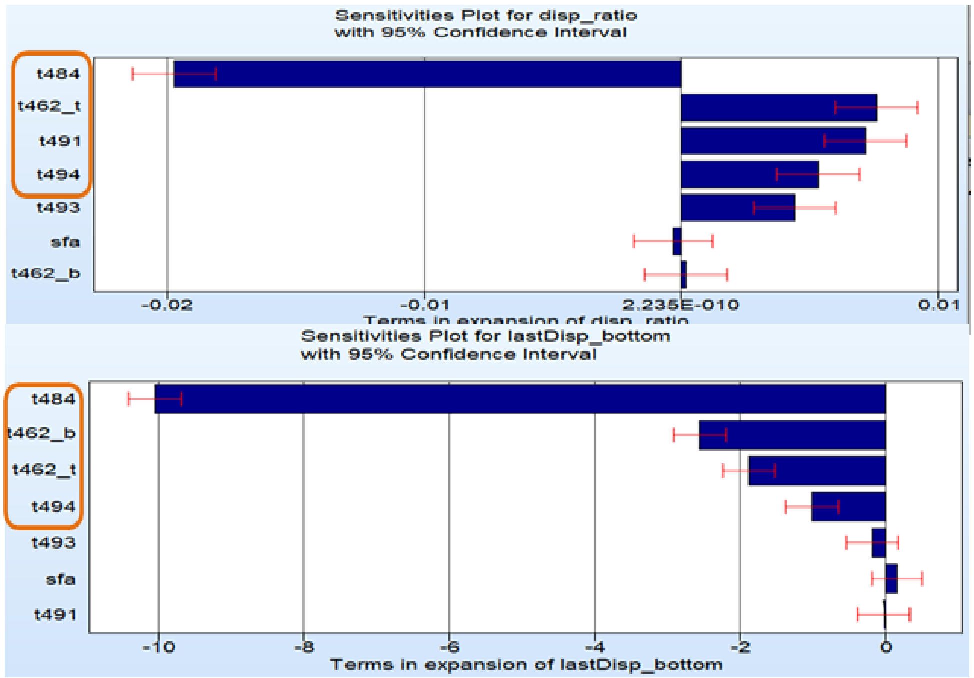

Fig.7: ANOVA B-pillar response sensitivity results for the DOE task. The four most important variables for the objective function and constraint are marked with the brown boxes. The important variables for the objective function and the constraint function are not identical.

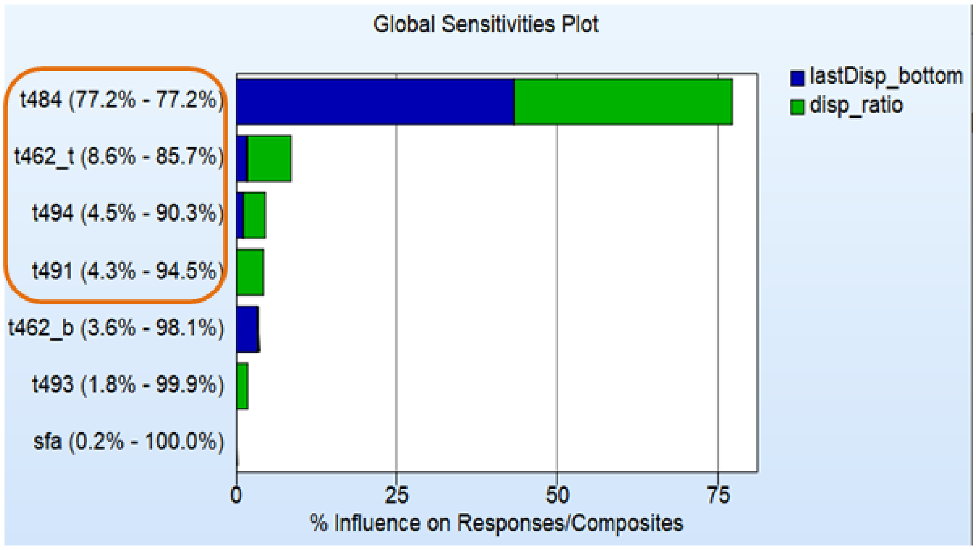

Fig.8: GSA/Sobol B-pillar response sensitivity results. The four most important variables with highest combined contribution to the objective function and constraint are in the brown box.

Looking at GSA sensitivity results, the influence of ***t484*** and ***t462_t*** is again visible. If variables ***t494*** and ***t491*** are kept, the 4 variables already represent 94.5% of the result variance which seems to be a reasonable choice for the optimization.

4.1.2 History results

Now looking at the sensitivity of the histories ($Ydisp_{bottom}$and $Ydisp_{diff}$), the importance of ***t484*** on both histories (70.57% and 80.80%) is again very clear. Variable ***t462_t*** (7.87% and 1.74%) seems actually to be less important than ***t462_b*** (15% and 11.32%). Variable ***t494*** appears as the 4th important variable excluding ***t491***.
Again, ***t493*** and ***sfa*** variables seem to not be important to the model and should be removed for the optimization.
If we add the percentages importance, variables ***t484***, ***t462_t***, ***t462_b***, ***t494*** already represents more than 95% of the total variance of the histories and is a reasonable choice for the optimization.

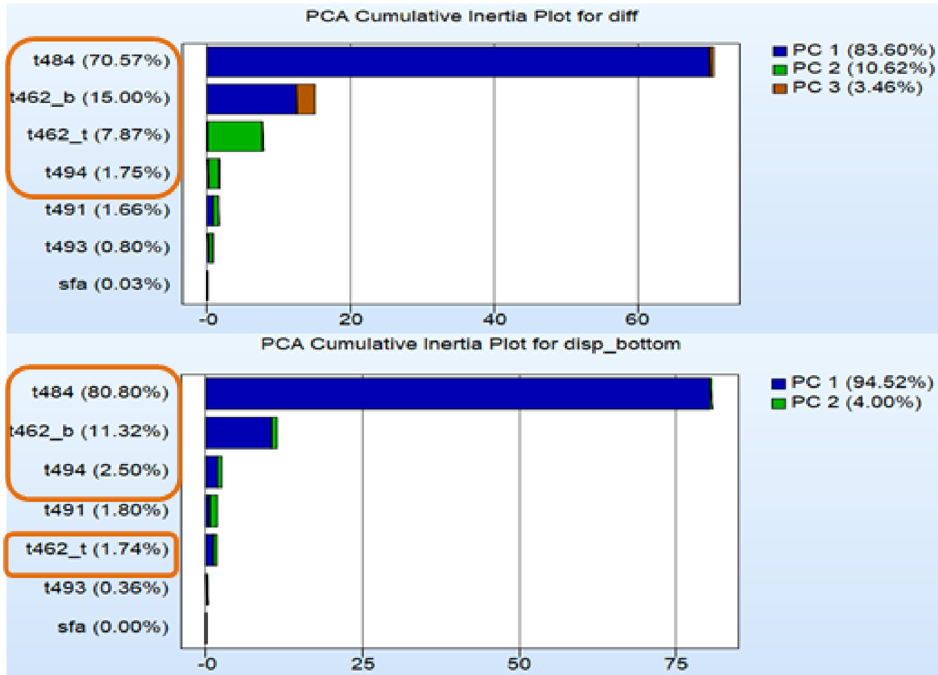

Fig.9: PCA history sensitivity results for the DOE task. The four most important variables for the displacement difference and bottom displacement histories are not identical. The final four

most important variables for optimization (in brown boxes) are selected based on the sum of the two influence measures.

4.1.3 Multi-response results

Now looking at the sensitivity of the multi-responses ($Ydisp_{bottom\_part}$ and $Ydisp_{top\_part}$), there is again no doubt on the importance of variable ***t484*** on both multi-responses (95.03% and 93.43%). Variable ***t462_b*** (2.83% and 4.05%) is more important in this case than ***t462_t*** (0.09% and 0.25%). Variable ***t491*** is the 4th important variable in this case, excluding ***t494***.
If we add the percentages importance, variables ***t484***, ***t462_t***, ***t462_b***, ***t491*** already represents more than 98% of the total variance of the multi-responses and is a reasonable choice for the optimization.

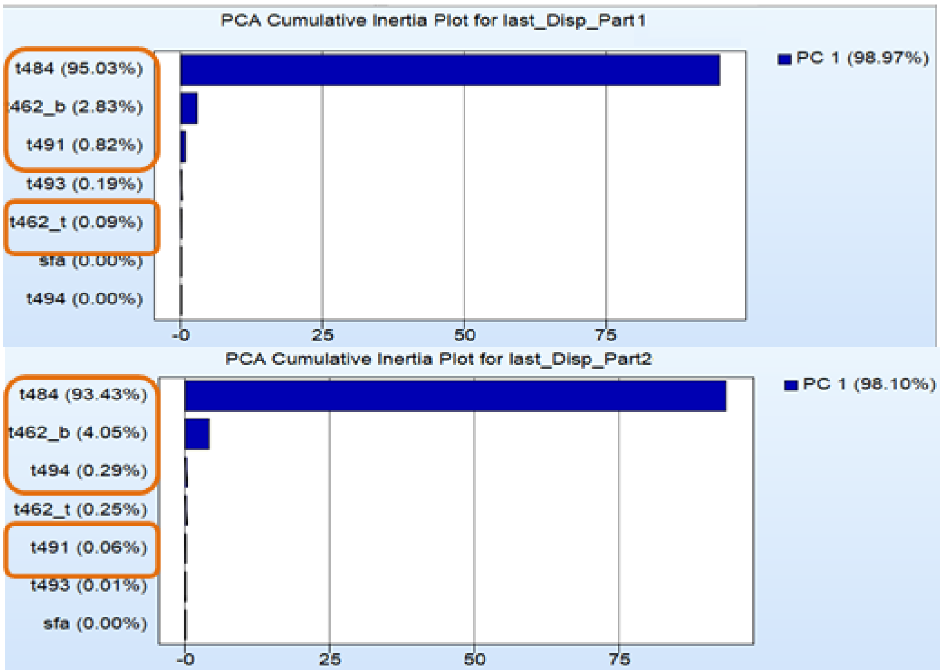

Fig.10: PCA multi-response sensitivity results for the DOE task. The four most important variables for the top and bottom part displacement multi-responses are not identical. The final four most important variables for optimization (in brown boxes) are selected based on the sum of the two influence measures.

4.1.4 Summary

Variables	**ANOVA**	**GSA/Sobol**	**PCA-history**	**PCA-Multi-response**
Sfa				
t462_t	√	√	√	√
t462_b	?		√	√
t484	√	√	√	√
t491	?	√		√
t493				
t494	?	√	√	

Table 3: Design variable selection after DOE

Depending on the applications, if the user knows the nodal point that is interesting to look at and the point in time (the final time or the maximum or the minimum point) that should be considered, then looking at GSA and ANOVA might be enough. But if the user is not really sure where or when the interesting information is, then he might be interested in looking at a bigger picture and use PCA.
In the context of this study, we decided to look at everything and defined 4 design variables sets:

- Full: containing all design variables;
- GSA-based: containing the 4 design variables highlighted by the GSA sensitivity (***t484***, ***t462_t***, ***t494***, and ***t491***);

- PCA-history based: containing the 4 design variables highlighted by the PCA sensitivity on histories (***t484***, ***t462_t***, ***t462_b*** and ***t494***);
- PCA-Multi-response based: containing the 4 design variables highlighted by the PCA sensitivity on multi-responses (***t484***, ***t462_t***, ***t462_b***, and ***t491***).

Two types of optimization are also performed for this study:
- A single iteration optimization with 55 samples
- A SRSM (sequential with domain reduction) optimization with 5 maximal iterations.

4.2 Single iteration optimization

To perform a fair comparison of the different sensitivity analysis methods, the same number of most important variables are selected for each method (GSA-based response sensitivity/PCA-based history sensitivity/PCA-based multi-response sensitivity). Additionally, the same number of samples are used to perform an optimization using each variable set. The results are also compared to those obtained using all the variables. The results are listed in Table 1: (note that the values displayed in black in the table correspond to the variables set to constant for the regarding optimization).
The GSA-based set of design variables gives worse results than the full set. Whereas the PCA-based sets gave similar results as the optimization with all the variables. We can also see that for all the variables, the optimum ***t484*** value is 1.7. As this variable is the most important, it dominates the optimization and hits the upper bound in this case. The other optimal variable combinations, which are based on different important variable choices depending on the analysis method, lead to differences in the optimum objective function value.
In general, once a DOE is performed to obtain sensitivity information and an initial estimate of the optimum, a sequential optimization is performed using a reduced set of design variables (runs from the DOE can be included in the optimization to avoid waste). Single iteration optimization was performed in this section to demonstrate the influence of variable choice using the same sample set.

Values	**Full**	**GSA-based**	**PCA-history based**	**PCA-Multi-response based**
Sfa	1	***1***	***1***	***1***
t462_t	1.8	1.8	1.8	1.47866
t462_b	1.8	***0.9855***	1.8	1.8
t484	1.7	1.7	1.7	1.7
t491	2.568	3.05	***2.56***	3.1
t493	2.527	***2.12***	***2.12***	***2.12***
t494	2.2	2.02	2.2	***2.12***
$\boldsymbol{Ydisp_{bottom}(final\ time)}$ (objective function)	66.992	68.15	66.885	66.952
$\boldsymbol{Ydisp_{ratio}}$ (constraint)	0.327714	0.327008	0.32561	0.328487

Table 4: Optimal results of single iteration optimization for the 4 different sets of variables. The variables in the black boxes with white text are set to the baseline values and were not considered as important for optimization using the different sensitivity criteria.

4.3 Sequential response Surface Method (SRSM) based Optimization

This section provides a comparison of the optimal solutions obtained using iterative optimization. This is what a user will normally do after a DOE. The results are listed in **Error! Reference source not found.** (note that the values displayed in black in the table correspond to the variables set to constant for the regarding optimization).
Here, again, the results obtained for the GSA-based set are worse than with the full set or with the PCA-based sets. The optimization with the PCA-history based set of design variables gave similar results than the full but converged faster (2 iterations instead of 5), required less runs per iterations (8 instead of 13) and only took 2hours and 26 minutes instead of 14hours and 22 minutes for the full optimization. For some reason, the optimization using the variable set based on PCA-based multi-response sensitivity did not converge and gave relatively slightly worse results compared to the PCA-history based set. The history sensitivity is in fact expected to give the best results as the objective function is defined for a single node, which is the same node considered during history sensitivity analysis. Additionally, the history sensitivity considers the difference of top and bottom displacements

at all time states unlike the scalar response defined at the last time that can be noisy at times due to different top and bottom recovery start times.

Values	**Full**	**GSA-based**	**PCA-history based**	**PCA-Multi-response based**
sfa	1	***1***	***1***	***1***
t462_t	1.8	1.8	1.8	1.8
t462_b	1.8	***0.9855***	1.8	1.8
t484	1.7	1.7	1.7	1.7
t491	3.1	2.1	***2.56***	2.528
t493	1.777	***2.12***	***2.12***	***2.12***
t494	2.2	2.02	2.2	***2.12***
$\boldsymbol{Ydisp_{bottom}}(\boldsymbol{final\ time})$ (objective function)	66.8086	69.5322	66.8595	67.1323
$\boldsymbol{Ydisp_{ratio}}$ (constraint)	0.328146	0.326956	0.325617	0.325541
Experiments per iteration	13	8	8	8
Convergence	no	Yes	Yes	No
Number of iterations	5 (max)	2	2	5 (max)
Total execution time	14h22’04”	02h07’52”	02h26’53”	06h19’46”

Table 5: Optimal results of SRSM optimization for the 4 different sets of variables. The variables in the black boxes with white text are set to the baseline values and were not considered as important for optimization using the different sensitivity criteria.

5 Summary

Sensitivity analysis is an important feature for understanding the behavior of a model’s response. It highlights the important variables and helps in selecting the most important variables for optimization. Variable screening offers a significant gain in terms of time and resources when dealing with large models with several design variables. Even by performing the DOE first and then running the optimization, the gain is still significant.

In this work, ANOVA and GSA measures are used to obtain sensitivities of responses. While ANOVA measures only the linear relation between variables and responses, GSA/Sobol is a non-linear measure giving more reliable and accurate information as long as the meta-model is non-linear. If the user has prior knowledge of the location and time for quantities of interest and the responses are well behaved, using response sensitivity is sufficient to analyze the importance of variables for such scalar quantities. However, if the user wants to have a more general analysis, using response may sometimes lead to loss of critical information.

In contrast to the methods employing responses only, PCA is a more general approach which offers an analysis from a field point of view. It is a new feature that was added in LS-OPT 7.0 and it enables performing sensitivity analysis of histories and multi-responses. For certain problems, such as the one in this paper involving combination of results at different spatial points that may start recovery at different times, analysis of a history may involve lesser noise compared to the corresponding responses.

In the future, we are looking to extend the PCA measure to multi-histories to make the implementation more flexible.

6 Literature

[1] Lamboni, M., Makowski, D., & Monod, H: “Multivariate global sensitivity analysis for discrete-time models”, PhD dissertation, auto-saisine, 2008, 2015.

[2] Williams, L.J.: “Principal component analysis”, Wiley interdisciplinary reviews: computational statistics, 2(4), 2010, 433-459.

[3] Missoum, S.: “Probabilistic optimal design in the presence of random fields”, Structural and Multidisciplinary Optimization, 35(6), 2008, 523-530.

[4] Turk, M., & Pentland, A.: “Eigenfaces for recognition”, Journal of cognitive neuroscience, 3(1), 1991, 71-86.