

Statistics and Non-Linear Sensitivity Analysis with LS-OPT[®] and D-SPEX

H. Müllerschön, M. Liebscher
DYNAmore GmbH, Germany

W. Roux, N. Stander
LSTC (Livermore Software Technology Corporation)

U. Reuter
Department of Civil Engineering, TU Dresden, Germany

Abstract

For stochastic simulations usually many simulations are performed, therefore much information is available for the simulation engineer. In order to evaluate this information and to assess the results of stochastic investigations software tools such as LS-OPT and D-SPEX are available. Good and clearly arranged presentation of the results is important, so that the engineer really benefit from the data mining.

D-SPEX is intended to provide features that are not currently implemented in the LS-OPT viewer. Therefore, it is a complement to the visualization capabilities of LS-OPT. Its primary focus is on the visualization of meta-models although it also provides features to visualize stochastic results. D-SPEX is also thought of as a testing platform for new features that might evolve in LS-OPT. By opening the command file of LS-OPT, D-SPEX reads all data of the optimization or robustness project. The current version of D-SPEX is fully compatible with LS-OPT version 3.3. For more information on the features of D-SPEX, see [12].

Introduction

In combination with LS-OPT, D-SPEX supports CAE engineers in post-processing of stochastic analysis results. In the field of stochastic analysis D-SPEX is a valuable and easy to use tool. Tasks in stochastic analysis that can be solved with LS-OPT and D-SPEX can be arranged in the following groups:

Uncertainty propagation within a stochastic analysis of a mechanical model the randomness of structural parameters (input parameters) is propagated through the computational model. Randomly distributed responses (model output) are obtained as result, see Figure 1. Together with the obtained results statistical methods allow now, e. g., the computation of the mean value and variance and the determination of distribution types of the considered random responses.

Sensitivity analysis determines the significance/contribution of the structural parameters onto to considered responses. The identification of non-relevant and relevant structural parameters for model reduction purposes is one of the major tasks. The sensitivity analysis may also improve the understanding of the model behavior and may clarify interactions among input parameters.

Structural reliability analysis with respect to constraints (failure, damage, requirements ...).

Visualization of response variations by fringing statistical results on the FE-Model. This is possible by utilizing the DYNASTats capabilities in LS-OPT.

This paper focuses in particular on the statistical analysis / post-processing of random results and the provided sensitivity analysis facilities. The sensitivity analysis is considered in view of the nonlinearity of the underlying computational model (FE model).

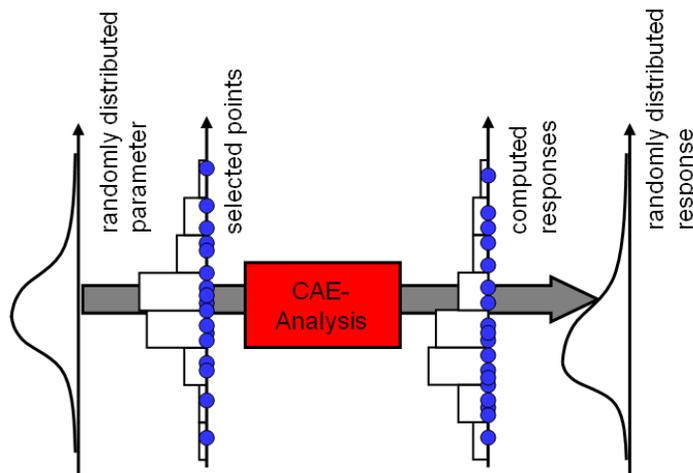


Figure 1 Uncertainty propagation in probabilistic mechanics

Statistical post-processing in D-SPEX

Several statistical techniques summarize and describe an obtained data collection. These techniques are notionally aggregated to the *descriptive statistics techniques*. They form the basis of virtually every quantitative analysis of data.

The most frequently used and probably most beneficial statistical measures are introduced briefly and can be evaluated with D-SPEX. They are presented to the user in a clear and visually attractive form.

Mean Value, Standard Deviation / Variance, Anthill Plots

Mean value and standard deviation are basic features of the data set. They give a feeling what can be expected and with which spreading. The variance V is determined by

$$\text{Var}(X) := V(X) := E((X - \mu)^2) \quad (1)$$

where $\mu = E(X)$ denotes the expected value of the random variable X . The standard deviation is defined as square root of the variance

$$\sigma = \sqrt{\text{Var}(X)} = \sqrt{E((X - \mu)^2)} \quad (2)$$

Anthill plots provide an overview of the relation between two random variables. The kind of relation linear, non-linear ... (if there is any) can be identified sometimes visually. Mean values, standard deviations and anthill plots of previously selected data are presented for convenience in one single plot (Figure 2).

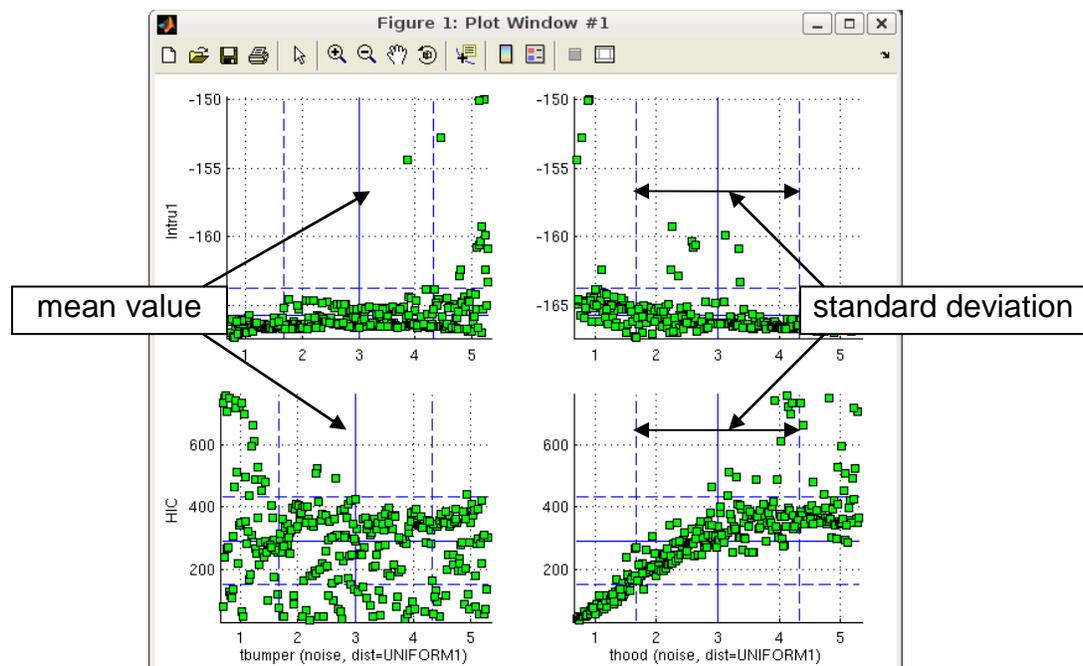


Figure 2 Mean values, standard deviations and anthill plots are visualized exemplarily for structural parameters (thood, tbumper) and responses (HIC, Intru1). Green color of points indicate feasible runs, red would be infeasible (violated constraints)

Correlation analysis

Correlation, (often measured as a correlation coefficient), indicates the strength and direction of a linear relationship between two random variables [13]. A number of different coefficients are used for different situations. The best known is the Pearson product-moment correlation coefficient, which is obtained by dividing the covariance of the two variables by the product of their standard deviations. The correlation coefficient $\rho_{X,Y}$ between two random variables X and Y is defined as

$$\rho_{X,Y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)} \sqrt{E(Y^2) - E^2(Y)}} \tag{3}$$

again E denotes the expected value. For a result of a numerical experiment e.g. obtained through a Monte Carlo Simulation the Pearson product-moment correlation coefficient can be used to estimate the correlation X and Y . The Pearson correlation coefficient is then the best estimate of the correlation of X and Y . The Pearson correlation coefficient is written:

$$r_{xy} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}} \tag{4}$$

The values of the correlation coefficients after Eqs. (3) and (4) are located in the interval $[-1, 1]$. Where values near to 0 indicate a weak (linear) relationship and values near to -1/1 a strong (linear) relationship between the investigated random variables X and Y .

The correlation is a very popular and powerful measure to summarize relationships between variables but in the case of outliers or (weak) non-linearity it may lead to wrong conclusions. Therefore a correlation analysis does not replace the individual examination of the data (e.g. through anthill plots).

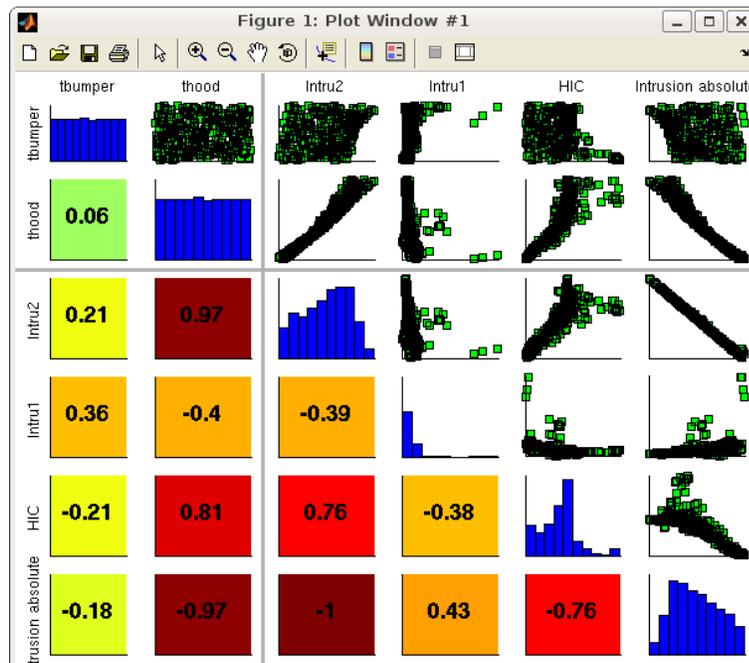


Figure 3 Correlation coefficients, histograms and anthill plots of an example data set

Determination of the probability distribution type

For further investigations it might be necessary to determine the distribution type and the corresponding distribution parameters of a random response. Several points in the parameter space have been selected with in a previously performed stochastic analysis and structural responses have been computed. The obtained set of computed responses (Figure 1) represents a random sample set. In order to determine a possible distribution type of the random response statistical tests are applied onto this sample set. The Kolmogorov-Smirnov test is one of the most frequently used. The empirical distribution $F_0(x)$ of the sample set is compared with a given / proposed cumulative distribution function (cdf) $F_X(x)$. In the notation of statistical testing theory hypothesis $H_0 : F_0(x) = F_X(x)$ is tested against $H_1 : F_0(x) \neq F_X(x)$.

The empirical distribution function $F_0(x)$ of observations X_i is defined as

$$F_0(x) = \frac{1}{n} \sum_{i=1}^n I_{X_i \leq x} \quad (5)$$

where $I_{X_i \leq x}$ represents the indicator function. The goodness of fit is evaluated with Kolmogorov-Smirnov statistic

$$D_n = \sup_x |F_n(x) - F_0(x)| \quad (6)$$

The null hypothesis H_0 is rejected at level α (often $\alpha = 0.05$) if

$$D_n > \Delta_{n;1-\alpha/2} \quad (7)$$

where $\Delta_{n;1-\alpha/2}$ is a threshold value that is computed on the basis of the Kolmogorov distribution. That means the result is – with respect to the level α – binary (true/false). It is frequently misunderstood that if the null hypothesis H_0 is not rejected it does not mean that H_0 necessarily true (but might be). Literally it does not mean $F_0(x) = F_X(x)$. That is unfortunately not specific to the Kolmogorov-Smirnov test, but for all statistical tests procedures. Besides the Kolmogorov-Smirnov test D-SPEX provides also a visual representation of the histogram vs. proposed pdf and empirical distribution function vs. proposed cdf, see Figure 4.

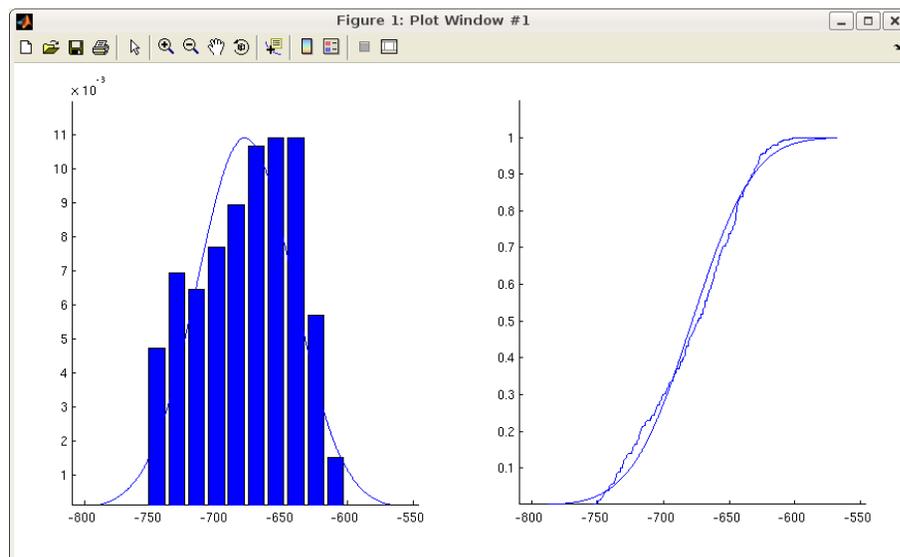


Figure 4 Visualization of the histogram/ proposed pdf and the empirical distribution / proposed cdf

Linear and Non-linear Sensitivity Analysis

The structural optimization and uncertainty analysis of moderate to complex computational models is costly and often requires additionally effort due to the high number of design variables that might be considered in the design process. Responses can depend on all variables, but often the significance onto the variance of the structural responses differs greatly. In many cases a structural response is dominated by only a few variables. Objectives of a sensitivity analysis might be [1]:

- Identification of significant and insignificant variables. Possibly reduction of the dimension (number of design variables) of the optimization problem.
- Improvement of understanding of the model behavior (highlight interactions among variables, find combinations of variables that result in high, or low, values for the model output)

In this sense a suitable definition of global sensitivity can be found in [2]

“Sensitivity analysis (SA) is the study of how the variation in the output of a model (numerical or otherwise) can be apportioned, qualitatively or quantitatively, to different sources of variation.”

Most of the global sensitivity analysis approaches can be classified into two groups [3]:

- *Regression-based methods:* The standardized regression coefficients (SRC) are based on a linear regression of the output on the input vector. In a linear regression, the value of the standardized regression coefficient is exactly the same as the correlation coefficient. The Pearson correlation coefficients measure the effect of each design variable by the correlation it has with the structural response. In the case of monotone but nonlinear structural behaviour rank based measures may be used, leading to the so-called SRRC (standardized rank regression coefficients). Based on these measures only limited conclusion can be drawn in the case of general nonlinear non-monotonic models. This conclusion is also appropriate for quadratic correlation approaches.
- *Variance-based methods:* The observed variance of the structural responses is partitioned into components induced by the respective structural variables. This decomposition of the variance is usually called ANOVA techniques for “ANalysis Of VAriance”. The Fourier amplitude sensitivity test (FAST) indices [6],[7] and the Sobol’ indices [5] are intended to represent the sensitivities for general (non-linear) models.

LS-OPT and D-SPEX provide the ability to evaluate the correlation coefficient and to perform a linear ANOVA. The Sobol’s variance-based approach, the Sobol’ indices, has received much attention in literature in the last years since they provide accurate sensitivity information for most models [3]. Sobol’s approach is implemented in LS-OPT and is called *Stochastic Contributions*.

It is proposed that the application of Sobol’s approach, correlation analysis and (polynomial) regression based ANOVA will lead to comparable results for the linear models. The obtained ranking of variables in terms of sensitivity will be almost equal. However, in the general nonlinear case it is assumed that Sobol’s approach results in more satisfactory sensitivity measures.

A brief overview about the decomposition of variance after Sobol is given in the following section. The (regression-based) correlation coefficient, the (variance-based) linear ANOVA, and Sobol’ indices (Stochastic Contributions) are compared at first by means of analytical examples followed by a simple crash analysis.

Stochastic Contributions / Sobol's Approach

The Sobol' indices are sensitivity measures for arbitrary complex computational models. They estimate the effect of the random input variables onto the model output [8]. This section just provides a short introduction.

The models under investigation are described by a function $\underline{Y} = f(\underline{X})$, where $\underline{X} = (X_1, X_2, \dots, X_n)$ is the random input vector consisting of n random variables (i.e. structural parameters) and where $\underline{Y} = (Y_1, Y_2, \dots, Y_m)$ denotes the random output vector consisting of m random variables (i.e. structural responses).

Each random model response Y_j ($j = 1, 2, \dots, m$) is characterized by its variance D^j . It is possible to decompose each variance D^j into partial variances associated with the single random input variables X_1, X_2, \dots, X_n as follows:

$$D^j = \sum_{i=1}^n D_i^j + \sum_{1 < i < k \leq n} D_{i,k}^j + \dots + D_{1,2,\dots,n}^j \quad (8)$$

and to relate each partial variance to one Sobol' index

$$S_{i_1, \dots, i_s} = \frac{D_{i_1, \dots, i_s}^j}{D^j} \quad \text{with} \quad 1 \leq i_1 < \dots < i_s \leq n, \quad s = 1, 2, \dots, n \quad (9)$$

Each of the Sobol' indices S_{i_1, \dots, i_s} represent a sensitivity measure that describes which amount of each variance D^j is caused due to the randomness of the single random input variables and its mapping onto the output variables.

In practice, all partial sensitivity indices involving the single input variable X_i ($i = 1, 2, \dots, n$) are summed up to the total sensitivity index S_{T_i} in order to evaluate the total effect of X_i [9]. The total Sobol' indices consider thus interaction among the input variables.

Each Sobol' index S_{T_i} represents the sensitivity of the computational model regarding its input variable X_i without any assumption of linearity or monotonicity in the model. Two essential types of problems can be studied with the aid of the Sobol' indices:

1. Ranking of input variables in $\underline{Y} = f(X_1, X_2, \dots, X_n)$
2. Identification of insignificant input variables in $\underline{Y} = f(X_1, X_2, \dots, X_n)$.

The approach is to estimate the Sobol' indices $S_{T_1}, S_{T_2}, \dots, S_{T_n}$ and to order the variables according to these values.

In order to quantify which amount of each variance D^j is caused due to a single input variable X_i the corresponding Sobol' index S_{T_i} may be normalized according to Eq. (10).

$$\text{norm} S_{T_i} = \frac{S_{T_i}}{\sum_{k=1}^n S_{T_k}} \quad (10)$$

That is, input variable X_i has a share in variance D^j of $\text{norm} S_{T_i}$.

The Sobol' indices are practically computed using Monte Carlo simulation [10], this means that for computationally demanding models, e.g. finite element models in engineering mechanics, efficient meta models should be applied. Using Monte Carlo simulation, the total effect of an input variable X_i may be computed as

$$\hat{S}_{T_i} = 1 - \hat{S}_{\sim i} \quad \text{with} \quad \hat{S}_{\sim i} = \frac{\hat{D}_{\sim i}^j}{\hat{D}^j} \tag{11}$$

Considering the effects not containing X_i :

$$\hat{D}_{\sim i}^j = \frac{1}{N} \sum_{k=1}^N f(\underline{X}_{\sim ik}^{(1)}, X_{ik}^{(1)}) f(\underline{X}_{\sim ik}^{(1)}, X_{ik}^{(2)}) - \hat{f}_0^2 \tag{12}$$

The superscripts (1) and (2) indicate that two different samples are generated and mixed. $\underline{X}_{\sim ik}^{(1)}$ denotes the k th sample point with

$$\underline{X}_{\sim ik}^{(1)} = (X_{1k}^{(1)}, \dots, X_{(i-1)k}^{(1)}, X_{(i+1)k}^{(1)}, \dots, X_{nk}^{(1)})$$

For further explanation see [4].

Analytical Examples

Two analytical functions

$$Y_1 = f(X_1, X_2) = 10X_1 + 0,5X_2^3 \tag{13}$$

$$Y_2 = f(X_1, X_2) = 2X_1 - X_2^2 \tag{14}$$

are investigated. Function Eq. (13) is non-linear monotonic, Eq. (14) possess a non-linear non-monotonic behavior.

The input variables are uniformly distributed in the interval $[-5, 5]$. The function graphs are shown in Figure Figure 5 and Figure 6.

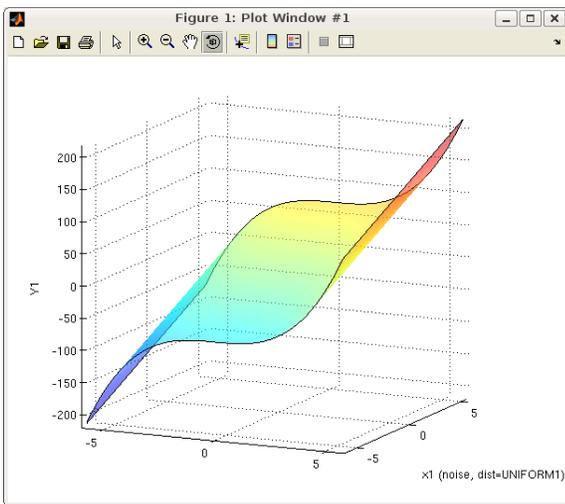


Figure 6 Function plot of Y1

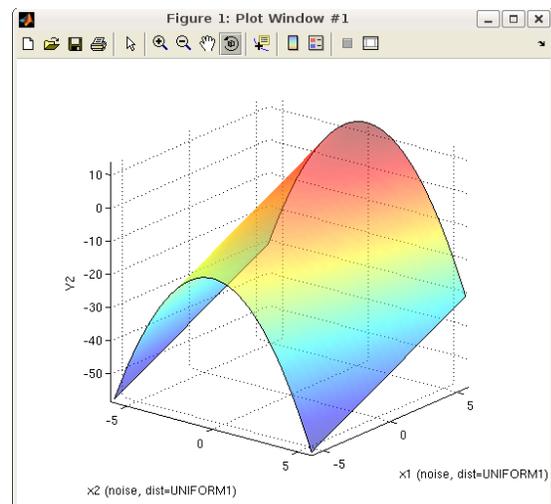


Figure 5 Function plot of Y2

The histograms of the stochastic results Y_1 and Y_2 are visualized with the aid of D-SPEX in Figure 7 and Figure 8.

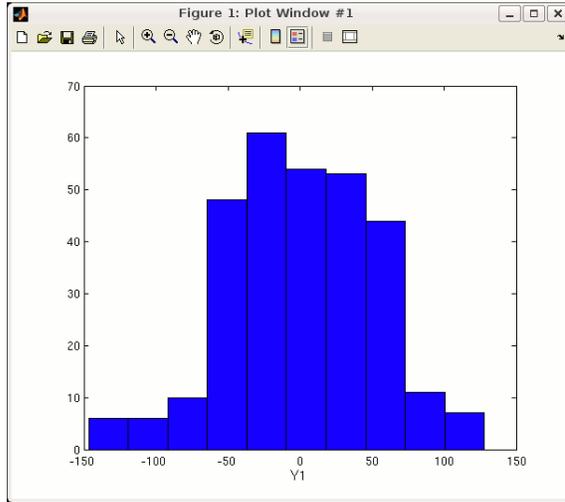


Figure 7 Histogram of Y1

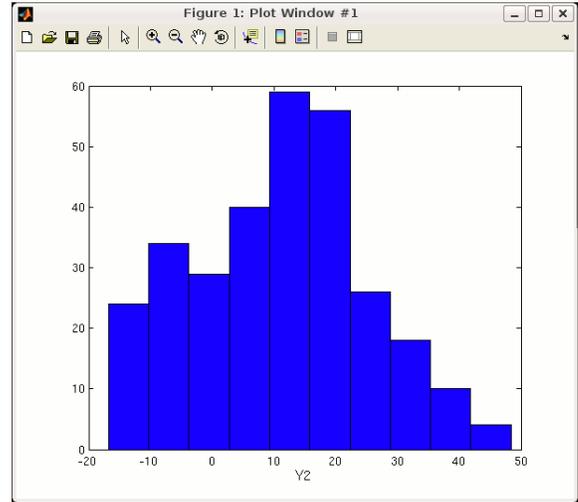


Figure 8 Histogram of Y2

In this preliminary sensitivity analysis relevant input variables should be identified and ranked for each of the functions Eqs. (13) and (14). The non-relevant variables that have none or only a small effect should be removed to reduce the model size for subsequent investigations (i.e. an optimization).

The regression-based correlation coefficient r , the variance-based linear ANOVA, and Sobol' indices S_T are evaluated as sensitivity measures and compared.

It is obvious from the function plots Figure 5 and Figure 6 that there is no non-relevant input variable. Both variables contribute to the variance of the results Y_1 and Y_2 . The ranking of the variables can also already be gained from the plots. For Y_1 the input variable X_1 and for Y_2 the input variable X_2 contribute most the respective variances of the results Y_1 , Y_2 . The results of a numerical sensitivity analysis should support this visual interpretation. But in contrast to the Sobol' indices the correlation does not, see. Table 1.

Table 1 Results of the sensitivity analysis

	Y_1			Y_2		
	r_{X_i, Y_1}	lin. ANOVA ¹	S_{T_i}	r_{X_i, Y_1}	lin. ANOVA ²	S_{T_i}
X_1	0.77	11.60	0.60	0.60	24.50	0.37
X_2	0.58	11.40	0.40	0.00	0.00	0.63

¹ Terms in expansion of $Y_1 (\cdot 10^4)$

² Terms in expansion of Y_2

Function Y_1

Both sensitivity measures r_{X_i, Y_1} and S_{T_i} will lead to the same ranking of the input variables X_1 and X_2 . Also the ratio of importance between X_1 and X_2 is almost comparable (0.77/0.58 vs. 0.60/0.40). However, the linear ANOVA might be misleading.

The numerical value of the Sobol' indices $S_{T_1} = 0.6$ and $S_{T_2} = 0.4$ can also be expressed in terms of variance or standard deviation. The input variable X_1 is responsible for 60% and X_2 for 40% of the variance/standard deviation of the result variable Y_1 . Whereas the correlation coefficient measures the linear relation between X_1 and Y_1 . Here also X_1 has a strong (linear) effect onto the result Y_1 . As proposed in the introduction for non-linear monotone computational models correlation coefficient and Sobol' indices lead to the same conclusion. However, the Sobol' indices provide a clear idea about the absolute and relative partial effect of each input variable.

Function Y_2

Here the effect of the quadratic term X_2^2 is neglected by the correlation coefficient and the linear ANOVA. That is obvious not the case. This term contributes significantly to the result Y_2 . The Sobol' index $S_{T_2} = 0.63$ indicates that. The input variable X_2 is responsible for the major part of the total variance (in this case 63%).

Simple Crash Analysis

A front crash of a simple car model against a pole is investigated. For the sensitivity analysis one loadcase is considered where the car hits the pole on the right hand side. The design parameters are t_{bumper} , t_{hood} that is the thickness of the bumper and the hood.

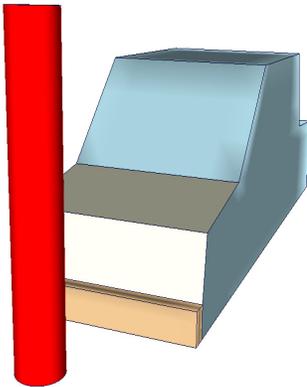


Figure 9 Simple car model

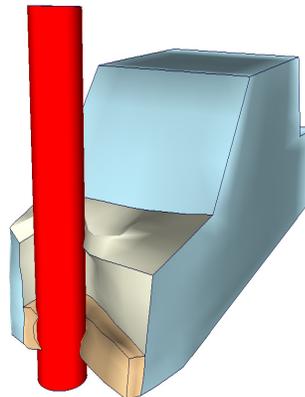


Figure 10 Car hits the pole on the right hand side

The considered responses are

- the minimal x displacement of node 167 (Intru1) and 432 (Intru2)
- the HIC value at node 432 and
- the absolute intrusion (Intrusion absolute) that is calculated from the difference of Intru1 and Intru2

as highlighted in the figures below.

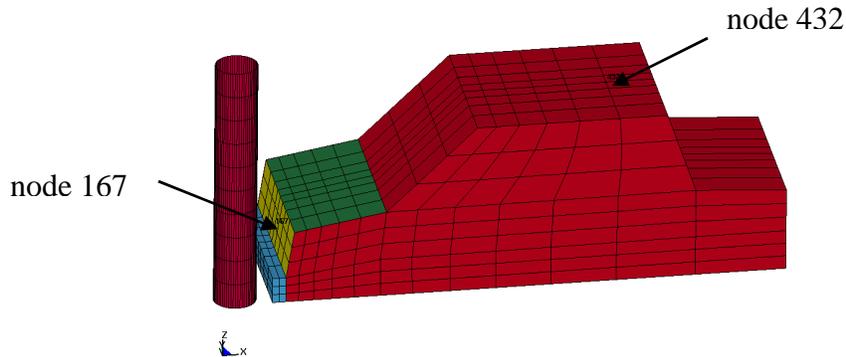


Figure 11 Evaluated nodes of the finite element model

There is no information about the probability distribution functions (pdf's) of the design parameters. In that case as straightforward approach a uniform probability distribution of the parameters is assumed. It has to be kept in mind that those possible existing probability distributions that normally differ from the uniform pdf could influence the results of the sensitivity analysis significantly.

The correlation coefficients together with histograms and anthill plots can be visualized with D-SPEX in one single plot window see Figure 12.

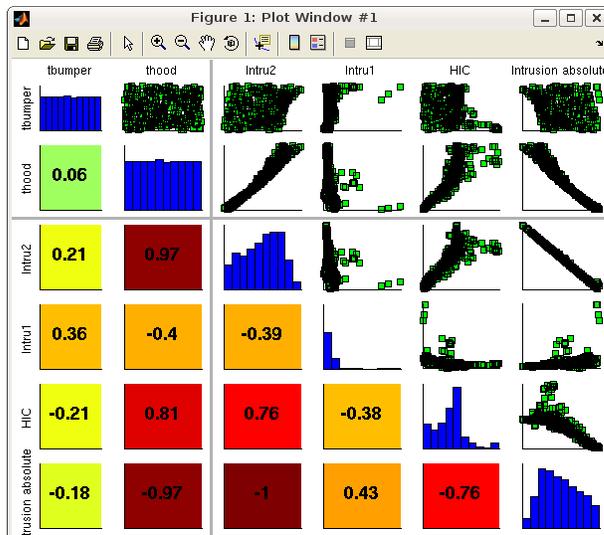


Figure 12 Correlation Coefficients, Histograms, and function plots

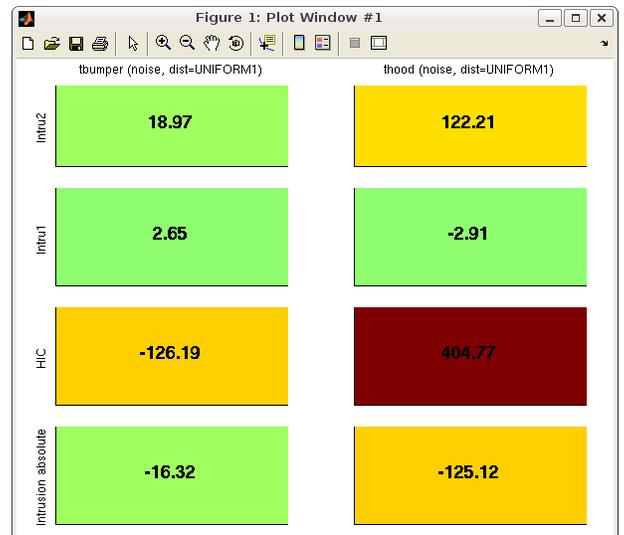


Figure 13 Linear ANOVA Results

In all cases the correlation coefficient indicates that the parameter `thood` has a strong effect onto the responses (except for `Intru1` where the effect of the parameters is similar). This also corresponds to the results of the linear ANOVA (Figure 13) and to the Stochastic Contributions/Sobol' indices (Figure 14).

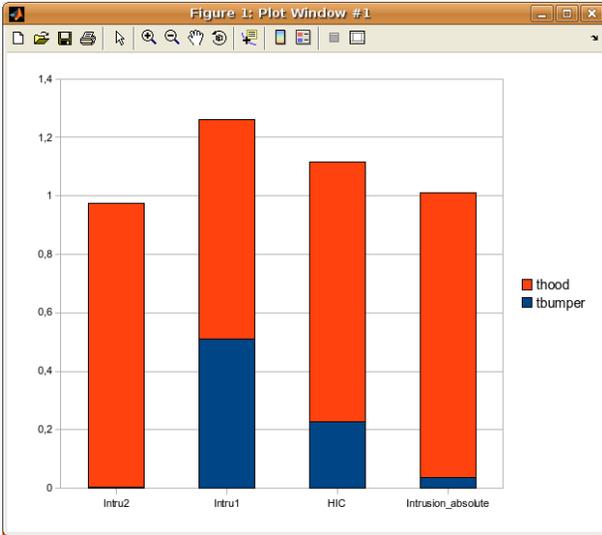


Figure 14 Plot of Sobol' indices S_{T_i} (Stochastic Contributions)

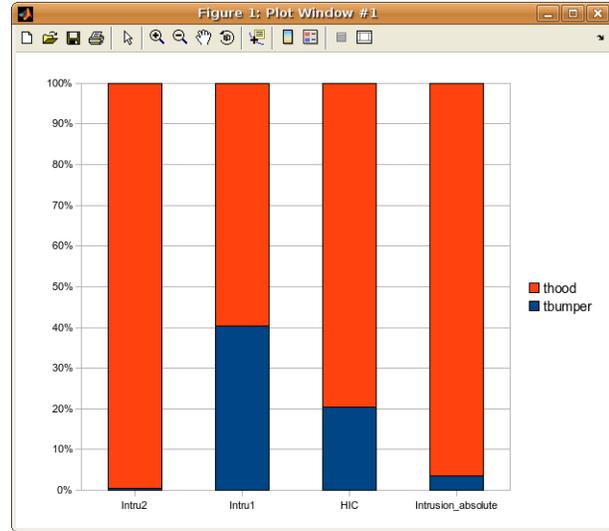


Figure 15 Plot of normalized Sobol' indices $norm S_{T_i}$ (Stochastic Contributions)

In the Sobol' indices/Stochastic Contributions plots the length of each section in one stacked bar represents the contribution of the specific parameter (`tbumper`, `thood`) to the total variance of the respective result. Figure 15 provides a normalized view on these results. For example, the contribution to the response `HIC` of the variable `tbumper` is 20% and of the variable `thood` it is 80%.

Conclusions

LS-OPT and D-SPEX provide the capability of evaluation and visualization of stochastic analysis results. Advanced methodologies are implemented which can be applied by direct FEA-simulations or, what is usually more cost effective, by utilizing Meta-Model based Monte Carlo Analysis. By opening the command file of LS-OPT, D-SPEX reads all data of the optimization or robustness project. The current version of D-SPEX is fully compatible with LS-OPT version 3.3.

In D-SPEX analysis results can be displayed by ant-hill plot matrices, where the user can select/highlight point groups and get direct access to the results of all simulations associated to the points of this group by any post-processor. Correlation matrices and ANOVA bars can be displayed as well, to get a clearly arranged view on the relationships between input and output variables. In addition, plotting of Sobol' indices are possible, which provides an advanced correlation measure for non-linear responses. An additional new feature is plotting of proposed probability and cumulative distribution functions (PDF/CDF) associated to empirical histogram data. On this the Kolmogorov-Smirnov test is utilized.

References

- [1] Joint Research Centre (JRC) of the European Commission: <http://sensitivity-analysis.jrc.ec.europa.eu/>
- [2] Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D. Saisana, M., and Tarantola: *Global Sensitivity Analysis. The Primer*, John Wiley & Sons, 2007.
- [3] Sudret B.: *Global sensitivity analysis using polynomial chaos expansions*. Reliab Eng Syst Safety (2007), doi:10.1016/j.ress.2007.04.002
- [4] Stander, N., Roux, W., Eggleston, T., Craig, K.: *LS-OPT User's Manual*, Version 3.2, Livermore Software Technology Corporation, Livermore, 2007
- [5] Sobol I.M.: *Sensitivity estimates for nonlinear mathematical models*. Math Modeling Comput Exp, 1:407–14, 1993.
- [6] Cukier H., Levine R.I., Shuler K.: *Nonlinear sensitivity analysis of multiparameter model systems*. J Comput Phys 26:1–42, 1978.
- [7] Saltelli A., Tarantola S., Chan K.: *A quantitative, model independent method for global sensitivity analysis of model output*. Technometrics, 41(1):39–56, 1999.
- [8] Sobol I.M.: *Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates*. Mathematics and Computers in Simulation 55:271–280, 2001.
- [9] Homma T., Saltelli A.: *Importance measures in global sensitivity analysis of nonlinear models*. Reliab Eng Syst Safety 52:1–17, 1996.
- [10] Sudret B.: *Global sensitivity analysis using polynomial chaos expansions*. Reliab Eng Syst Safety 93(7):964-979, 2008.
- [11] Wikipedia: <http://en.wikipedia.org/wiki/Correlation>, 2008.
- [12] Witowski, K., Thiele, M., Seer G., Mühlhuber, W., van den Hove, M.: *New Features in D-SPEX with Applications*, Proceedings 6th LS-DYNA User Forum, Frankenthal, Germany, 2007

