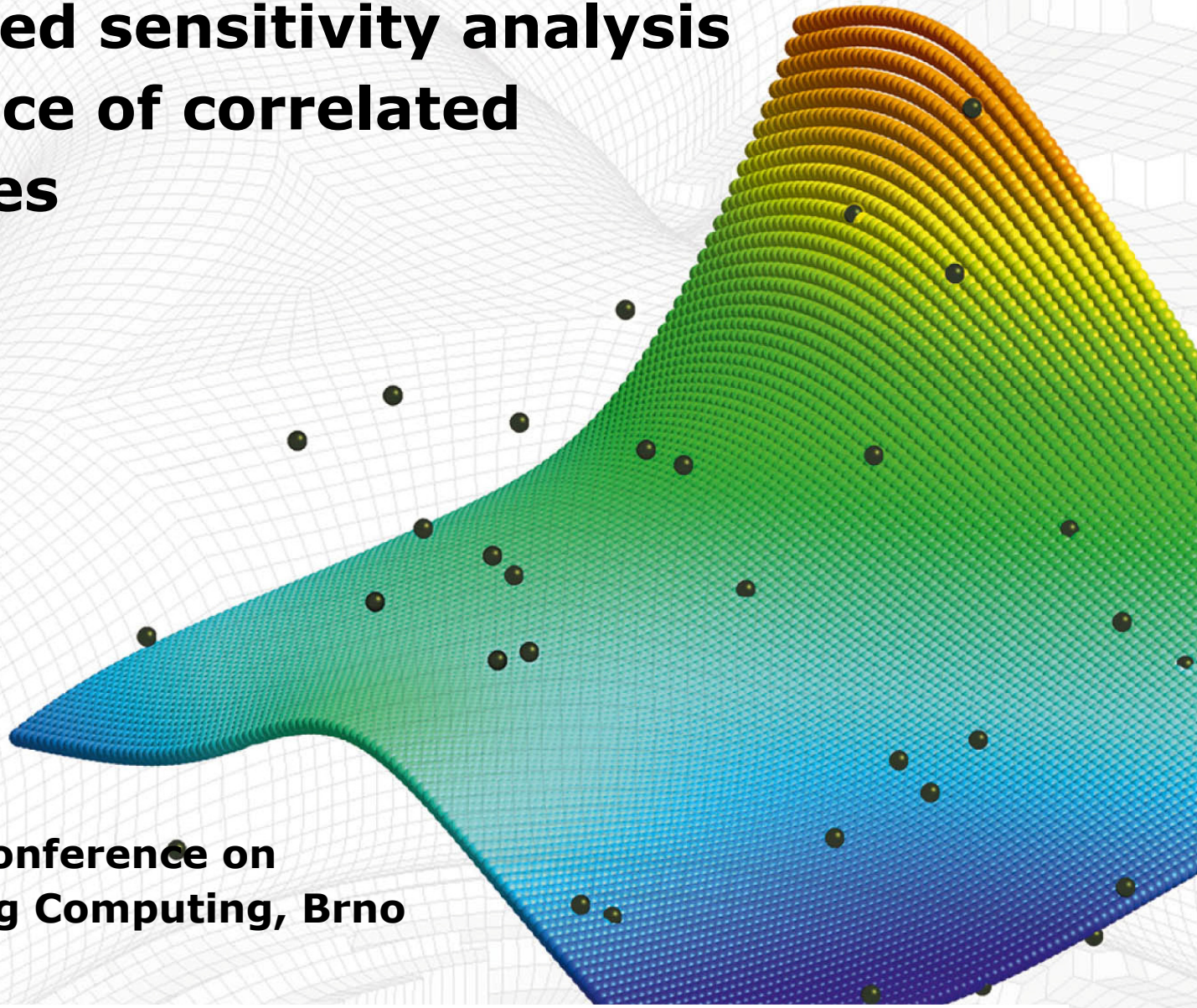
# Variance-based sensitivity analysis in the presence of correlated input variables

**Thomas Most**
**Dynardo GmbH**
**Weimar, Germany**



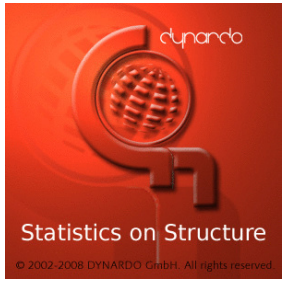
**5th International Conference on Reliable Engineering Computing, Brno**

# Introduction

# Dynardo

- Founded: 2001 (Will, Bucher, CADFEM International)
- More than 35 employees, offices at Weimar and Vienna
- Leading technology companies Daimler, Bosch, Eon, Nokia, Siemens, BMW are supported

## Software Development

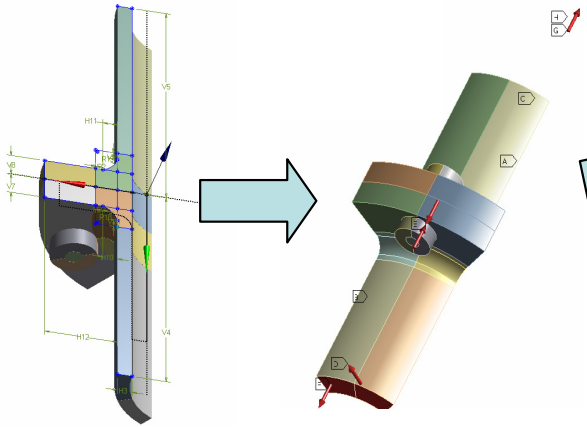




Dynardo is engineering specialist for CAE-based sensitivity analysis, optimization, robustness evaluation and robust design optimization

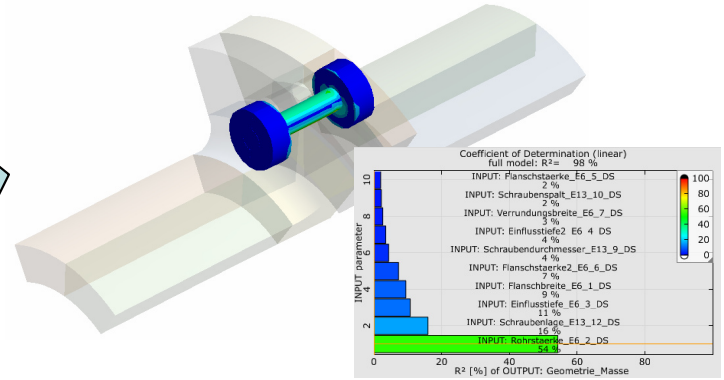
## CAE-Consulting

- Mechanical engineering
- Civil engineering & Geomechanics
- Automotive industry
- Consumer goods industry
- Power generation

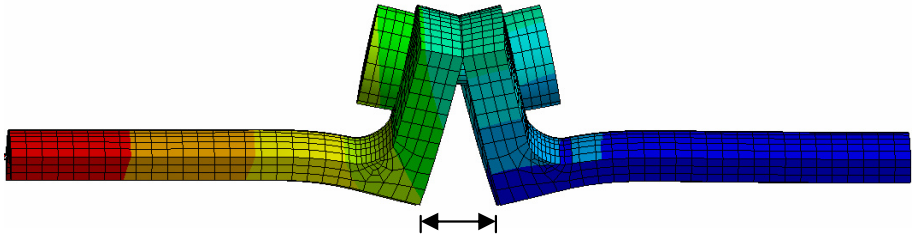
# Multidisciplinary Optimization



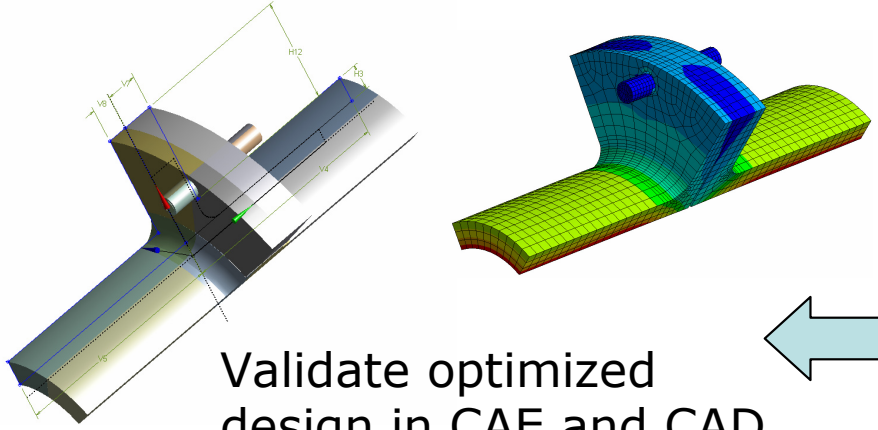
CAD and CAE Parameter definition




Sensitivity study – identify the most important parameters and check variation of response values



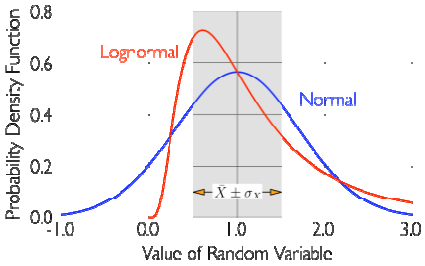

Define optimization goal and optimize

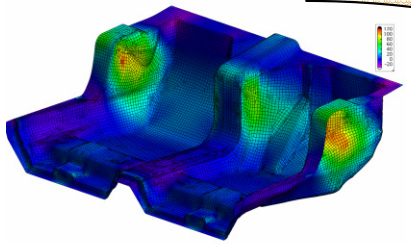


Validate optimized design in CAE and CAD

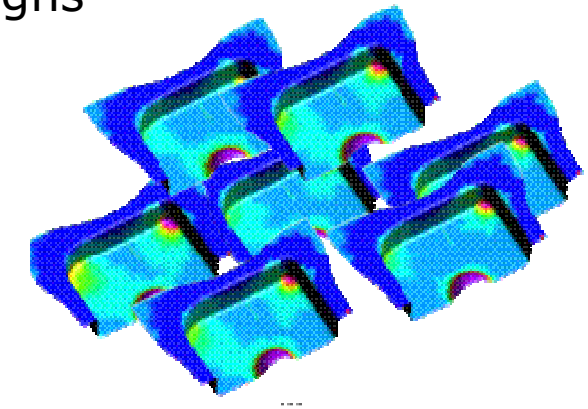
# Variance based robustness analysis

1) Define the robustness space using scatter range, distribution and correlation

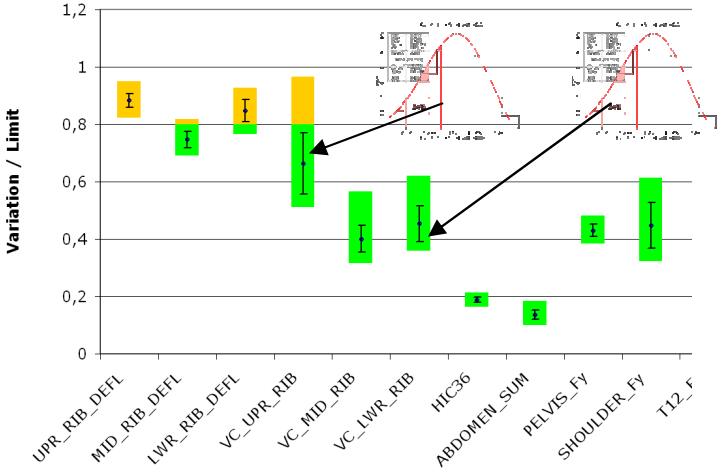


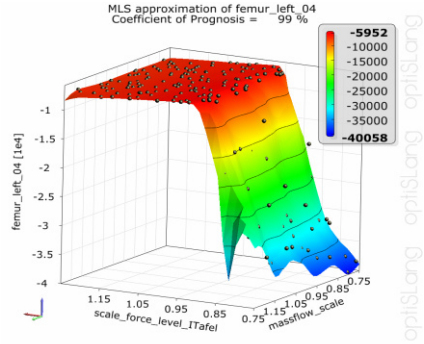
2) Scan the robustness space by producing and evaluating *n* designs

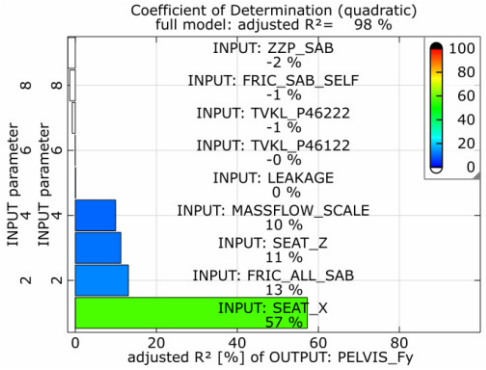


3) Check the variation



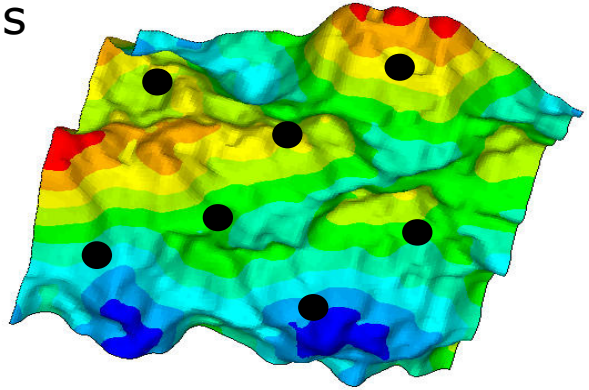

5) Identify the most important scattering variables



# Sensitivity analysis

- Sensitivity analysis scans the design/random space and measures the sensitivity of the inputs with statistical measures



- Application as pre-investigation of an optimization procedure or as part of an uncertainty analysis

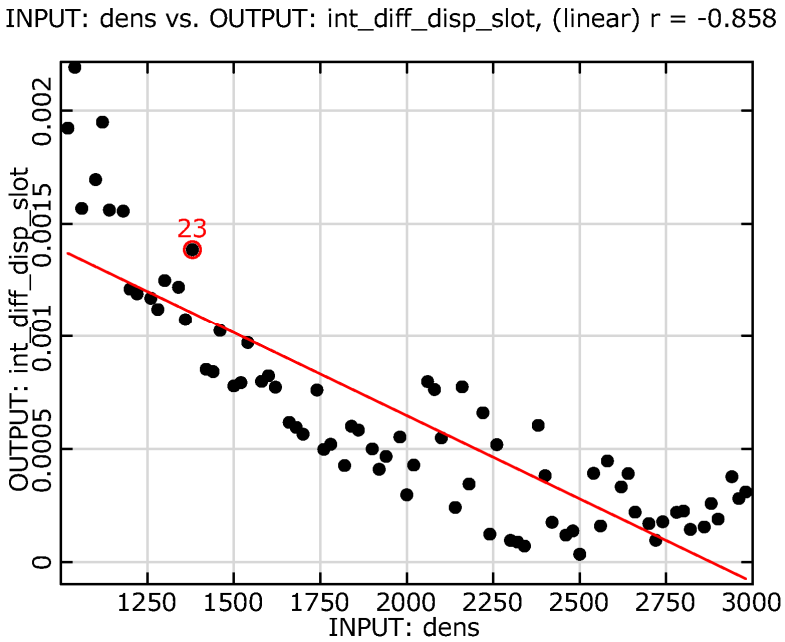
- Results of a global sensitivity study are:
  - **Sensitivities** of inputs with respect to important responses
  - **Estimate** the variation of responses
  - **Estimate** the noise of an underlying numerical model
  - **Better understanding** and verification of dependences between input and response variation

- Requirements for industrial applications:
  - Treatment of a large number of inputs
  - Consideration of strongly nonlinear dependencies
  - Manageable numerical/experimental effort

# Methods for sensitivity analysis

**Local methods**

- Local derivatives
- Standardized derivatives



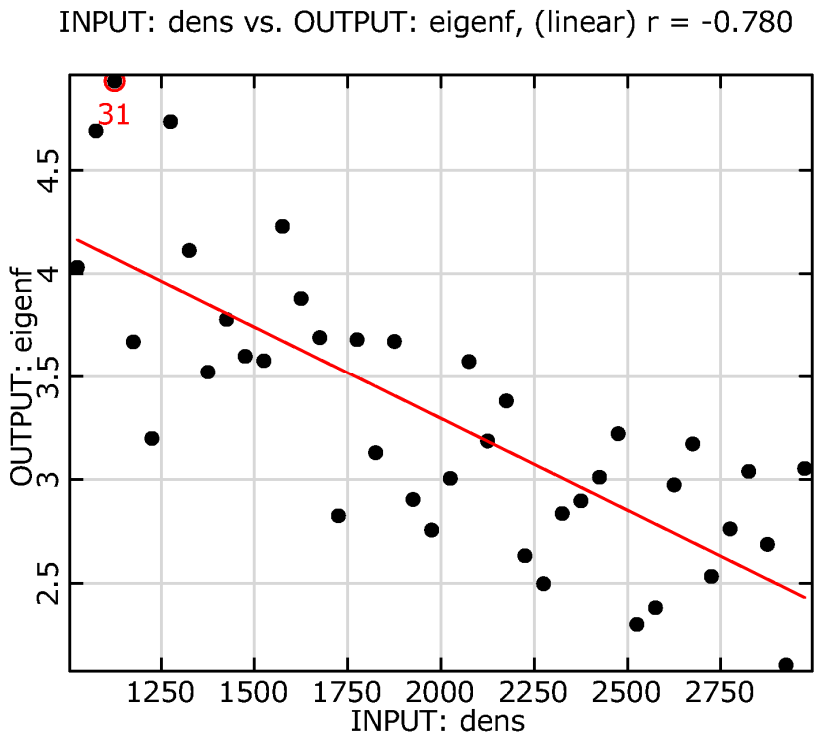

**Global methods**

- Anthill plots
- Coefficients of correlation
- Rank order correlation

- Standardized regression coefficients
- Stepwise polynomial regression
- Varianced-based analysis: Sobol' indices

- **Advanced surrogate models including prediction analysis and optimal subspace detection: Metamodel of Optimal Prognosis**

# Coefficient of correlation




$$\rho(X,Y) = \frac{COV(X,Y)}{\sigma_X \sigma_Y}$$

$$\approx \frac{1}{N-1} \frac{\sum_{i=1}^{N}(x_i - \hat{\mu}_X)(y_i - \hat{\mu}_Y)}{\hat{\sigma}_X \hat{\sigma}_Y}$$

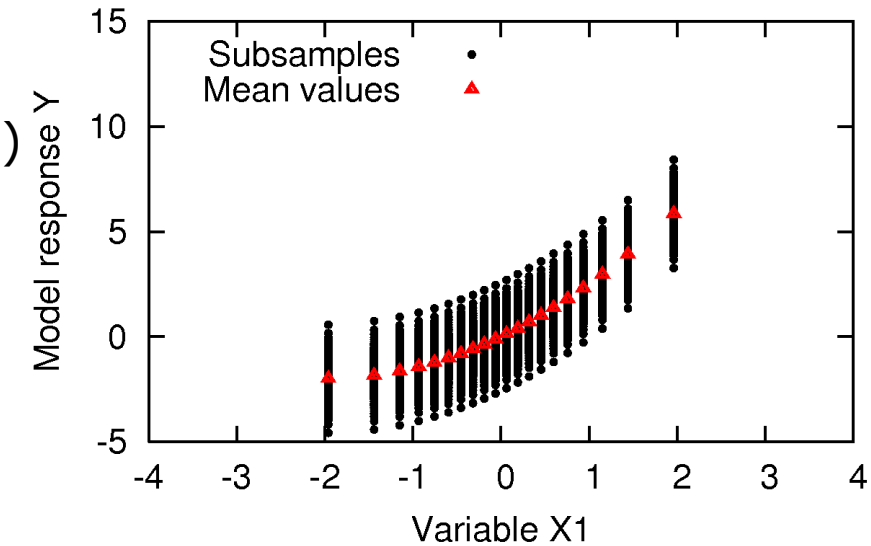
- Defined as standardized covariance of two variables
- Coefficient of correlation is always between -1 and 1
- Defines degree of linear dependence
- Only linear (quadratic, monotonic) dependencies without interactions

# Variance-based sensitivity analysis

# Variance-based sensitivity analysis

- Apportion of the variation of the model output to the sources of the variation of the model input (Saltelli 2008)
- First order sensitivity index (Sobol' 1993) describing the decoupled influence of a single variable $X_i$ on the model output $Y$

$$S(X_i) = \frac{V_{X_i}(E_{\mathbf{X}_{\sim i}}(Y|X_i))}{V(Y)}$$




- Total sensitivity index describing the decoupled and coupled (total) influence of a single variable (Homma & Saltelli 1996)

$$S_T(X_i) = 1 - \frac{V_{\mathbf{X}_{\sim i}}(E_{X_i}(Y|\mathbf{X}_{\sim i}))}{V(Y)}, \quad \sum S_T(X_i) \geq 1$$

# Variance-based sensitivity analysis

- Variance based sensitivity analysis quantifies only the input sensitivities around their mean
- Existing estimators require a large number of model evaluations for complex systems and are often not applicable
- ANOVA (Analysis of Variances) Decomposition is unique only for independent inputs

$$V(Y) = \sum_{i=1}^{k} V(f_i(X_i)) + \sum_{i=1}^{k} \sum_{j>i}^{k} V(f_{ij}(X_i, X_j)) + \ldots$$
$$+ V(f_{1\ldots k}(X_1, \ldots, X_k))$$

➢ Sobol and other estimators use a recombination of matrices **A** and **B** in $\mathbf{C}_i$ and can be applied only for independent inputs

$$\hat{S}_i = \frac{\mathbf{y_A}^T \mathbf{y_{C_i}} - n(\bar{y}_\mathbf{A})^2}{\mathbf{y_A}^T \mathbf{y_A} - n(\bar{y}_\mathbf{A})^2}, \quad \hat{S}_{T_i} = 1 - \frac{\mathbf{y_B}^T \mathbf{y_{C_i}} - n(\bar{y}_\mathbf{B})^2}{\mathbf{y_B}^T \mathbf{y_B} - n(\bar{y}_\mathbf{B})^2}$$

# Estimator for correlated inputs

- Decomposition of each input variable in a correlated and an uncorrelated part

$$X_i = f(X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_m) + X_i^{U,\mathbf{X}_{\sim i}}$$
$$= X_i^{C,\mathbf{X}_{\sim i}} + X_i^{U,\mathbf{X}_{\sim i}}$$

- In case of linearly correlated variables (e.g. Nataf model), the correlated part can be represented as linear combination

$$X_i^{C,\mathbf{X}_{\sim i}} = \beta_{X_i,0} + \sum \beta_{X_i,j} X_j$$

- Matrix assembling by considering the correlated and uncorrelated part

$$\tilde{\boldsymbol{C}}_i^C = \boldsymbol{B}^{U,X_i} + \boldsymbol{A}^{C,X_i}, \quad \tilde{\boldsymbol{C}}_i^U = \boldsymbol{A}^{U,\mathbf{X}_{\sim i}} + \boldsymbol{B}^{C,\mathbf{X}_{\sim i}}$$

➢ First order and total effect indices are obtained for the correlated and the uncorrelated part of each input variable

# Example: additive linear model

- Three normally distributed inputs with zero mean:

$$Y = X_1 + X_2 + X_3, \quad \mathbf{C_{XX}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 2\rho \\ 0 & 2\rho & 4 \end{bmatrix}$$

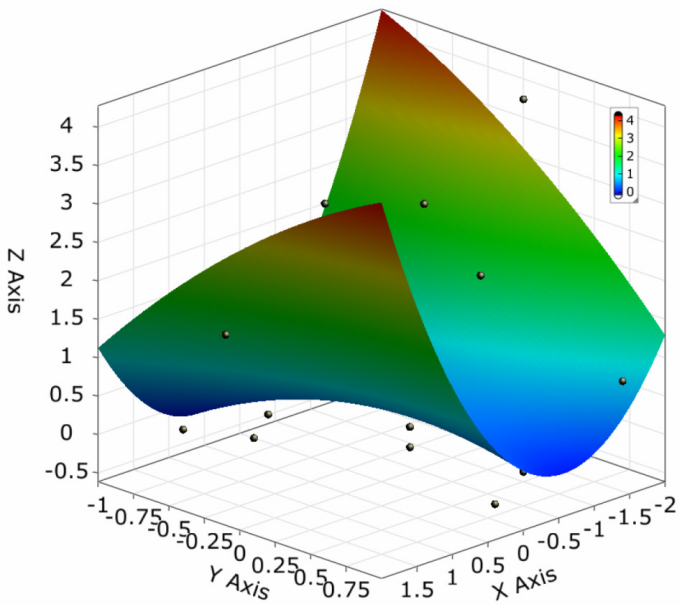
- First order and total effect sensitivity indices:

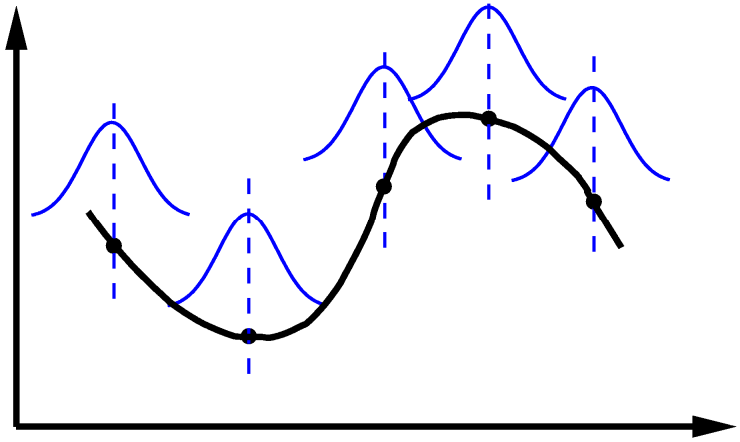
| | | $\rho$=0.0 | $\rho$=0.8 | $\rho$=1.0 | $\rho$=-0.5 | $\rho$=-0.8 |
|---|---|---|---|---|---|---|
| *Correlated* | $X_1$ | 0.167 | 0.109 | 0.100 | 0.250 | 0.357 |
| | $X_2$ | 0.167 | 0.735 | 0.900 | **0.000** | 0.129 |
| | $X_3$ | 0.667 | 0.852 | 0.900 | 0.563 | 0.514 |
| Uncorrelated | $X_1$ | 0.167 | 0.109 | 0.100 | 0.250 | 0.357 |
| | $X_2$ | 0.167 | 0.039 | **0.000** | 0.188 | 0.129 |
| | $X_3$ | 0.667 | 0.157 | **0.000** | 0.750 | 0.514 |

# Response Surface Method

# Response Surface Method

- Approximation of response variables as explicit function of input variables
- Approximation function can be used for sensitivity analysis and/or optimization
- Global methods (Polynomial regression, Neural Networks, …)
- Local methods (Spline interpolation, Moving Least Squares, Radial Basis Functions, Kriging, …)
- Approximation quality decreases with increasing input dimension
- Successful application requires objective measures of the prognosis quality



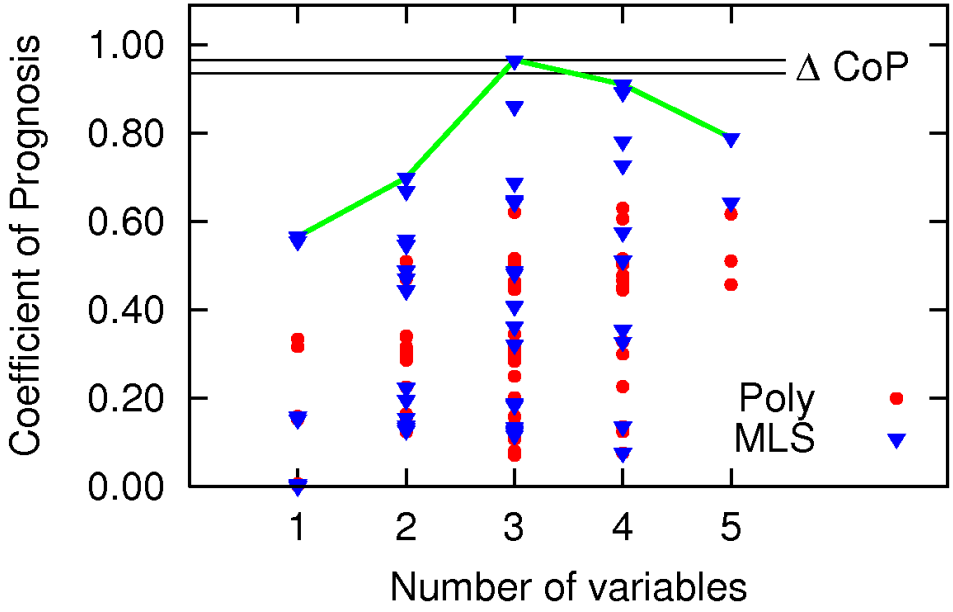



# Metamodel of Optimal Prognosis (MOP)

- Approximation of solver output by fast surrogate model
- Reduction of input space to get best compromise between available information (samples) and model representation (number of inputs)
- Advanced filter technology to obtain candidates of optimal subspace
- Determination of optimal approximation model (polynomials, MLS, …)
- Assessment of approximation quality (CoP)

**MOP solves 3 important tasks:**

- Best variable subspace
- Best meta-model
- Estimation of prediction quality

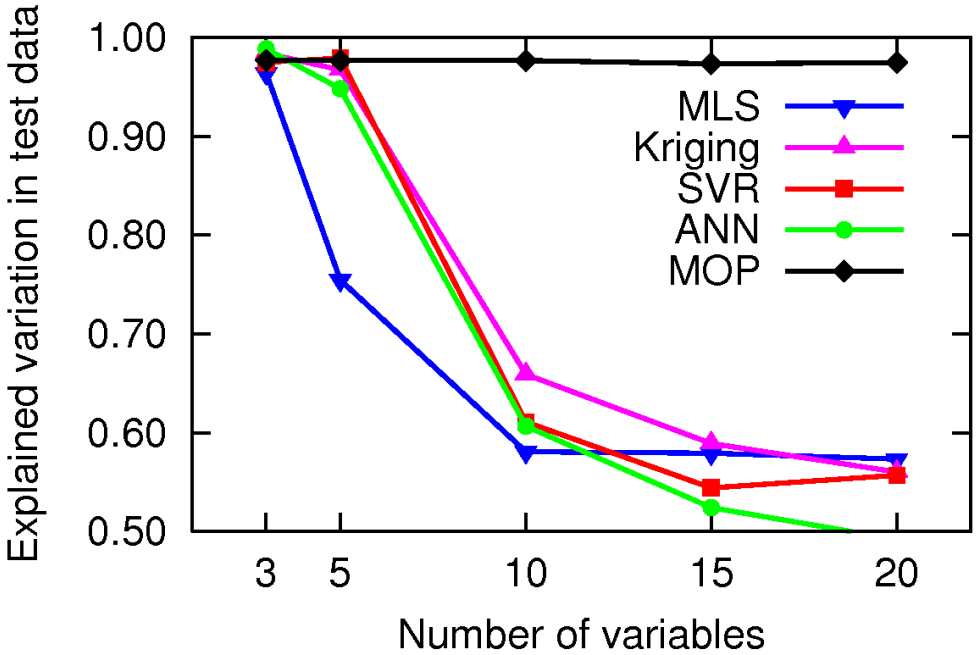


# Example: Analytical nonlinear function

## MOP vs. advanced meta-models

- 100 LHS support points
- Explained variation of independent test data set with 100 samples as error measure

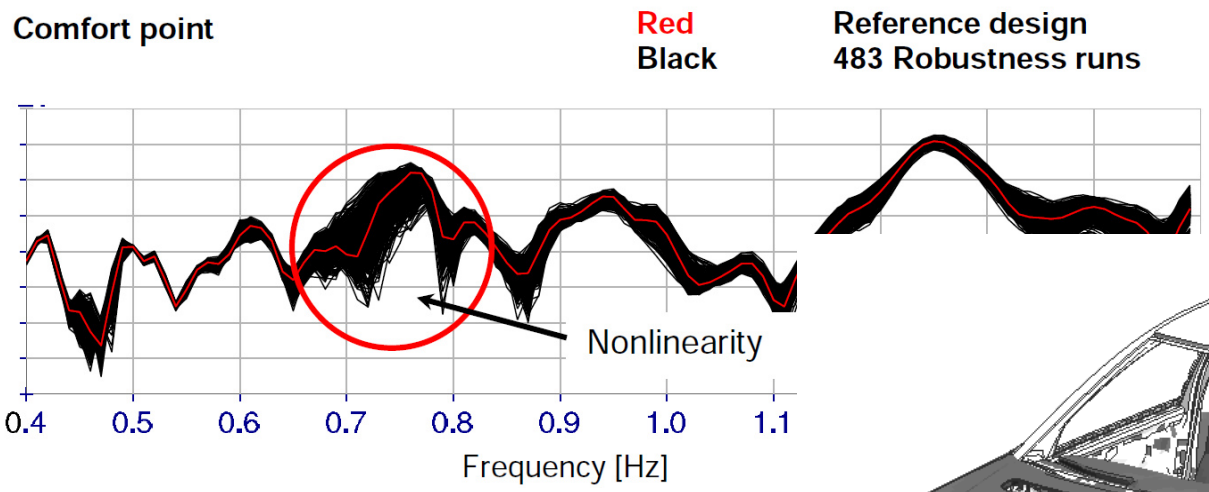
$$\epsilon_{Test} = 1 - \frac{SS_E^{Test}}{SS_T^{Test}}$$

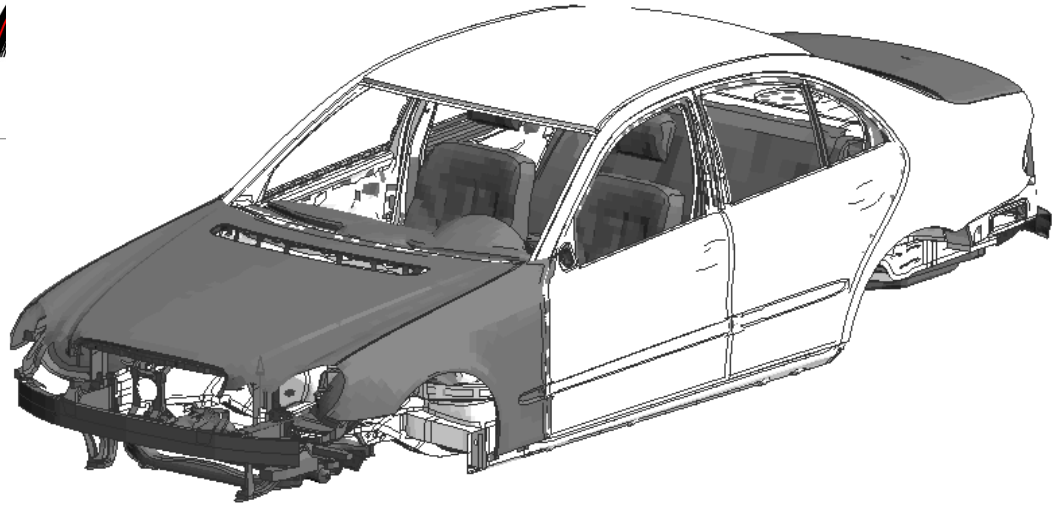



- With increasing number of unimportant variables the approximation error increases for MLS, Kriging, Support Vector Regression (SVR) and artificial neural networks (ANN)
- MOP detects 3 important variables and enables best approximation

# Application

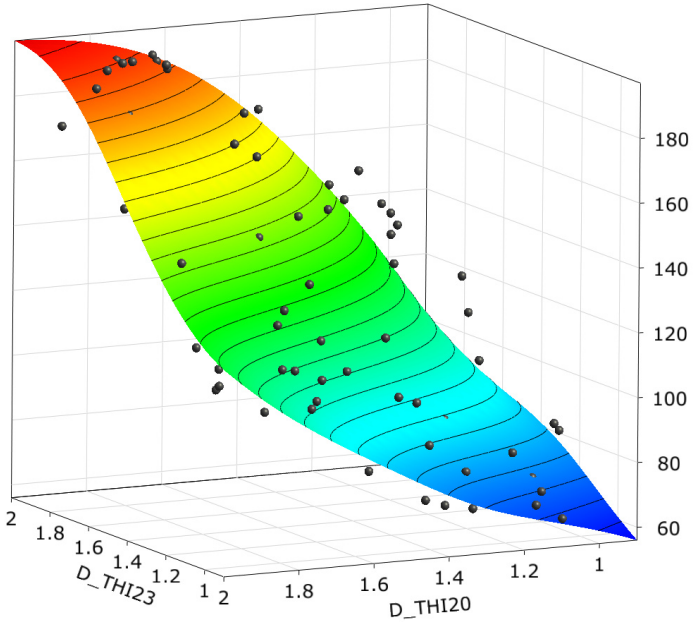
# Application: Noise Vibration Harshness

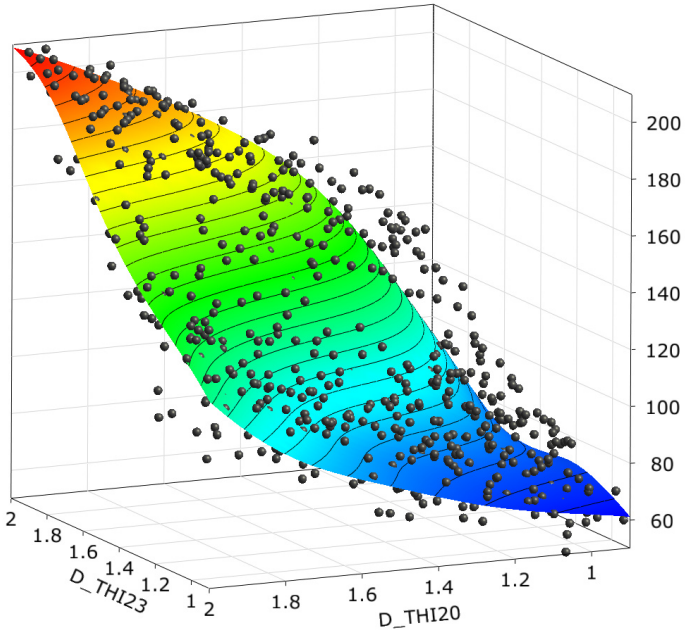


- Input parameters are 46 sheet thicknesses of a car body
- Variation of inputs within a +/- 20% interval
- Output values are sound pressure levels at certain frequencies
- Already single solver run is very time consuming

# Application: Noise Vibration Harshness




| Samples | 100 | 200 | 400 | 600 | 800 |
|---|---|---|---|---|---|
| Full model | 90.9% | 91.7% | 95.7% | 96.3% | 96.9% |
| D_THI5 | - | - | 2.4% | 2.3% | 2.7% |
| D_THI6 | 6.0% | 5.3% | 8.2% | 8.3% | 8.7% |
| D_THI20 | 41.3% | 42.7% | 42.3% | 43.4% | 42.2% |
| D_THI23 | 49.1% | 48.0% | 50.7% | 51.0% | 53.8% |

# Summary

- Sensitivity analysis gives advanced information about design parameters and helps to simplify the optimization problem
- Variance-based sensitivity indices can quantify nonlinear dependencies and variable interactions
- Presented decomposition of variables in correlated and uncorrelated parts can distinguish between the output variance sources
- Meta-modeling makes Sobol estimators applicable for complex systems
- MOP serves optimal meta-model in best subspace
- Coefficient of Prognosis give reliable estimate of approximation quality
- Small CoP indicates insufficient number of samples or unexplainable solver behavior/problems
- MOP-approach was proven to be successful in many applications in virtual product development