An intuitive variance based variable screening method (for multidisciplinary vehicle design exploration)

R. Sala, N. Baldanzini, M. Pierini

Author Affiliation

Dipartimento di Ingegneria Industriale
Università degli Studi di Firenze

Abstract

The goal of this poster is to highlight an simple and intuitive approach that could be used for variable screening. It is targeted at a limited sampling budged range where established Sensitivity Analysis (SA) methods don't perform well, and usually qualitative variable screening methods are applied. A simple explanation of the concept is given, and the performance is compared with several variable screening methods and a sensitivity index estimation method. Although the approach was implemented for the purpose of variable screening for vehicle design related problems, it could be applied to other problems with high dimensionality, nonlinear responses and expensive function evaluations.

Multidisciplinary design exploration and optimization of full vehicle structures

Within the Industrual vehicle design proces nummerical simmulation methods are commonly used. Where previously model understanding was an implicit part of the modeling process, the past and ongoing developments of numerical simulation techniques allows the creation models that are beyond modeler intuition. The challenge in the vehicle design process is to asses the responses of the simulation models for multiple disciplines, in order to find effective trade-off solutions for an efficient overall design. Variable screening methods can be a useful aid within this process.

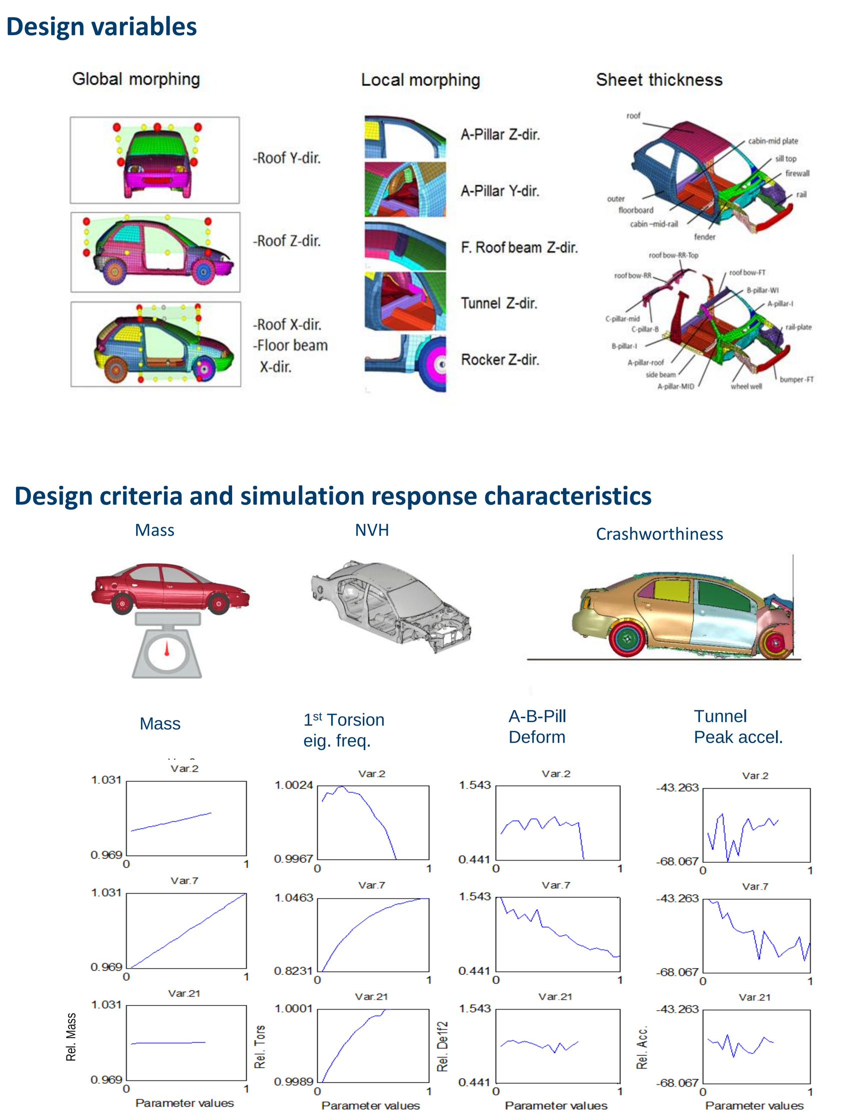
Typical problem

Minimize the mass of the vehicle while satisfieng the constraints on vibro-accoustic and crashworthiness criteria.

The challanges are:

- High dimensionality (The number of design variables ranges between 20 and 200).
- Computationaly expensive simulations leading to a tight sampling budget
- Nonlinear to highly nonlinear responses.

The approach presented in this poster is applied on the combination of Mass-, vibro-accoustic- and crashworthiness-respones, that are obtained from simulations using mockups of the Finite element models that are provided by the National Crash Analysis Center [NCAC].



Figures 1A and 1B, An overview of typical design variables, and response characteristics

The intuitive approach

The approach highlighted in this poster, is based on the partitioning of scatterplot projections in to several discrete intervals along the input variables X_i, and to use the means and variances of output Y of points within those intervals in order to estimate importance measures for variable screening. The importance measures can be interpreted as low resolution estimators of Sobol' indices [Sbl93] (also called global sensitivity indices [Sbl01]).

If in each of the scatter plot projections over the function variables the domain of each variable is divided in k divisions. Let M_{ik} be the matrix with means within the k devisions of the ith variable. The variable screening measure or Discrete Interval Sensitivity index (DISi) can be defined as:

$$DIS_i = VAR(M_{ik})/VAR(Y) \qquad (1)$$

The concept is taken from the definition of the Sobol indices [Sbl93, Sbl01]. Wich are defined as:

$$S_i = V_i/VAR(Y) \qquad (2)$$

Where Vi is the variance contribution of variable i defined as:

$$V_i = VAR_{X_i}(E_{X\sim i}(Y|X_i)) \qquad (3)$$

Graphical representation

A graphical representation of the approach can be seen in the diagrams below, where for a relativly small number of samples the estimates for the mean and variances are plotted, based on a course partitioning grid.

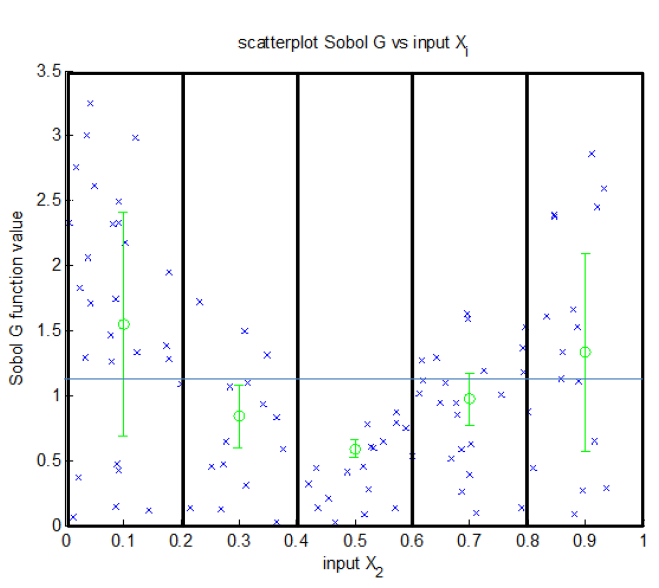


Figure 2 Scatter plot and partition estimates of means and variances for the sobol g function.

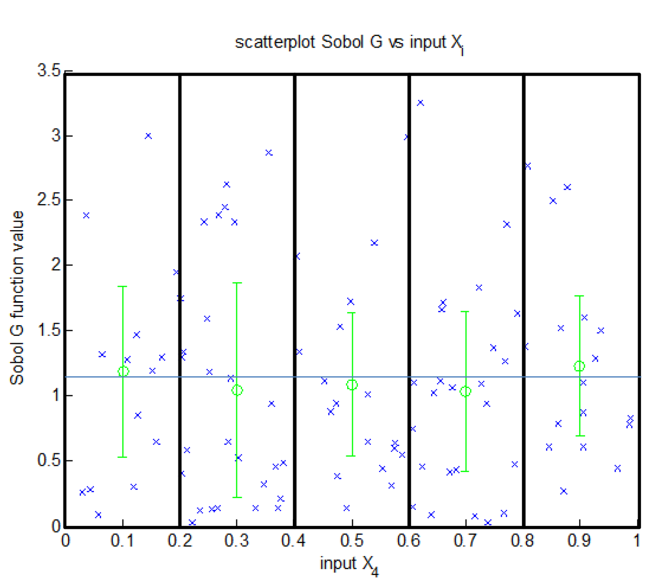


Figure 3 Scatter plot and partition estimates of means and variances for the sobol g function.

In this approach the variables with a higher variance of the means are estimated to have a higher first order effect. Using recursive partitioning also interactions and thus higher order indices can be estimated analogously. The estimation of higherorder interactions requires however more samples such that in practice the applicibility is limited to second order interactions.

The concept of the partitioning approaches, is in literature sometimes referred to as: "a gridding based approach"[Hlt06], and is very similar to the concepts found in the analysis of variance (ANOVA) and "partition of sums of squares approaches" such as found in [Sch59]. In those settings the partitioning procedure is however used to obtain different type of importance measures such as the F-statistic, or p-values. and are more focused to analyze sources of variation within particular variables as opposed to among different variables. In the literature study preceding the investigations of multidisciplinary design exploration and optimization, several review papers regarding variable screening and sensitivity analysis were consulted [Hft89, Bhl12, Hlt06, Slt10]. It was surprising that the application of the gridding based approach for Sobol' index estimation method was not mentioned. In the extensive review of [Hlt06] the concepts of both gridding based methods and variance decomposition based sensitivity indices were described in relative detail, but as separate approaches, and the obvious step to combine them was absent, and seems absent or at least rear in the literature in general. A possible explanation for this could be that: when a larger number of samples can be used more efficient estimation methods are available. However for small sampling budgets, this simple approach seems to provide useful results compared to other screening and SA methods, and therefore the by this approach obtained Discrete interval sensitivity indices might be suitable as an importance measure for variable screening.

Comparisons

In figure 4 the approach is compared with other variable screening methods that are recomended in the literature [Slt10,Kch09,Sbl09,Cmp07,Mor91].

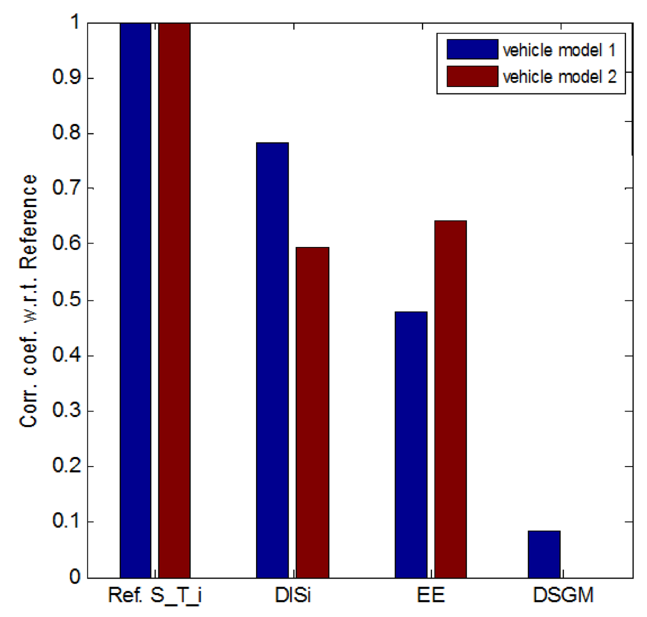


Figure 4 Comparison of variable screening methods w.r.t. Total sensitivity index for a small number of samples (about 4times the dimension) , the dimension (d) of the test problems were 32 and 72 respectively. The functions investigated are the peak accelration responses of 2 vehicle models. The total sensitivity is estimated on a surrogate model based on 2000 samples.

Compared are:

- Elementary Effect/Modified Morris (EE) by [cmp07,Mor91]

$$\mu_i = \int_{H^n} \left|\frac{\partial f}{\partial x_i}\right| dx \qquad (4)$$

- Derivative-based Global Sensitivity Measures (DGSM) [Kch09, Sbl09]

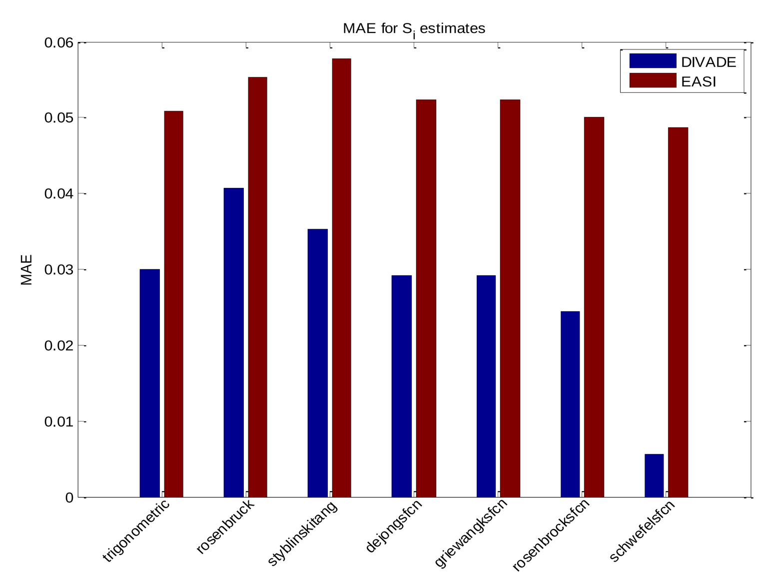
$$\nu_i = \int_{H^n} \left(\frac{\partial f}{\partial x_i}\right)^2 dx \qquad (5)$$

The screening measures or sensitivity estimators are related to eachother or the Total sensitivity by:

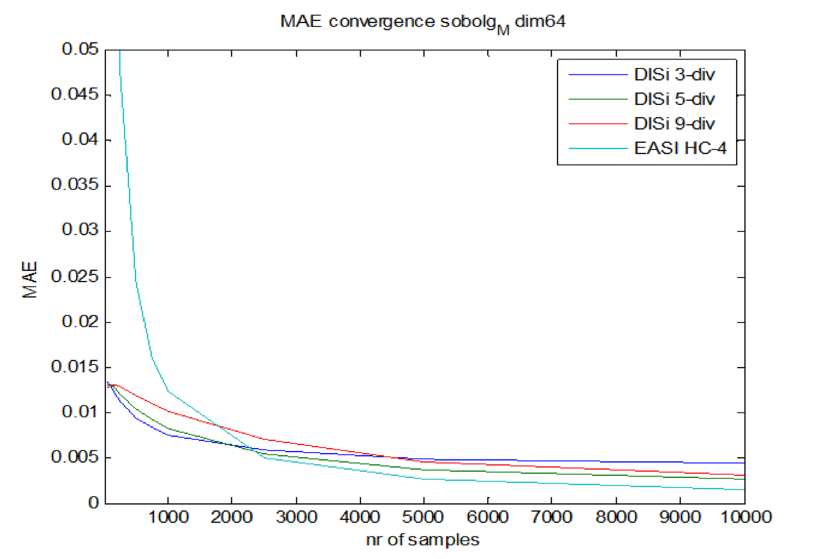
$$S_i^{tot} \leq \frac{\nu_i}{\pi^2 VAR(Y)} \quad (6), \qquad \mu_i \leq \sqrt{\nu_i} \quad (7) \text{ and}$$

$$DIS_i \leq S_i \leq S_i^{tot} \qquad (8).$$

For the example investigated the DISi indicator seems to perform similar as the EE indicator. An advantage of the DISi indicator is that no special sampling plan is required, and that the samples can be «recycled» as starting points in a meta-heuristic optimization. In figure 6 the approach is compared to the «EASI» sobol index estimation method [Pls10], for an instance of the sobol g function. Figure 5 shows the results of a comparison on several different test functions for a fixed number of samples. In terms of Mean Absolute Error (MAE) w.r.t. a reference solution.



For a collection of test functions that are scalable in dimension . The Mean Absolute Error (MAE) between sensitivity index estimates based on 200 samples and a reference estimate based on 10^7 sampels is compared (d=16) .



A comparison of the estimates obtained by the partitioning aproach (DISi) and the EASI method is made. The MAE w.r.t. a reference estimate based on 10^7 samples are plotted against the number of samples used in the estimates. The coefficients of the sobol g function instance are chosen such that they approximatly follow a normal distribution with an standard deviation set as of ¼ of the function dimensionality (d=64).

The diagram in figure 6 shows that the Sobol index estimation by the EASI method is more accurate when more samples are available. The gridding based approach is however a more accurate estimator for smaller sample sets.

Summary

A simple approach was highlighted that combined the concepts of gridding based methods and variance decomposition in the setting of variable screening. Although for problems with large sampling budgets, more effective Sobol' index estimation methods exist, the estimation resulting from this approach provides useful results for problems with expensive small sample budgets, which is relevant for tasks related to multi disciplinary design exploration for vehicle models, or other tasks involving expensive simulators or experiments.

Acknowledgements

This work is performed in the scope of the GRESIMO project, and funded as an Marie Curie ITN fellowship within the 7th European Community Framework Program (Grand agreement 290050).

References

[Bhl12] Bühlmann, Peter, and Jacopo Mandozzi. "High-dimensional variable screening and bias in subsequent inference, with an empirical comparison." *Computational Statistics* (2012): 1-24
[Cot79] Cotter, Sarah C. "A screening design for factorial experiments with interactions." *Biometrika* 66.2 (1979): 317-320.
[Cmp07] Campolongo, Francesca, Jessica Cariboni, and Andrea Saltelli. "An effective screening design for sensitivity analysis of large models." Environmental modelling & software 22.10 (2007): 1509-1518.
[Hft89] Haftka, Raphael T., and Howard M. Adelman. "Recent developments in structural sensitivity analysis." *Structural optimization* 1.3 (1989): 137-151.
[Hlt06] Helton, Jon C., et al. "Survey of sampling-based methods for uncertainty and sensitivity analysis." *Reliability Engineering & System Safety* 91.10 (2006): 1175-1209.
[Mor91] Morris, Max D. "Factorial sampling plans for preliminary computational experiments." Technometrics 33.2 (1991): 161-174.
[Kch09] Kucherenko, S., et al. "Monte Carlo evaluation of derivative-based global sensitivity measures." *Reliability Engineering & System Safety* 94.7 (2009): 1135-1148.
[NCAC] National Crash Analysis Center, Finite element model Archive http://www.ncac.gwu.edu/vml/models.html
[Pls10] Plischke, Elmar. "An effective algorithm for computing global sensitivity indices (EASI)." *Reliability Engineering & System Safety* 95.4 (2010): 354-360.
[Slt10] Saltelli, Andrea, and Paola Annoni. "How to avoid a perfunctory sensitivity analysis." *Environmental Modelling & Software* 25.12 (2010): 1508-1517.
[Sbl93] Sobol' , I. Sensitivity analysis for non-linear mathematical models. *Mathematical Modeling & Computational Experiment (Engl. Transl.)*, 1993, **1**, 407–414. Translated from Sobol', I. (1990). Sensitivity estimates for nonlinear mathematical models. *Matematicheskoe Modelirovanie* **2**, 112–118.
[Sbl01] Sobol', Ilya M. "Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates." *Mathematics and computers in simulation* 55.1-3 (2001): 271-280.
[Sbl09] Sobol, Ilya M., and Sergei Kucherenko. "Derivative based global sensitivity measures and their link with global sensitivity indices." *Mathematics and Computers in Simulation* 79.10 (2009): 3009-3017.