Advanced Surrogate Models for Multidisciplinary Design Optimization

Alberto Serna[1]*, Christian Bucher[2,3]

[1] Daimler AG, GR/ACS, 71059 Sindelfingen

[2] DYNARDO – Dynamic Software and Engineering GmbH, Weimar

[3] Center of Mechanics and Structural Dynamics, Vienna University of Technology, Karlsplatz 13 / E2063, 1040 Vienna

In the context of Multidisciplinary Design Optimization (MDO) for vehicle development, concepts and processes are subjected to continuous changes. Complex and computational expensive FEA models need to be integrated into a common workflow in order to produce valid design improvements. This turns into almost impracticable optimization, robust design and/or reliability analyses.

Alternative to direct optimization, surrogate models offer fast model predictions which might facilitate such costly analyses. Nevertheless, still many challenges hide behind the application of these techniques. When considering vehicle design, passive safety simulations reveal many of the problems which can be faced when looking for metamodels over the basis of high dimensional nonlinear data.

In this work we try to further extent the applicability of surrogate models to derive model sensitivities and improve prediction capabilities for such cases. We believe that optimization can only succeed over the basis of high fidelity simulations for which the surrogate models may act as filters. This filtering process gives us information which could even be lately used to construct reduce order FE-models which feature the most important relationships from the original models.

Leaving aside the local or global approximation issue, we concentrate on learning as much as possible at the first stages of the analyses, keeping our optimization problem flexible in order to refine our predictions sequentially and adaptively by shrinking space and dimensions. A framework for nonlinear modeling with simultaneous model selection is presented, managing surrogate modeling and sensitivity analysis for complex problems. This main focus is on fully Bayesian estimation applied to anisotropic Gaussian Processes (GP).

Keywords: Crash, Optimization, MDO, Kernels, Gaussian Processes, Bayesian, Classification, Clustering.

*Contact: Dipl.-Ing. Alberto Serna, Daimler AG, 059/X603, GR/ACS, 71059 Sindelfingen, E-Mail: alberto.serna_esteban@daimler.com

1. Multidisciplinary Design Optimization

In the last years, an increasing effort for Multidisciplinary Design Optimization (MDO) in the automotive industry can be tracked. This trend results from the fact that design improvements at different disciplines (Crash, NVH, Durability, Aerodynamic, etc.) often counteract among them. The integration of different CAE processes into a common optimization algorithm which allows a systematic formulation of the problem becomes a main issue. It is worth to remind that in spite of the increasing computational resources, simulation models still grow continuously in their complexity to offer more accurate physical representations. Hence, we rely on high fidelity models, but MDO remains as an extremely computationally intensive task.

Even though optimization is by itself a challenging task, the requirement for robust design in many situations further difficulties the solution. The experience shows that our deterministic optima tend to be non-robust, and this is the main issue which makes the inclusion of robustness criteria in the optimization high interesting. In order to include uncertainty in the analysis, stochastic, fuzzy or combination of both approaches can be used to mathematically model the problem.

So called Robust Design Optimization (RDO) includes robust measures in the formulation of the optimization problem. Hence, iterative procedures where robustness of the optimum is proved after optimization step are hardly applicable, and integrated schemes become necessary. In some cases, robustness evaluation at each optimization iteration is proposed, which can be cumbersome and is inefficient, as long as noise effects are not independent in the design space. The assumption that both noise and design variables spaces are not related is also not always adequate.

State of the art solutions for MDO in nonlinear systems are summarized here:

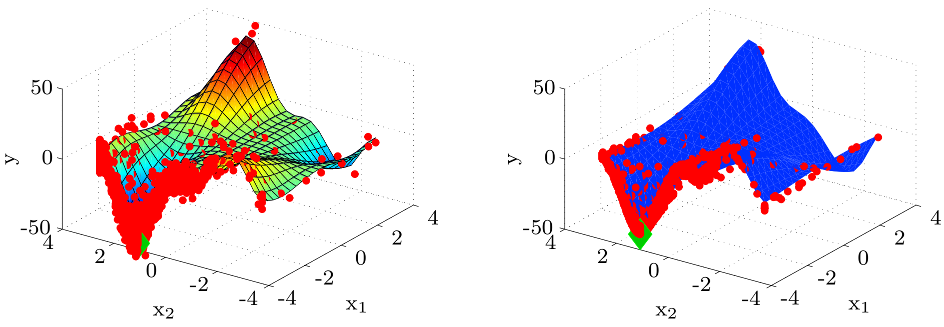
Direct Optimization with Genetic/Evolutionary Algorithms (EA) Stochastic optimization methods are known to handle noisy, non-smooth responses and are able to operate efficiently in large problems [Duddeck (2008)]. One of the most important inconvenience is the fact that this strategy cannot be fully paralleled, as it involves an stepwise evolution process. Moreover, these algorithms require a calibration process, which limits their application to computational intensive systems. Even though the solution can be speed up by tuning evolution strategies, most of the benefit of such methods resides on the exponential convergence after clusters of potential optima are found. Therefore, the stochastic search for first populations provides not much advantages compared with other sampling strategies. Additionally, the solution is inflexible to variations in the objective or constraint function.

Surrogate Model based Optimization Under this framework, model responses are substitute by global approximates which computation is much faster than the original ones. They try to represent the underlying model based on sampling data, which can be a challenging task. The potential of this approach relies on the possibility to carry out expensive uncertainty (UA), sensitivity (SA) or optimization analysis in the global design space. Sequential sampling can also be carried out to reduce approximation error once the location of the expected optimum is found [Donald u. a. (1998); Huang u. a. (2006); Jurecka u. a. (2007)]. Surrogate models allow the user to better understand the behavior of the system, which is extremely important in order to optimize complex models. This strategy also benefits from a similar evaluation time for all disciplines, easing their integration into a multidisciplinary framework.

MAM Multipoint Approximation Method [Toropov (1992)]) introduces the use of adaptive surrogate model for optimization, in which sequential models are built in subspaces of the design, panning the space until convergence is achieved. They are thought for nonlinear noisy problems. Originally, first order Response Surface Methods (RSM) are employed to keep the iterative process efficient. Convergence in the global design space can sometimes be an issue, although robustness of the method is claimed in various references [Kurtaran u. a. (2002)].

ESL Alternatively to nonlinear methods, linear optimization advantages apply if the dynamic loads can be efficiently computed for different states in the path dependent solution. As a result, it is possible to track the nonlinear optimal design, while the number of involved variables do not significantly increase computational demands. Approaches based on Equivalent Static Loads (ESL) describe important advances in the efficient optimization in the context of structural optimization [Kang u. a. (2006); Jeong u. a. (2008)]. Nevertheless, the formulation of the optimization problem as well as the level of nonlinearity strongly affect the convergence of the iterative process, becoming limited on its application.

Fig. 1 compares a direct approach against surrogate or metamodel based optimization, showing the potential reduction in the number of analyses needed to cope with the ultimate solution. The term metamodel is used in this context as a basic model structure which is able to copy the behavior of completely different high fidelity models by adaptation to sampled data. We want to remark that optimization algorithms are also extremely sensitive to problem dimensionality. For the given nonlinear example with 20 dimensions, around 2000 simulations where carried out by the EA in order to converge to the actual optimum. This result was also obtained by applying the same EA optimization on a valid surrogate model generated from a sample of 100 realizations.



(a) Direct (b) Surrogate Model Based

Figure 1: Response optimization.

In the context of vehicle virtual prototyping, and focused on the structural behavior, studies conclude that crash discipline, with emphasis in highly nonlinear frontal crash simulations, becomes critical when trying to make prediction by means of regression metamodels [Kögl u. a. (2008)]. The inherent nonlinear physical behavior of such load cases combines with underlying numerical scatter coming from parallel processing, explicit integration and modeling assumptions. This leads to noiy or unstable results, which can still be sometimes used for qualitative prediction with the model. Quantitative predictions fail in many cases, especially as far as high order responses (e.g. acceleration) are involved, fo which precision turns out not to be enough to represent model causality. This limits the application of MDO to full vehicle concepts.

2. Approach

Several different surrogate models are available in literature and implemented in commercial software, and their use for optimization has been rapidly extended. However there is still a huge problematic due to the wide range of application of such methods, and the complexity involved in modeling real high dimensional data. Besides, problems tend to be over-parameterized and usually few data are available for learning. In the ambit of virtual vehicle prototyping, in statistics widely used term "curse of dimensionality" condemns many modeling efforts.

In absence of previous knowledge, stochastic sampling methods are usually preferred to collect model information. In most applications, optimal Latin Hypercube Sampling (LHS) serves to deliver correlation measures which are used as basic sensitivities. However, this information is not valid for a large class of problems, where nonlinear patterns arise. Moreover, correlation measures in high dimensional problems require also from large datasets to test for variable significance. In fact, dimension reduction (PCA, PLS) or shrinkage methods (e.g. Ridge Regression) can provide for the necessary screening in a more efficient way.

Prediction based on small samples sizes requires additional analysis effort. First, optimal sampling methods which provide space filling and surrogate model oriented designs are desired. Surrogate model flexibility needs to be maximized, while overfitting remains under control. For that purposes, kernel based expansions and model ensembles provide flexible regression schemes, while model building can be achieved with the help of a full Bayesian approach which combines variable selection and parameter estimation.

We approach Robust Design just by smoothing simulation data, which is not always expected to be deterministic. This problem arises as soon as variables are filtered out to increase resolution, as they immediately become noise factors. We search for robust configurations in terms of general stability, whereas very low probability events subject of reliability analysis might only be addressed after the problem has been targeted and bounded precisely. Additionally, we make use of probabilistic models, which can deliver confidence bounds for the predictions.

Besides, classification surrogate models for their use as optimization constraints are introduced. A different perspective which becomes necessary as long as selected model responses are driven by discontinuous modes. Related methods are analogous to the regression case, while the probabilities for class membership are modeled instead of a single conditional distribution. This can be achieved by using softmax expressions for the regression models and alternative inference procedures.

Fig. 2 proposes a basic workflow to handle problematic responses for which linear methods did not succeed, introducing nonparametric extensions with variable selection. If required, derived sensitivities will lead an adaptive or sequential sampling strategy later on. Throughout the procedure, validation measures are considered to decide whether found models can be transferred into the MDO problem definition. Main decision measures shown below are based on lost functions for regression and classification problems

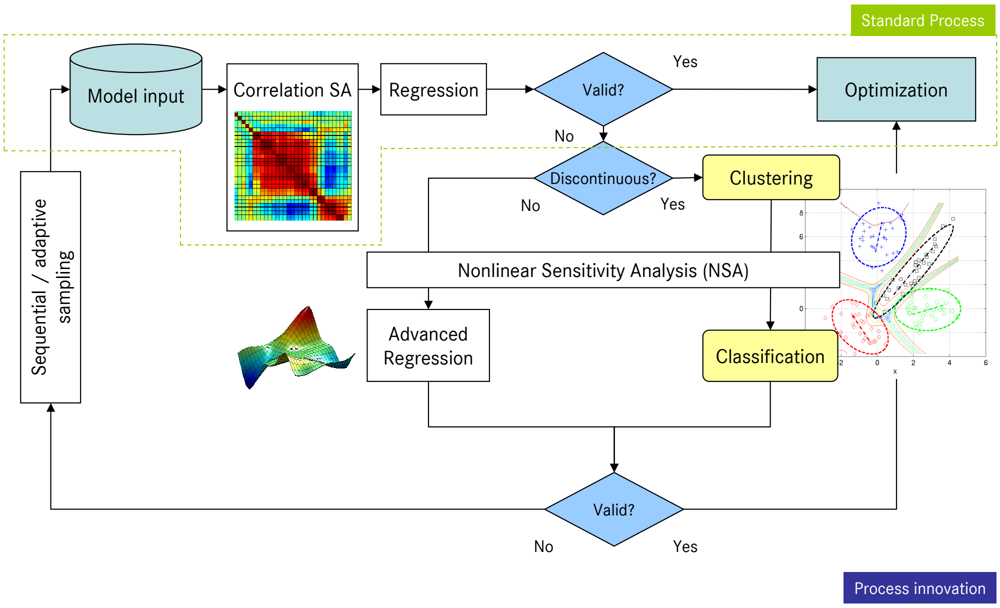


Figure 2: Approach for surrogate modeling of highly nonlinear systems.

and calculated in a k-fold cross validation scheme. Basically, residuals coming from the actual value (y) and the predicted one ($\hat{y}$) are considered.

$$\textit{Mean Square Error (MSE)} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad 0 \leq MSE \leq \infty \tag{1}$$

$$\textit{Coefficient of Prediction (COP)} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad 0 \leq COP \leq 1 \tag{2}$$

$$\textit{Classification Rate (CLASS)} = \frac{1}{n}\sum_{i=1}^{n} I(y_i, \hat{y}_i) \quad I = \begin{cases} 1 & y_i = \hat{y}_i \\ 0 & else \end{cases} \quad 0 \leq CLASS \leq 1 \tag{3}$$

3. Identified Problem Classes

To gain understanding of the different problems to be faced when looking for surrogate models, a explicit division between discontinuous and continuous responses is introduced here. Two different case studies are considered: a B-pillar concept design for side crash in passenger car and the subframe rail design of a commercial vehicle.

The first one involves high nonlinearities mainly due to material plasticity and joint failure. Nevertheless, the intrusion levels at the FE-model in Fig. 3 show continuous behavior with respect to the varied design parameters.

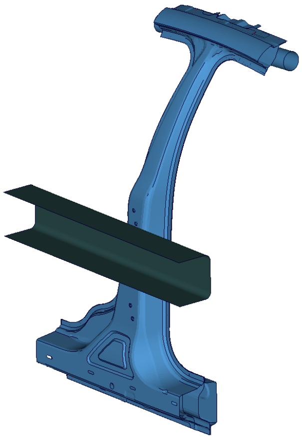
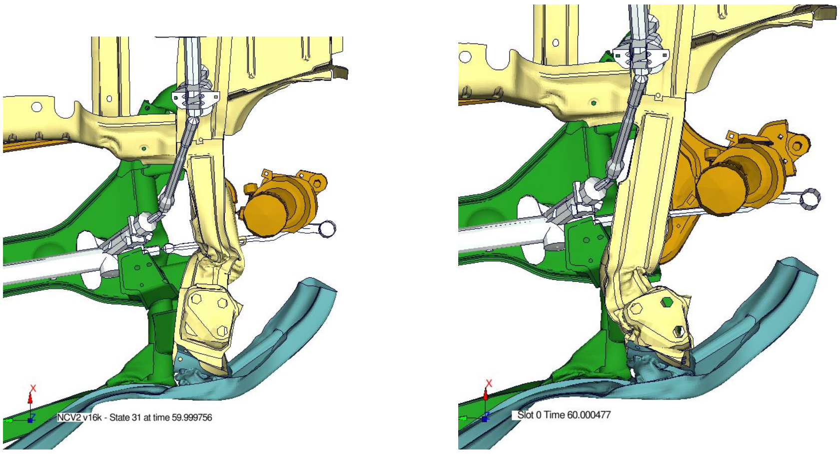


Figure 3: B-pillar model in original and deformed state.

In contrast, in the early development phase of the subframe rail, buckling modes which induce unstable response behavior in the structure might arise. Sensitivity analysis needs to be carried out to understand how a robust design could be achieved. Besides, different variable relationships might exist depending on the resulting deformation modes depicted in Fig. 4, which makes ordinary regression models not suitable.



(a) Outer

(b) Inner

Figure 4: Contradictory subframe rail buckling modes.

In such discontinuous cases, the main goal should be to model the buckling condition based on critical design variables. Moreover, it is highly probable that no continuous model can be found for each deformation mode (left,right), but only noisy responses. Forecasting with the FE-model at that stage might also no longer be valid.

Unfortunately, for many load cases it becomes difficult to make such clear classifications, and compromise solutions need to be adopted. In that sense, additional studies from crash FEA data with high noise levels are summarized in the Appendix to complete a first overview. Some of the considered frontal load cases are known to be most challenging for metamodeling purposes.

4. Kernel Methods

When trying to derive model input-output relationships, linear models often provide good approximations. It is important to remark that in spite of nonlinearity occurring on dynamic responses, variable and response relationships might still be linear. Even for many nonlinear models the linearity assumption holds under certain assumptions.

Unfortunately, this is not always the case for complex systems. In order to cope with different problems in a flexible way, kernel methods provide an optimal basis for maximizing model learning capabilities, as a generalization of the linear model via a transformation into nonlinear feature spaces. In this paper, we present several examples which require from such methods to learn from our simulation results. In addition, methods for inferring model structure in high dimensional spaces are introduced.

Thanks to such a transformation, a Reproducing Kernel Hilbert Space (RKHS) is proposed ($\mathcal{H}$), based on nonlinear mapping functions, which reduce dimensionality problems to learning capacity problems:

$$\begin{aligned} \Phi : \quad \mathcal{X} &\rightarrow \mathcal{H} && (4) \\ x &\mapsto \Phi(x) && (5) \end{aligned}$$

A Hilbert Space is a generalization of an Euclidean Space into n-dimensional, providing important mathematical properties like existence of an inner product. This operations can be used for choosing similarity functions for regression or classification methods.

It is possible to derive the kernel function from the RKHS, thanks to the existence of all point function evaluations:

$$\begin{aligned} p_x : \quad \mathcal{H} &\rightarrow \mathbb{R} && (6) \\ f &\mapsto p_x(f) = f(x) && (7) \end{aligned}$$

Continuity is built in the mapping, meaning that whenever f and f' are close in $\mathcal{H}$, then $f(x)$ and $f'(x)$ are close in $\mathbb{R}$. This can be thought of as a topological prerequisite for generalization ability [Scholkopf und Smola (2002)]. After mapping to higher dimensional feature spaces learning might become much easier based on low dimensional similarity measures, like shown in Fig. 5.

Linear methods are still applicable in the nonlinear projected space. This property is also called kernel dual representation [Bishop u. a. (2006)], as it provides a linear decomposition of eigenvalues and modes, as described in Mercer's theorem.

Theorem 4.1 (Mercer's Theorem)**.** *If k is a continuous kernel of a positive definite integral operator on $L_2(\chi)$ (where χ is some compact space),*

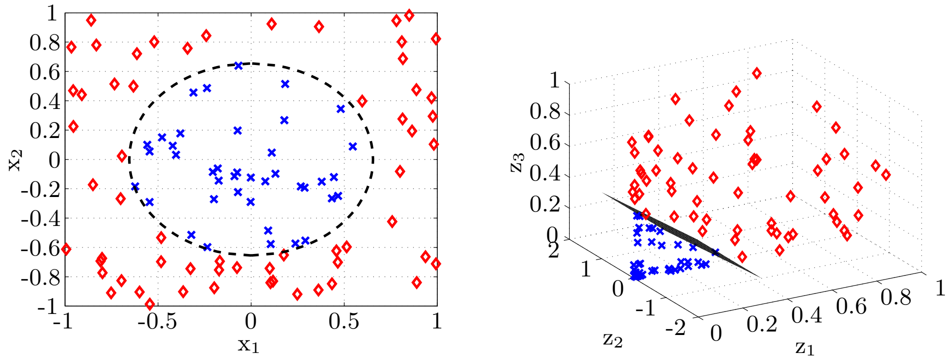
$$\int_{\chi} k(x,\chi) f(x) f(\chi)\, dx\, d\chi \geq 0 \tag{8}$$

It can be expanded as,

$$k(x,\chi) = \sum_{i=1}^{\infty} \lambda_i \phi_i(x) \phi_i(\chi) \tag{9}$$

using eigenfunctions ϕ_i and eigenvalues $\lambda_i \geq 0$.

$$
\begin{aligned}
\Phi: \quad \mathbb{R}^2 &\rightarrow \mathbb{R}^3 \\
(x_1, x_2) &\mapsto (z_1, z_2, z_3) := (x_1^2, \sqrt{2}x_1x_2, x_2^2)
\end{aligned}
$$



(a) Nonlinear classification in original space

(b) Linear classification in nonlinear feature space

Figure 5: Expansion to RKHS feature space.

Many well-known data-driven models for supervised learning tasks are based on RKHS, like Gaussian Processes (GP), Supported Vector Machines (SVM) or Relevance Vector Machines (RVM). In contrast, other classic nonparametric regression methods like Weighted or Moving Least Squares (W/MLS) suffer from support problems due to its formulation, where model shape depends on the predicted values. Non robustness and extrapolation issues make MLS not adequate, due to the inability to cover space boundaries in high dimensions with additional support points.

Below, general kernel regression expressions are introduced, with the example of the Gaussian kernel function. The last expression shows a model predictor, as a function of the Gram matrix K and input observations, including regularization terms λ. The question for the different kernel methods resides on how linear weights and kernel shape functions should be estimated.

$$k(x, \chi) = \phi(x)^T \phi(x'_n) \tag{10}$$

$$k(x, \chi) = \exp\left(-\frac{\|x - \chi\|}{2\sigma^2}\right) \quad \text{Gaussian Kernel} \tag{11}$$

$$\widehat{y} = w^T \phi(x) = k(x)^T (K + \lambda I_n)^{-1} x \tag{12}$$

4.1. Gaussian Processes

In the mathematical context of kernel methods, the probabilistic perspective given by stochastic processes proved to be the best to learn from our complex simulations in terms of generalization and computing time. Roughly speaking a stochastic process is a generalization of a probability Gaussian process distribution (which describes a finite-dimensional random variable) to functions. From the Bayesian point of view, we may define prior functions for the model to be learned. Gaussian processes provide several computational advantages and can be shown to be the solution of many adaptive schemes like neuronal networks, saving the adaptation process necessary in their architectures.

They aim an extension of linear model to infinite nonlinear basis, whose complexity is related to the amount of available data.

From the kernel perspective, a Gaussian Process can be specified by its mean and covariance function:

$$f(\mathbf{x}) \sim \mathcal{GP}\{\mathbb{E}[f(\mathbf{x})], k(\mathbf{x}, \chi)\} \tag{13}$$

Parameter estimation of GP can be very similar to other kernel methods like Supported Vector Machines (SVM), for which a convex problem can formulated. However some conditions regarding likelihood concavity need to be met. This is to say, that the MAP (Maximum A Posteriori) estimation for GP has a close correspondence with the SVM solution [Rasmussen und Williams (2006)].

For a general stochastic process, marginal likelihood can be computed as the integral of likelihood times the prior distribution as:

$$p(\mathbf{y}|X) = \int p(\mathbf{y}|f, X)p(f|X)\,\mathrm{d}f \tag{14}$$

Considering a Gaussian prior, the following expression holds for the regression result:

$$p(\mathbf{y}|X) = -\frac{1}{2}\mathbf{y}^T(K + \Sigma_n^2 I)^{-1}\mathbf{y} - \frac{1}{2}|K + \Sigma_n^2 I| - \frac{n}{2}\log 2\pi \tag{15}$$

Coming back to the their application in MDO, these methods maximize generalization ability for small samples, providing basis for nonlinear sensitivity analysis via so called Automatic Relevance Determination (ARD) covariance functions, while keeping computational efforts affordable. Besides, probabilistic models can be used to provide prediction bounds.

In this work, we show the advantages of introducing more complex kernel or covariance functions, in order to learn hidden or latent features. Interpretability is also a main benefit of stochastic processes when compared with full black box solutions like neuronal networks or other kernel methods like SVM.

5. Bayesian Approach

Data-driven metamodels require from parameter or function estimation. The Bayesian perspective describes a general framework for estimation, in which the aim is to approximate the posterior distribution of the probabilistic distribution from the prior and likelihood expressions (16). Likelihood functions need to be assumed for the underlying processes. Analytical inference is often mathematically intractable, and approximations for the posterior distribution by means of Likelihood type II, Laplace and Expectation Propagation (EP) are widespread. In the past years, a trend towards Monte Carlo Markov Chains (MCMC) can be followed.

$$\pi(\theta|X) = \frac{\pi(X)l(\theta, X)}{\int \pi(X)l(\theta, X)d\theta} \tag{16}$$

MCMC is considered one of the 10 most important mathematical algorithms. It is based on random walks, therefore Monte Carlo, but following a Markov chain. A Markov chain is a mathematical model for stochastic processes, in which the current state only depends on most of the previous results. The chain ensures that the walk follows high density regions in the probability space, thus reducing the effort for estimation of integrals or optimization. In contrast to Monte Carlo sampling, observations are correlated. The main advantage is the convergence properties of the chain, although it is not possible to determine beforehand the amount of realizations needed. The mixing properties of the chain highly depend on the model complexity and chosen sampling method. A main issue of the method is to ensure convergence of the chain, for example by means of parallel chains. Usually, first chain iterations are rejected (so called "burn in"), in order to remove biased realizations from the whole sample.

A compact introduction to most common methods can be found in [Andrieu u. a. (2003)], whose main characteristics are briefly introduced next:

Metropolis-Hastings algorithm Generates a random walk using a proposal density (q, e.g. Gaussian) and a criterion for accepting or rejecting proposed moves,

$$\alpha < \min\left\{\frac{p(x^{new})q(x^{current}|x^{new})}{p(x^{current})q(x^{new}|x^{current})}, 1\right\} \tag{17}$$

and became very popular due to its wide applicability.

Gibbs sampling A special case of the former, more efficient but limited to the cases where all the conditional distributions of the target distribution can be sampled exactly.

Slice sampling A general version off the Gibbs sampler, depends on the principle that one can sample from a distribution by sampling uniformly from the region under the plot of its density function. This method alternates uniform sampling of an auxiliary variable in the vertical direction with uniform sampling from the horizontal "slice" defined by the current vertical position. This is an alternative to approach sequential integration of multivariate distributions.

Adaptive MCMC Variations of the Metropolis-Hastings algorithm in order to improve converge properties by tuning proposal distribution parameters during estimation. Provides improved mixing in large dimensions.

Hybrid Monte Carlo (HMC) This method tries to avoid random walk behavior by introducing an auxiliary momentum vector and implementing Hamiltonian dynamics where the potential function is the target density. Gradient information is used at then discarded ar each state. The end result of Hybrid MCMC is that proposals move across the sample space in larger steps with improved convergence and space filling properties.

Reversible Jump method (RJMCMC) It is a variant of Metropolis-Hastings that allows proposals that change the dimensionality of the space, which can be useful for feature identification purposes.

5.1. Bayesian Modeling for Feature Identification

The task of finding nonlinear relevant components can be accomplish by combining flexible metamodels with variable selection, performing estimation simultaneously. Kernel methods provide the necessary flexibility, especially if ensembles of kernels with anisotropic properties are used. Bayesian methods allow the inclusion of penalized likelihood functions or even latent variable vectors to find the optimal model structure.

Even though maximum likelihood minimization provides good results in many applications, its tendency to fail under circumstances (multimodal, high dimensional distributions) makes the full Bayesian approach a valuable alternative. Moreover, variable selection can also be included in the Bayesian modeling (18), defining discrete priors for model structure ($\mathcal{H}_i$, not to confound with RKHS definition), linear coefficient parameter ($\mathbf{w}$) and kernel hyperparameters (θ). By integrating with Monte Carlo Markov Chains in a fully Bayesian formulation, posteriors might be inferred in a relative efficient manner by means of sampling strategies like Reversible Jump MCMC or the Gibbs Sampler.

For example, marginal distributions can be sequentially integrated, first for model parameters:

$$p(\mathbf{w}|\mathbf{y}, X, \theta, \mathcal{H}_i) = \frac{p(\mathbf{y}|X, \theta, \mathcal{H}_i)p(\mathbf{w}|\theta, \mathcal{H}_i)}{p(\mathbf{y}|X, \theta, \mathcal{H}_i)} \tag{18}$$

Then, the posterior over the hyperparameters, including the previous marginal likelihood (also called evidence) as the new likelihood function,

$$p(\theta|\mathbf{y}, X, \mathcal{H}_i) = \frac{p(\mathbf{y}|X, \mathcal{H}_i)p(\theta|\mathcal{H}_i)}{p(\mathbf{y}|X, \mathcal{H}_i)} \tag{19}$$

and finally, conditional distribution for the model structure can be elaborated based on found evidence $p(\mathbf{y}|X) = \sum_i p(\mathbf{y}|X, \mathcal{H}_i)p(\mathcal{H}_i)$:

$$p(\mathcal{H}_i|\mathbf{y}, X) = \frac{p(\mathbf{y}|X, \mathcal{H}_i)p(\mathcal{H}_i)}{p(\mathbf{y}|X)} \tag{20}$$

Most kernel regression models only focus on learning of kernel location parameters (sparse approximations) but not the kernel scale parameters. Kernels with a single scale parameter, such as the Gaussian kernel whose precision matrix is a scale multiplied by the identity matrix. These kernels assume homogeneity across all covariates, which is usually not true in modern applied problems, particularly when the number of covariates is large. ARD kernels provide a first attempt to improve flexibility, with diagonal precision matrix, which assigns a scale parameter for each covariate. This facilitates a variety of structures that can be used in feature selection.

In order to maximize learning capacity, ensembles of anisotropic, stationary kernels are defined:

$$y_i = \sum_j \beta_j K(\mathbf{x}, \chi_j) + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \phi^{-1}) \tag{21}$$

$$K(\mathbf{x}, \chi) = \exp\left\{-\sum_{l=1}^{p} \lambda_l (x_l - \chi_l)^2)\right\} \tag{22}$$

Besides, we propose compositions of kernel by applying kernel operations to further tune the model. It is possible to operate on unique kernels via summation, multiplication, convolution, scaling and more without altering their necessary mathematical properties. Also, considering different similarity functions for noise parameters might help to precise in the approximation.

To conclude, we use such complex inference approach for the determination of important hidden features in high dimensional data, which otherwise could not be discovered. This methodology is only applied at those cases for which variable screening methods did not succeed.

6. Regression based Surrogate Models

Having a look to continuous problems from section 3, we might employ regression theory to construct surrogate models which can infer the conditional density distribution, and hence the process underlying noisy data.

The following list summarizes some of the most popular methods which can be found in the literature, including nonlinear versions, which are crucial to fit selected model responses:

- Linear Regression or Response Surface Models (RSM)
- Moving Least Squares (MLS)
- Regression Trees (CART)
- Artificial Neuronal Networks (ANN)
- Generalized Additive Models (GAM)
- Supported Vector Regression (SVR)
- Projection Pursuit Regression (PPR)
- Multivariate Adaptative Regression Splines (MARS)
- Gaussian Process (GP)
- Kernel Partial Least Squares (KPLS)

The build many branches and research areas which are often strongly linked. Some of them share many similarities. Kernel regression variants include different regularization strategies like Partial Least Squares, Lasso or Ridge regression, till full Bayesian approach.

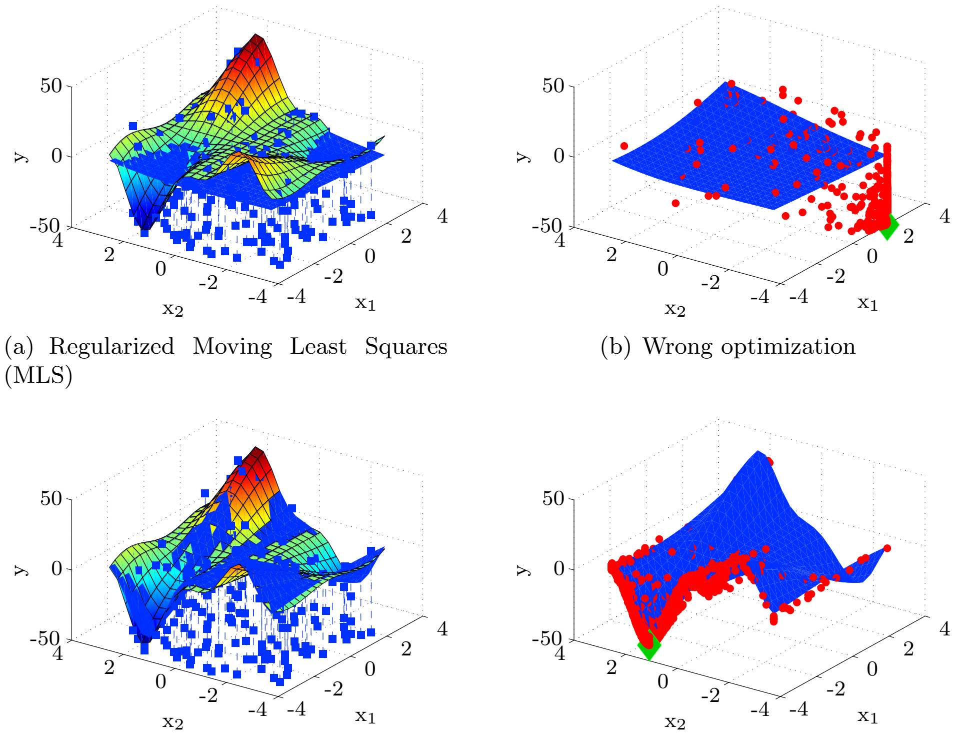
After following a best surrogate model approach, we found kernel methods, and particularly Gaussian Processes the most convenient strategy for generalization in complex problems with small datasets and more than 20 variables. In the following, we will discuss examples where these models provide optimal solutions.

6.1. Example: Continuous Function

To illustrate the curse of dimensionality problem, we make use of a highly nonlinear function, which will be fitted via different regression methods. This example reproduces the case of a low dimensional manifold carrying most of the variation of the response. First we sample the minimal two dimensional base, and then increase the dimensionality up to twenty variables. The last do actually not have any contribution.

The model is based on the Ishigami function, related to research in nonlinear sensitivity methods:

$$y(x_1, \ldots, x_{20}) = sin(x_1) + 7sin^2(x_2) \cdot x_1 \cdot (x_2 + 0.5) + \mathcal{N}(0, 5); \quad x_i \in [-\pi, \pi] \tag{23}$$



(a) Regularized Moving Least Squares (MLS)

(b) Wrong optimization

(c) ARD Gaussian Process (GP)

(d) Succesful optimization

Figure 6: Optimization based on regression models in optimal subspace of the high dimensional Ishigami function.

Fig. 6 attempts to address the risk of overlooking important trends of the underlying model in the surrogate, which turns into a completely wrong optimization results. On the left side, surrogate models with 100 support points are overlaid to the original model. On the right side, EA optimization history is plotted, with a biased localization of the optimum for the MLS metamodel [Most und Will (2008)]. This is related to the insufficient model sensitivity, with merely estimation of linear trends.

This example reflects the need for flexible approximations schemes, for example by including anisotropic designs which allow for directional sensitivity in the model. Only defining kernels which can accurately reproduce the underlying variogramm can the variable selection succeed. Nevertheless, complexity in kernel design extends to the parameter estimation process, and hence Bayesian methods tend to provide an optimal methodology for model building.

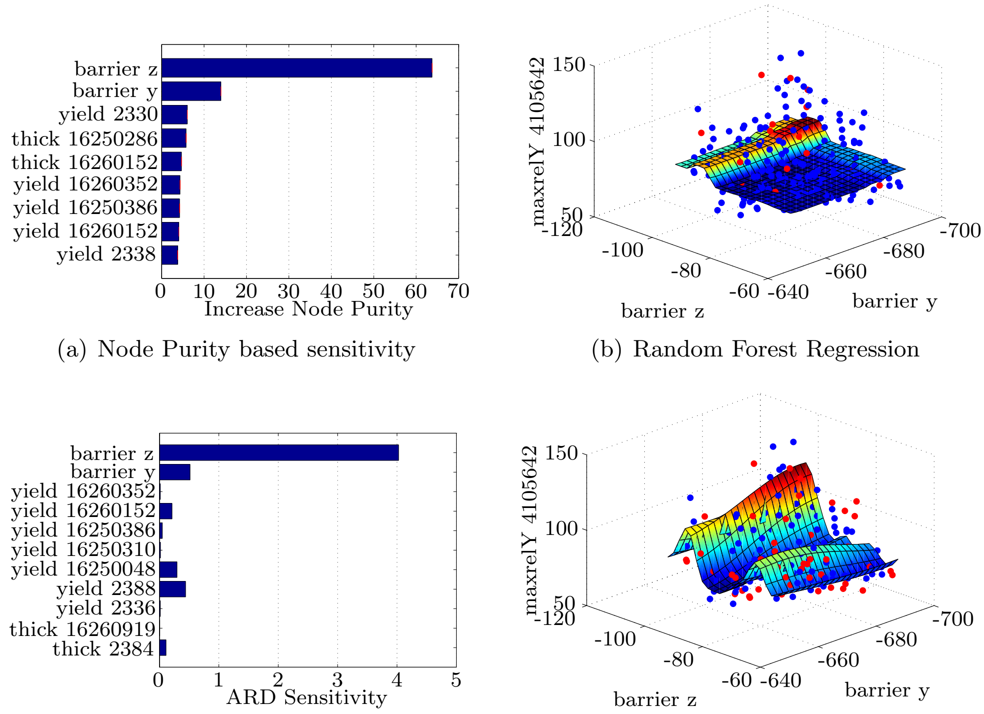
Noise form assumptions play also a central role to increase the precision in the estimation. Heteroscedasticity can be managed by the metamodel.

6.2. Application: High Dimensional Nonlinear Continuous Problem

Next, we present a real case which shows great similarities with the previously introduced example. It deals with one of the multiple responses obtained from a frontal crash FEA for a roadster model.

The problem is high dimensional with 99 variables and only 196 simulations are available. Such load cases usually deliver high noise levels in the solution. Again, we want to filter relevant trends in the data, which can help to understand model behavior and decide about the possible inclusion of the response inside the MDO process. We derive sensitivity indexes from fitted nonparametric models.

Three representative solutions are overviewed. First, we use Random Forest implementation (RF) [Liaw und Wiener (2002)] in order to show its potential to screen important variables, especially if their variation is nonlinear. In our implementation, different Pareto rule inspired strategies are managed to select the optimal subspace. However, this approach might sometimes fail and often suggests the need for iterative search solutions, which might be too expensive.



(a) Node Purity based sensitivity

(b) Random Forest Regression

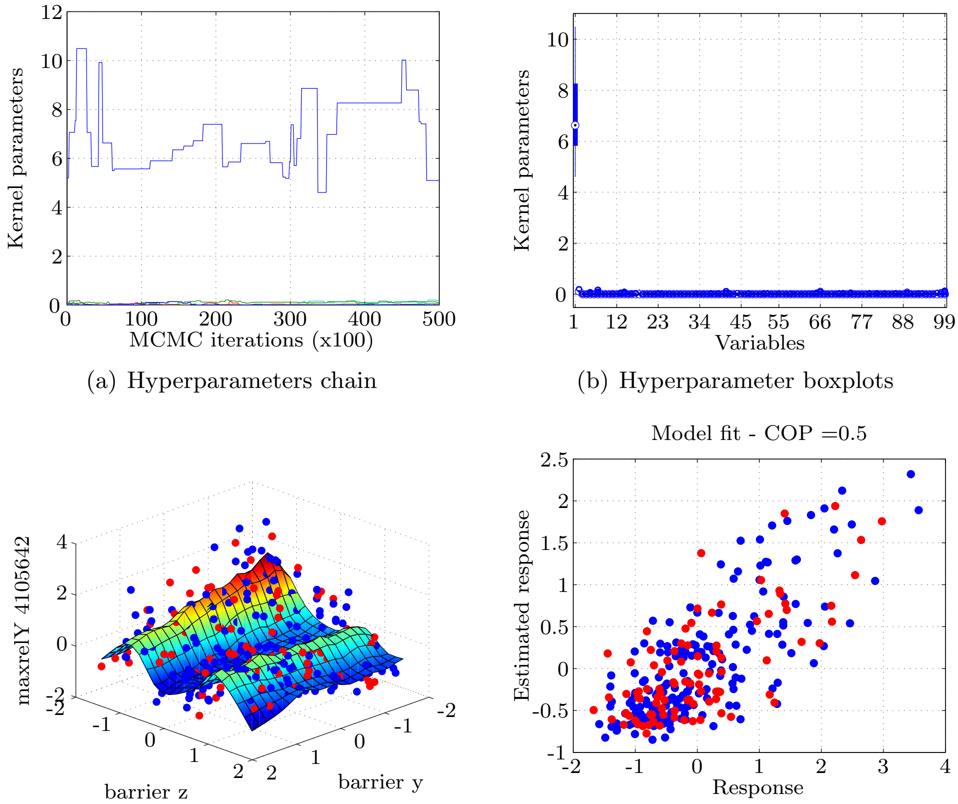
(c) ARD based sensitivity

(d) GP Regression

Figure 7: Selected nonlinear response from frontal crash of a roadster model (see Appendix, Table 7).

The drawbacks coming from the use of local linear regression by the RF are overcome by the use of Gaussian Process, whose kernel or covariance functions can better copy the variation of the unknown process. The last can be used to determine sensitivities in the data more precisely like shown in Fig. 7, although variable screening through RF results is performed previously in the 99 dimensional space.

Finally, to overcome the dimensionality problem while keeping flexibility in the approximation, we evaluate the performance of kernel ensembles in a fully Bayesian implementation. Combining structure and parameter estimation by means of an implementation of the RJMCMC sampling method, we identify the optimal regression subspace (Fig. 8). Despite its potential, the method still shows classical convergence problems in the estimation process, mainly owed to overlapping in the kernel ensemble. Robust estimation is granted only at the cost of roundabout 50,000 MCMC iterations (20 minutes in Intel DualCore® at 3,0 Ghz).



(a) Hyperparameters chain

(b) Hyperparameter boxplots

(c) Low dimensional nonlinear manifold (standarized)

(d) Regression

Figure 8: Feature identification with GP and fully Bayesian approach.

To conclude, we can claim that screening with Random Forest provides robust enough criteria in order to apply anisotropic GP in the pre-filtered design space, which parameters can efficiently computed for small sample size. Downscaling of the algorithm is also possible thanks to the last advances in sparse approximations [Rasmussen und Williams (2006)], which can decrease computational effort from $\mathcal{O}(n^3)$ to $\mathcal{O}(nd^2)$. Nevertheless still exist cases where Random Forest might overlook important variables.

7. Classification based Surrogate Models

Taking over the problem class definition from section 3, we faced now an alternative to continuous responses. This alternative needs to be confronted when learning from highly nonlinear simulation models. The objective is to find natural clusters in data, and decision models which can predict the group membership for new design configurations. This feature can also be fitted within MDO, using found probabilistic models as optimization constraints. Hence, it would be possible to tackle with noisy responses, providing that instabilities or differentiated patterns can be found.

A similar approach is presented in [Missoum (2007); Basudhar u. a. (2008)], however no solutions for high dimensional data are offered. This is an essential aspect for its use in real applications.

In this paper, High Dimensional Discriminant Analysis (HDDA) [Bouveyron u. a. (2007)] and classification based on Random Forest [Liaw und Wiener (2002)] are used at the first stage. Both methods provide quite flexible solutions which can deal with high dimensional spaces by means of adaptive Gaussian density estimation and stochastic model space partitioning respectively. Kernel based classification and full Bayesian extensions are postponed for posterior investigations.

7.1. Clustering Discontinuous Responses

Assigning simulations to different groups can be done is several ways. Knowledge or design based constraints can serve to the task of splitting data into groups, in spite of the fact that the transition between them might be obscure.

Instead, it would be desirable to find physical evidence in the data. We first attempt to do that by applying clustering algorithms. In the machine learning field the process is called unsupervised learning, because only feature data is available to investigate relationship between different simulations. Data splitting is normally carried out based on distance measures. Usually, data clustering supports later data evaluation by means of classification, outlier detection, subset regression, within more complex analysis [Steinbach u. a. (2004)].

Different Clustering strategies can be practiced. A common division is carried out between hierarchical (K-means, K-medoids) and density based methods like Gaussian Mixture Models (GMM). In the last years, many new algorithms arose to cope with dimensionality challenges (Bayesian Approach [Tadesse u. a. (2005)], CLIQUE, etc.)

Several schemes are proposed for dimensionality reduction in the last years. In contrast to the classification or regression case, we are modeling full densities and not only conditional distributions. As a consequence, the curse of dimensionality becomes even more critical, and projection methods almost unavoidable [Parsons u. a. (2004)].

In our implementation, we leave aside high dimensional strategies at a first stage due to the lack of robustness, and content ourselves with pairwise scatterplot based clustering. Clustering is performed on the variation of the target response for each variable. For such purposes, we employ density based Gaussian Mixture modeling, which estimation is carried out using Expectation Maximization (EP). This method combined with low correlated sampling sets allows us to recognize differentiated groups in some of the analyzed case studies. The validity of such clusters is checked with the help of time series or spatial information when available.

7.2. Example: Discontinuous Function

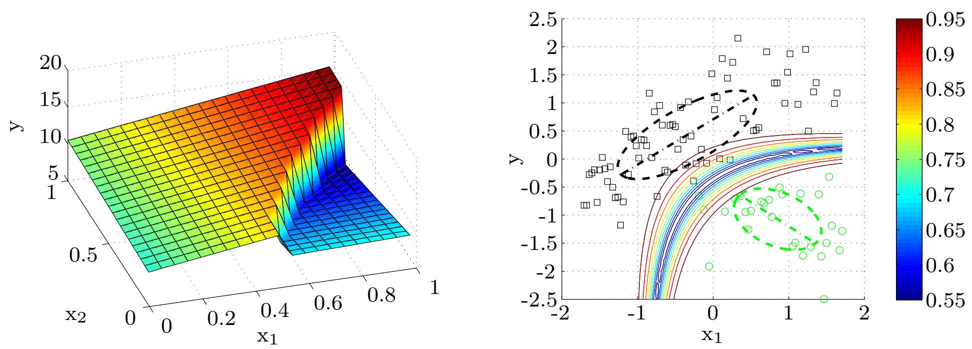
The following basic example pretends to represent the class of problems aimed with this approach. In the ideal case, we face a model which has a switching condition as a function of the design variables. The response alternates from two different states. This unstable behavior is quite common in many physical systems. The variational characteristics for each state can be completely different, and hence, no unique model can efficiently predict the response of the system.

In the current example, only 3 out of 20 variables control the unstable behavior of the system. This assumption seems quite likely to us, because physical discontinuities tend to depend on few design parameters in a nonlinear manner. Also variation within each state follows an ordinary linear model, for which the contribution of the three classification variables varies for each switching state. Additionally, normal distributed noise is added to each expression:

$$\mathbf{y} = \begin{cases} X\beta + \epsilon & \text{if } x_1 \geq \sqrt{x_2^2 + x_3^2} \\ X\gamma + \zeta & \text{if } x_1 < \sqrt{x_2^2 + x_3^2} \end{cases} \tag{24}$$

Our objective is to find such alternating models, be able to predict the transition between them, and finally try to characterize system behavior for each state. We sample 100 model realizations via Latin Hypercube with minimal input correlation.

Clustering for each variable forming the design space with the target response leads us to the projection which seems to split data optimally. We allow a maximum of three clusters to be identified. Different criteria like Maximum Likelihood (ML), Partition Coefficient (PC) or Classification Entropy (CE) are evaluated in order to choose the optimal number of clusters.



(a) Discontinuous example function (noiseless)

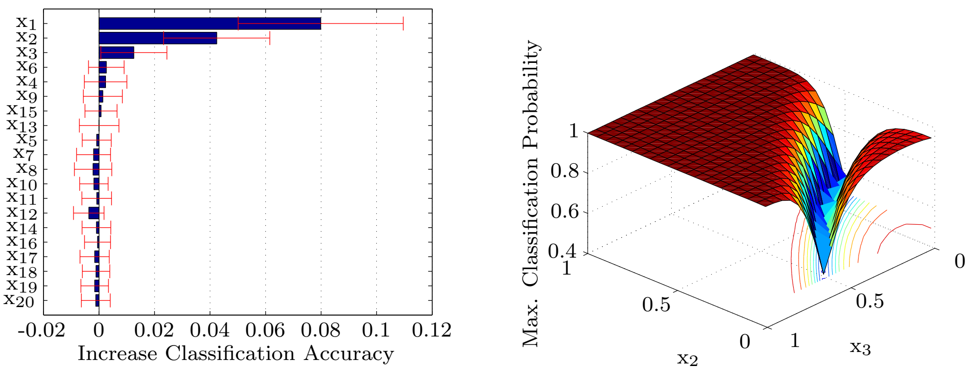
(b) Found optimal cluster densities

Figure 9: Pairwise Clustering with Gaussian Mixture Models (GMM).

After splitting results into clusters, we make use of Random Forest classification to find the optimal subspace for classification. We screen off non-significant variables and proceed with classification in a contracted dimensional space.

For this example, it is possible to find switching states in the pairwise projections of the original design space. Once this is realized, classification methods can accurately delimit the transition between both models (Fig. 10). Classification rates over 0.9 achieved in a 5-fold cross validation setting are collected into Table 1, with the number of clusters appended to the models (c:2). In this case, linear regression on the classified data performs accurately (COP over 0.9), as long as all variables are considered again in the estimation process.

Real application examples (see Appendix) show to us that groups tend to overlap themselves and usually noisy process are hidden in each cluster. In fact, these are the cases where we can get more advantage of modeling group membership probabilities accurately. If defined groups are actually relevant, high dimensional classification methods allow representing complex decision boundaries, which can be used as surrogate models in MDO.



(a) Variable Screening with Random Forest (b) Transition between group memberships

Figure 10: Classification Model (HDDA).

7.3. Application: High Dimensional Discontinuous Problem

The process described in section 7.2 is now applied to a sample data from a Roof Impact Test based on FEA. Among seven response variables, only one is selected to reproduce here the modeling process.

In Fig. 11 we can observe two identified clusters corresponding to the minimal value of the displacement. This division is supported by the time evolution of the response, which indicates some discontinuity, which can be caused for example by a contact setting. This is an extreme case which could also be determined visually. However, in many cases not every response can be analyzed in detail, and process automation becomes essential.

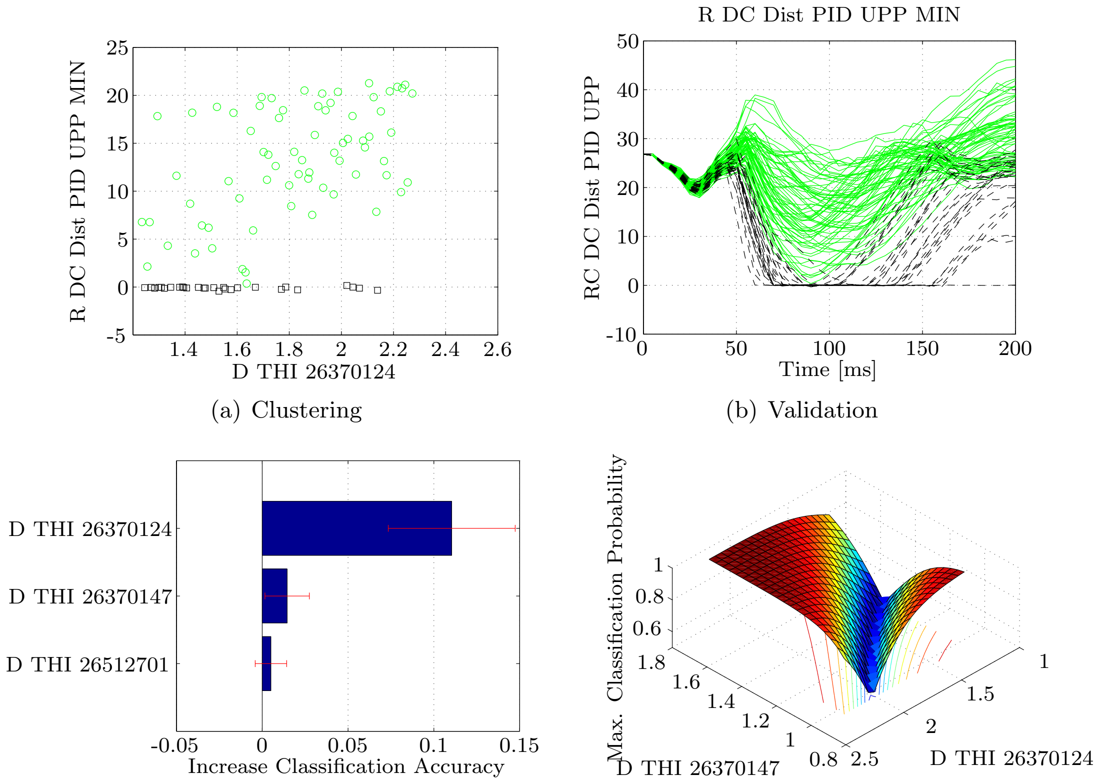
Once we group the different cases, classification is performed on labeled data, including variable selection. This is carried out based on node purity criterion for the Random Forest model, to find surrogate models on the optimal subspace. It is important to remark that sensitivities might be completely different to correlation values or regression based contributions. Decision boundaries can be use to predict valid design combinations.

8. Conclusions

This paper pretends to highlight the challenges related to the search for surrogate models when highly complex models are involved. We are aware of the limitations to reproduce all effects encountered in our high fidelity FE-models, but among several possibilities for understanding and optimizing such designs, we believe that a combined strategy with regression and classification models can be extremely useful.

However, for such purpose, more flexible metamodels which enhance identification of nonlinear trends are introduced in our processes. Simultaneous model building and inference of model parameters turns to be critical for large sample spaces with few data available.

Advanced regression models need to be considered for problems including nonlinear behavior. Kernel methods offer a generalization of linear models with robust estimation methods. FE-model validation can be supported with the findings from surrogate models, and not limited to correlation measures. As a result, numeric problems might be identified and controlled, also reducing uncertainty in the metamodels.



(a) Clustering

(b) Validation

(c) Variable screening with Random Forest

(d) Transition between group memberships

Figure 11: Response classification for Roof Impact Test (see Appendix, Table 4).

The same context applies for the discontinuous models, where classification provides an alternative to cope with unstable and/or noisy responses, which can be defined as optimization constraints within MDO strategies. High dimensional classification delivers encouraging results, although some cases still require from larger sample sizes or more clear definition of the physical response under study.

Hence, we approach optimization for our problems in a stepwise manner, carrying out a sequential learning process flexible to modifications in the formulation of the optimization problem. The aim is to maximize gained information for each step.

9. Outlook

Derived from the results presented, it turns that robust clustering in such high dimensional problems might be extremely valuable. Nevertheless, we believe that the integration of available time and spatial data needs to be prioritize in order to improve surrogate model prediction quality.

A parallel topic for us consist on merging data-driven and mechanistic surrogate models, so that sample resolution can be further reduced based on additional physical considerations.

We face now an integrated effort to implement and further develop some of the methods covered in this article and push them forward until we can prove their usability in the design development phase. The importance of variable selection needs to be further extended into the parameterization of vehicle geometry.

References

[Andrieu u. a. 2003] Andrieu, C. ; De Freitas, N. ; Doucet, A. ; Jordan, M.I.: An introduction to MCMC for machine learning. In: *Machine learning* 50 (2003), Nr. 1-2, S. 5–43

[Basudhar u. a. 2008] Basudhar, Anirban ; Missoum, Samy ; Sanchez, Antonio H.: Limit state function identification using Support Vector Machines for discontinuous responses and disjoint failure domains. In: *Probabilistic Engineering Mechanics* 23 (2008), Nr. 1, S. 1 – 11. – URL http://www.sciencedirect.com/science/article/B6V4M-4PN05WW-2/2/26bd4dff94fd38c1d44a3ec3b89333f3. – ISSN 0266–8920

[Bishop u. a. 2006] Bishop, C.M. u. a.: *Pattern recognition and machine learning*. Springer New York, 2006

[Bouveyron u. a. 2007] Bouveyron, C. ; Girard, S. ; Schmid, C.: High-dimensional data clustering. In: *Computational Statistics and Data Analysis* 52 (2007), Nr. 1, S. 502–519

[Donald u. a. 1998] Donald, RJ ; Matthias, S. ; William, JW: Efficient Global Optimization of Expensive Black-Box Function. In: *Journal of global optimization* 13 (1998), S. 455–492

[Duddeck 2008] Duddeck, F.: Multidisciplinary optimization of car bodies. In: *Structural and Multidisciplinary Optimization* 35 (2008), Nr. 4, S. 375–389

[George und McCulloch 1997] George, E.I. ; McCulloch, R.E.: Approaches for Bayesian variable selection. In: *Statistica Sinica* 7 (1997), S. 339–374

[Huang u. a. 2006] Huang, D. ; Allen, T.T. ; Notz, W.I. ; Zeng, N.: Global Optimization of Stochastic Black-Box Systems via Sequential Kriging Meta-Models. In: *Journal of Global Optimization* 34 (2006), Nr. 3, S. 441–466

[Jeong u. a. 2008] Jeong, SB ; Yi, SI ; Kan, CD ; Nagabhushana, V. ; Park, GJ: Structural optimization of an automobile roof structure using equivalent static loads. In: *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering* 222 (2008), Nr. 11, S. 1985–1995

[Jurecka u. a. 2007] Jurecka, F. ; Ganser, M. ; Bletzinger, K.U.: Update scheme for sequential spatial correlation approximations in robust design optimisation. In: *Computers and Structures* 85 (2007), Nr. 10, S. 606–614

[Kang u. a. 2006] Kang, B.S. ; Park, G.J. ; Arora, J.S.: A review of optimization of structures subjected to transient loads. In: *Structural and Multidisciplinary Optimization* 31 (2006), Nr. 2, S. 81–95

[Kögl u. a. 2008] Kögl, M. ; Klimetzek, F. R. ; Pletschen, B. ; Möller, J. ; Fülbier, K ; Pfeiffer, M.: Multidisziplinäre Optimierung des Fahrzeugrohbaus. In: *SIMVEC Berechnung und Simulation im Fahrzeugbau 2008* VDI, 2008, S. 721. – ISBN 978–3–18–092031–3

[Kurtaran u. a. 2002] Kurtaran, H. ; Eskandarian, A. ; Marzougui, D. ; Bedewi, NE: Crashworthiness design optimization using successive response surface approximations. In: *Computational Mechanics* 29 (2002), Nr. 4, S. 409–421

[Liaw und Wiener 2002] Liaw, Andy ; Wiener, Matthew: Classification and Regression by randomForest. In: *R News* 2 (2002), Nr. 3, S. 18–22. – URL http://CRAN.R-project.org/doc/Rnews/

[Missoum 2007] Missoum, S.: Controlling structural failure modes during an impact in the presence of uncertainties. In: *Structural and Multidisciplinary Optimization* 34 (2007), Nr. 6, S. 463–472

[Most und Will 2008] Most, T. ; Will, J.: Meta-model of Optimal Prognosis-An automatic approach for variable reduction and optimal meta-model selection. (2008). – URL http://www.dynardo.de

[Parsons u. a. 2004] Parsons, L. ; Haque, E. ; Liu, H.: Subspace clustering for high dimensional data: a review. In: *ACM SIGKDD Explorations Newsletter* 6 (2004), Nr. 1, S. 90–105

[Rasmussen und Williams 2006] Rasmussen, C.E. ; Williams, C.K.I.: *Gaussian processes for machine learning*. Springer, 2006

[Scholkopf und Smola 2002] Scholkopf, B. ; Smola, A.J.: *Learning with kernels*. Citeseer, 2002

[Steinbach u. a. 2004] Steinbach, M. ; Ertöz, L. ; Kumar, V.: The challenges of clustering high dimensional data. In: *New directions in statistical physics: econophysics, bioinformatics, and pattern recognition* (2004), S. 273

[Tadesse u. a. 2005] Tadesse, M.G. ; Sha, N. ; Vannucci, M.: Bayesian Variable Selection in Clustering High-Dimensional Data. In: *Journal of the American Statistical Association* 100 (2005), Nr. 470, S. 602–618

[Toropov 1992] Toropov, VV: Multipoint approximation method in optimization problems with expensive function values. In: *Proceedings of the 4th international symposium on Systems analysis and simulation table of contents* Elsevier Science Inc. New York, NY, USA, 1992, S. 207–212

A. Case Studies

This appendix contains the results of surrogate modeling for selected responses coming from different vehicle models and load cases.

5-fold cross validation is performed to avoid overfitting and obtain reliable prediction quality measures. Variable screening for regression (Rank correlation) as well as for classification cases (Random Forest derived sensitivities) is performed. A combination of both strategies is used to reduce risk of filtering important variables (i.e. Comb.) and additionally optimal subspace is sought through nonlinear sensitivity indexes (GP hyperparameters, i.e. Hyper.)

For the cases where regression does not perform well, classification models based on clustered data are gathered. The number of found clusters is appended to the name of the classifier. Negative COP values, as well as classification based on invalid clusters results are not removed from the tables.

Finally, prediction measures are stored for each model for comparison and selection before integration into the MDO problem formulation. Results are promising and induce the application of discontinuous or noisy responses with high classification rates ($CLASS > 0.85$) as constraints. For many of the listed Crash FEA based responses, nonlinear regression does not offer better results than linear methods in terms of COP. In spite of the low COP values, sometimes it is possible to find out relevant nonlinear trends which can be used to improve surrogate model quality by sampling sequentially in reduced spaces.

Response	Metamodel	Screening	Train/Test	Variables	COP	MSE	$CLASS$
	RSM (linear)	None	(100/100)	20	0.51	4.09	···
	RSM (linear)	Comb.	(100/100)	8	0.46	4.47	···
	GP	Comb.	(100/100)	8	0.47	4.39	···
	RF-reg	Comb.	(100/100)	8	0.38	5.16	···
	GP	Hyper.	(100/100)	4	0.53	3.91	···
y	RSM (quadratic)	None	(100/100)	20	0.46	4.46	···
	GP-class-cluster c:2	Comb.	(100/100)	5	···	···	0.95
	RF-class-cluster c:2	Comb.	(100/100)	5	···	···	0.88

Table 1: Metamodel summary for the discontinuous function

Response	Metamodel	Screening	Train/Test	Variables	COP	MSE	$CLASS$
	RSM (linear)	None	(100/100)	20	···	284.16	···
	RSM (linear)	Comb.	(100/100)	7	···	206.04	···
	GP	Comb.	(100/100)	7	1.00	0.01	···
	RF-reg	Comb.	(100/100)	7	0.30	141.42	···
	GP	Hyper.	(100/100)	2	1.00	0.01	···
y	RSM (quadratic)	None	(100/100)	20	···	268.14	···
	GP-class-cluster c:2	Comb.	(100/100)	5	···	···	0.00
	RF-class-cluster c:2	Comb.	(100/100)	5	···	···	0.00

Table 2: Metamodel summary for the high dimensional Ishigami function

Response	Metamodel	Screening	Train/Test	Variables	COP	MSE	$CLASS$
	RSM (linear)	None	(100/100)	20	···	215.33	···
	RSM (linear)	Comb.	(100/100)	10	0.07	179.86	···
	GP	Comb.	(100/100)	10	0.79	41.21	···
	RF-reg	Comb.	(100/100)	10	0.07	178.33	···
	GP	Hyper.	(100/100)	2	0.82	34.84	···
y	RSM (quadratic)	None	(100/100)	20	···	231.83	···
	GP-class-cluster c:2	Comb.	(100/100)	10	···	···	···
	RF-class-cluster c:2	Comb.	(100/100)	10	···	···	···

Table 3: Metamodel summary for the high dimensional Ishigami function with additional noise

Response	Metamodel	Screening	Train/Test	Variables	COP	MSE	$CLASS$
	RSM (linear)	None	(98/98)	41	0.15	43.93	···
	RSM (linear)	Comb.	(98/98)	14	0.48	26.76	···
	GP	Comb.	(98/98)	14	0.40	30.65	···
	RF-reg	Comb.	(98/98)	14	0.24	39.00	···
DC Dist PID LOW END	GP	Hyper.	(98/98)	8	0.47	27.20	···

continued on next page ...

... continued from previous page

Response	Metamodel	Screening	Train/Test	Variables	*COP*	*MSE*	*CLASS*
	RSM (quadratic)	None	(98/98)	41	0.05	48.77	···
	HDDA-cluster c:2	Comb.	(98/98)	17	···	···	···
	RF-class-cluster c:2	Comb.	(98/98)	17	···	···	···
DC Dist PID LOW MIN	RSM (linear)	None	(98/98)	41	0.64	47.19	···
	RSM (linear)	Comb.	(98/98)	10	0.71	38.41	···
	GP	Comb.	(98/98)	10	0.72	37.17	···
	RF-reg	Comb.	(98/98)	10	0.48	68.51	···
	GP	Hyper.	(98/98)	6	0.74	34.58	···
	RSM (quadratic)	None	(98/98)	41	0.60	53.20	···
	HDDA-cluster c:3	Comb.	(98/98)	10	···	···	0.71
	RF-class-cluster c:3	Comb.	(98/98)	10	···	···	0.70
DC Dist PID UPP END	RSM (linear)	None	(98/98)	41	0.07	50.75	···
	RSM (linear)	Comb.	(98/98)	13	0.52	26.34	···
	GP	Comb.	(98/98)	13	0.42	31.70	···
	RF-reg	Comb.	(98/98)	13	0.22	42.44	···
	GP	Hyper.	(98/98)	8	0.52	26.52	···
	RSM (quadratic)	None	(98/98)	41	···	63.32	···
	hdda-cluster c:2	Comb.	(98/98)	1	···	···	···
	RF-class-cluster c:2	Comb.	(98/98)	1	···	···	···
DC Dist PID UPP MIN	RSM (linear)	None	(98/98)	41	0.46	31.94	···
	RSM (linear)	Comb.	(98/98)	10	0.68	18.89	···
	GP	Comb.	(98/98)	10	0.69	18.27	···
	RF-reg	Comb.	(98/98)	10	0.49	30.23	···
	GP	Hyper.	(98/98)	6	0.70	18.14	···
	RSM (quadratic)	None	(98/98)	41	0.32	40.55	···
	HDDA-cluster c:2	Comb.	(98/98)	8	···	···	0.87
	RF-class-cluster c:2	Comb.	(98/98)	8	···	···	0.82
DC Intru RoofSideMember N4000026 Z rel mi107mm MAX	RSM (linear)	None	(98/98)	41	0.63	35.35	···
	RSM (linear)	Comb.	(98/98)	10	0.72	26.30	···
	GP	Comb.	(98/98)	10	0.70	28.20	···
	RF-reg	Comb.	(98/98)	10	0.48	49.00	···
	GP	Hyper.	(98/98)	7	0.75	23.73	···
	RSM (quadratic)	None	(98/98)	41	0.65	33.35	···
	hdda-cluster c:2	Comb.	(98/98)	1	···	···	···
	RF-class-cluster c:2	Comb.	(98/98)	1	···	···	···
DC Intru RoofSideMember N4000027 Z rel mi107mm MAX	RSM (linear)	None	(98/98)	41	0.67	41.03	···
	RSM (linear)	Comb.	(98/98)	10	0.80	25.08	···
	GP	Comb.	(98/98)	10	0.80	24.96	···
	RF-reg	Comb.	(98/98)	10	0.47	67.42	···
	GP	Hyper.	(98/98)	6	0.76	29.71	···
	RSM (quadratic)	None	(98/98)	41	0.70	37.30	···
	HDDA-cluster c:2	Comb.	(98/98)	5	···	···	···
	RF-class-cluster c:2	Comb.	(98/98)	5	···	···	···
DC Intru RoofSideMember N4000056 Z rel mi107mm MAX	RSM (linear)	None	(98/98)	41	0.78	31.38	···
	RSM (linear)	Comb.	(98/98)	11	0.86	20.15	···
	GP	Comb.	(98/98)	11	0.86	19.42	···
	RF-reg	Comb.	(98/98)	11	0.48	74.01	···
	GP	Hyper.	(98/98)	6	0.82	25.86	···
	RSM (quadratic)	None	(98/98)	41	0.78	31.78	···
	HDDA-cluster c:2	Comb.	(98/98)	4	···	···	···
	RF-class-cluster c:2	Comb.	(98/98)	4	···	···	···

Table 4: Metamodel summary in a roof impact test for a passenger car

Response	Metamodel	Screening	Train/Test	Variables	*COP*	*MSE*	*CLASS*
Occupant load criterion	RSM (linear)	None	(91/91)	46	0.70	61.52	···
	RSM (linear)	Comb.	(91/91)	9	0.75	50.29	···
	GP	Comb.	(91/91)	9	0.71	59.73	···
	RF-reg	Comb.	(91/91)	9	0.53	95.97	···

continued on next page ...

... *continued from previous page*

Response	Metamodel	Screening	Train/Test	Variables	*COP*	*MSE*	*CLASS*
	GP	Hyper.	(91/91)	5	0.70	61.81	···
	RSM (quadratic)	None	(91/91)	46	0.74	52.03	···
	HDDA-cluster c:2	Comb.	(91/91)	12	···	···	0.89
	RF-class-cluster c:2	Comb.	(91/91)	12	···	···	0.89
	RSM (linear)	None	(91/91)	46	0.35	5.73	···
	RSM (linear)	Comb.	(91/91)	13	0.58	3.72	···
	GP	Comb.	(91/91)	13	0.30	6.10	···
	RF-reg	Comb.	(91/91)	13	0.32	5.96	···
Door relative displacement	GP	Hyper.	(91/91)	7	0.40	5.22	···
	RSM (quadratic)	None	(91/91)	46	0.12	7.69	···
	HDDA-cluster c:2	Comb.	(91/91)	15	···	···	0.85
	RF-class-cluster c:2	Comb.	(91/91)	15	···	···	0.84
	RSM (linear)	None	(91/91)	46	0.67	0.55	···
	RSM (linear)	Comb.	(91/91)	10	0.76	0.40	···
	GP	Comb.	(91/91)	10	0.70	0.49	···
	RF-reg	Comb.	(91/91)	10	0.51	0.81	···
Frontal panel intrusion	GP	Hyper.	(91/91)	5	0.59	0.67	···
	RSM (quadratic)	None	(91/91)	46	0.75	0.41	···
	HDDA-cluster c:2	Comb.	(91/91)	13	···	···	0.89
	RF-class-cluster c:2	Comb.	(91/91)	13	···	···	0.89
	RSM (linear)	None	(91/91)	46	0.60	597.45	···
	RSM (linear)	Comb.	(91/91)	5	0.84	231.08	···
	GP	Comb.	(91/91)	5	0.91	135.86	···
	RF-reg	Comb.	(91/91)	5	0.63	553.14	···
Subframe rail lower displacement	GP	Hyper.	(91/91)	3	0.91	128.95	···
	RSM (quadratic)	None	(91/91)	46	0.55	667.09	···
	HDDA-cluster c:2	Comb.	(91/91)	5	···	···	0.90
	RF-class-cluster c:2	Comb.	(91/91)	5	···	···	0.89
	RSM (linear)	None	(91/91)	46	···	111.77	···
	RSM (linear)	Comb.	(91/91)	10	0.53	40.95	···
	GP	Comb.	(91/91)	10	0.25	64.84	···
	RF-reg	Comb.	(91/91)	10	0.33	58.37	···
Subframe rail upper displacement	GP	Hyper.	(91/91)	6	0.34	57.70	···
	RSM (quadratic)	None	(91/91)	46	···	146.55	···
	HDDA-cluster c:2	Comb.	(91/91)	4	···	···	0.90
	RF-class-cluster c:2	Comb.	(91/91)	4	···	···	0.85
	RSM (linear)	None	(91/91)	46	0.60	0.84	···
	RSM (linear)	Comb.	(91/91)	14	0.75	0.52	···
	GP	Comb.	(91/91)	14	0.69	0.66	···
	RF-reg	Comb.	(91/91)	14	0.40	1.26	···
Impact velocity	GP	Hyper.	(91/91)	8	0.59	0.86	···
	RSM (quadratic)	None	(91/91)	46	0.39	1.28	···
	HDDA-cluster c:2	Comb.	(91/91)	2	···	···	···
	RF-class-cluster c:2	Comb.	(91/91)	2	···	···	···

Table 5: Metamodel summary for a frontal crash in a passenger car

Response	Metamodel	Screening	Train/Test	Variables	*COP*	*MSE*	*CLASS*
	RSM (linear)	None	(48/48)	7	···	···	···
	RSM (linear)	Comb.	(48/48)	7	···	···	···
	GP	Comb.	(48/48)	7	···	···	···
Subframe rail endtime displacement	RF-reg	Comb.	(48/48)	7	0.01	···	···
	GP	Hyper.	(48/48)	5	···	···	···
	RSM (quadratic)	None	(48/48)	7	0.15	···	···
	hdda-cluster c:3	Comb.	(48/48)	1	···	···	···
	RF-class-cluster c:3	Comb.	(48/48)	1	···	···	···
Subframe rail maximum lateral displacement	RSM (linear)	None	(48/48)	7	···	···	···
	RSM (linear)	Comb.	(48/48)	5	0.11	997.01	···
	GP	Comb.	(48/48)	5	···	···	···

continued on next page ...

... continued from previous page

Response	Metamodel	Screening	Train/Test	Variables	*COP*	*MSE*	*CLASS*
	RF-reg	Comb.	(48/48)	5	0.17	919.77	···
	GP	Hyper.	(48/48)	4	···	···	···
	RSM (quadratic)	None	(48/48)	7	0.17	921.08	···
	hdda-cluster c:2	Comb.	(48/48)	1	···	···	···
	RF-class-cluster c:2	Comb.	(48/48)	1	···	···	0.92

Table 6: Metamodel summary in a frontal crash for a commercial vehicle

Response	Metamodel	Screening	Train/Test	Variables	*COP*	*MSE*	*CLASS*
Maximum relative displacement X 4105573	RSM (linear)	None	(196/196)	99	···	701.65	···
	RSM (linear)	Comb.	(196/196)	14	0.42	266.51	···
	GP	Comb.	(196/196)	14	0.36	291.72	···
	RF-reg	Comb.	(196/196)	14	0.32	312.54	···
	GP	Hyper.	(196/196)	7	0.37	286.63	···
	RSM (quadratic)	None	(196/196)	99	···	858.26	···
	HDDA-cluster c:2	Comb.	(196/196)	21	···	···	0.85
	RF-class-cluster c:2	Comb.	(196/196)	21	···	···	0.85
Maximum relative displacement X 504	RSM (linear)	None	(196/196)	99	···	228.37	···
	RSM (linear)	Comb.	(196/196)	11	0.34	68.34	···
	GP	Comb.	(196/196)	11	0.21	81.82	···
	RF-reg	Comb.	(196/196)	11	0.34	67.96	···
	GP	Hyper.	(196/196)	7	0.22	80.59	···
	RSM (quadratic)	None	(196/196)	99	···	285.01	···
	HDDA-cluster c:2	Comb.	(196/196)	9	···	···	0.90
	RF-class-cluster c:2	Comb.	(196/196)	9	···	···	0.92
Maximum relative displacement X 515	RSM (linear)	None	(196/196)	99	···	144.93	···
	RSM (linear)	Comb.	(196/196)	15	0.12	49.63	···
	GP	Comb.	(196/196)	15	···	74.89	···
	RF-reg	Comb.	(196/196)	15	0.12	49.77	···
	GP	Hyper.	(196/196)	9	0.06	53.16	···
	RSM (quadratic)	None	(196/196)	99	···	172.45	···
	HDDA-cluster c:2	Comb.	(196/196)	17	···	···	0.89
	RF-class-cluster c:2	Comb.	(196/196)	17	···	···	0.90
Minimum relative displacement Y 4105573	RSM (linear)	None	(196/196)	99	···	463.72	···
	RSM (linear)	Comb.	(196/196)	17	0.27	214.92	···
	GP	Comb.	(196/196)	17	0.03	286.07	···
	RF-reg	Comb.	(196/196)	17	0.22	228.95	···
	GP	Hyper.	(196/196)	10	0.13	255.68	···
	RSM (quadratic)	None	(196/196)	99	···	523.77	···
	HDDA-cluster c:2	Comb.	(196/196)	12	···	···	0.78
	RF-class-cluster c:2	Comb.	(196/196)	12	···	···	0.79
Maximum relative displacement Y 4105573	RSM (linear)	None	(196/196)	99	···	412.43	···
	RSM (linear)	Comb.	(196/196)	20	0.30	161.30	···
	GP	Comb.	(196/196)	20	0.04	221.16	···
	RF-reg	Comb.	(196/196)	20	0.27	167.82	···
	GP	Hyper.	(196/196)	7	0.24	175.78	···
	RSM (quadratic)	None	(196/196)	99	···	603.30	···
	HDDA-cluster c:2	Comb.	(196/196)	8	···	···	···
	RF-class-cluster c:2	Comb.	(196/196)	8	···	···	···
Minimum relative displacement Y 4105642	RSM (linear)	None	(196/196)	99	···	98.59	···
	RSM (linear)	Comb.	(196/196)	19	···	38.90	···
	GP	Comb.	(196/196)	19	···	47.99	···
	RF-reg	Comb.	(196/196)	19	0.08	35.52	···
	GP	Hyper.	(196/196)	10	···	44.84	···
	RSM (quadratic)	None	(196/196)	99	···	131.72	···
	HDDA-cluster c:2	Comb.	(196/196)	15	···	···	0.81
	RF-class-cluster c:2	Comb.	(196/196)	15	···	···	0.79
Maximum relative displacement Y 4105642	RSM (linear)	None	(196/196)	99	···	634.24	···

continued on next page ...

... *continued from previous page*

Response	Metamodel	Screening	Train/Test	Variables	COP	MSE	$CLASS$
	RSM (linear)	Comb.	(196/196)	18	0.24	182.42	$\cdots$
	GP	Comb.	(196/196)	18	0.44	135.56	$\cdots$
	RF-reg	Comb.	(196/196)	18	0.36	155.35	$\cdots$
	GP	Hyper.	(196/196)	4	0.52	114.88	$\cdots$
	RSM (quadratic)	None	(196/196)	99	$\cdots$	796.04	$\cdots$
	HDDA-cluster c:2	Comb.	(196/196)	19	$\cdots$	$\cdots$	$\cdots$
	RF-class-cluster c:2	Comb.	(196/196)	19	$\cdots$	$\cdots$	$\cdots$

Table 7: Metamodel summary in a frontal crash for a roadster model