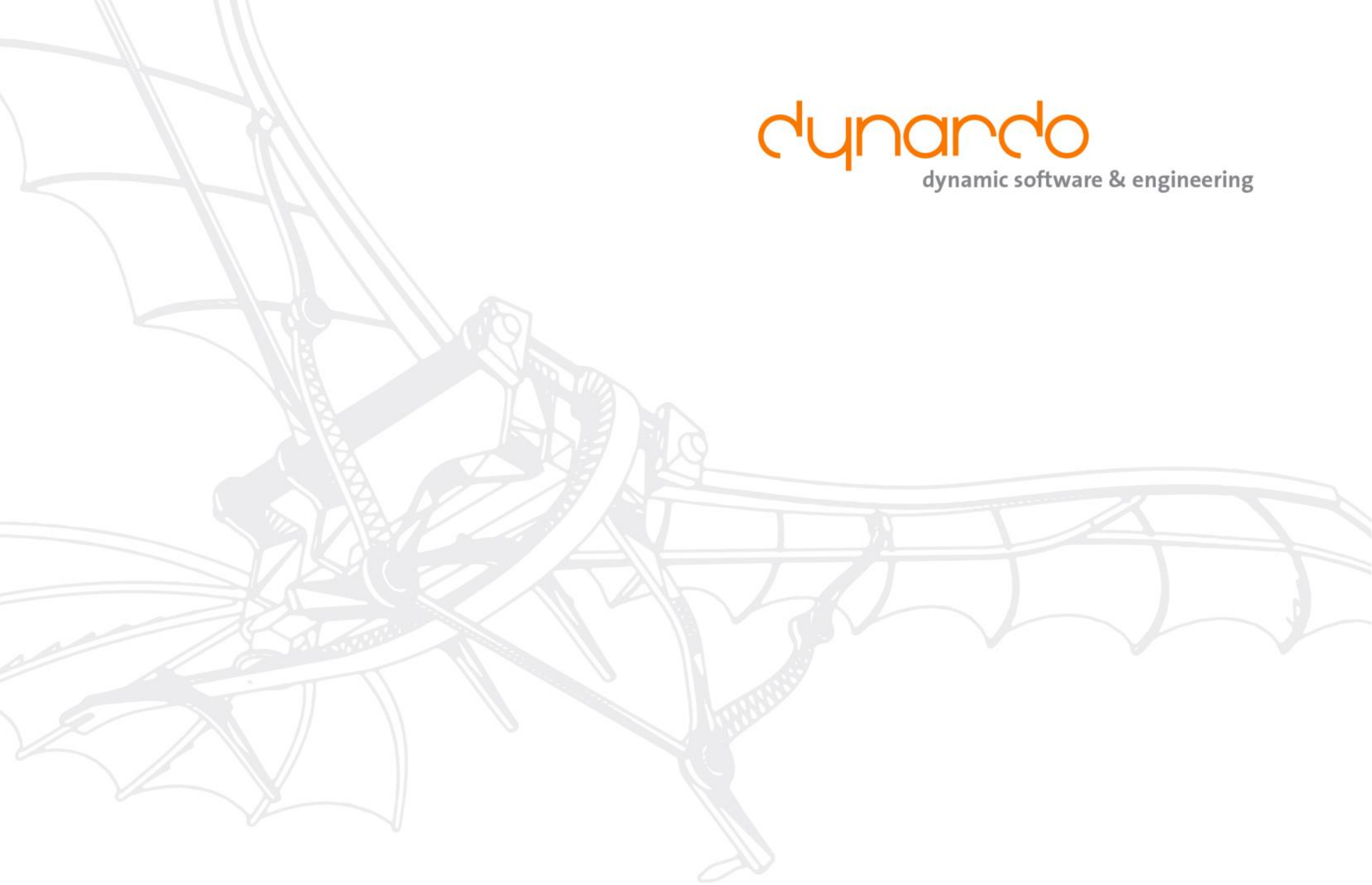

Lectures

Metamodell of Optimal Prognosis (MOP) - an Automatic Approach for User Friendly Parameter Optimization

Johannes Will & Thomas Most

presented at the Weimar Optimization and Stochastic Days 2009
Source: www.dynardo.de/en/library

Metamodell of optimized Prognosis (MoP) - an Automatic Approach for User Friendly Parameter Optimization

Johannes Will*, Thomas Most**

[1] *DYNARDO – Dynamic Software and Engineering GmbH, Weimar, Germany

**Research Training Group 1462, Bauhaus Universität Weimar, Germany

Abstract

In real case applications of CAE-based optimization tasks within the virtual prototyping process, every single numerical simulation may takes hours or even days. To perform CAE-based optimization, hence efficient surrogate models to replace the costly design runs would be an interesting alterative. Generally the available metamodel techniques show several advantages and disadvantages. Usually they are limited to a small number of optimization variables and the quality of prognosis is not known. In this paper we present an automatic approach for the selection of the optimal suitable metamodel as well as test the prognosis quality of the metamodel. We introduced the coefficient of prognosis (CoP) which enables an objective assessment of the metamodel prognosis based on an additional test data set. Therefore we call the selected metamodel with the best prognosis quality the metamodel of optimized prognosis (MoP). Together with an automatic reduction of the variable space using advanced filter techniques an efficient approximation is enabled also for high dimensional problems. We could verify the approach for several weakly and highly nonlinear examples with low and high dimensional input variable spaces. The approach can identify the required variables efficiently. After the generation of MoP the prognosis quality of important responses can be investigated and the MoP can be used for optimization purpose. Having all pieces together we set up a fully automatic, user friendly flow starting with a Latin Hypercube scan of the design space, generating the MoP and running optimization algorithms on MoP. After validating the “optima” from MoP at the real design space the user can decide to stop or enter further optimization procedures.

Keywords: surrogate models, regression analysis, variable reduction, Coefficient of prognosis (CoP), metamodel of optimized prognosis (MoP), optiSLang

* Contact: Dr.-Ing. Johannes Will, DYNARDO – Dynamic Software and Engineering GmbH, Luthergasse 1d, D-99423 Weimar, E-Mail: johannes.will [@] dynardo.de

1 Introduction

Meta modelling is one of the most popular strategies for design exploration within nonlinear optimization and stochastic analysis (see e.g. [1, 2, 3]). Moreover, the engineer has to calculate the general trend of physical phenomena or would like to re-use design experience on different projects. Due to the inherent complexity of many engineering problems it is quite alluring to approximate the problem and to solve other design configurations in a smooth sub-domain by applying a surrogate model ([4, 5]). Starting from a reduced number of simulations, a surrogate model of the original physical problem can be used to perform various possible design configurations without computing any further analyses. In one of our previous publications [13] we investigated several meta-model types and variable reduction techniques by means of different examples. In this previous paper we summarized that no universal approach exists and the optimal filter configurations can not be chosen generally. Therefore we developed an automatic approach for this purpose based on a library of available efficient meta-models and tools for variable reduction. This approach serves us based on a new measure for the approximation quality the Meta-model of Optimal Prognosis – the coefficient of Prognosis.

This paper is constructed as follows: first we present several meta-model approaches which are used later in our investigations. Then we introduce different filter techniques for variable reduction. Afterwards we present the framework of the meta-model selection and we finish this paper by validating the presented methodology by means of several numerical examples.

2 Metamodel approaches

2.1 Polynomial least square approximation

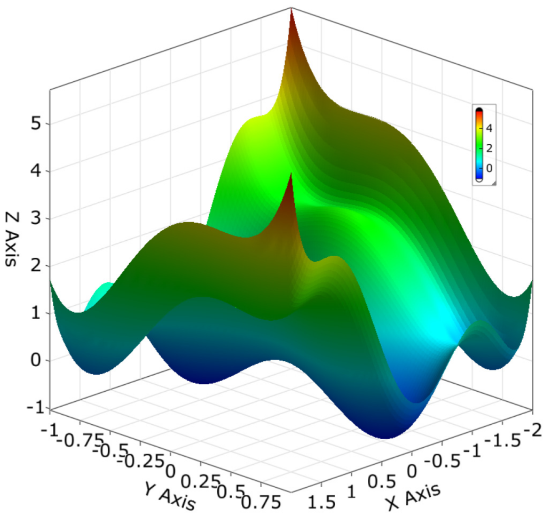

Figure 1: Original model response function $z(x, y)$

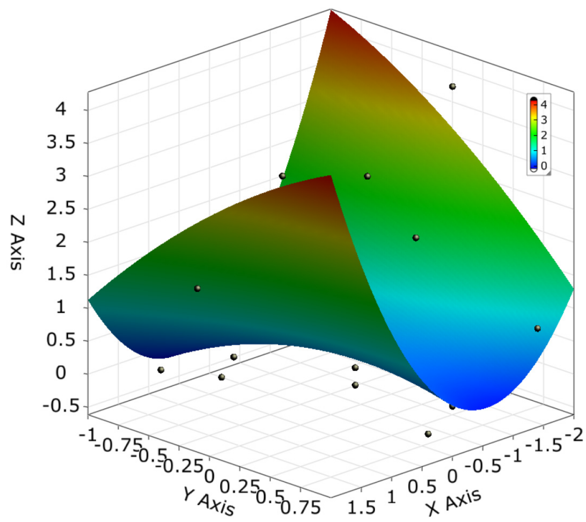

Figure 2: Quadratic polynomial least square approximation $\hat{z}(x, y)$

A commonly used approximation method of model responses, objectives, constraints and state functions

$$y(\mathbf{x}) \mapsto \hat{y}(\mathbf{x})$$

is the regression analysis. Usually, the approximation function is a first or second order polynomial [6, 7, 8] as shown in Figure 2. Based on the definition of the polynomial basis

$$\mathbf{p}^T(\mathbf{x}) = \left[1\ x_1\ x_2\ \ldots\ x_1^2\ x_2^2\ \ldots\ x_1 x_2\ \ldots\right] \tag{1}$$

the approximation function reads

$$y(\mathbf{x}) \approx \hat{y}(\mathbf{x}) = \mathbf{p}^T(\mathbf{x})\hat{\boldsymbol{\beta}}. \tag{2}$$

The approximate coefficients $\hat{\boldsymbol{\beta}}$ can be calculated as follows: using a defined number m of function values

$$\mathbf{y} = \left[y_1\ y_2\ \ldots\ y_m\right]^T \tag{3}$$

which can be approximated by the polynomial as

$$\mathbf{y} = \mathbf{P}^T\hat{\boldsymbol{\beta}} + \boldsymbol{\varepsilon} \tag{4}$$

where $\boldsymbol{\varepsilon}$ and $\mathbf{P}$ contain the approximation errors and the base polynomials of each support point, respectively, with

$$\boldsymbol{\varepsilon} = \left[\varepsilon_1\ \varepsilon_2\ \ldots\ \varepsilon_m\right]^T \tag{5}$$

and

$$\mathbf{P} = [\mathbf{p}(\mathbf{x}_1)\ \mathbf{p}(\mathbf{x}_2)\ \ldots\ \mathbf{p}(\mathbf{x}_m)]\,. \tag{6}$$

Together with the least square postulate

$$S = \sum_{k=1}^{m} \varepsilon_k^2 = \boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} \to \min \tag{7}$$

we obtain the following relation depending on the coefficients $\hat{\boldsymbol{\beta}}$

$$S(\hat{\boldsymbol{\beta}}) = (\mathbf{y} - \mathbf{P}^T\hat{\boldsymbol{\beta}})^T(\mathbf{y} - \mathbf{P}^T\hat{\boldsymbol{\beta}}) \to \min\,. \tag{8}$$

The solution of Eq. (8) yields to the well-known formulation

$$\hat{\boldsymbol{\beta}} = (\mathbf{P}\mathbf{P}^T)^{-1}\mathbf{P}\mathbf{y}. \tag{9}$$

Of course the accuracy of the approximation compared to the real problem has to be checked and verified. For reasonably smooth problems, the accuracy of response surface approximations improves as the number of points increases. However, this effect decreases with the degree of oversampling. An attractive advantage of the response surface methodology is the smoothing by approximating the sub problem. Especially for noisy problems like crash analysis, for which the catch of global trends is more important and the local noise may not be meaningful, a smoothing of the problem may be advantageous. However, linear and quadratic functions are possibly weak approximations near and far from certain support points. Using polynomials of higher than second order may only result in higher local accuracy with many sub-optima. Because of that in the last years, different advanced surrogate models have been developed to improve the accuracy and predictability of surrogate models.

2.2 Moving Least Squares approximation

The Moving Least Squares (MLS) approach was introduced by [9] and can be understood as an extension of the polynomial regression. Similarly the basis function can contain every type of function, but generally only linear and quadratic terms are used. This basis function can be represented exactly by obtaining the best local fit for the actual interpolation point. The approximation function is defined as

$$\hat{y}(\mathbf{x}) = \mathbf{p}^T(\mathbf{x})\mathbf{a}(\mathbf{x}) \tag{10}$$

with changing (``moving'') coefficients $\mathbf{a}(\mathbf{x})$ in contrast to the global coefficients of the polynomial regression.

Again the number of supporting points m exceeds the number of coefficients n, which leads to an over determined system of equations. This kind of optimization problem is solved by using the least squares approach

$$\mathbf{P}\hat{y}(\mathbf{x}) = \mathbf{P}\mathbf{P}^T\mathbf{a}(\mathbf{x}). \tag{11}$$

In order to obtain a local regression model in the MLS method distance depending weighting functions $w = w(s)$ have been introduced, where s is the standardized distance between the interpolation point and the considered supporting point

$$s_i = \frac{\|\mathbf{x} - \mathbf{x}_i\|}{D} \tag{12}$$

and D is the influence radius, which is defined as a numerical parameter. All types of functions can be used as weighting function $w(s)$ which have their maximum in $s = 0$ and vanish outside of the influence domain specified by $s = 1$. Mostly the well known Gaussian weighting function is used

$$w_{exp}(d) = exp\left(-\frac{s^2}{\alpha^2}\right) \tag{13}$$

where the definition of the influence radius D influences directly the approximation error. A suitable choice of this quantity enables an efficient smoothing of noisy data. In our work the influence radius D is chosen automatically based on an additional test data set.

The final approximation scheme reads

$$\hat{y}(\mathbf{x}) = \mathbf{p}^T(\mathbf{x})\mathbf{A}^{-1}(\mathbf{x})\mathbf{B}(\mathbf{x})\mathbf{y} \tag{14}$$

with

$$\begin{aligned} \mathbf{A}(\mathbf{x}) &= \mathbf{P}\mathbf{W}(\mathbf{x})\mathbf{P}^T \\ \mathbf{B}(\mathbf{x}) &= \mathbf{P}\mathbf{W}(\mathbf{x}) \\ \mathbf{W}(\mathbf{x}) &= diag(w(d_1), \ldots, w(d_m)) \end{aligned} \tag{15}$$

In Figure 3 and 4 the approximation functions for deterministic and noisy data with automatically determined D are shown.

In [10] a new weighting function was presented which enables the fulfilment of the MLS interpolation condition with high accuracy.

$$w_R(s_i) = \frac{\widetilde{w}_R(s_i)}{\sum_{j=1}^{m} \widetilde{w}_R(s_j)} \tag{16}$$

with

$$\widetilde{w}_R(s) = (s^2 + \varepsilon)^{-2}; \quad \varepsilon << 1 \,. \tag{17}$$

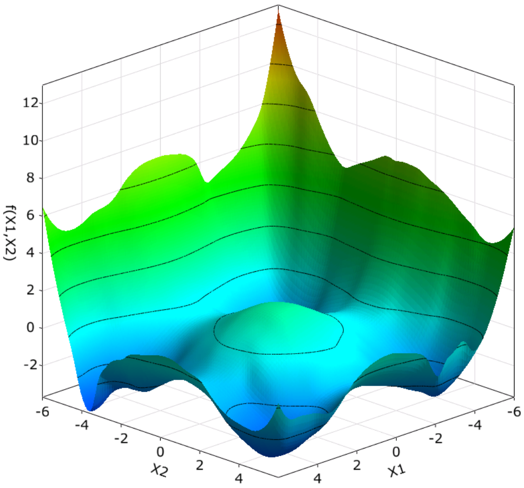

Figure 3: MLS approximation of deterministic data (exponential weighting)

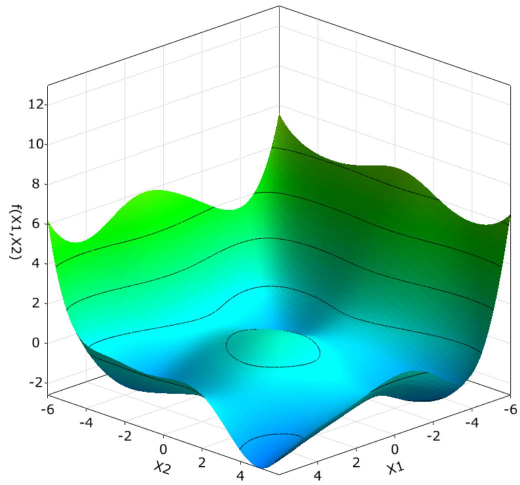

Figure 4: MLS approximation of noisy data (exponential weighting)

The regularization parameter ε has to be chosen small enough to fulfil the support point values with a certain accuracy. This approach is very suitable for problems where an interpolating meta-model is required.

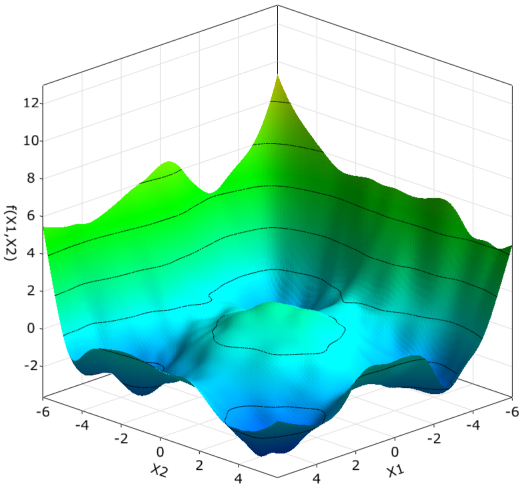

Figure 5: MLS interpolation of deterministic data (regularized weighting)

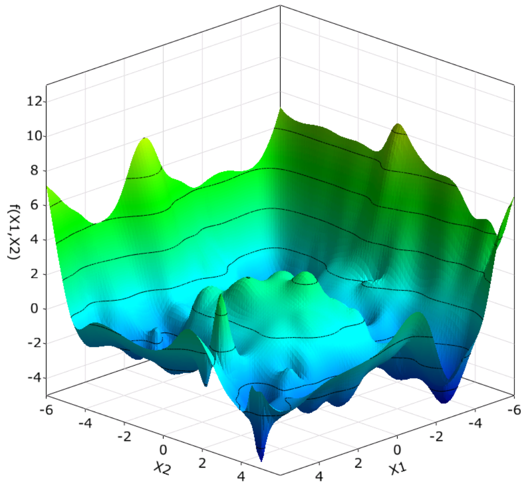

Figure 6: MLS interpolation of noisy data (regularized weighting)

For problems of noisy input data the noise is represented by the approximation function and thus the classical MLS approach with exponential weighting function is more suitable. In Figure 5 and 6 the interpolating MLS results are given for comparison. The main advantage of the MLS approach compared to the polynomial regression is the possibility to represent arbitrary complex nonlinear (but still continuous) functions. By increasing the number of support points the approximation function will always converge to the exact formulation.

2.3 Box-Cox transformation

The approximation quality of the presented meta-model approaches can be generally improved by a transformation of the response values which reduces nonlinear effects. A very suitable transformation for this purpose is the Box-Cox transformation proposed by [11] which is a family of power transformations covering a wide range of transformation functions. This transformation is defined based on a transformation parameter λ as

$$\hat{y}^{BC}(\lambda) = \begin{cases} \dfrac{\hat{y}^{\lambda} - 1}{\lambda \cdot \bar{y}^{\lambda-1}} & \lambda \neq 0 \\ \bar{y}\, ln\, \hat{y} & \lambda = 0 \end{cases}; \quad \bar{y} = \left(\prod_{i=1}^{m} \hat{y}_i \right)^{\frac{1}{m}} \tag{18}$$

where the scaling with the geometrical mean $\bar{y}$ of the m sample values is necessary to obtain comparable error values for different values of λ. In Figure 7 different transformation functions are shown. Based on the formulation in Eq. (18) we search for the optimal λ which minimizes the approximation error

$$E(\lambda) = \sum_{i=1}^{m} (y_i - \hat{y}_i^{BC}(\lambda))^2 \rightarrow min. \tag{19}$$

The approximation function $\hat{y}$ is carried out using polynomial regression on the whole data set. For the MLS approach the basis of this polynomial is assumed always as quadratic without mixed terms. Once the optimal value λ_{opt} is obtained the support point values are transformed and the meta-model can be build up. The approximations of the meta-model at the interpolation points have to be back-transformed afterwards. This back-transformation reads

$$\hat{y}(\mathbf{x}) = \begin{cases} (\hat{y}^{BC}(\mathbf{x}) \cdot \lambda_{opt} \cdot \bar{y}^{\lambda_{opt}-1} + 1)^{1/\lambda_{opt}} & \lambda_{opt} \neq 0 \\ exp\left(\dfrac{\hat{y}^{BC}(\mathbf{x})}{\bar{y}} \right) & \lambda_{opt} = 0 \end{cases} \tag{20}$$

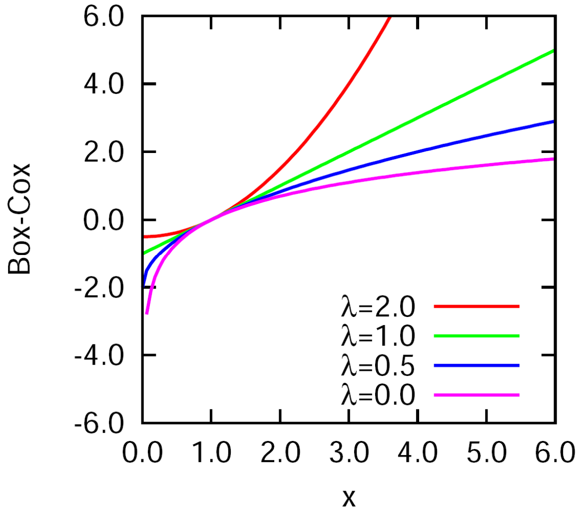

Figure 7: Box-Cox transformation for different values of λ

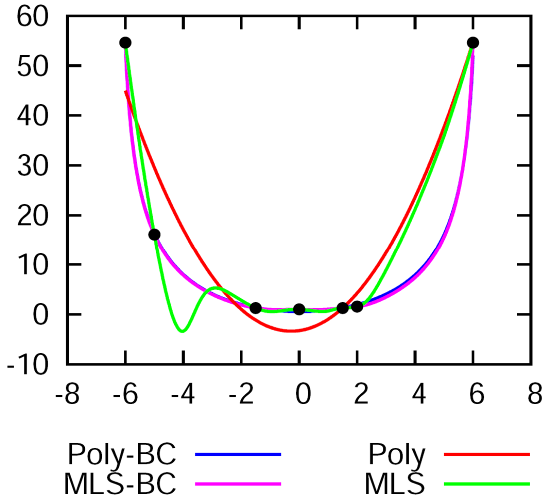

Figure 8: Improvement of an approximation function by the Box-Cox transformation

From Eq. (20) it can be seen, that for arbitrary $\lambda_{opt} \neq 0$ the back-transformation can only be performed if

$$\hat{y}^{BC}(\mathbf{x}) \cdot \lambda_{opt} \cdot \bar{y}^{\lambda_{opt}-1} + 1 \geq 0 \quad (21)$$

this can not be assured for every type of meta-model and approximated function. Thus we define the reciprocal of λ to be a natural number, and then the back-transformation can always be performed. This limits the set of possible values for λ to the following

$$\lambda \in [-1, -\frac{1}{2}, -\frac{1}{3}, -\frac{1}{4}, -\frac{1}{5}, \ldots 0, \ldots \frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{1}{2}, 1]. \quad (22)$$

In Figure 8 the MLS and polynomial approximation functions are shown for a simple example. The figure indicates that on the one hand the approximation error in the support points is reduced due to the transformation and on the other hand the smoothness of the approximation function is increased which is a quite useful property for a later analysis on the approximation function.

3 Variable reduction

3.1 Significance filter

Various statistical analysis procedures are available for the subsequent evaluation of correlation of input parameters and the responses. For example, the coefficients of correlation ρ_{ik} are calculated from all pair wise combinations of both input variables and response according to:

$$\rho_{ij} = \frac{1}{N-1} \frac{\sum_{k=1}^{N} \left(x_i^{(k)} - \mu_{x_i}\right)\left(x_j^{(k)} - \mu_{x_j}\right)}{\sigma_{x_i}\sigma_{x_j}} \tag{23}$$

The quantity ρ_{ij}, called the linear correlation coefficient, measures the strength and the direction of a linear relationship between two variables. The linear correlation coefficient is sometimes referred to as the Pearson product moment correlation coefficient. The quadratic coefficients of correlation

$$\rho_{ij} = \frac{1}{N-1} \frac{\sum_{k=1}^{N} \left(\hat{y}^{(k)}(x_i) - \mu_{\hat{y}(x_i)}\right)\left(x_j^{(k)} - \mu_{x_j}\right)}{\sigma_{\hat{y}(x_i)}\sigma_{x_j}} \tag{24}$$

is defined as the linear coefficient of correlation (see Equation (23)) between the least-squares fit of a quadratic regression $\hat{y}(x_i)$ of the variable x_j on the samples $x_i^{(k)}, x_j^{(k)}$. A correlation greater than 0.7 is generally described as strong, whereas a correlation less than 0.3 is generally described as weak. These values can vary based upon the type of data being examined. All pair wise combinations (i, j) can be assembled into a correlation matrix C_{XX}, as shown in Figure 9.

The computed correlation coefficients between the input variables vary from the assumed values depending on the sample size. This deviation is used to judge in a first step, which variables are significant concerning there influence on the output variables. We define an error quantile which is chosen between 90% and 99% and compute the corresponding correlation error in the input-input correlations. This is done for linear and quadratic correlations simultaneously. Based on these quantile values we assume only these input variables to be significant concerning an output variable if there correlation values are above the given error values. For the most practical cases this leads to a reduced number of input variables which is shown in Figure 10. All values in gray are assumed to be insignificant.

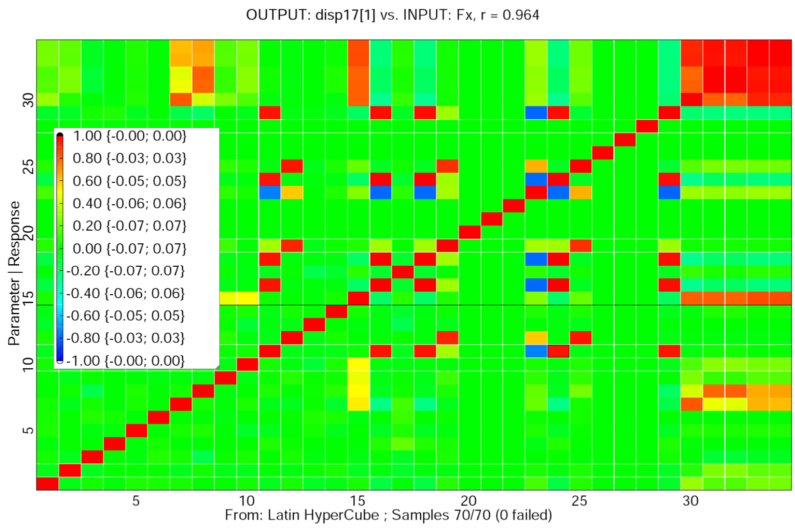

Figure 9: Matrix C_{XX} of the linear correlation coefficients

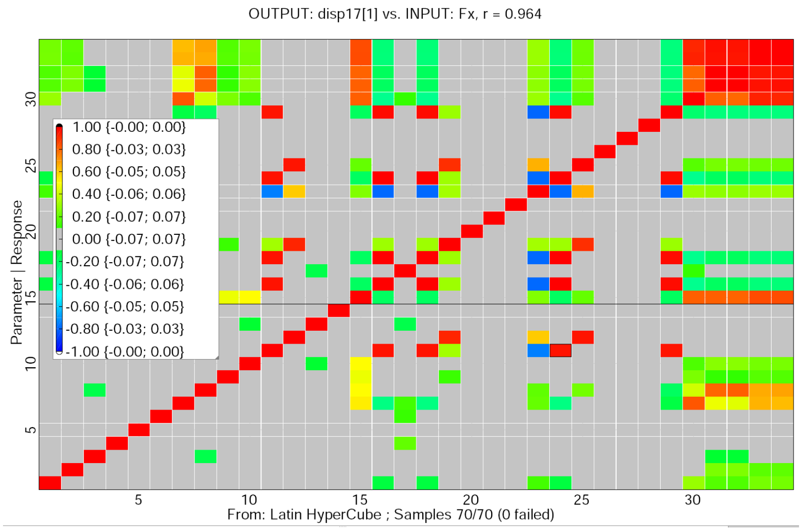

Figure 10: Matrix C_{XX} of the most significance linear correlation coefficients

3.2 Importance filter

Generally the remaining variable set still contains variables which are not needed for an approximation. With the importance filter we identify the important variables for the approximation model as described as follows: Based on a polynomial regression using the remaining variables of the significance filter we estimate the quality of the model representation by the coefficient of determination (CoD):

$$R^2 = \frac{\sum_{k=1}^{N} (\hat{y}(k) - \mu_y)^2}{\sum_{k=1}^{N} (y(k) - \mu_y)^2}. \tag{25}$$

In order reduce the influence of an increasing number of variables the adjusted coefficient of determination was introduced

$$R^2_{adj} = 1 - \frac{N-1}{N-p}\left(1 - R^2\right) \tag{26}$$

where p is the number of coefficients used in the polynomial regression. Based on this quantity the influence of each variable is studied by leaving this variable out of the regression model and compute the modified CoD. The difference between the CoD of the full and the reduced model is defined as the coefficient of importance introduced by [12]. This coefficient of importance (CoI) reads for the variable i

$$CoI_i = CoD(X_1, \ldots, X_n) - CoD(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n). \tag{27}$$

Based on a given value of the minimum required CoI_{min} only the variables having

$$CoI_i \geq CoI_{min} \tag{28}$$

are considered in the final approximation. Generally the value CoI_{min} is taken between 1% and 9%.

4 Meta-model of Optimal Prognosis (MoP)

4.1 Coefficient of prognosis

The selection of the optimal filter configuration and the best suitable meta-model for a specific problem is difficult as shown in [13]. In order to develop an automatic approach we need to define a measure for the characterization of the approximation quality. For this purpose we use the generalized coefficient of determination

$$R^2 = \left(\frac{\mathbf{E}[Y \cdot \hat{Y}]}{\sigma_Y \sigma_{\hat{Y}}} \right)^2 = \left(\frac{\sum_{k=1}^{N} \left(y^{(k)} - \mu_y\right)\left(\hat{y}^{(k)} - \mu_{\hat{y}}\right)}{(N-1)\sigma_y \sigma_{\hat{y}}} \right)^2 \qquad (29)$$

which results for the special case of pure polynomial regression in the formulation of Eq. (25). The generalized CoD is applicable for all types of meta-models and is equivalent to the square of the linear correlation coefficient between the true sample values and the model predictions. In order to judge the quality of an approximation we have to evaluate the prognosis quality. For this purpose we use an additional test data set. The agreement between this real test data and the meta-model estimates is measured by the so-called coefficient of prognosis CoP

$$CoP = \left(\frac{\mathbf{E}[Y_{Test} \cdot \hat{Y}_{Test}]}{\sigma_{Y_{Test}} \sigma_{\hat{Y}_{Test}}} \right)^2 ; \quad 0 \leq CoP \leq 1. \qquad (30)$$

The advantage of the CoP compared to other existing error measures, for example the mean squared error, is the automatic scaling of the CoP, where we can derive that for example a CoP equal to 0.8 is equivalent to a meta-model prediction quality of 80% for new data points.

4.2 Determination of the optimal meta-model

Based on the definition of the coefficient of prognosis we can derive the optimal meta-model with corresponding variable space as follows: For each meta-model type we investigate all possible significance and filter configurations by varying the significance quantile from 99% down to a given minimal value. Then a polynomial regression is built up and the coefficients of importance are calculated for each variable. The threshold CoI_{min} is varied from 0.01 to a given value and based on the remaining variables the meta-model is built up and the coefficient of prognosis is computed. The configuration with the maximum CoP is finally taken as optimal meta-model with corresponding variable space for each approximated response quantity. While for the meta-model construction the training data set is used for the meta-model itself and the test data set for the calculation of the CoP the correlations for the significance filter and the regression for the importance filters are obtained by using the merged data set from training and test data.

If no additional test data set is available the initial data set is split into training and test data. The samples are selected in that way that in each data set the response ranges are represented with maximum conformity to the entire data set.

4.3 Ranking of variable importance

Similar to the CoI we want to formulate a measure for the importance of a single variable on the overall prognosis quality, the CoP. For this purpose we utilize variance based sensitivity indices [14] on the optimal meta-model approximation

$$CoP_i = CoP \cdot S_{Ti}^{MOP}, \qquad S_{Ti}^{MOP} = 1 - V^{MOP}(Y|X_i)/V^{MOP}(Y) \tag{31}$$

where the sensitivity indices considering total effects S_{Ti}^{MOP} are evaluated using the conditional variances $V^{MOP}(Y|X_i)$ on the optimal meta-model using the original input variable distributions. If we sum up all single CoP_i we should end up with the total CoP for a purely additive model. Higher values indicate interaction terms between input variables.

5 Numerical examples

5.1 Weakly nonlinear problem

In this example we investigate a weakly nonlinear problem with 50 input variables. For our procedure eight sets of 100 Latin Hypercube samples are available. We use 100, 200, 300 and 700 samples as training data and one 100 sample set as test data. For the cases of 100, 200, 300 training samples we use the still available 400 samples as verification data. From a number of several response quantities we have chosen two representatives. In Figure 11 the anthill-plots for both responses depending on the variable with the highest influence is shown. The figure indicates, that for both response the influence is almost linear. In Table 1 the obtained results for the optimal meta-model and variable space is given.

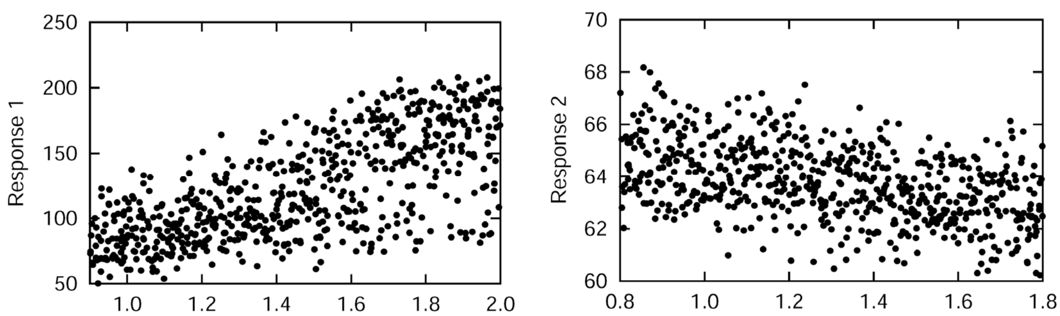

Figure 11: Anthill-plots for the weakly nonlinear problem

The results clearly indicate that an excellent prognosis quality can be obtained by the optimal meta-model and the estimated prognosis from the CoP is close to this of the verification data set. For the first response the variable space can be dramatically reduced and for the second case the most variables remain in the final space. In this example the benefit of a sophisticated meta-model as MLS is not huge compared to classical polynomial regression but still remarkable. Another interesting result is that with increasing number of training samples the optimal basis order increases. Furthermore we observe for the second response that the CoD of the training data set may judge the approximation as a very good regression but the prognosis for new data is of less quality. This indicates the requirement of a measure like the CoP for a realistic assessment of the prognosis quality of the meta-model.

In Figure 12 a three-dimensional plot of both responses depending on the two most important input variables is shown. A good agreement of approximation and training and test data can be recognized for the first response due to the small number of remaining variables. For the second response where the final variable number is much larger the subspace-plot does not provide very much information. However, the weakly nonlinear behaviour of the response function and the agreement of approximation and available samples are apparent.

	Response 1				Response 2			
	Number of training samples				Number of training samples			
	100	200	300	700	100	200	300	700
Polynomial								
No. variables	4	5	3	3	25	31	32	38
Optimal basis	linear	quadr.	quadr.	quadr.	linear	linear	quadr.	quadr.
CoD Training	0.903	0.899	0.901	0.900	0.879	0.820	0.852	0.835
CoP Test	0.899	0.896	0.902	0.905	0.786	0.815	0.831	0.834
CoP Verification	0.894	0.892	0.897	-	0.691	0.738	0.765	-
MLS								
No. variables	3	4	4	4	25	31	32	32
Optimal basis	linear	linear	quadr.	quadr.	linear	linear	quadr.	quadr.
CoD Training	0.970	0.977	0.996	0.989	0.879	0.918	0.886	0.917
CoP Test	0.920	0.929	0.956	0.969	0.786	0.818	0.832	0.841
CoP Verification	0.927	0.917	0.946	-	0.658	0.735	0.768	-
CoD_{full}	0.900	0.894	0.899	0.893	0.860	0.819	0.825	0.814

Table 1: Results for the weakly nonlinear example with 50 input variables

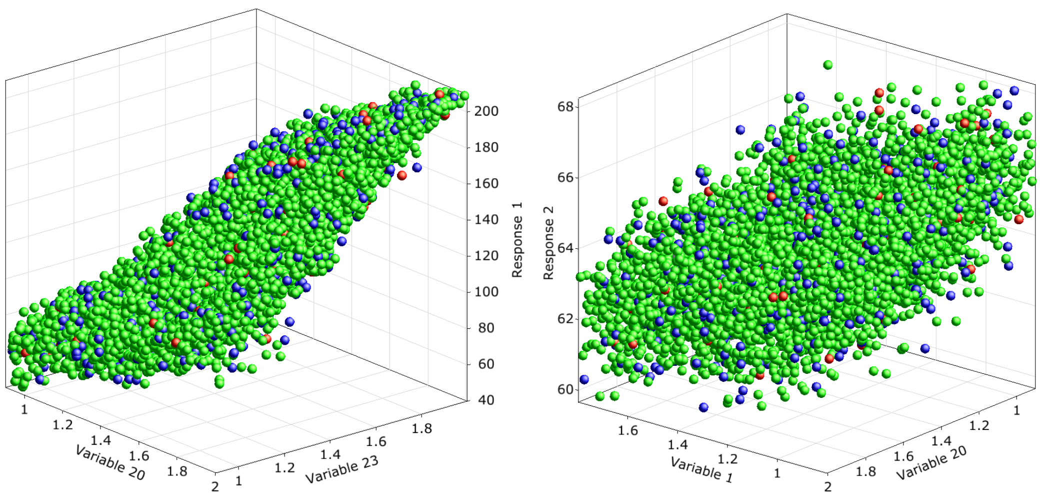

Figure 12: Training (blue) and test (red) data and MLS approximation (green) of the weakly nonlinear problem for both responses depending on the two most significant variables

5.2 High dimensional instability problem

In this example a high dimensional problem is investigated where the global trend is disturbed by an instability. This problem has 100 input variables and three sets of 100 Latin Hypercube sampling are available. Figure 13 shows the anthill-plots for the two most important variables. The figure indicates that the response shows a strongly nonlinear behaviour depending on variable 99. We have applied our approach by using 200 samples as training data and the remaining 100 samples as test data. Based on a polynomial regression the *CoD* of the model used for the importance filter is quite small as indicated in Table 2. The *CoP* of the test data set confirms this results which shows that this problem can not be represented sufficiently by a classical polynomial. If more sophisticated metamodels are used the MoP approach results in significant better results detecting two variables in final variable space. These results and the belonging approximation function are given additionally in Figure 2. For this example it is quite clear that more complex metamodels are not only necessary to improve the approximation quality but also to detect the important variables correctly.

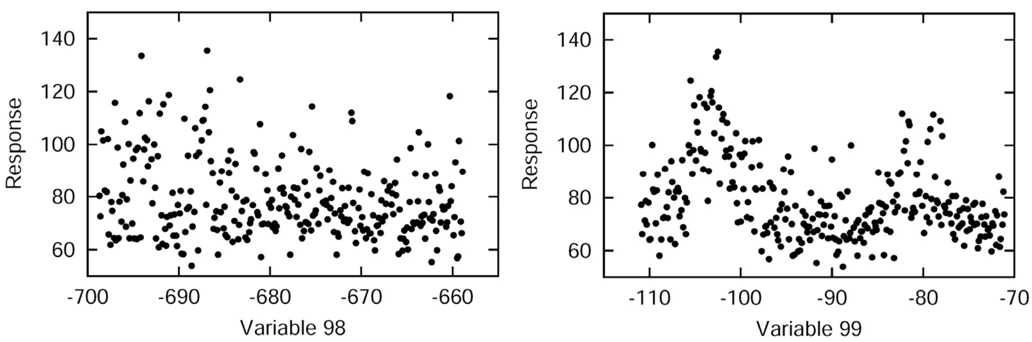

Figure 13: Anthill-plots for the high dimensional instability problem

	Polynomial	MLS
No. variables	13	2
CoD Training	0.289	0.568
CoP Test	0.205	0.462
CoD_{full}	0.312	0.312

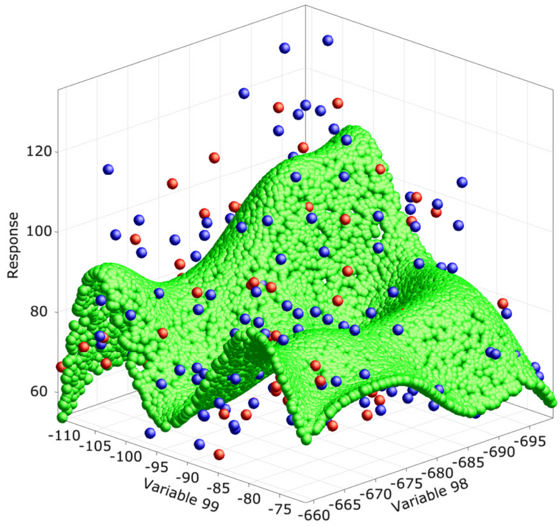

Table 2: Three-dimensional plot of the training (blue) and test (red) data and MLS approximation (green) for the high dimensional instability problem and corresponding numerical results

5.3 Low dimensional instability problem

This example is quite similar to the previous one but the investigated initial variable space consists of 25 input variables which are much less as before. Again an instability leads to a highly nonlinear dependence of the response on the input variables. For our investigation only 100 Latin Hypercube samples are available which are split into the training and test data set by using 50/50, 70/30 and 80/20 percentage fractions.

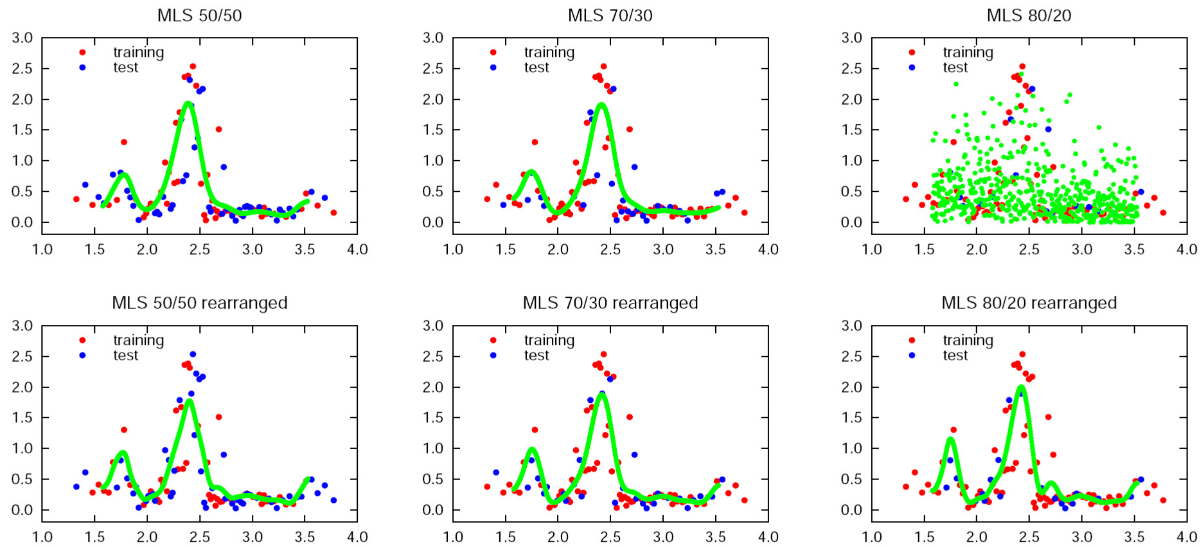

Figure 14: Anthill-plot of the response depending on the most important variable 19 with corresponding optimal approximation for different sample splitting

Sample splitting		50/50	70/30	80/20
MLS	No. variables	1	1	9
	CoD Training	0.735	0.753	0.929
	CoP Test	0.587	0.462	0.668
MLS	No. variables	1	1	1
(rearranged samples)	CoD Training	0.705	0.692	0.705
	CoP Test	0.543	0.690	0.760
	CoD_{full}	0.273	0.273	0.273

Table 3: Results for the low dimensional instability problem for different sample splitting from two different arrangements of the entire data set

The results are shown in Figure 14 and Table 3. The figures indicate a similar approximation function for the two investigated sample arrangements with 50/50 and 70/30 sample splitting. For the first arrangement with 80/20 splitting the test data are not suitable to lead to the optimal approximation model. The CoP values shown significant deviations for almost similar approximation functions as shown in Figure 14. This is also a result of the small number of test samples. This clarifies that a certain amount of test samples is require for a stable application of the MoP approach. Nevertheless for this example the approach serves a very good approximation of the nonlinear problem.

5.4 Very high dimensional problem

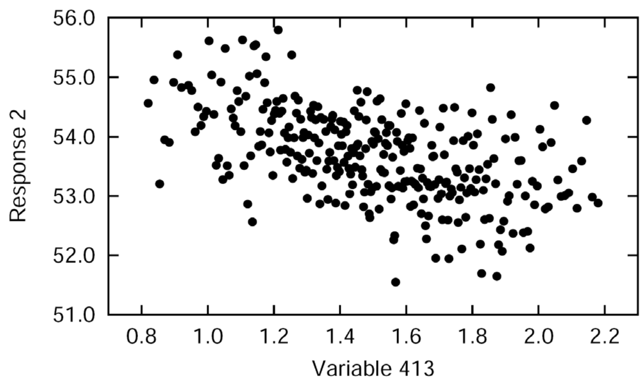

Figure 15: Anthill-plot for the very high dimensional problem

In the final example we investigate a very high dimensional problem with 500 input variables. 500 Latin Hypercube samples are used with a 80/20 splitting for the training and test data. Three different responses are investigated with the MoP approach. In Figure 15 the second response is shown depending on the most important variable. The MoP approach leads for all responses to a remarkable reduction of the variable space and a good prognosis of the optimal meta-model. The results in Table 4 show only a small difference between the optimal polynomial and MLS approximations which indicates a weak nonlinearity in the MoP. Clearly the number of support points in combination with the number of important variables restrict the ability of the MoP to represent response function non linearity.

		Response 1	Response 2	Response 3
Polynomial	No. variables	33	11	26
	CoD Training	0.872	0.888	0.872
	CoP Test	0.642	0.831	0.791
MLS	No. variables	33	11	26
	CoD Training	0.872	0.927	0.984
	CoP Test	0.642	0.836	0.811
	CoD_{full}	0.819	0.895	0.904

Table 4: Results for the very high dimensional problem

6 Implementation of CoP/MoP in optiSLang

Since version 3.1 the CoP/MoP approach is included in the commercial software optiSLang [15]. There is a flow created which can evaluate CoP/MoP for any set of samples or test data. The user can modify the data split, the metamodel type which will be evaluated, the different filter settings and can force to reduce (Delta CoP) the number of important variables in the final MoP model.

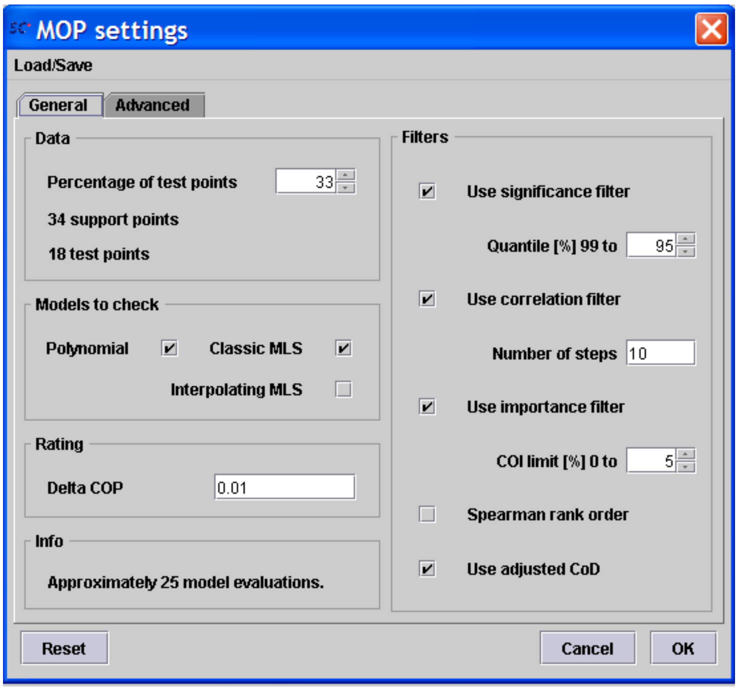

Figure 16: Settings of the MoP generation flow in optiSLang

Since it is available in optiSLang the generation of the metamodel of optimized prognosis and the calculation of coefficient of Prognosis was successfully applied at several problem types. The following example shows a noise non-linear problem having 8 optimization variables and 200 samples. Running traditional correlation analysis using Spearman ranked data two important variables could be identified and a CoI of 73% (Fig. 17) for the full model was measured. Running CoP/MoP we also find the two important variables, but with very good representation of the nonlinear response function we can achieve a CoP of 99% (Fig. 17/18).

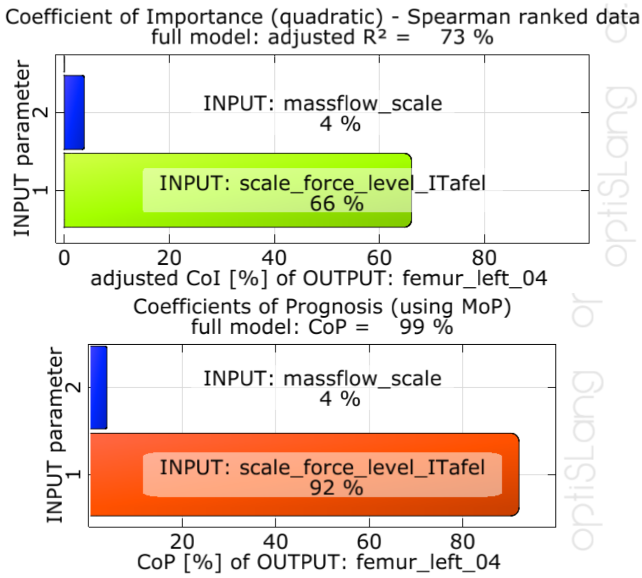

Figure 17: left Coefficient of Importance using traditional correlation analysis, right: Coefficient of Prognosis using new CoP/MoP approach

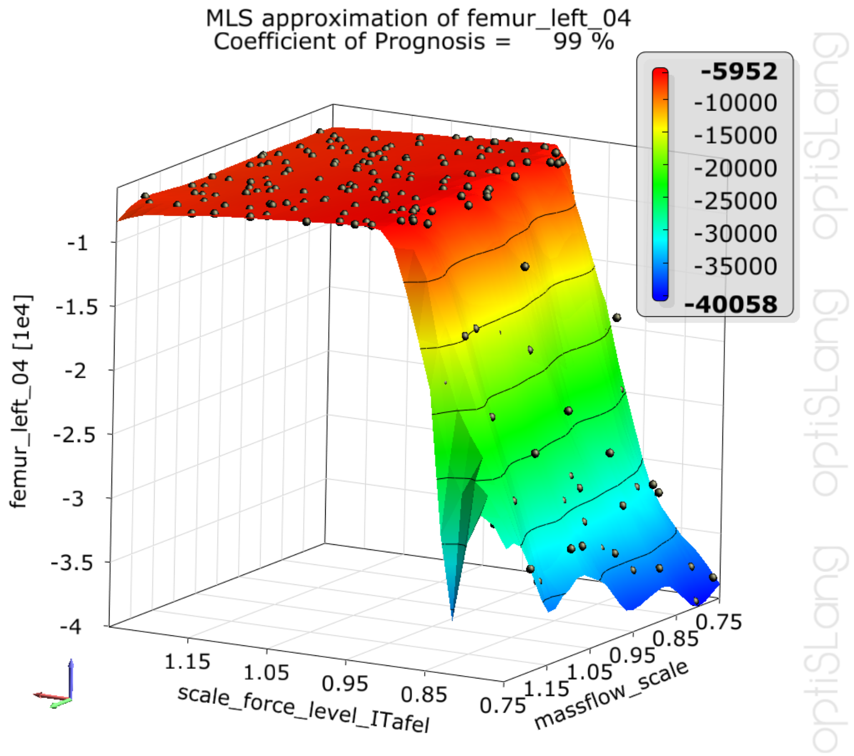

Figure 18: Vizualisation of the MoP, black dots represent the sample set (regression and test data set)

From our experience so fare we can state and recommend:

- Compare CoI of correlation analysis and CoP, the differences between the two should be verified.
- Check plausibility and prognosis quality by plotting the MoP with the two most important input variables.
- If the CoP/MoP approach can reduce the set of important to a very small number (<5) very good representation of nonlinearities are achives even with small number of samples (100).
- If the CoP/MoP approach cannot reduce the set of important input parameters smaller than 10 ..20 the sample set must be large to represent nonlinear correlation.
- The CoP measurement of the full model is more reliable than the CoI measurement of the full model.

7 Conclusion

In this paper we presented an approach for an automatic selection of the optimal metamodel for the investigated problem. We introduced the coefficient of prognosis which enables an objective assessment of the metamodel prognosis based on an additional test data set. We could verify the approach for several weakly and highly nonlinear examples with low and high dimensional input variable spaces.

The approach can identify the required variables efficiently and the obtained optimal metamodel can be used afterwards for a fast optimization. The only restriction is a required minimum number of test samples to represent the variable space sufficiently. In our future work we will address following improvements:

- include cross correlation techniques to become independent of the splitting of test data set
- improve CoP measurement of single variable importance in case of small/medium number of sample points
- implement adaptive sampling strategies

After generation of MoP the next step will be to use the MoP for optimization purpose within optiSLang. Then a black box algorithm combining high end sensitivity study and optimization will be available for medium and high dimensional non linear problems.

References

[1] BOOKER, A. J., DENNIS, J. E., JR., FRANK, P. D., SERAFINI, D. B., TORCZON, V. AND TROSSET, M. A.: "RIGOROUS FRAMEWORK FOR OPTIMIZATION OF EXPENSIVE FUNCTIONS BY SURROGATES." STRUCTURAL OPTIMIZATION, 17(1):1 - 13, 1999.

[2] GIUNTA, A. AND WATSON, L. T. A.: "COMPARISON OF APPROXIMATION MODELING TECHNIQUE: POLYNOMIAL VERSUS INTERPOLATING MODELS." 7TH AIAA/USAF/NASA/ISSMO SYMPOSIUM ON MULTIDISCIPLINARY ANALYSIS & OPTIMIZATION, PAGES 392 - 404. AIAA, ST. LOUIS, MO, 1998.

[3] SIMPSON, T. W., BOOKER, A. J., GHOSH, S., GIUNTA, A., KOCH, P. N. AND YANG, R. J.: "APPROXIMATION METHODS IN MULTIDISCIPLINARY ANALYSIS AND OPTIMIZATION: A PANEL DISCUSSION." STRUCTURAL AND MULTIDISCIPLINARY OPTIMIZATION, 2003.

[4] SACKS, J., WELCH, W. J., MITCHELL, T. J. AND WYNN, H. P.: "DESIGN AND ANALYSIS OF COMPUTER EXPERIMENTS." STATISTICAL SCIENCE, 4(4):409 - 435, 1989.

[5] SIMPSON, T. W., PEPLINSKI, J., KOCH, P. N. AND ALLEN, J. K.: "METAMODELS FOR COMPUTER-BASED ENGINEERING DESIGN: SURVEY AND RECOMMENDATIONS." ENGINEERING WITH COMPUTERS, 17(2):129 - 150, 2001.

[6] BOX, G. E. P. AND DRAPER, N. R.: "EMPIRICAL MODEL BUILDING AND RESPONSE SURFACES." JOHN WILEY & SONS, NEW YORK, USA, 1987.

[7] MYERS, R. H.: "RESPONSE SURFACE METHODOLOGY." ALLYN AND BACON INC., BOSTON, USA, 1971.

[8] MYERS, R. H. AND MONTGOMERY, D. C.: "RESPONSE SURFACE METHODOLOGY - PROCESS AND PRODUCT OPTIMIZATION USING DESIGNED EXPERIMENTS." JOHN WILEY & SONS, INC., NEW YORK, 1995.

[9] LANCASTER, P. AND SALKAUSKAS, K.: "SURFACE GENERATED BY MOVING LEAST SQUARES METHODS." MATHEMATICS OF COMPUTATION, 37:141-158, 1981.

[10] MOST, T. AND BUCHER, C.: "A MOVING LEAST SQUARES WEIGHTING FUNCTION FOR THE ELEMENT-FREE GALERKIN METHOD WHICH ALMOST FULFILLS ESSENTIAL BOUNDARY CONDITIONS." STRUCTURAL ENGINEERING AND MECHANICS, 21(3):315-332, 2005.

[11] BOX, G. E. P. AND COX, D. R.: "AN ANALYSIS OF TRANSFORMATIONS." JOURNAL OF THE ROYAL STATISTICAL SOCIETY. SERIES B (METHODOLOGICAL), 26:211-252, 1964.

[12] BUCHER, C.: "BASIC CONCEPTS FOR ROBUSTNESS EVALUATION USING STOCHASTIC ANALYSIS." IN K.-U. BLETZINGER AND OTHERS, EDITORS, EFFICIENT METHODS FOR ROBUST DESIGN AND OPTIMISATION - EUROMECH COLLOQUIUM 482, 2007. 2007.

[13] ROOS, D., MOST, T., UNGER, J. F. AND WILL J.: "ADVANCED SURROGATE MODELS WITHIN THE ROBUSTNESS EVALUATION." PROCEEDINGS OF THE WEIMARER OPTIMIERUNGS- UND STOCHASTIKTAGE 4.0, WEIMAR, GERMANY, NOVEMBER 29-30. 2007.

[14] SALTELLI, A. ET AL.: "GLOBAL SENSITIVITY ANALYSIS. THE PRIMER." JOHN WILEY & SONS, LTD, CHICHESTER, ENGLAND, 2008

[15] OPTISLANG - THE OPTIMIZING STRUCTURAL LANGUAGE, VERSION 3.1, DYNARDO, WEIMAR, 2009, WWW.DYNARDO.COM