Approximation Techniques for Response Surface Method

JANIS AUZINS and ROLANDS RIKARDS

Institute of Materials and Structures, Riga Technical University,

Kalku St. 1, LV-1658, Riga, LATVIA

email: auzinsjp@latnet.lv, rikards@latnet.lv

Abstract. For metamodel building the approximation method on the basis of orthogonal multivariate polynomials of discrete variable is proposed. Latin hypercube type experimental designs with minimal mean distance between regular *s*-dimensional grid points and experimental points are used. The best functions are selected using augmented dataset, obtained by local approximation. Optimal number of terms in regression function is chosen by two-stage holdout method. Ten artificial and one practical problem with 2 to 8 input variables are tested. Results for artificial problems are compared with the results obtained by other authors using Radial Basis Functions, Kriging, Multivariate Adaptive Regression Splines, Quasi-regression and local polynomial approximation. In most cases the proposed method gives predicted accuracy comparable with the best results obtained by other methods.

Keywords: response surface, approximation, experimental design, orthogonal polynomials

1. Introduction

In structural optimization the computational time for some problems is too high to use conventional methods of minimization. For example, it can take several hours to get a finite element solution of one variant in order to obtain a response of structure. It is necessary to perform calculations of several thousand variants for the optimum design problem.

In order to reduce computational efforts methods based on approximation concepts can be used. Nowadays, these methods take a dominant position in structural optimization (Barthelemy and Haftka, 1993). The development of approximation functions has become a separate problem in optimum structural design (Toropov 1989). The approximating models can be built in different ways. Empirical model building theory is discussed in (Box and Draper, 1987). To construct a more general model of the original function the method of experiment design (Audze and Eglais, 1977; Krug and Sosulin, 1977) together with approximate model building (Eglais, 1981; Myers and Montgomery, 1995; Khuri and Cornell, 1996) can be employed. A simplified model,

called metamodel, is built using results of a numerical experiment in the points of experiment design. Response analysis using the simplified model is computationally much less expensive than solution using the original model. Although there is wide literature about experiment design and building of approximating functions it should be noted that there are some special features present in experiment design but not present in the physical experiment. The main features are as follows:

1) The results obtained in the numerical experiment are deterministic without a statistical error. Repetition of the results is 100%. It means that there is no statistical dispersion of the model parameters. However computer models produce numerical noise as a result of the incomplete convergence of iterative processes, round-off errors, and the discrete representation of continuous physical phenomena when different number of calculation steps or a different finite element grid being generated (see Giunta et al., 1994). In deterministic computer experiments, replication at a sample point is meaningless; therefore, the points should be chosen to fill the design space.

2) Mathematical model of the object is unknown, i.e., the form of the regression equation is not known. Therefore, well-known criteria for experiment design optimality, for example, D-optimality, cannot be used. Such criteria can be used only in the case when the form of regression equation is known.

The first space filling design for computer experiment was proposed by Audze and Eglais, 1977. They firstly proposed designs in which the number of levels for each variable is equal to the total number of runs. Audze and Eglais have firstly used also the space filling criterion for such designs based on function similar to potential energy of gravity. The publication (Audze and Eglais, 1977) was in Russian. In the papers (Rikards, 1993 and Rikards et al., 1992) the space filling criterion proposed by Eglais was outlined in English. These experimental designs was popular by scientists in former Soviet Union and are used so far not only in East Europe, see for example Venter et al., 2000. Later the same kind of experimental designs (without any optimization) was proposed as a Monte Carlo integration technique by McKay et al. (1979) and the name *Latin hypercube samplings* was introduced. Numerous space filling experimental designs have been developed in an effort to provide more efficient and effective means for sampling deterministic computer experiments based on Latin hypercubes. Later a lot of space filling criteria for Latin hypercube designs was proposed by many authors: Maximin Latin hypercubes (Morris and Mitchell, 1995; Johnson et al., 1990), Minimal Integrated Mean Square Error designs (Sacks et al., 1989), Orthogonal array-based Latin hypercube designs (Tang, 1993), Orthogonal Latin hypercubes (Ye, 1997), Integrated Mean Square Error (IMSE) optimal Latin hypercubes (Park 1994).

The approach of experiment design and approximation proposed by Eglais (Audze and Eglais, 1977; Eglais 1981), and a corresponding program RESINT can give good results for the problems based on numerical experiment. This approach was employed for solution of optimal design and identification problems (Rikards, 1993; Rikards et al., 2001). However, sometimes the results of the approximation are not satisfactory.

In last years the so called non-parametric approximation methods becoming widely used for the design and analysis of computer experiments: local polynomial approximation (Cleveland and Devlin, 1988; Koehler and Owen, 1996), Kriging (Sacks et al., 1989; Booker et al., 1999). Finally, other statistical techniques such as

Multivariate Adaptive Regression Splines (Friedman, 1991) and Radial Basis Functions (Hardy, 1971; Dyn et al., 1986, Powell, 1987) are beginning to draw the attention of many researchers. However, these methods are computationally expensive not only for metamodel building, but also for using of metamodels for prediction. In addition sometimes these methods are very sensitive to the noise (Jin et al., 2000) and can give overfitting with many false local extremes. Thus, the finding of accurate global approximation for use in constrained optimization is still actual problem.

2. Identification of the model

Based on experiment design approach the stages of model-building are as follows.

2.1 Elaboration of the plan (design) of experiment

There is wide literature on elaboration of plan (design) of experiment (Audze and Eglais, 1977; Krug and Sosulin, 1977; Myers and Montgomery, 1995; Khuri and Cornell, 1996; Haftka et al., 1998). However, as was mentioned above, the criterion for experiment design is related to the form of the regression equation. Good results can be obtained using experiment designs proposed in (Audze and Eglais, 1977), where for each variable x_i, the number of different values is equal to the number of experiments. Usually, it is assumed that numerical values of x_i are uniformly distributed in the interval [-1, 1]. For these designs, the points are distributed according to the criterion of minimum of potential energy of repulsive forces for a set of unit mass points. The magnitude of these repulsive forces is inversely proportional to the cubed distance between the points. Although such criterion has no mathematical validation, good results can be obtained for approximating functions using the experiment designs (Audze and Eglais, 1977).

A different approach, which was proposed in (Kreinin et al., 1976), is the so-called *L*-optimum continuous experimental design. Here pseudo optimum designs are formed by adding new points to the existing plan. Similar approaches of sequential experiment are the Hammersley sequence (Niederreiter, 1992), the Sobol' sequence (Sobol', 1976), Halton sequence (Halton, 1960) an the Faure sequence (Faure, 1982). However, for the same number of points, a fixed design is usually better than the continuous design. Actually, the number of experiment points is always selected by evaluating computational efforts of the calculation of original function for one point.

2.2 Selection of a set of regression functions

A mathematical model is built by linear combinations of some class of functions. Therefore, it is essential that a wide range of functions can be selected. Usually, a set of

finite number of functions is employed. In this case, there is no such completeness of the set of functions as it is for approximations by infinite series.

In (Eglais, 1981), the set of regression functions is formed as a product of variables raised to the power of positive or negative integer in a way that the total sum of functions does not exceed 400. There has been considerable interest on the topic of experimental designs for model discrimination (function selection) (Atkinson and Fedorov, 1975; Pukelsheim and Rosenberger, 1993; Meyer et al., 1995), discussed in (Srivastava, 1996). So called MEP+*P* plans involve candidate models in which all candidates include all first order terms and the candidates differ by the combination of *P* terms from a chosen list of higher order terms. This approach needs very large computation time for relatively small number (up to 10) of candidates (Allen and Yu, 2000).

In one-dimensional case, the orthogonal functions (trigonometric, Legendre or Chebyshev polynomials (see Abramovitz and Stegun 1964)) are often used, for example, Efromovich (1992) proposes a version using orthogonal series of functions on [0,1]. Employing the orthogonal functions it is easy to select the optimal functions, since each function is tested separately. However, such selection can be performed only in the case when the functions are orthogonal to the discrete plan of experiment, i.e., when the information matrix is diagonal. In the case of small number of experiments for the integrally orthogonal functions, the information matrices are not diagonal. Therefore, there is no advantage of using the Chebyshev polynomials in comparison with simple polynomial functions. The diagonal information matrix means that the number of orthogonal functions cannot exceed the number of experiments. An and Owen, 2001, propose quasi-regression method, in which products of univariate Legendre polynomials are used as basis functions, coefficients of these functions are obtained by Monte Carlo integration instead of solving the regression equations. Quasi-regression approach needs very large number of observations (experimental runs) in order to obtain acceptable accuracy. This paper will show how to complete the experiment in order to employ several thousand orthogonal functions using low number of experimental runs.

2.3 Testing for goodness of fit

Using the bank of functions, it is possible to build regression functions with a different number of terms. In (Eglais, 1981), it was proposed to find the best regression function (with minimum of standard deviation) with number of terms *r*, where *r* is larger than the number of terms expected for the optimum regression equation. Then term reduction is performed to eliminate the "worst" functions. Thus, step-by-step, shorter regression functions are formed. By drawing a graph of standard deviation as function of number of terms, the break in the graph can be observed. According to (Eglais, 1981), the break corresponds to the optimum number of terms. By using such a method, satisfactory results can be obtained. However, the method has some defects:

1) There is no reason for choice of the initial number of terms *p*.

2) Employing non-orthogonal functions makes it difficult to find the combination of the "best" p functions from about 400 possible functions (400 is not a large number in the case of an experiment with 10 or more variables).
3) The choice of the break in the diagram of term reduction is subjective, and the optimum number of terms cannot always be obtained.

The adequacy of the mathematical model can be checked by employing additional experimental points, which leads to a more objective method. In this method, two alternative variants of the correlation function are compared by calculating the standard deviation in additional points, which are not used in determination of the parameters (coefficients) of the correlation function. Such an index characterizes the prediction ability of the mathematical model. The disadvantage of this method is the necessity to use larger number of experimental points, because the number of additional points is about the same as the initial number of experimental points. For refinement there are different variants of *holdout, cross-validation* and *bootstrap* methods. These methods do not add a new measurement (which requires computer time), but eliminate k experimental runs and re-estimate the regression model from the remaining N - k runs. Such elimination is repeated l times and the average of root mean square error (RMSE) for estimating of accuracy of prediction (or for comparing prediction accuracy of alternating models) is employed. A well known statistical details are discussed in Kleijnen and Van Groenendaal (1992); also see Kohavi (1995). In the present paper *leave-one-out* cross validation and *k-fold cross validation* is used and these methods are explained below.

3. Orthogonal multivariate polynomial functions

Nomenclature:

N – number of experiments (experimental runs);
s – number of input variables (number of arguments of equation of regression);
x_i – i-th input variable (i=1,.., s);
x_{ij} – numerical value of i-th variable in the j-th experiment;
F_i – value of target function, measured (or calculed) at the i-th experiment;
$\mathbf{X}_i$ – matrix-column containing numerical values of all variables for i-th experiment;
$\mathbf{F}$ – matrix-column containing values of target function in all experimental points;
z_i – the i-th point of the interpolated experimental table (augmented experiment design);
G_i – approximated value of target function in the i-th point z_i of augmented experiment;
N_F – total number of the base functions;
$f_i(x)$ – the i-th base function;
M – number of equidistance points for each input variable in the augmented experimental set (s-dimensional grid size);
$p_i(y,M)$ – univariate (Chebyshev type) polynomial of i-th order from set of polynomials, orthogonal on M discrete points from interval [-1,1], (i=0,1,…, M-1);
$\hat{F}_D(x)$ – approximation of function $F(x)$, using sample set D;
t – number of nearest neighbors (bandwidth) used in local approximation;

$N_t(z_l)$ – set of indices of the nearest neighbors to the point z_l.

In this paper we are constructing the basis functions over $[-1,1]^s$ by taking tensor products of univariate basis functions. The purpose is to obtain orthogonal functions on the set of points from regular s-dimensional grid of size M. As univariate functions we are taking polynomials $p_i(y,M)$, which are orthogonal on discrete set of M equidistance points from interval [-1,1]. These polynomials can be obtained from the well known discrete Chebyshev polynomials $t_i(x,M)$, which are orthogonal on the set of integer values (0,1,…, M-1), (see Chebyshev 1948, Abramovitz and Stegun 1964; Nikiforov and V.B. Uvarov, 1988; Nikiforov et al., 1991) . Also the so called three-term recurrence relation is applicable. Thus, one can obtain:

$$p_0(y,M)=1\,, \tag{1a}$$

$$p_1(y,M)=y\,, \tag{1b}$$

$$p_i(y,M)=yp_{i-1}(y,M)-\frac{(i-1)^2\left[M^2-(i-1)^2\right]}{(M-1)^2\left[4(i-1)^2-1\right]}p_{i-2}(y,M)\quad i=2,...,M-1 \tag{1c}$$

These polynomials satisfy an orthogonality relation of the form

$$\sum_{k=0}^{M-1}p_i(y_k,M)p_j(y_k,M)=\delta_{ij}\frac{i!^2\,(M+i)!)}{(2i+1)(N-i-1)!\,(i+1)_i^2}\left(\frac{2}{M-1}\right)^{2i} \tag{2}$$

where δ_{ij} is the Kronecker symbol and $(a)_i=a(a+1)\cdots(a+i-1)$ denotes the Pochhammer symbol and $y_i=(2i-M)/M$ are discrete equidistance points from interval [-1,1]. These polynomials are similar to the Legendre polynomials and become equal to Legendre polynomials in the case when M tends to infinity. Moreover, they form a very important special subclass of the Hahn polynomials $h_i^{\alpha,\beta}(y,M)$, (Nikiforov and Uvarov, 1988; Nikiforov et al., 1991): that with $\alpha=\beta=0$.

In the case of grid size M=7, the polynomials are:

$$\begin{aligned}
&p_0(y,7)=1\,,\quad p_1(y,7)=y\,,\quad p_2(y,7)=y^2-0.4444444444\,,\\
&p_3(y,7)=y^3-0.7777777777y\,,\\
&p_4(y,7)=y^4-1.0634920634y^2+0.1269841269\,,\\
&p_5(y,7)=y^5-1.2962962962y^3+0.3080540858y\,,\\
&p_6(y,7)=y^6-1.4646464646y^4+0.4870931537y^2-0.0213777991
\end{aligned} \tag{3}$$

These polynomials presented in figure 1 are orthogonal at the discrete 7-point set.

In the method described below the orthogonality of base functions will be used for selection of perspective functions, but not for determination of coefficients of the regression equation. Selection will be performed not for the initial experimental points, but for the augmented set of experimental points. The augmented experiment set is created by local approximation (smoothing) of the experimental results. Therefore, the augmented experiment set can be chosen in order to easily obtain a corresponding set of orthogonal polynomials.

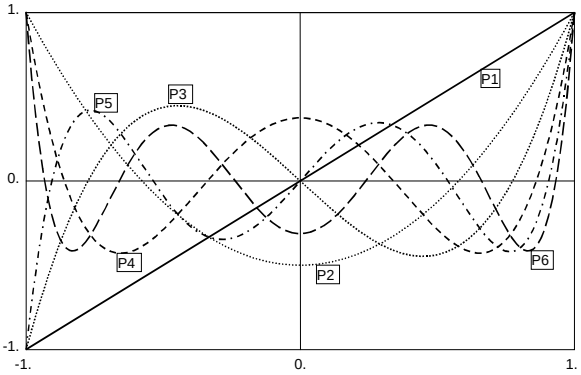

Figure 1. Orthogonal 7-point polynomials.

In this paper we are taking the *s*-dimensional basis functions to be tensor products of univariate orthogonal polynomials $p_i(y,M)$

$$f_k(x) = \prod_{j=1}^{s} p_{\alpha_{kj}}(x_j, M) \tag{4}$$

Here α_k is an *s*-dimensional vector containing degrees of factors of *k*-th function for each variable x_j. The sets that we are choosing for analysis are defined by vectors $\alpha_k=(\alpha_{k1},\ldots, \alpha_{ks})$ of nonnegative integers satisfying limitations

$$\sum_{j=1}^{s} \alpha_{kj} \le d_j \tag{5}$$

$$\max_{1\le j\le s} \alpha_{kj} \le m \tag{6}$$

Here d_j is maximum degree of univariate polynomial for *j*-th variable and *m* is maximum order of the basis functions.

When the maximum degree limit for all variables are identical and equal to the maximum allowed value *M*-1, then the total number of basis functions (all possible products of one-dimensional polynomials) are $N_F = M^s$. Therefore, in the augmented set there should be at least M^s experiments. It is easy to examine that all M^s functions f_i are mutually orthogonal to the set of M^s discrete points.

$$\sum_{k=1}^{N_F} f_i(z_k) f_j(z_k) = 0, \quad if\ i \ne j \tag{7}$$

The next problem is augmenting of the initial experiment to M^s observations.

4. Augmenting of experiment using local quadratic approximation

Several methods for data augmentation exist (See Tanner, 1991). Here the local quadratic approximations are used to obtain the approximated values in the M^s points of the regular *s*-dimensional grid (augmented experiment). We are using the well known nearest neighbor approach (see Kleijnen and Groenendaal, 1992; Cleveland and Devlin, 1988; Fan and Gijbels, 1996). In order to obtain the interpolated values of function in the grid point z_j, the point set $T(z_j)$ of *t* nearest experimental points in the *s*-dimensional space is found. Then, in the point z_j, a numerical value of the second order approximating polynomial is calculated using weighted least square method. Let $\|\cdot\|$ measure Euclidean norm; then the distance between the points z_j and x_i in *s*-dimensional space is given by

$$\|x_i - z_j\| = \sqrt{\sum_{k=1}^{s} (x_{ik} - z_{jk})^2} \tag{8}$$

The number of nearest points *t* often called *bandwidth* should be larger or equal to the number $1+s(s+3)/2$, which is the number of terms in the second order polynomial of *s* variables. From the experience with different approximations, it can be concluded that *t* should be considerably larger. In order to select the optimum value of *t*, we are using the *leave-one-out cross validation* (Efron, 1983; Mitchell and Morris, 1992) method, in which local approximations for initial points of the experiment are tested. In this approach, each sample point used to fit the model is removed one at a time, the model is rebuilt without that sample point, and the difference between the model without the sample point and actual value at the sample point is computed for all sample points. The cross validation root mean square error (CVRMSE) is computed

$$CVRMSE_1 = \sqrt{\frac{\sum_{i=1}^{N} (F_i - \hat{F}_i)^2}{N}} \tag{9}$$

The bandwidth value, which gives minimum of CVRMSE is selected. We are using the least square estimators

$$\Phi = \sum_{i \in T(z_j)} w_i(z_j, x_i) \left(F_i - \beta_0 - \sum_{k=1}^{s} \beta_k x_{ik} - \sum_{k=1}^{s} \sum_{l=1}^{k} \beta_{kl} x_{ik} x_{il} \right)^2 \tag{10}$$

with weighting function

$$w(z_j, x_i) = (1-u)^4 \ , \tag{11}$$

where $u = \dfrac{\|z_j - x_i\|}{\max\limits_{k \in T(z_j)} \|z_j - x_k\|}$, so that u=1 for farthest neighbor.

We also tried the *tricube* function $w(z_j, x_i) = (1-u^3)^3$, as suggested by Cleveland, as well as locally unweighted approximation, but the function (11) as proved empirically is to be slightly better than others.

Similarly, the local linear approximation was employed. However, results were not as good as for the local quadratic approximation (see Table 1 below).

5. Selection of experimental design

Knowing that in approximation it will be necessary to obtain the values of interpolated function in the points of the *s*-dimensional grid, it is essential that points of initial experiment are distributed so that locally approximated values will be as accurate as possible. By performing approximation with the second order polynomials, the error of interpolation is increasing proportionally to the product of distance between the points used in interpolation, if the points are distributed close enough. In the neighborhood of the grid point, there are only few points of the real experiment (usually only one to two points). Therefore, the quality of experiment plan is characterized by a sum of the distance raised to the power of four between all grid points and the closest experimental point

$$\Gamma = \sum_{i=1}^{N_F} \Delta_i^4 , \tag{12}$$

where Δ_i is the distance between the grid point z_i and the closest experimental point. When in (12) the power of four instead of the power of two is employed, the criterion is brought closer to the criterion of maximum distance. In such a way it is not allowed to form the plans of experiment with empty domains in some places. The minimum of criterion (12) without constraints can be easily found by employing the Newton's method. However, many local minimums were found with about the same value of the cost function Γ. In the left plan of figure 2, one projection of the 36-point five-dimensional design is shown (the value of the cost function $\boldsymbol{\Gamma}$ =1739 is best that was found for this design).

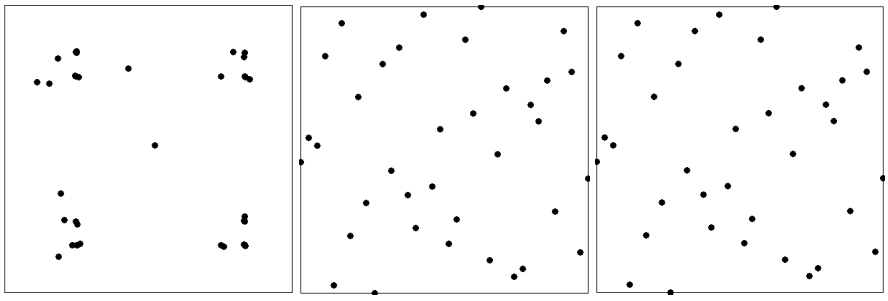

Figure 2. 36-point designs in plane 1-2. From left to right: optimized unconstrained design, Γ=1739; Eglais' Latin hypercube type design , Γ =3322; optimized Latin hypercube design, Γ =3318.

By employing the Eglais' criterion good values of the cost function (12) can also be obtained (see figure 2, middle plan). In the right plan of figure 2 the design optimized according to the criterion (12) is shown, using as initial estimate the Eglais' plan. It can

be seen that the improvement is negligible. By distributing regularly the levels of experiment in the interval [-1,+1] but not requiring that extreme levels should be exactly at –1 and +1, a slightly better value of cost function Γ =3271 was obtained.

6. Selection of regression functions

Selection of the regression functions is performed in the following three stages.

1. Selection of N_p best orthogonal functions using augmented set of experiment. In this stage the basis functions f_k are ranged according to deviation in the points of regular grid

$$\Phi_k = \sum_{i=1}^{N_F} \left(a_k f_k(z_i) - G_i\right)^2 \tag{13}$$

These functions are orthogonal and coefficients of approximation a_k can be easily determined

$$a_k = \frac{\sum_{i=1}^{N_f} G_i f_k(z_i)}{\sum_{i=1}^{N_f} f_k^2(z_i)}. \tag{14}$$

It is not necessary to employ all of the N_F functions. Also, maximum degree d and maximum order m can be restricted. For example, if it is known that in the mathematical model the powers of the variables are not higher than three, then the limitation d=3 and m=3 can be used. By setting d=2 and m=2 the well known second order polynomial approximation can be obtained. The number of functions N_p to be selected can exceed several thousands, however, this parameter has no significant influence on computational efforts of the selection process.

2. Ranking of the best functions by using the points of initial experiment. In this stage ranking of the functions selected in the first step is performed. This selection is more complicated since the functions f_k are not orthogonal to the initial experiment. In distinction from algorithm (Eglais, 1981), where a term reduction is performed, in this step the number of terms l will be increased. We start with one term l=1, and from the previously selected N_p functions as the best the function with minimum standard deviation in the points of initial experimental design is selected

$$\Phi_1 = \sum_{i=1}^{N} \left(a_k f_k(x_i) - F_i\right)^2 \tag{15}$$

where the coefficient a_k is calculated by using the formula

$$a_k = \frac{\sum_{i=1}^{N} F_i f_k(x_i)}{\sum_{i=1}^{N} f_k^2(x_i)} \tag{16}$$

Further, the regression function is built increasing the number of terms *l*, i.e., adding new terms. In this step the function chosen as the best (added from the basis functions) is the function selected in the first stage with the minimum standard deviation

$$\Phi_l = \sum_{i=1}^{N} \left(\sum_{j=1}^{l} a_{k_j} f_{k_j}(x_i) - F_i \right)^2 \tag{17}$$

Note that there is a difference from the case of augmented experiment. Now the coefficients a_j , $j = 1,2,..,l$ are obtained from solution of linear algebraic equations with a symmetric positive definite information matrix $\mathbf{A}_l$ (see e.g. Krug and Sosulin, 1977). It is a time consuming process because N_p–l+1 functions, which could be used in the mathematical model, must be examined. The process of selection can be considerably accelerated since for all N_p–l+1 variants in the information matrix only the *l*-th row and column are changed.

The total number of terms in the regression equation is usually limited to *N*/2.

3. Determination of the optimum number *r* of terms in the regression equation.

Estimating the accuracy of metamodels it is important not only to predict its future prediction accuracy, but also for choosing the best metamodel from a given set – this is equal to the selection of optimal number of terms in regression function. The best estimator for accuracy of prediction of metamodels and for selecting the best metamodels is calculation of prediction error on an additional sample of size *B*, which is not used by metamodel building. Then the Root Mean Square Error (*root MSE*)

$$rootMSE = \sqrt{\frac{\sum_{N+1}^{N+B} \left[F_i - \hat{F}(x_i) \right]^2}{B}}, \tag{18}$$

Average Absolute Error (*avg.abs.error*)

$$avg.abs.error = \frac{1}{B} \sum_{i=N+1}^{B} \left| F_i - \hat{F}(x_i) \right|, \tag{19}$$

Maximum absolute error (*max.abs.error*)

$$max.abs.error = \max_{N<i\le N+B} \left| F_i - \hat{F}(x_i) \right| \tag{20}$$

can be used for selection of the best metamodel, or optimal number *r* of terms in regression function, respectively.

Unfortunately, this method is frequently too expensive, because we must perform additional calculations or carry out new natural experiments. In practice the dataset size is usually smaller than we would like to use.

V. Eglais (Eglais, 1981) proposed approach, in which only residual root mean square error is used

$$residual\,RMSE = \sqrt{\frac{\sum_{i=1}^{N}\left(F(x_i) - \hat{F}(x_i)\right)^2}{N - r + 1}} \tag{21}$$

By drawing a graph residual *RMSE* versus the number of term *r* the break in the graph can be observed. According to (Eglais, 1981), the break corresponds to the optimum number of terms. It will be shown in demonstration examples, that using of this simple approach is limited.

Two most common methods for accuracy estimation without using an additional experiments are cross validation and bootstrap. It is known that no accuracy estimation can be correct all the time (Kohavi, 1995), therefore we are interested in identifying a method that is well suited for the biases and variance in real datasets. For example, leave-one-out cross validation is almost unbiased, but has high variance (Efron, 1983). Holdout, cross validation and bootstrap methods are widely used in statistics (Kohavi, 1995; Efron, 1993; Moore and Lee, 1994; Kleijnen and Van Groenendaal,1992).

We are using a version of holdout method for estimation of prediction accuracy of metamodels. The holdout method, sometimes called test sample estimation, partitions the data into two mutually exclusive subsets called test set, or holdout set D_h and training set $D \backslash D_h$. It is common to designate 2/3 of the data as the training set and the remaining 1/3 as the test set (Kohavi, 1995). The estimation of model coefficients is performed using the training set of size *N-k* and the holdout estimated accuracy is defined as root mean square error calculated in the *k* test points

$$RMSE_k = \sqrt{\sum_{i \in D_h}\left(f(x_i) - \hat{f}_{D \backslash D_h}(x_i)\right)^2 \Big/ k}\ . \tag{22}$$

The holdout estimate is a random number that depends on the division into a training set and a test set. The holdout method is repeated *L* times, and the estimated accuracy is derived by averaging the runs. The standard deviation can be estimated as the standard deviation of the accuracy estimations from each holdout run and the number of holdout runs *L* can be determined for a given confidence interval (Kohavi, 1995). The holdout method is a pessimistic estimator because only a portion of the data is given for building the approximation (Kohavi, 1995). The more experimental points we leave for the test set, the higher is the bias of our estimate. However, fewer test set points means that the confidence interval for the accuracy will be wider as shown in (Kohavi, 1995).

Kohavi, 1995 proposed using of ten-fold cross validation, which need smaller number of validation runs. Our practice shows, that 1/3 holdout as well as ten-fold cross validation can give worse suggestions for the choice of *r*. This will be showed in demonstration examples. In practice, we are using a two-stage holdout validation. At first, the 1/3 holdout is carried out and the optimal number of terms *r*, which gives minimal RMSE, is chosen. In questionable cases second holdout test with $k=N-2r$ is executed to get more accurate value.

The number of trials L to obtain the predicted *RMSE* with the necessary standard deviation from which the optimum number of terms can be drawn is about 10N to 50N. This is the most time consuming stage of the mathematical model identification for problems with the small (2 to 4) number of independent variables. It is noteworthy to emphasize that the final parameters of the mathematical model - the coefficients a_j, (j=1,2,..,r) - are obtained employing all of the N experimental points.

7. Demonstration of the method

The proposed method of identification and the performance of corresponding software APROX are demonstrated using several examples – 11 artificial test problems and one practical problem from the robot design.

7.1. Example 1

In the first test example the original function is given in a form of function with five arguments

$$F = (x_4 + 1)\sin(x_1) + x_2 x_3 \tag{23}$$

It can be seen that the argument x_5 is not presented in the function. For the experiment a 36-point 5-dimensional plan proposed in (Audze and Eglais, 1977) is considered. Since the function is known the standard deviation in 10000 points distributed regularly in the domain of approximation $-1 \le x_i \le 1$ is used as the criterion of the accuracy of the mathematical model. Ranking of the basis functions is as follows

$$x_1,\quad x_2x_3,\quad x_1x_4,\quad p_3(x_1),\quad p_3(x_1)x_4,\quad p_5(x_1),\ p_5(x_1)x_4,\quad p_3(x_2)x_4x_5\,.$$

Predicted and actual standard deviations are presented in figure 3. From the graph of predicted standard deviations it can be seen that optimum number of terms in the mathematical model is seven. It should be mentioned that regression equation (mathematical model) does not contain the argument x_5. The actual standard deviation is less than 10^{-5} % of the function range.

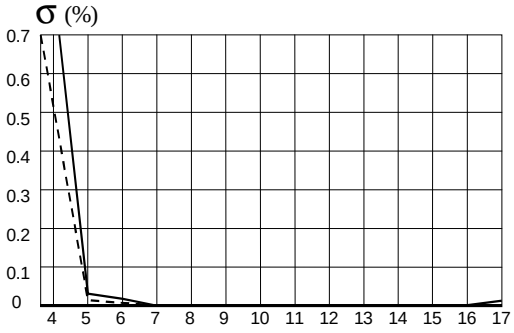

Figure 3. Predicted (____) and actual (-----) standard error versus number of terms in equation of regression.

7.2. Example 2: Identification of mathematical model of the composite plate

In the second example an actual identification problem is solved. The task is to build a mathematical model for eigenfrequencies of the cross-ply laminated composite plate. The arguments (variables) are five elastic constants of material (Rikards et al., 2001)

$$x_1 = \frac{E_1}{1-\nu_{12}\,E_2/E_1}, \qquad x_2 = \frac{E_2}{1-\nu_{12}\,E_2/E_1},$$
$$x_3 = \frac{\nu_{12}E_1}{1-\nu_{12}\,E_2/E_1}, \quad x_4 = G_{23}, \quad x_5 = G_{12} \tag{24}$$

where E_1, E_2 are the longitudinal and transverse elastic moduli for the single layer of the cross-ply laminated composite plate, ν_{12} is Poisson's ratio of the layer and G_{23}, G_{12} are the shear moduli of the layer.

The eigenfrequencies of a plate with free-free boundary conditions were calculated by the finite element method (FEM) in the 36 reference points the experiment plan. A 36-point 5-dimensional plan proposed in (Audze and Eglais, 1977) was used. The original numerical functions were calculated for the first 30 non-zero eigenvalues corresponding to the first 30 lowest eigenfrequencies of the plate. In addition, the FEM solution was obtained in the other 35 points in order to verify the adequacy of the models obtained. For this, a 35-point 5-dimensional plan of experiment (Audze and Eglais, 1977) was employed. In Table 1, the relative Root Mean Square Error values for different approximations are presented (r is number of terms in the model). The Relative Root Mean Square Error E_r is calculated in percents by formula

$$E_r = \frac{100}{STD}\sqrt{\frac{1}{N_C}\sum_{i=1}^{N_C}\left[\hat{\lambda}(x_i)-\lambda_i\right]^2} \tag{25}$$

Here λ_i is a value of the original function in the i-th point of additional experiment, $\lambda(x_i)$ is a value of the approximating function in the same point, *STD* stands for standard deviation of original function $STD = \sqrt{\frac{1}{N_C-1}\sum_{i=1}^{N}\left(\lambda_i-\bar{\lambda}\right)^2}$ and N_C =35 is number of confirmation points.

In Table 1, it is shown that approximation by orthogonal polynomials is better than the approximation by using the code RESINT (Eglais, 1981). An exception is mode 17 for which a more accurate result is obtained by the code RESINT. Using the program APROX the optimum number of terms in the regression equation can be determined more accurately. However, the best approximations were obtained using the local quadratic approximation.

Table 1. Relative Root Mean Square Errors for different eigenfrequencies in 35 verification points.
* - optimum number of terms (see figure. 4, mode 5).

		Mode 4	Mode 5	Mode 7	Mode 8	Mode 13	Mode 14	Mode 17
RESINT (Eglais, 1981)	r	3	7	4	5	6	8	7
	E_r	1.752	0.600	2.040	1.860	1.890	1.530	1.050
Orthogonal polynomials APROX	r	3	7	4	5	6	8	7
	E_r	1.572	0.600	1.986	1.932	1.692	1.494	1.212
	r*	8	10	7	10	9	11	7
	E_r	0.456	0.474	1.173	0.471	0.513	1.104	1.212
Local linear approximation	E_r	0.60	2.04	1.155	0.60	0.894	1.545	0.963
Local quadratic approximation	E_r	0.21	0.39	0.927	0.24	0.429	1.011	0.774

Possibly, this is because good approximations of the original functions (eigenvalues of vibrating plate) can be obtained by the second order polynomials. In the first example (23) the standard deviation for the local quadratic approximations is 1.59%, which is much higher than for the orthogonal polynomials.

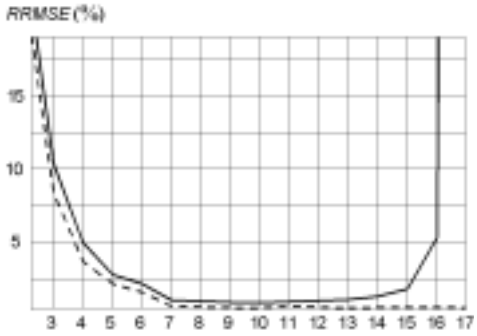

Figure 4. Predicted (___) and actual (----) standard *RRMSE* versus number of terms in equation of regression for mode 5.

7.3. Examples from Hock and Schittkowski

Eight other mathematical problems are utilized in our study. These test problems are chosen from (Hock and Schittkowski, 1981) which offers 180 problems for testing nonlinear optimization algorithms. We choose some problems which are studied in (Jin et al., 2000) in order to compare the results with methods used by other authors – Kriging (Sacks et al., 1989; Booker et al., 1999), Multivariate Adaptive Regression Splines (MARS), (see Friedman, 1991), Radial Basis Functions (RBF) (Powell, 1987; Dyn et al., 1986) as well as well known Second Order Polynomial Regression (PR)

often used as synonym of Response Surface method (Giunta et al., 1994 and many other researchers). We are using linear variable substitution in the original test functions in order to transform the definition region of variables to [-1,1]. Here the indexing of functions used in Jin et al., 2000 is employed.

A 100 point regular grid is used for the two-variable problems and a 125 point optimized Latin hypercube designs for the three variable problems. Present results are compared with results given in (Jin et al., 2000). Grid size M=10 for examples with two variables and M=20 for example F_{10} with 3 variables is employed. Additional n=100000 confirmation points was used for calculation of measures for prediction of accuracy:

R Square:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(F(x\ ) - \hat{F}(x_i)\right)^2}{\sum_{i=1}^{n} \left(F(x_i) - \overline{F}\right)^2}\ , \tag{26}$$

Relative Average Absolute Error (RAAE):

$$RAAE = \frac{\sum_{i=1}^{n} \left|F(x_i) - \hat{F}(x_i)\right|}{n * STD}, \tag{27}$$

Relative Maximum Absolute Error (RMAE):

$$RMAE = \max_{1 \le i \le n} \frac{\left|F(x_i) - \overline{F}(x_i)\right|}{STD}, \tag{28}$$

where STD stands for standard deviation

$$STD = \sqrt{\frac{\sum_{i=1}^{n} \left(F(x_i) - \overline{F}\right)^2}{n}} \tag{29}$$

7.3.1. Function F_6

$$F_6(x) = \left[30 + (5x_1 + 5)\sin(5x_1 + 5)\right]\left[4 + \exp\left(-(2.5x_2 + 2.5)^2\right)\right] \tag{30}$$

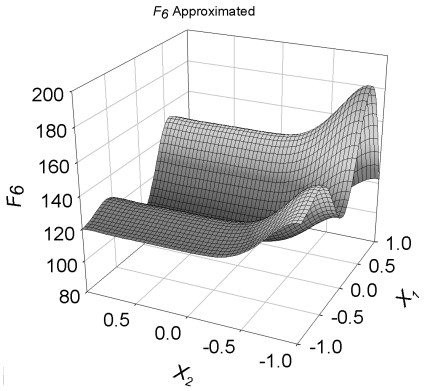

Figure 5. Approximated function F_6.

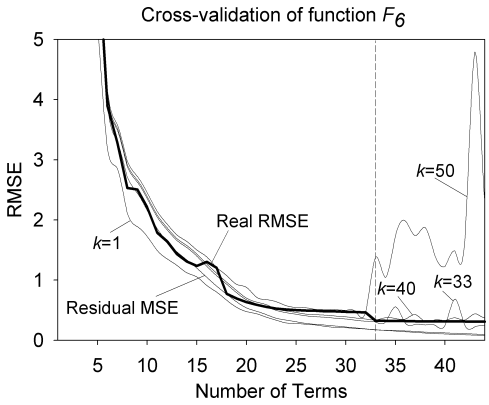

Figure 6. Cross validation scores of approximation F_6.

Table 2. Accuracy of methods used for test function F_6. * - Results for MARS, RBF and Kriging are taken from (Jin et al., 2000), OP – orthogonal polynomials.

	MARS*	RBF*	Kriging*	PR	OP
R^2	0.9924	0.9610	0.9999	0.2738	0.9998
RAAE	0.0471	0.118	0.00417	0.675	0.0083
RMAE	0.921	1.40	0.0994	2.48	0.0683

This function is defined as high order nonlinear. Optimum number of terms r=33 obtained by 33-holdout. It is seen that the OP approximation is very accurate, however, Kriging gives slightly better result.

7.3.2. Function F_7

$$F_7(x) = \sin(5\pi x_1/6)\cos(0.125\pi x_2) \tag{31}$$

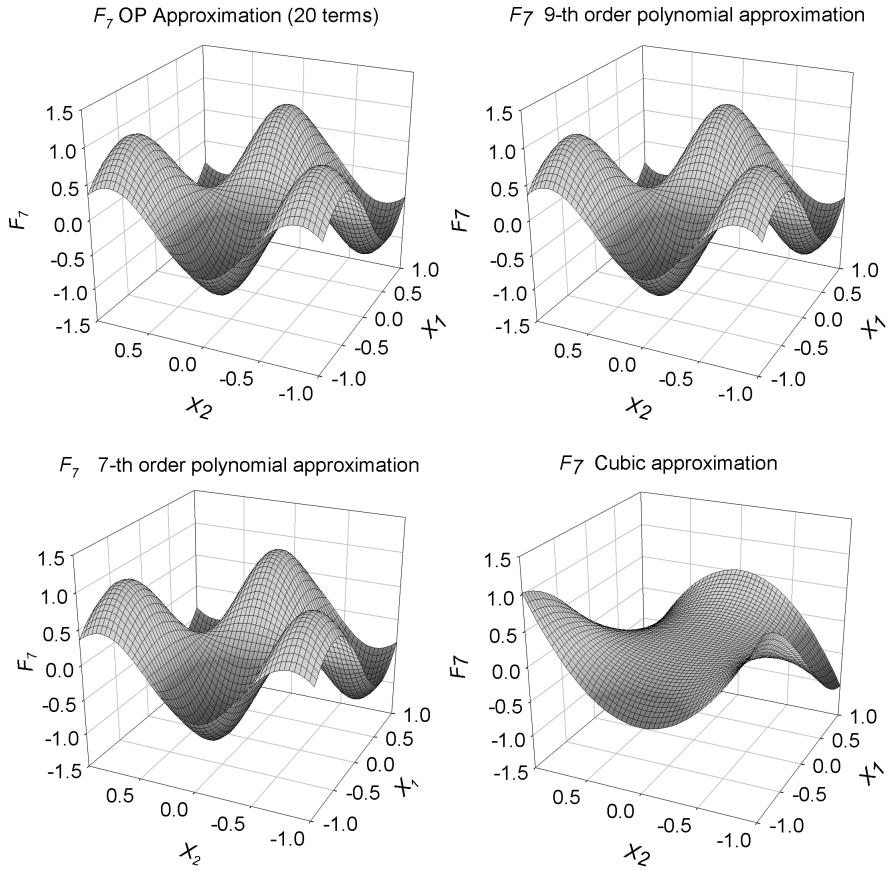

Figure 7. Original function F_7 and its approximation polynomials of various order. OP – 20 terms, RMSE=0.0005; 9-th order polynomial - 55 terms, RMSE=0.00319; 7-th order approximation - 36 Terms, RMSE=0.0284; cubic approximation - 10 terms, RMSE=0.3829

Table 3. Accuracy of methods used for test function F_7. * - Results for MARS, RBF and Kriging are taken from (Jin et al., 2000).

	MARS*	RBF*	Kriging*	PR	OP
R^2	0.6605	0.9892	>0.9999	-0.0204	>0.99999
RAAE	0.451	0.0700	0.00168	0.846	0.00043
RMAE	3.180	0.653	0.0702	2.13	0.0038

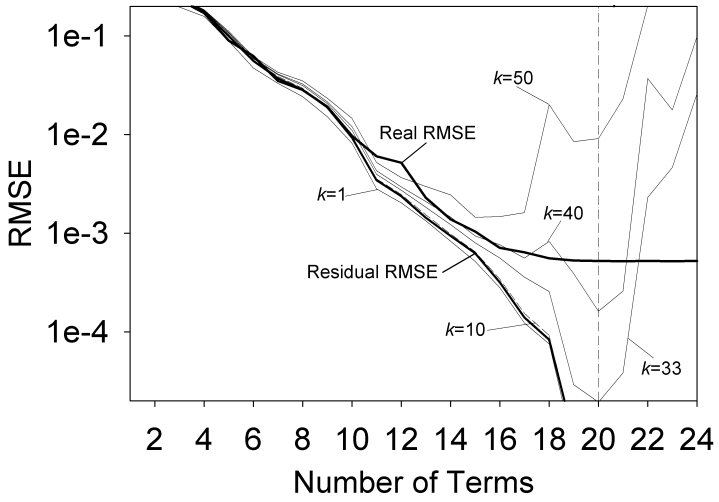

Figure 8. Cross validation scores of approximation F_7.

This function is defined as high order nonlinear. Optimal number of terms r=20, obtained by 33-holdout. The OP method gives the best approximation.

7.3.3. Function F_8

$$F_8 = \sin(3x_1 + 3x_2 + 2) + 9(x_1 - x_2)^2 - 4.5x_1 + 7.5x_2 + 2 \tag{32}$$

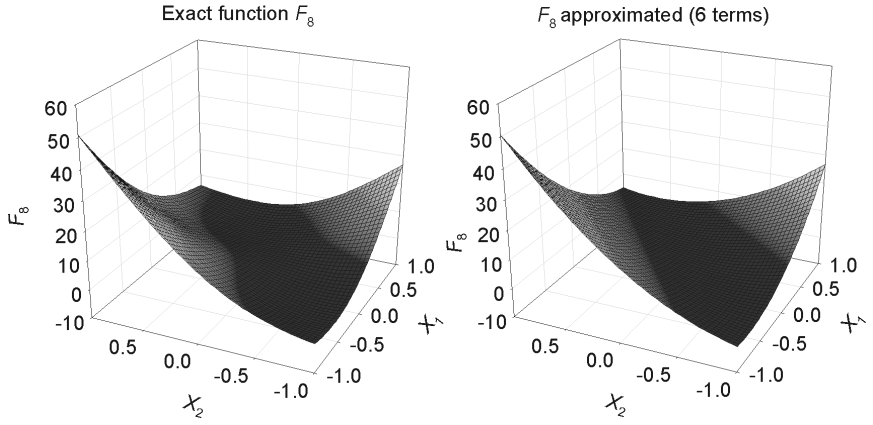

Figure 9. Exact and approximated (with 6 terms) function F_8. R^2=0.9947, *RAAE*=0.0644, *RMAE*=0.153

Table 4. Accuracy of methods used for test function F_8. * - Results for MARS, RBF and Kriging are taken from (Jin et al., 2000).

	MARS*	RBF*	Kriging*	PR	OP
R^2	0.9954	0.9842	0.9997	0.9947	>0.9999
RAAE	0.0416	0.0443	0.00312	0.0644	0.00247
RMAE	0.593	1.21	0.321	0.153	0.00918

OP with six best terms for approximation gives the same result as second order polynomial (PR) approximation. This is enough accuracy for practical applications, although OP with 50 terms gives more accurate approximation.

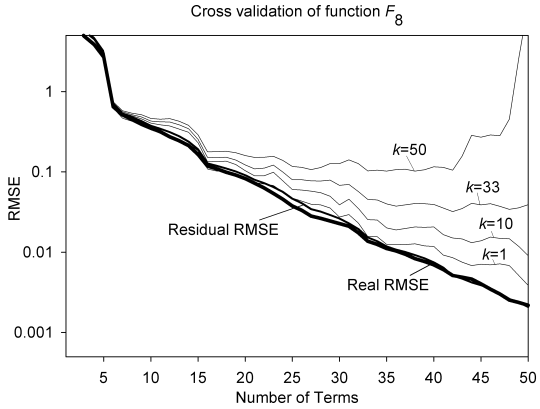

Figure 10. Cross validation scores of approximation F_8.

7.3.4 Function F_{10}

$$F_{10}(x) = \sum_{i=1}^{99} f_i(x)^2 \,, \tag{33}$$

where $f_i(x) = -0.01i + \exp\left(-\frac{(u_i - 12.8x_2 - 12.8)^{(2.5x_3+2.5)}}{49.95x_1 + 50.05} \right)$ and

$u_i = 25 + (-50\ln(0.01i))^{2/3}$, i=1,…,99

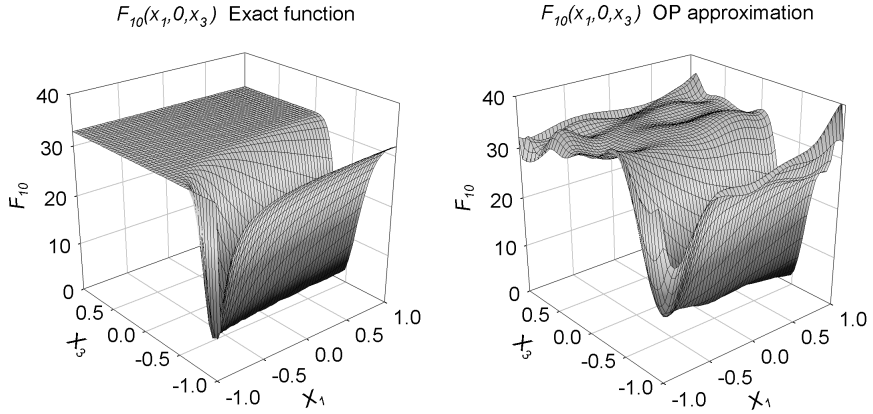

Figure 11. Original function F_{10} and its OP approximation (x_2 fixed equal to zero)

Table 5. Accuracy of methods used for test function F_{10}. * - Results for MARS, RBF and Kriging are taken from (Jin et al., 2000). (OP with 50 terms from 8000 base functions)

	MARS*	RBF*	Kriging*	PR	OP
R^2	0.7434	0.7441	0.7628	0.1868	0.9175
RAAE	0.340	0.324	0.229	0.763	0.2083
RMAE	2.44	2.51	3.50	2.08	2.353

OP gives the best result excluding RMAE, which is slightly better for PR. This function seems to be too complex for approximation with only 125 point experimental design.

7.3.5. Noisy function F_{13}

$$F_{13}(x) = 12.5x_1^2 + 25x_2^2 - 25x_1x_2 - 35x_1 - 35x_2 + \varepsilon(x_1, x_2), \quad (34)$$

where $\varepsilon(x_1,x_2)$ is a normal noise with mean equal to zero and standard deviation σ=3.5426

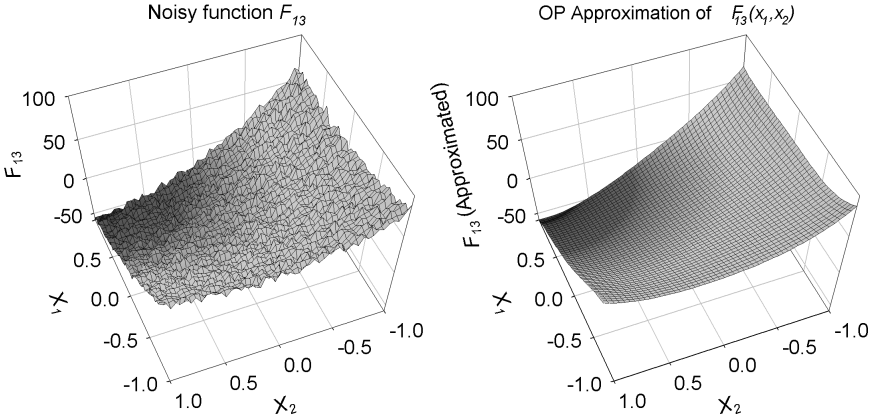

Figure 12. Noisy function F_{13} and its OP approximation

Table 6. Accuracy of methods used for noisy test function F_{13}. * - Results for MARS, RBF and Kriging are taken from (Jin et al., 2000).

	MARS*	RBF*	Kriging*	PR	OP
R^2	0.9982	0.9849	0.8758	0.9997	0.9997
RAAE	0.0554	0.0884	0.400	0.0119	0.0119
RMAE	0.237	0.710	2.74	0.0676	0.0676

PR and OP approximations here are identical and both give the best result, because OP method found a second order polynomial with 6 terms to be the best approximation.

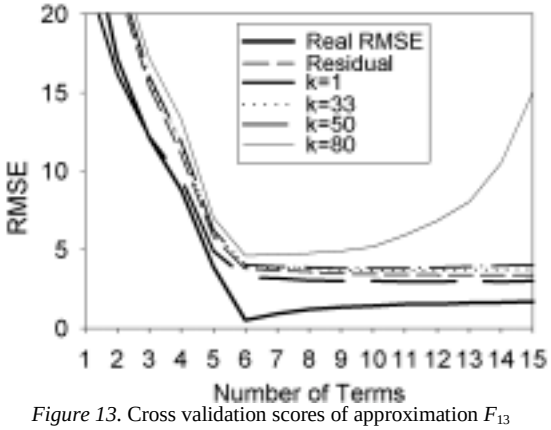

Figure 13. Cross validation scores of approximation F_{13}

7.3.6. Some other functions

Other functions of two and three variables, tested by Jin et al., 2000

$$F_9(x) = (10x_1 - 1)^2 + 100(x_1 - x_2)^2 + 10000(x_2 - x_3)^2, \tag{35}$$

$$F_{11}(x) = 5.3578547(20x_1 + 90)^2 + 0.8356891(20x_1 + 90)(10x_3 + 35) + 37.293239(20x_1 + 90) - 40792.141 \quad , \tag{36}$$

$$F_{12}(x) = 12.5x_1^2 + 25x_2^2 - 25x_1x_2 - 35x_1 - 35x_2 \tag{37}$$

are not interesting for approximation because the OP approximation for these functions give the exact solution. These functions can be built with given base functions. These examples show only that selection algorithm works correctly.

7.4. Two-bar structure problem

A two-bar structure design problem adopted from (Jin et al., 2001) is employed as an example in our study. Here three functions Y_1, Y_2, Y_3 should be approximated. A 73-run optimized Latin hypercube experimental design and grid size 20 was used for the OP approximation. These functions are as follows

$$\begin{aligned} Y_1(x) &= 5 * \pi \sqrt{(150x_3 + 750)^2 + (400x_2 + 600)^2} \\ Y_2(x) &= \frac{150000\sqrt{(150x_3 + 750)^2 + (400x_2 + 600)^2}}{5\pi(30x_1 + 50)(400x_2 + 600)} \\ Y_3(x) &= \frac{\pi^2 210000\left[6.25 + (30x_1 + 50)^2\right]}{8\left[(150x_3 + 750)^2 + (400x_2 + 600)^2\right]} \end{aligned} \tag{38}$$

Table 7. Accuracy of methods used for two-bar structure problem. * - Results for RBF and Kriging are taken from (Jin et al., 2001).

	RBF*		Kriging*		PR		OP	
	R^2	RMAE	R^2	RMAE	R^2	RMAE	R^2	RMAE
Y_1	0.9994	0.2460	0.9999	0.0041	0.9996	0.0903	1.0000	0.0002
Y_2	0.9747	1.7729	0.9985	0.4488	0.9036	1.5027	0.9983	0.2724
Y_2-Y_3	0.9957	0.6703	0.9998	0.1190	0.9832	0.6065	0.9994	0.1784

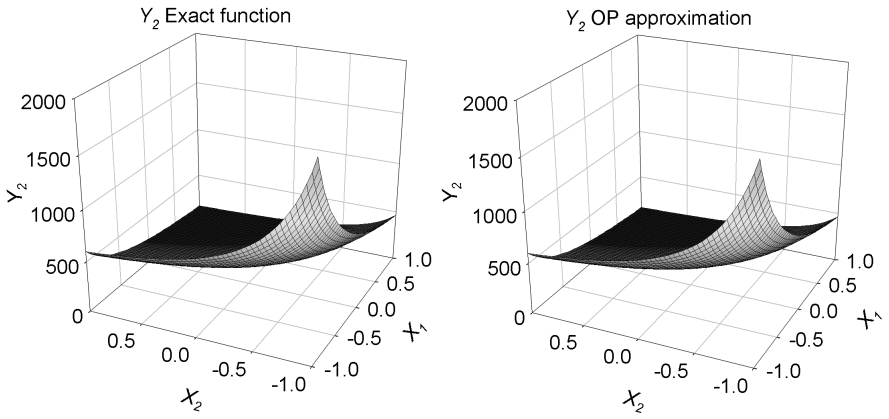

Figure 14. Exact function Y_2 and its OP approximation (x_3 value fixed equal 0)

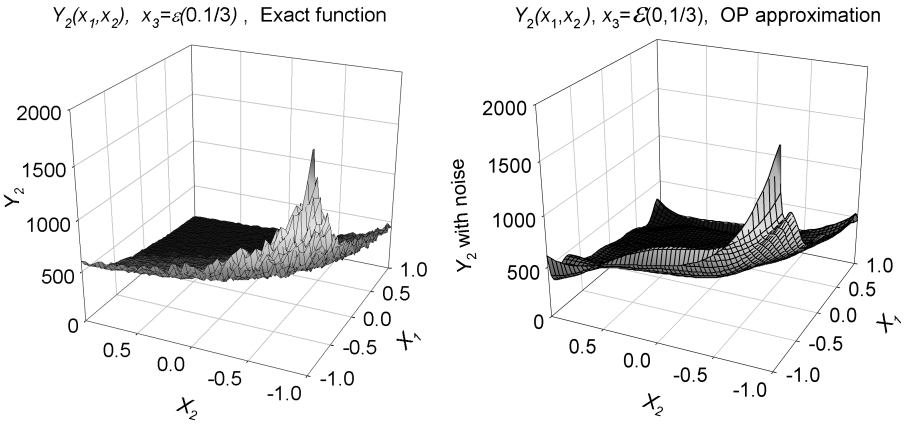

Figure 15. Approximation Y_2 as a two-variable function, when third variable are normal noise with mean equal to zero and standard deviation σ=1/3 (a 100-point two-dimensional experimental design is used).

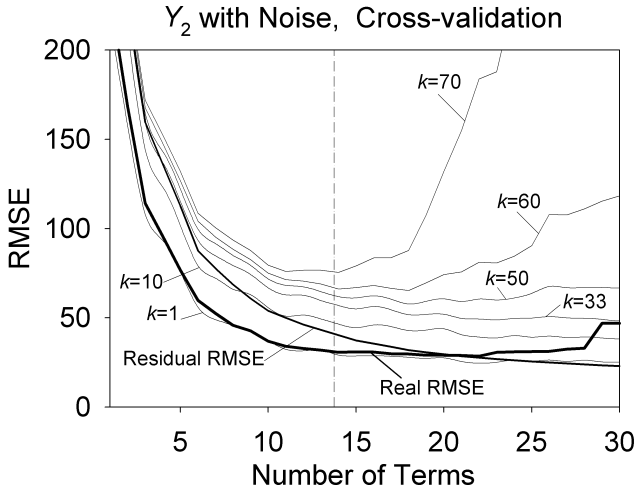

Figure 16. Cross validation scores of approximation Y_2.

In this case the optimal number of terms r=14. As can be seen in figure 16 the curve of residual RMSE used by approach of Eglais has not a distinct break. The leave-one-out cross validation (k=1) and ten-fold bootstrap (k=10) give wrong suggestions for choosing of r (too large values). A 1/3 holdout (k=33) suggests r to be about 20. Repeated holdout with 60=100-2×20 points left for confirmation (k=60) gives correct value r=14.

7.5. Robot arm function

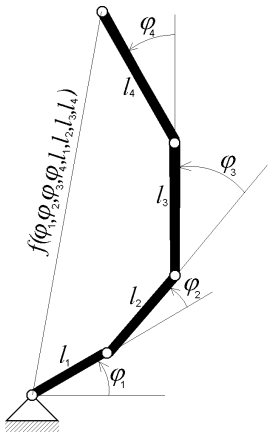

Figure 17. Robot arm

A function commonly used in the neural network literature is the robot arm function. A robot arm with 4 segments is considered. The shoulder of the arm is fixed at the origin in the (u,v)-plane. The segments of this arm have lengths l_1, l_2, l_3, and l_4. The first segment is at angle φ_1 with respect to the horizontal coordinate axis of the plane. For k=2,3,4, segment k makes angle φ_k with segment k-1. The end of the robot arm is at (see figure 17):

$$u = \sum_{j=1}^{4} l_j \cos\left(\sum_{k=1}^{j} \varphi_k\right), \qquad v = \sum_{j=1}^{4} l_j \sin\left(\sum_{k=1}^{j} \varphi_k\right), \tag{39}$$

and the response function F_r is the distance from the end of the arm to the origin expressed as a function of 8 variables φ_k ranging over [0,2π] and l_k ranging over [0,1], (An and Oven, 2001):

$$F_r(\varphi_1,\varphi_2,\varphi_3,\varphi_4,l_1,l_2,l_3,l_4,) = (u^2+v^2)^{1/2}. \tag{40}$$

Here we are using the *Lack of Fit* (*Lof*) estimation of the accuracy proposed by An and Owen, (2001)

$$Lof_{N,B} = \frac{\frac{1}{B}\sum_{i=N+1}^{N+B}\left(F_r(x_i)-\hat{F}_r(x_i)\right)^2}{(1/B)\sum_{i=N+1}^{N+B}\left(F(x_i)-\overline{F}\right)^2}, \tag{41}$$

where *B* is the number of additional confirmation points. Unlike to estimation, proposed in (an and Owen, 2001), we are employing additional points, not last of *N* points what use An and Owen. Therefore we should obtain more pessimistic estimation of accuracy of prediction.

This estimate can in unfavorable cases exceed one. When this happens, the interpretation is that a simple model predicting of the function by its global average is more accurate than our approximation.

An and Owen uses products of univariate Legendre polynomials of continuous variable as base functions and coefficients for approximation are obtained by Monte Carlo integration method, which needs a large number (>100000) observations. Any way, the base functions are not ranked. Using 4065 base functions An and Owen obtained *Lof*=19.2% and give the interpretation, that the chosen domain is too large for such a local approximation with Legendre polynomials. An and Owen, 2001 give conclusion that *"Polynomial basis functions do not seem to be well suited for the robot arm functio, over such a large range. Some failure of this type are inevitable for a high dimensional approximation method, but at least the quasi-regression method gives a clear indication of such failure having happened."*.

Here will be shown that ranking of orthogonal base functions built from discrete orthogonal polynomials can give much better result. So, figure 18 shows the *Lack of fit* versus *Number of Terms* in regression function.

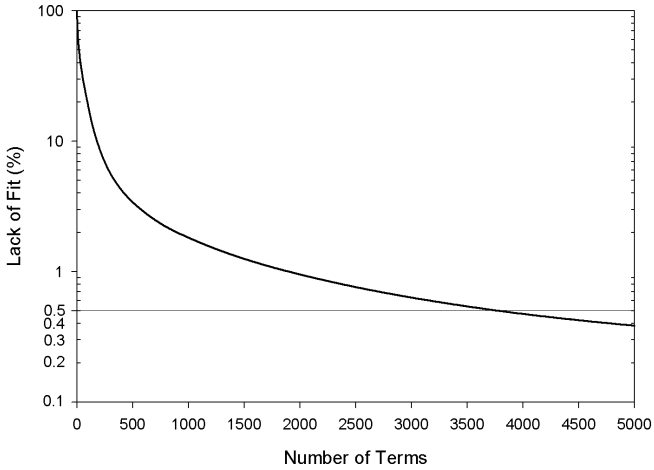

Figure 18. Lack of Fit versus Number of Terms for robot arm function approximation.

The first approximation attempt showed that F_r are independent of first variable – angle φ_1 (it is clearly seen also from figure 17), therefore, later we are using only 7 variables. Results presented in figure 18 was obtained by using base functions with limitations: maximum degree d=6 for variables φ_2, φ_3, φ_4, and d=2 for others variables; maximum order m=26. Total number of base functions N_f=27783, total number of regular grid points – 823543. Here the values of original function in grid points was calculated directly without local approximation. Value of *Lof* obtained with best 5000 functions was 0.385%. To solve the problem about 2 hours computer time on 800 MHz Pentium PC was necessary. It is clear that approximation with thousands of terms is not very good result, but this example shows that ranking of orthogonal polynomials of discrete variable can be more effective than use of large base of unranked functions.

Conclusions

The proposed method for metamodel building using tensor products of univariate orthogonal polynomials of discrete variable as base functions combined with function selection on the basis of augmented experiment (obtained with local smoothing methods) can be efficient for problems with small and medium number of independent variables (2 to 7). The method gives very accurate results in comparison with known approximation methods: Multivariate Adaptive Regression Splines, Kriging, Radial Basis Functions, Second Order Polynomial Approximation and Approximation with Rational functions. Proposed space filling designs of experiments also are effective for augmenting of experiments as well as for direct approximation.

Disadvantage of this approach is that it is relatively time-consuming for model building in the case when number of independent variables is more than 5-6. However it is very fast in response prediction when the model is built. It is very important for using the metamodel in optimization.

For the choice of optimal number of terms in the regression function the suggested from many authors leave-one-out cross validation, 1/3-holdout and ten-fold bootstrap methods can give miscalculation. Therefore, after first attempt (1/3-holdout or ten-fold bootstrap test) suggested number of terms r, the repeated holdout test with N-2r confirmation points is recommended.

For future improvements of this approach the methods to speed-up the model selection in the case of large number of variables should be developed. Example of *Robot arm function* shows that ideas of quasi-regression (An and Owen, 2001) using orthogonal polynomials of discrete variable (instead of Legendre polynomials of continuous variable) and Monte Carlo type estimation of required coefficients could be promising.

Acknowledgments

The authors gratefully acknowledge professor Wei Chen and graduate researcher Ruichen Jin at University of Illinois, Chicago for providing the authors with the test results obtained by Kriging, Radial Basis Functions and MARS methods. Thanks are also to the three anonymous reviewers for the valuable suggestions and comments which greatly helped in improving the first version of the paper.

References

T. Allen and L. Yu, “Low Cost Response Surface Methods for and from Simulation Optimization,” in *Proceedings of the 2000 Winter Simulation Conference*, J.A. Joines, R. R. Barton, K. Kang and P. A. Fishwick eds., Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, vol. 2, 2000, pp. 704-714.

J. An and A. Owen, “Quasi-regression,” *Journal of Complexity*, 2001, published electronically August 30, 2001, pp.1-20.

A. C. Atkinson and V. V. Fedorov, “Optimal Design: Experiments for Discriminating Between Several Models,” *Biometrika*, 62, 1975, pp. 289-303.

P. Audze and V. Eglais, “New approach to planning out of experiments,” in *Problems of Dynamics and Strength* vol. 35, E. Lavendel ed., Zinatne, Riga, 1977, pp. 104-107 (in Russian).

J.-F.M. Barthelemy and R. T. Haftka, “Approximation concepts for optimum structural design – a review,” *Structural Optimization*, 5, 1993, pp. 129-144.

A. J. Booker, J. E. Dennis Jr., P. D. Frank, D. B. Serafini, V. Torczon and M. W. Trosset, "A Rigorous Framework for Optimization of Expensive Functions by Surrogates," Structural Optimization, 17 (1), 1999, pp. 1-13.

G. E. P. Box and N. R. Draper, *Empirical model-building and response surfaces*, New-York, John Wiley & Sons, 1987.

W. S. Cleveland and S. J. Devlin, “Locally weighted regression: An approach to regression analysis by local fitting,” *Journal of the American Statistical Association*, 83, 1988, 596-610.

N. Dyn, D. Levin and S. Rippa, "Numerical Procedures for Surface Fitting of Scattered Data by Radial Basis Functions," SIAM Journal of Scientific and Statistical Computing, Vol. 7, No. 2, 1986, pp. 639-659.

S. Efromovich, “On orthogonal series estimators for random design nonparametric regression,” in *Computing Science and Statistics. Proceedings of the 24th Symposium on the Interface*, 1992, pp. 375-379.

B. Efron, "Estimating the error rate of a prediction rule: improvement on cross-validation," *Journal of the American Statistical Association*, 78 (382), 1983, 316-330.

V. Eglais, “Approximation of data by multi-dimensional equation of regression,” in *Problems of Dynamics and Strength* vol. 39, E. Lavendel ed., Zinatne, Riga, 1981, pp. 120-125 (in Russian).

M. Abramowitz and I. A. Stegun, eds., *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*, vol. 55 of National Bureau of Standards Appl. Math. Series, U.S. Government Printing Office, Vashington D. C., 1964.

P. L. Chebyshev, "Interpolation of equidistant values," *Complete Works*, USSR Academy of Sciences Publishing House, Moscow-Leningrad, 1948; French Transl. Oeuvres, Chelsea Pub Co., New York 1962.

J. Fan and I. Gijbels, *Local Polynomial Modelling and its Applications*, Chapman and Hall, London, 1996.

H. Faure, "Discrépance de suites associées à un système de numération (en dimensions)," *Acta Arith.* , 41, 1982, pp. 337-351.

J. H. Friedman, "Multivariate Adaptive Regression Splines," *The Annals of Statistics*, Vol. 19, No. 1, 1991, pp. 1-67.

A. A. Giunta, J. M. Dudley, R. Narducci, B. Grossman, R. T. Haftka, W. H. Mason, and L. T. Watson, "Noisy Aerodynamic Response and Smooth Approximations in HSCT Design," *Proceedings of the 5th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, Panama City Beach, FL ,Sept. 1994, pp.1117–1128.

R. T. Haftka, E. P. Scott, and J. R Cruz, "Optimization and experiments: A survey," *Appl. Mech. Rev*., 51 (7), 1998, pp. 435-448.

J. H. Halton, "On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals," *Numer. Math*., 2, 1960, pp. 84-90.

R. L. Hardy, "Multiquadratic Equations of Topography and Other Irregular Surfaces," J. Geophys. Res., Vol. 76, 1971, pp. 1905-1915.

W. Hock and K. Schittkowski, *Test Examples for NonlinearPprogramming Codes*, New York, Springer-Verlag, 1981.

R. Jin, W. Chen and T. Simpson, "Comparative Studies of Metamodeling Techniques under Multiple Modeling Criteria", in press, *Journal of Structural & Multidisciplinary Optimization* (accepted January 2001). also presented in *8th AIAA/NASA/USAF/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, paper no. AIAA-2000-4801, Long Beach, CA, September 6-8, 2000.

R. Jin, X. Du and W. Chen, "The Use of Metamodeling Techniques for Optimization under Uncertainty", accepted by 2001 *ASME Design Automation Conference*, Paper No. DAC21039, Pittsburgh, PA, September 9-12, 2001.

M. E. Johnson, L. M. Moore and D. Ylvisaker, "Minimax and Maximin Distance Designs," *Journal of Statistical Planning and Inference*, Vol. 26, No. 2, 1990, pp. 131-148.

A. I. Khuri and J. A. Cornell, *Response Surfaces: Designs and Analyses*, New York, Marcel Dekker, 1996.

Kleijnen J.P.C. and W. Van Groenendaal (1992), *Simulation: a Statistical Perspective*, Wiley, Chichester.

J. R. Koehler and A. B. Owen, "Computer Experiments," Handbook of Statistics (Ghosh, S. and Rao, C. R., eds.), Elsevier Science, New York, 1996, pp. 261-308.

R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," *Proceedings of IJCAI-95*, C. S. Mellish, (ed.), Morgan Kaufmann, 1995, pp 1137-1143.

G. V. Kreinin, V. I. Segeyev, I. N. Statnikov, and I. T. Chernyavski, "Experiment design based on LP-grid approach for solution of optimum design problems," in: *Computer Modeling in Mechanical Engineering*, Nauka, Moscow, 1976 (in Russian).

G. K. Krug and Yu. A. Sosulin, *Experiment Design in Problems of Identification and Extrapolation*, Nauka, Moscow, 1977 (in Russian).

M. D. McKay, W. J. Conover and R. J. Beckman, (1979). "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code," *Technometrics*, 21(2), 1979, pp. 239–245.

R. D. Meyer, D. M. Steinberg and G. Box, "Follow-Up Designs to Resolve Confounding in Miltifactor Experiments," *Technometrics*, 38, 1995, pp. 303-313.

T. J. Mitchell and M. D. Morris, "Bayesian Design and Analysis of Computer Experiments: Two Examples," *Statistica Sinica*, Vol. 2, 1992, pp. 359-379.

A. W. Moore, M. S. Lee, "Efficient algorithms for minimizing cross validation error," in *Proceedings of the Eleventh International Conference on Machine Learning*, New Brunswick, NJ: Morgan Kaufmann, 1994, pp. 190-198.

R. H. Myers and D. C Montgomery, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, New York, Johm Wiley & Sons, 1995.

H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*, SIAM, Philadelphia, 1992.

A. F. Nikiforov and V.B. Uvarov, *Special Functions of Mathematical Physics*, Birkhäuser-Verlag, Basel, 1988.

A. F. Nikiforov, S. K. Suslov and V. B. Uvarov, *Classical orthogonal polynomials of a discrete variable*, Springer-Verlag, Berlin-Heidelberg-New York, 1991.

J.-S. Park, "Optimal Latin-Hypercube Designs for Computer Experiments," *Journal of Statistical Planning and Inference*, Vol. 39, No. 1, 1994, pp. 95-111.

M. J. D. Powell, "Radial Basis Functions for Multivariable Interpolation: A Review," *Algorithms for Approximation (Mason, J. C. and Cox, M. G., eds.)*, Oxford University Press, London, 1987, pp.

F. Pukelsheim and J. L. Rosenberger, "Experimental Designs for Model Discrimination," *Journal of the American Statistical Association*, 88, 1993, pp. 642-649.

R. Rikards, "Elaboration of Optimal Design Models for Objects from Data of Experiments," in *Optimal Design with Advanced Materials*, Proceedings of the IUTAM Symposium, Lungby, Denmark, 18 - 20 August, 1992, P.Pedersen ed., Elsevier Science Publishers, 1993, pp. 148-162.

R. Rikards, A. K. Bledzki, V. Eglais, A. Chate, and K. Kurek, "Elaboration of Optimal Design Models for Composite Materials from Data of Experiments." *Mechanics of Composite Materials* (published from Mekhanika Kompozitnykh Materialov by Plenum Press Corporation, Consultants Bureau, New York), 28 (4), 1992, pp. 295-304.

R. Rikards, A. Chate, and G. Gailis, "Identification of elastic properties of laminates based on experiment design", *Int. J. Solids and Structures*, Vol. 38, 2001, pp. 5097-5115.

I. M. Sobol', "Uniformly distributed sequences with an additional property of uniformity," *Zh. Vychisl. Mat. i Mat. Fiz.* 16 ,1976, pp. 1332-1337, 1375 (Russian); English translation in U.S.S.R. *Comput. Math. and Math. Phys.* 16, 1976, pp. 236-242.

J. N. Srivastava, "A Critique of Some Aspects of Experimental design," in *Handbook of Statistics*, S. Ghosh and C. R. Rao eds., 13, 1996, pp. 309-342.

B. Tang, "Orthogonal Array-Based Latin Hypercubes," *Journal of the American Statistical Association*, Vol. 88, No. 424, 1993, pp. 1392-1397.

M. A. Tanner. *Tools for statistical Inference. Observed data and data augmentation methods*, Springer-Verlag, New York, N.Y., 1991.

V. V. Toropov, "Simulation approach to structural optimization," *Structural Optimization*, 1, 1989, pp. 37-46.

G. Venter, V. Balabanov and J. H. Garcelon, "Constructing High Quality Response Surface Approximations Using a Genetic Algorithm" in *4th EUROMECH, Book of Abstracts I*, Solid Mechanics Conference, Metz, France, June 26-30, 2000, p. 207.