Research paper

Struct Multidisc Optim 28, 146–155 (2004)
DOI 10.1007/s00158-004-0438-4

A method for using legacy data for metamodel-based design of large-scale systems

A. Srivastava, K. Hacker, K. Lewis, and T.W. Simpson

Abstract Despite a steady increase in computing power, the complexity of engineering analyses seems to advance at the same rate. Traditional parametric design analysis is inadequate for the analysis of large-scale engineering systems because of its computational inefficiency; therefore, a departure from the traditional parametric design approach is required. In addition, the existence of legacy data for complex, large-scale systems is commonplace. Approximation techniques may be applied to build computationally inexpensive surrogate models for large-scale systems to replace expensive-to-run computer analysis codes or to develop a model for a set of nonuniform legacy data. Response-surface models are frequently utilized to construct surrogate approximations; however, they may be inefficient for systems having with a large number of design variables. Kriging, an alternative method for creating surrogate models, is applied in this work to construct approximations of legacy data for a large-scale system. Comparisons between response surfaces and kriging are made using the legacy data from the High Speed Civil Transport (HSCT) approximation challenge. Since the analysis points already exist, a modified design-of-experiments technique is needed to select the appropriate sample points. In this paper, a method to handle this problem is presented, and the results are compared against previous work.

Key words approximation methods, design of experiments, kriging, response surface methodology

Received: 24 April 2002
Revised manuscript received: 7 April 2004
Published online: 8 July 2004

A. Srivastava[1], K. Hacker[1], K. Lewis[1,✉], and T.W. Simpson[2]

[1] Department of Mechanical and Aerospace Engineering, University at Buffalo, 1010 Furnas Hall, Buffalo, NY 14260
e-mail: kelewis@eng.buffalo.edu
[2] Department of Mechanical & Nuclear Engineering and Department of Industrial & Manufacturing Engineering, Pennsylvania State University, 329 Leonhard Building, University Park, PA 16802
e-mail: tws8@psu.edu

Nomenclature

DOE – design of experiments

MLE – maximum-likelihood estimate

MSE – mean square error

n_s – number of sample points

$\hat{\sigma}^2$ – variance estimate

$\mathbf{x}$ – vector of design variables

$\hat{y}$ – predicted response value at untried $\mathbf{x}$

1 Introduction

Practical engineering of multidisciplinary, large-scale systems analysis often requires integrating and running complex, computationally expensive computer analysis and simulation codes such as structural mechanics, aerodynamic analysis, and fluid dynamics models. Despite continuing increases in computing power, the complexity of these codes seems to keep pace with computing advances. Analysing such systems requires running computational parametric analyses. These analyses are often employed during optimization, making multi-objective, multidisciplinary optimization and concept exploration time consuming to say the least (Simpson *et al.* 1998, 2002). Also, since these codes are so computationally expensive to execute, whenever they are run, their data is typically collected into a legacy database without much, if any, structure. The motivation behind collecting the legacy data is typically to prevent engineers from having to run the same or similar analyses over and over again. An additional use of this legacy data could be to construct metamodels of the large-scale system, which is the primary motivation in this paper.

A departure from traditional parametric design analysis and single objective optimization approaches is needed in the early stages of design to facilitate the effective solution of multidisciplinary and multi-objective design problems. The limitations of traditional parametric design and analysis become apparent when applied to sys-

tems with high dimensionality. The problem is referred to as the problem of size whereby the approach sufficient for small, simple problems becomes inefficient and inappropriate as the size of the problem is increased. There are two specific reasons identified relating to the problem size that cause traditional approaches to fail: the number of variables and responses and the computational expense (Koch *et al.* 1999).

1. ***Number of variables and responses***
   Comprehensive parametric analyses becomes extremely time intensive as the number of variables in the design problem increases. Commonly, a one-at-a-time approach is used, during which one of the variables is varied while the others are held fixed. This approach is inefficient for a large number of variables and does not provide sufficient insight into possible interaction effects between the variables. In addition, as the number of constraints and objectives increases, it becomes increasingly difficult to identify optimal or good settings for the design variables, and parametric analysis becomes increasingly inefficient.
2. ***Computational Expense***
   The analysis of complex, multidisciplinary systems such as aircraft or automobiles often requires high fidelity computer analyses that can be computationally expensive to perform. For instance, engineers at Ford Motor Company report that a crash simulation of a full passenger car takes 36–160 hours to compute (Gu 2001), and engineers at The Boeing Company frequently use simulation codes that can take 15 to 18 hours for analysis of some design applications (Simpson and Meckesheimer 2004). This expense is the motivation to store information from the analysis runs as much as possible into legacy databases so that other users can use the data without having to run the analysis again (Bennett *et al.* 1998; Feynes *et al.* 2002). In addition, there are legacy data applications where the amount of computation per data item is relatively small, but the scale of the database makes the total computational requirement equivalent to that of typical computationally intensive applications. Examples of large databases would include the United States Environmental Protection Agency's environmental data collection efforts, data associated with the human genome effort, and NASA's earth observing system, which is projected to create ½ terabyte of primary data per day. In these instances, surrogate approximation models may be constructed and used in lieu of the actual analysis codes providing many benefits (Simpson *et al.* 2001):

   – They yield insight into the relationship between (output) responses, y, and (input) design variables, **x**.
   – They provide fast analysis tools for optimization and design space exploration since the cheap-to-run approximations replace the expensive-to-run computer analyses.
   – They facilitate the integration of discipline-dependent analysis codes.

There are five major steps in the application of a surrogate approximation model: 1) Identify design space, 2) Select an experimental design, 3) Sample design space, 4) Build predictive model, and 5) Explore the design space. These five steps lead to a solution, which may be further optimized or used for constructing a more enhanced approximation if the accuracy is not sufficient. Figure 1 illustrates this process, and the steps are elaborated as follows.

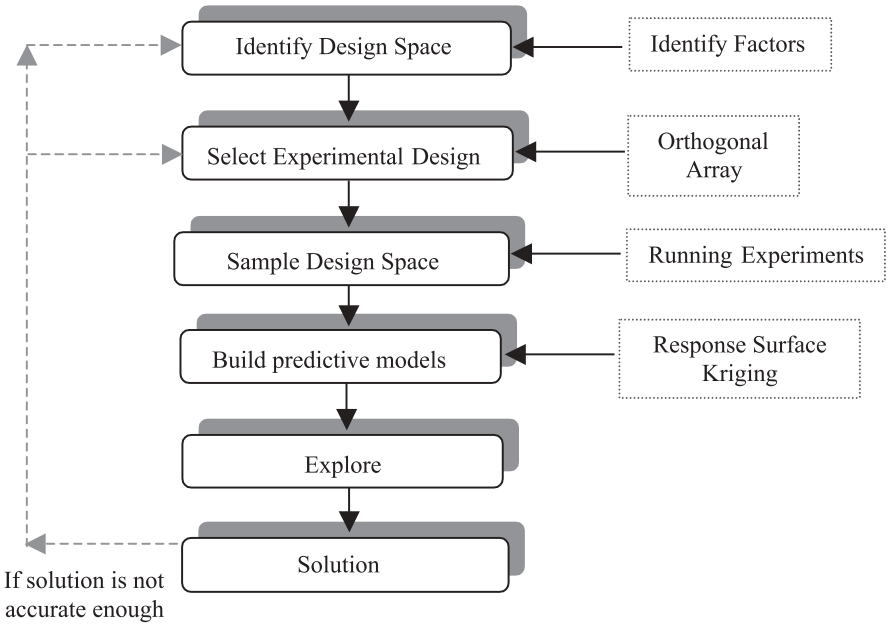

Fig. 1 Steps to construct approximation model

1. ***Identifying design space***
   In this step the designer identifies the area of interest and selects the design factors to study. Often, if there are numerous design variables, experiments are performed to "screen out" the less important effects (see, Myers and Montgomery 1995).
2. ***Selecting an experimental design***
   Sample points are chosen in the design space to investigate efficiently the relationship between the design factors and the responses and generate a predictive model. Details are provided in Sect. 2.2.
3. ***Sample design space***
   In this step, the computer analysis or simulation code that is being approximated is performed at each of the sample points identified in Step 2.
4. ***Build predictive models***
   Using the data gathered in Step 3, a predictive model is constructed using a number of methods (e.g. response surfaces, kriging, neural networks, splines, etc.).
5. ***Exploring the design space***
   The design space can then be explored to find regions of good design or optimized to improve the performance of the system using the predictive surrogate approximation models instead of the computationally expensive analysis code, resulting in large computational savings.

The most common surrogate approximation models are linear and quadratic polynomials created by performing ordinary least-squares regression on a set of analysis data. These polynomial models are known as response-surface models, (see Myers and Montgomery (1995) for further details). These methods are popular because they are simple to create and provide compact and explicit functional relationships between the responses and independent variables over the range of interest.

Response-surface methods were initially introduced to create smooth approximations of response data in the presence of random error. Therefore, response-surface approximations may not be accurate when approximating *deterministic* computer analyses due to lack of random error in the computer model (Simpson *et al.* 2001). In addition, first- and second-order response-surface models have limited capability to model non-linear and complex functions of arbitrary shape. Higher-order response surfaces may be used to model more complex functions, but a large number of sample points may be necessary to estimate all of the polynomial coefficients. Hence, many researchers recommend the use of *sequential* response-surface modelling approaches using move limits (see e.g. Toropov *et al.* (1996)) or trust region approaches (see e.g. Rodriguez *et al.* (1998)). For example, the concurrent subspace optimization procedure uses data generated during concurrent subspace optimization to develop response-surface approximations of the design space that form the basis of the subspace coordination procedure (Renaud and Gabriele 1994, 1991; Wujek *et al.* 1996). Also, the hierarchical and interactive decision refinement method recursively decomposes the design space into subregions and fits each subregion with a separate model during design space refinement (Reddy 1996).

Many of the above mentioned sequential approaches have been developed for single objective optimization applications. Much of engineering design is multi-objective in nature, however, and it is often difficult, if not impossible, to isolate a small region of good design which can be accurately represented by a low-order polynomial response-surface model. In (Koch *et al.* 1997), the difficulties encountered when screening large numbers of variables in problems with multiple objectives as a part of the response-surface approach are discussed. In (Barton 1992), it is pointed out that the response region of interest will almost never be reduced to a "small neighbourhood" which is favourable for all objectives during multi-objective optimization. Hence, this leads to a need for alternative approximation techniques with sufficient flexibility to build accurate global approximations of the design space and which are suitable for modelling deterministic computer experiments (Simpson *et al.* 1998).

Kriging, also known as design and analysis of computer experiments (DACE), is an alternate approximation method that may be a more statistically accurate and consistent method for approximating deterministic computer experiments (cf., Sacks *et al.* 1989; Simpson *et al.* 2001). Originally developed for spatial statistics and geostatistics, kriging is an interpolative approximation based on an exponentially weighted sum of the sample data. Furthermore, kriging models are very flexible due to the wide range of correlation functions that can be chosen for building the model. Depending on the type of correlation function used, a kriging model can either "honour the data," providing an exact interpolation of the data, or "smooth the data" in the presence of numerical noise (Cressie 1993). In this paper, we are concerned with spatial prediction, and we assume that the data are not correlated temporarily. Also, the details of kriging are not the focus of this paper and are given elsewhere (see e.g. Sacks *et al.* 1989; Koehler and Owen 1996; Simpson *et al.* 2002).

The High Speed Civil Transport (HSCT) problem studied in this paper is an example of a large-scale problem that requires extensive computer analyses and therefore is a system where a significant amount of legacy data exists. This legacy data of past analysis runs is part of the NASA Multidisciplinary Design Optimization Test Suite: ⟨http://mdob.larc.nasa.gov/mdo.test/⟩. The HSCT approximation challenge involves applying different approximation techniques to solve the problem of approximating the objective and constraints in the most efficient manner. In this paper we develop a modified experimental design strategy for building kriging metamodels of legacy data like the HSCT example and compare the accuracy and efficiency of this method with response surface metamodels. In previous work, kriging has been compared to

a number of other metamodelling strategies using a series of small dimensional case studies (Jin *et al.* 2001). In this paper, we are investigating the effectiveness of kriging and other metamodelling strategies using a large-scale problem of very high dimensionality. We however use some of the same comparison metrics as used in (Jin *et al.* 2001).

The remainder of the paper is divided as follows. In Sect. 2, an overview of design of experiments (DOE) is presented, and in Sect. 3, the details of kriging are presented. Section 4 contains a description of the HSCT case study, followed by detailed results and closing remarks in Sects. 5 and 6, respectively.

2 DOE terminology

In the construction of an approximation model it is first necessary to sample the design space. DOE techniques allow a designer to select those points intelligently and in such a way as to produce an accurate and statistically meaningful approximation. The goal is to sample enough points in the design space to be able to capture the general behaviour of the objective function(s) and constraints while keeping the computational expense at a reasonable level.

There are two categories of experimental design, "classical" experimental designs and "space-filling" designs. Examples of commonly used classical experimental designs for constructing second-order response-surface approximations are central composite design (CCD) and Box-Behnken (BB) which are the most widely used experimental designs. Some space-filling designs are: random Latin hypercubes, orthogonal array (OA), and orthogonal array–based Latin hypercube, among others.

In the "classical" design and analysis of physical experiments, random variation is accounted for by spreading the sample points out in the design space and by taking multiple data points (replicates) using traditional central composite designs. However, the "classical" notions of blocking, replication and randomization are irrelevant when performing deterministic computer experiments (Sacks *et al.* 1989); thus, the extra sample points should be chosen to "cover" or "fill" the design space (Booker 1998). Unal (1997) also suggested that the response-surface methods using classical experimental designs may not result in a good representation of the response surface since most of the modern engineering applications are deterministic computer experiments with no random error. Thus, space-filling designs are often more appropriate because they provide better "coverage" of the design space. Simpson and Mistree (1998) verified that many space-filling experimental designs perform better than classical experiments design for problems having more than three design variables. In this work, randomized orthogonal array designs from (Owen 1992) are employed to construct the surrogate approximation models. An overview of kriging is presented in the next section.

3 Kriging methodology

As opposed to a response surface, which is global in nature, in kriging the predictive model has both global and local components. More specifically, kriging postulates a combination of a global model and departures of the following form:

$$y(\mathbf{x}) = f(\mathbf{x}) + Z(\mathbf{x}) \tag{1}$$

where $y(\mathbf{x})$ is the unknown function of interest, $f(\mathbf{x})$ is a known function of $\mathbf{x}$, and $Z(\mathbf{x})$ is the realization of a stochastic process with mean zero, variance σ^2, and non-zero covariance. The $f(\mathbf{x})$ term in (1) is similar to the polynomial model in a response surface and provides a "global" model of the design space. In many cases $f(\mathbf{x})$ is simply taken to be a constant term β where β is estimated from the n_s sample points (Sacks *et al.* 1989).

While $f(\mathbf{x})$ "globally" approximates the design space, $Z(\mathbf{x})$ creates "localized" deviations so that the kriging model interpolates the n_s sampled data points. The $Z(\mathbf{x})$ function is a measure of the variation in the model from the sample points. It is obtained using an iterative optimization procedure to determine the maximum-likelihood estimates of the parameters which are used to fit a kriging model; details of this optimization procedure can be found in (Simpson *et al.* 1998; Sacks *et al.* 1989). In our work, we use a simulated annealing algorithm from (Goffe *et al.* 1994) to perform the optimization of this variance. While slow for large numbers of design variables, the algorithm is very robust at finding the maximum-likelihood estimates.

Finally, a note on model validation is worthwhile. With response-surface models, R^2 values and residual plots can be used to assess model accuracy (cf., Myers and Montgomery 1995). Since kriging approximation models typically interpolate the data, there are no residuals, and alternative validation measures must be used. Additional data points can be used to check the validity of the approximation models by computing relevant error measures using (2), (3) and (4). The maximum absolute error, average absolute error, and root-mean-square error (MSE) can be computed from additional validation data as follows.

$$\text{max. abs. error} = \max.\{|y_i - \hat{y}_i|\}_{i=1,\ldots,n_{error}} \tag{2}$$

$$\text{avg. abs. error} = \frac{1}{n_{error}} \sum_{i=1}^{n_{error}} |y_i - \hat{y}_i| \tag{3}$$

$$\text{root MSE} = \sqrt{\frac{\sum_{i=1}^{n_{error}} (y_i - \hat{y}_i)^2}{n_{error}}} \tag{4}$$

In (2)–(4), y_i is the true value, $\hat{y}_i$ is the predicted value of y_i from the kriging model, and n_{error} is the number of extra points used to assess the error in the kriging model. Once the model has been fitted, validated, and deemed sufficiently accurate, it can be used for prediction. The application of the kriging methodology to a large-scale problem is described in the next section.

4
Case study: high speed civil transport approximation challenge

In order to demonstrate the effectiveness of kriging as an approximation method for unstructured legacy data of large-scale systems we consider the application of kriging (and response surfaces) to readily available legacy data for the High Speed Civil Transport (HSCT) aircraft. The design of the HSCT is a complex and challenging problem in multidisciplinary design optimization (MDO). The design goal of the High Speed Civil Transport is to minimize take-off gross weight for a range of 5500 nautical miles and cruise speed of mach 2.4 with a capacity of 251 passengers (Giunta *et al.* 1997). The model was created at the Multidisciplinary Analysis and Design (MAD) Center for Advanced Vehicles at Virginia Tech. Twenty-five design variables describe the geometry of the aircraft and can be loosely grouped into five categories: wing planform, airfoil shape, tail areas, nacelle placement, and fuselage shape. In addition, one of the variables, mission fuel, defines the cruise mission. There are sixty-seven geometry, structural, performance, and aerodynamic constraints. Due to the large number of design variables and the large computational expense associated with each analysis (hours on a supercomputer), it is infeasible to apply traditional derivative or pattern-search optimization methods to this problem.

The only practical way to perform optimization studies on a problem of this complexity is to create approximate metamodels for the objectives and constraints. Analyses are run at a small number of sample configurations, and response models are created from this data. These models are then used to perform MDO and arrive at an optimal design configuration.

Since configuration and response data for 2490 different legacy HSCT analyses are available on the NASA MDOB test suite, no additional analysis is necessary. The data from one analysis consists of values for the 26 independent design variables (X_1–X_{26}) and the corresponding values of the objective function (take-off gross weight (TOGW)) and 68 inequality constraints (Giunta *et al.* 1997); our focus here is on approximating the objective function. The challenge is to choose up to 500 analysis points to construct surrogate approximations and use the remaining points to check the accuracy of the approximation. Since our focus is on the approximation techniques as opposed to the actual optimization process, we do not present details as to the physical significance of the design variables or constraints; the reader is directed to (Giunta *et al.* 1997) for more details. We outline the specific steps we followed in the approximation of the desired response, namely, the objective function for the HSCT.

Selection of experimental design

The results of the 2490 experiments are on the NASA MDOB test suite, which is available online at: ⟨http://mdob.larc.nasa.gov/mdo.test/⟩. The authors are not aware of the process by which this legacy data set was created, as the large amount of data and the 25 dimensional design space make it difficult to identify the technique that was used to generate the data. It is unstructured and nonuniform, which is characteristic of most legacy data sets. This problem is overcome by using the modified experimental design method detailed next.

As mentioned above, 500 or fewer points are to be selected from the available 2490 points. A randomized orthogonal array (Owen 1992) is constructed. The size of orthogonal array is based upon the number of levels selected and the number of points. The larger the number of levels, the better the expected accuracy of the approximation. Obviously, the number of levels cannot exceed the number of levels actually present in the data. Three

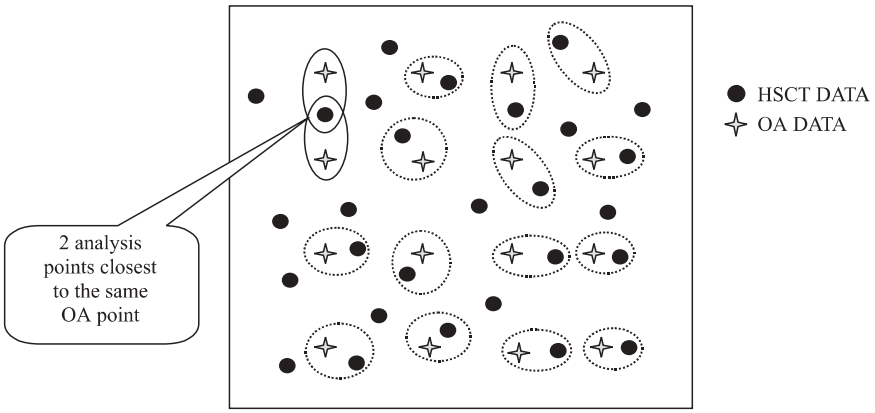

Fig. 2 Construction of pseudo-orthogonal array

Table 1 Different orthogonal array (OA) cases for 25 design variables

Case	Number of levels	Number of points
Case 1	5	250
Case 2	7	686
Case 3	9	1458

Table 2 Comparison of original and pseudo-orthogonal array

Case number	Number of points in original orthogonal array	Number of points in pseudo-orthogonal array
Case 1	250	126
Case 2	686	283
Case 3	1458	372

cases were considered with increasing numbers of levels; the sizes of the three orthogonal arrays are 250×25, 686×25, 1458×25 where the first parameter is the number of experiments required and the second parameter is the number of rows or design variables. The number of variable levels for each array is given in Table 1.

Since it is unlikely that the combination of levels in a particular row of the OA matches one of the analysis points exactly, the goal is to find the "nearest neighbour" of each point in the OA, i.e. the analysis point which is closest to each point in the 25-dimensional design space sampled by the OA as illustrated in Fig. 2. This nearest neighbour is then inserted into the OA in place of the original OA row.

There is a possibility, however, that one legacy analysis point may be the closet point to two rows in the OA, as shown in the top left portion of Fig. 2. This would mean that this analysis point would take the place of two OA rows. Since an OA cannot have two rows that are the same, duplicate rows are removed from the sample data set, and the remaining rows represent a "pseudo-orthogonal array" that is not full. The differences between the size of the original orthogonal arrays and pseudo-orthogonal arrays are summarized in Table 2.

The next and final step is to build approximation models using the sample data. The response surface and kriging models are constructed using three different sets of data shown in Table 2 (126, 283 and 372 points). The approximation models for the three cases are discussed next.

5 Results

Kriging approximation models were constructed for the three cases using the method described in Sect. 4. For comparison purposes, quadratic response-surface models were also constructed. The results for the three cases are presented in Figs. 3, 4, and 5. Figure 3 shows the percentage error for both the kriging and response surface metamodels when 126 analyses were used to create the approximation (Case 1). Note that the scale for the upper plot differs by two orders of magnitude from the scales in the middle and lower plots. From the upper plot of the error at all 2490 analysis points, it can be seen that

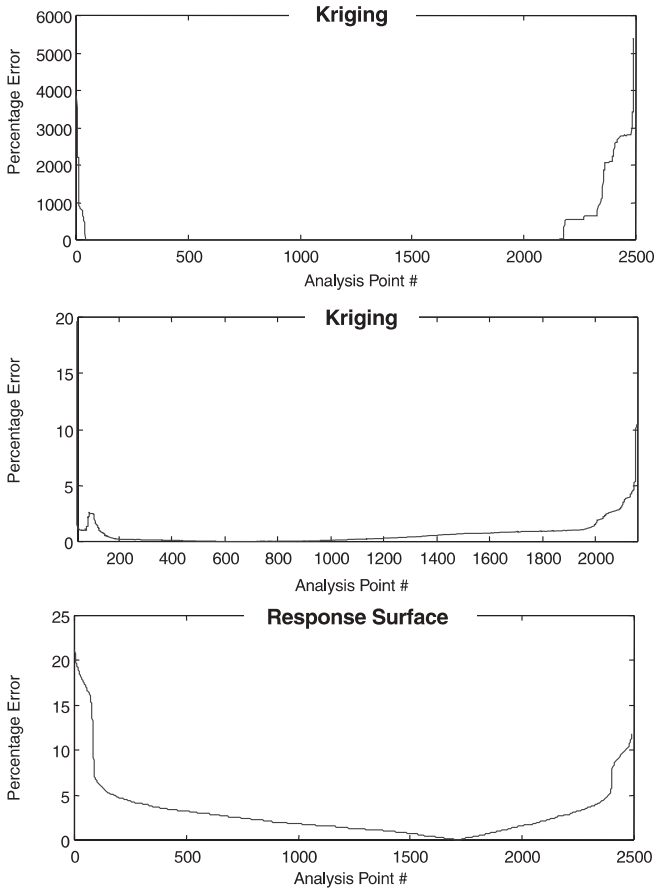

Fig. 3 Percentage-error analysis for Case 1

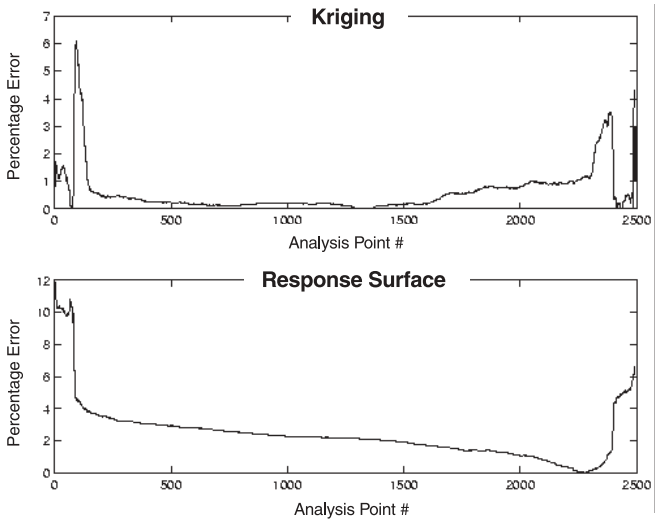

Fig. 4 Percentage-error analysis for Case 2

Table 3 Summary of results for HSCT problem

	Kriging			Response surface		
Number of points	Average percentage error	Max percentage error	Root-mean-square error percentage	Average percentage error	Max percentage error	Root-mean-square error percentage
126 points	207.83	5379.00	667.66	3.15	21.23	4.51
283 points	0.59	6.12	1.00	2.45	11.91	3.04
372 points	0.17	2.02	0.24	1.15	7.81	1.41

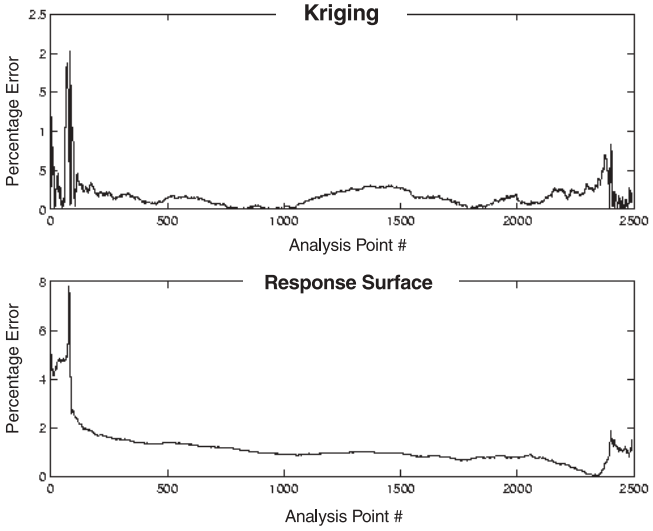

Fig. 5 Percentage-error analysis for Case 3

the errors in the kriging metamodel are very high for two regions. The middle plot shows the error in the kriging metamodel only for analysis points 42 through 2160, which is much lower. The high errors in portions of the kriging result from the large dimensionality of the design space and the small number of points used to sample it. The points in the original OA were not close to some of the points at the extremes of the legacy data set, resulting in large errors in these regions. Since kriging is an interpolating model, we would expect at least 126 points where the error is zero (the points used in the approximation), and a number of points with zero error are clear in the middle plot of Fig. 3. The response surface is more accurate over the entire data set, most notably at the extremes, because of the smoothing behaviour of the surface. Compared to the kriging metamodel, there appears to only be a small handful of points around analysis point 1700 where the error is zero. This can be seen in the lower plot.

In Case 2, the response-surface and kriging metamodels were constructed using 283 points. Figure 4 shows a plot of the percentage-error analysis. The errors for kriging are now on the same order with the response surface. The kriging metamodel slightly outperforms the response surface on average across all the legacy data points. The behaviour of the response-surface metamodel is smoother than the kriging metamodel, as would be expected. The oscillations in the kriging metamodel are due to the interpolation nature of kriging and the small number of points being used.

In Case 3 the number of points used in the approximation was further increased to 372. The percentage-error analysis is shown in Fig. 5. The increase in the number of sample points has further reduced the error for both the response-surface and kriging models. The kriging model still behaves a little more erratically, due to the nature of its construction. A number of points that have zero error in the kriging model are apparent as well, as expected. A summary of the results for all three cases is presented in Table 3. The response-surface models produce consistent results with a reasonable degree of accuracy. Kriging works well if a sufficient number of points are available to create the model. Note that 283 and 372 points are still a small fraction of the 2490 points in the overall legacy data set. In addition, the response surface produced a better approximation than the kriging for Case 1 when only a small number of data points were used. In the next section, some closing remarks are made based on these results.

6 Closing remarks

In this section, the results of the HSCT case study are discussed to draw some overall conclusions as to the appropriateness of both kriging and response surfaces in modelling a legacy system problem of this complexity and large dimensionality.

The results clearly show that kriging is a viable alternative to response-surface modelling. In previous studies, kriging has been used to form predictive models of relatively simple systems (< 10 design variables), see, e.g., (Sacks *et al.* 1989; Simpson 2002; Simpson *et al.* 1998; Giunta *et al.* 1998; Sasena 1998). In this work, however, kriging has been applied to a problem on a much larger scale that is more consistent with many practical problems of interest and has proven to be quite effective, particularly when relatively few sample points are available for use. Its effectiveness does diminish if the number of sample points gets too small. When only 126 points are used, there are areas of the data set that are not well mod-

Table 4 Comparison with previous work

Approximation technique	Average % error	Max % error	Number of points
Kriging	0.17	2.02	372
Response surface	1.15	7.81	372
Response surface RSD1 (Khatib et al. 1998)	0.49	2.89	500

elled. This is due to the interpolating nature and highly nonlinear behaviour of the resulting model.

In Table 4, the results from Case 3 (372 points) are compared with previous results for this problem (Khatib *et al.* 1998). In this previous work a specialized response surface was used, while in this paper a very generic response surface is used. Therefore, the better performance of the RSD1 response surface is not surprising (although 500 points are used in the RSD1 to construct the metamodel). With only 372 points, the pseudo-orthogonal approach with the kriging-based metamodel displays the lowest error levels.

The number of points used in the creation of an approximation plays an important role in determining the accuracy of the predicted response. It is noted in Tables 3 and 4 that the errors decrease significantly with an increase in the number of sample points, especially in the case of kriging; however, the response-surface approximation model gives very good performance in all three cases. Thus, response surfaces may be the preferred approximation technique when only a small number of analyses can be performed (or have been performed with legacy data), as in cases where there is large computational expense for each run, or when *sequential* experimentation is utilized to improve the accuracy of the approximation in regions of interest continuously. As the number of points increases, kriging models are more accurate at capturing the nonlinearity in the design space, at least in this case. For both methods, the accuracy of the approximations are very good, considering the number of points used and the fact that no *a priori* information was known about the shape of the underlying functions that were used in the original analyses. For large-scale legacy systems of this kind, huge computational cost savings can be realized by using surrogate-based metamodelling approaches.

The management of the data in any problem of this size and complexity is a difficult task. In the case study considered in this paper, the problem is exacerbated because of the inconsistency between the legacy data and the orthogonal arrays used to sample this data (i.e. the legacy data was nonuniform and not structured to match a particular OA). The use of the pseudo-orthogonal array concept was an effective approach to overcome the nonuniform nature of legacy data of complex system without a significant loss of accuracy.

While both approximations predict new points very quickly, an important consideration is the expense involved in constructing the approximations from the given data. The response-surface model took less than 50 seconds of CPU time on a 167 MHz Sun Ultra 1 to construct and predict the responses for all 2490 data points. In comparison, the kriging model required about 10 minutes per design variable to construct the model. Recall from Sect. 3 that an optimization problem must be solved to determine the maximum-likelihood estimates for the model parameters. For these experiments, a simulated annealing algorithm was employed to determine the parameters for the kriging models; it is a very robust but somewhat inefficient algorithm for this type of optimization (Simpson *et al.* 2001). It is also somewhat prone to generating an ill-conditioned correlation matrix, which tends to occur when points "pile up" near each other in the design space; however, that is not the case here (condition number $= 1.55 \times 10^{-7}$, which is small but not indicative of ill-conditioning, see (Martin and Simpson 2004) for details on when ill-conditioning problems arise). We believe, however, that the opposite problem is actually causing the poor fitting kriging model in Case 1, namely, there are too few sample points to cover the design space adequately. As a result, many of the theta parameters for the kriging model for Case 1 are large (> 5.0), which yields a relatively flat hyper-surface – in this case, the constant term β that is used for the f(x) portion of (1) – with many "spikes" to interpolate the data. The theta parameters for the other cases are not nearly as large, indicating a more "smoothly" fitting kriging model. Meanwhile, the response-surface model is able to capture the overall trend within the design space, be it linear or quadratic, even though only a few samples are used. Future studies should examine the process of ensuring that ample sample points are taken to achieve good coverage of the design space.

Finally, the results from this study indicate that neither of the methods clearly outperforms the other, both having advantages and disadvantages depending on the computational expense, the desired accuracy of the model and the degree of nonlinearity in the system. This work provides a support tool to designers who ultimately must decide the approximation method most appropriate for the problem under consideration.

Acknowledgements We gratefully acknowledge NSF Grant DMI-9800435 and NASA-Langley Grant NGT-152185 for their support of this work. We also thank the Center for Com-

putational Research at the University at Buffalo for their help with running the models on high-performance computing. Tim Simpson's work was supported, in part, by Dr. Kam Ng, ONR 333, through the Naval Sea Systems Command under Contract No. N00039-97-D-0042, Delivery Order No. 0242.

References

Barton, R.R. 1992: Approximation models for Simulation Input-Output Relations. In: Swain, J.J.; Goldsman, D. *et al.* (eds.) *Proceedings of the 1992 Winter Simulation Conference*, 289–299, Arlington, VA, IEEE

Bennett, J.A.; Fenyes, P.A.; Hearing, W.J.; Neal, M.O. 2002: Issues in Industrial Multidisciplinary Optimization. In *7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis & Optimization* **1**, 12–22 AIAA-98-4727, St. Louis, MO, AIAA

Booker, A.J. 1998: Design and Analysis of Computer Experiments. In *7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis & Optimization* **1**, 118–128 AIAA-98-4757, St. Louis, MO, AIAA

Cressie, N.A.C. 1993: *Statistics for Spatial Data.* New York: Wiley

Feynes, P.A.; Donndelinger, J.A.; Bourassa, J.-F. 2002: A New System for Multidisciplinary Analysis and Optimization of Vehicle Architectures. In *8th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis & Optimization*, AIAA-2002-5509, Atlanta, GA, AIAA

Giunta, A.A.; Balabanov, V.; Haim, D.; Grossman, B.; Mason, W.H.; Watson, L.T.; Haftka, R.T. 1997: Multidisciplinary Optimisation of a Supersonic Transport Using Design of Experiments Theory and Response Surface Modelling. *Aeronaut J* **101**, 347–356

Giunta, A.; Watson, L.T.; Koehler, J. 1998: A Comparison of Approximation Modeling Techniques: Polynomial Versus Interpolating Models. In *7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis & Optimization* **1**, 392–404, AIAA-98-4758, St. Louis, MO, AIAA

Goffe, W.L.; Ferrier, G.D.; Rogers, J. 1994: Global Optimization of Statistical Functions with Simulated Annealing. *J Econom* **60**, 65–100. Source code is available at http://netlib2.cs.utk.edu/opt

Gu, L. 2001: A Comparison of Polynomial Based Regression Models in Vehicle Safety Analysis. *ASME Design Engineering Technical Conferences – Design Automation Conference,* Pittsburgh, PA, ASME, Paper No. DETC2001/DAC-21063

Jin, R.; Chen, W.; Simpson, T. 2001: Comparative Studies of Metamodelling Techniques Under Multiple Modelling Criteria. *Struct Multidisc Optim* **23**, 1–13

Khatib, W.; Fleming, P.J. 1998: NASA High Speed Civil Transport Aircraft. In *7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis & Optimization,* AIAA-98-4760, St. Louis, MO, AIAA

Koch, P.N.; Allen, J.K.; Mavris, D.; Mistree, F. 1997: The Problem of Size in Robust Design. In *Advances in Design Automation, ASME Design Engineering Technical Conferences*, ASME DETC97DAC3983, Sacramento, CA, ASME

Koch, P.N.; Simpson, T.W.; Allen, J.K.; Mistree, F. 1999: Statistical Approximations for Multidisciplinary Design Optimization: The Problem of Size. *AIAA J Aircr* **36**, 275–286

Koehler, J.R.; Owen, A.B. 1996: Computer Experiments. In: Ghosh, S.; Rao, C.R. (eds.) *Handbook of Statistics.* New York: Elsevier Science, pp. 261–308

Martin, J.D.; Simpson, T.W. 2004: On the Use of Kriging Models to Approximate Deterministic Computer Models, In: Chen, W. (ed.) *ASME Design Engineering Technical Conferences – Design Automation Conference* (held in Salt Lake City, UT), under review. New York: ASME

Matheron, G. 1963: Principles of Geostatistics. *Econ Geol* **58**, 1246–1266

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

Owen, A.B. 1992: Orthogonal Arrays for Computer Experiments, Integration and Visualization. *Stat Sin* **2**, 439–452, source code available at: http://lib.stat.cmu.edu/designs/owen.small

Osio, I.G.; Amon, C.H. 1996: An Engineering Design Methodology with Multistage Bayesian Surrogates and Optimal Sampling. *Res Eng Des* **8**, 189–206

Reddy, S.Y. 1996: HIDER: A Methodology for Early-Stage Exploration of Design Space. In: Dutta, D. (ed.) *Advances in Design Automation* ASME, DETC96/DAC-1089, Irvine, CA

Renaud, J.E.; Gabriele, G.A. 1991: Sequential Global Approximation in Non-Hierarchic System Decomposition and Optimization. In: Gabriele, G. (ed.) *Advances in Design Automation – Design Automation and Design Optimization* **32**(1), 191–200, Miami, FL, ASME

Renaud, J.E.; Gabriele, G.A. 1994: Approximation in Nonhierarchic System Optimization. *AIAA J* **32**, 198–205

Rodriguez, J.F.; Renaud, J.E.; Watson, L.T. 1998: Trust Region Augmented Lagrangian Methods for Sequential Response Surface Approximation and Optimization. *ASME J Mech Des* **120**, 58–66, ASME

Sasena, M.J. 1998: *Optimization of Computer Simulations via Smoothing Splines and Kriging Approximation Models*. M.S. Thesis, Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI

Sacks, J.; Welch, W.J.; Mitchell, T.J.; Wynn, H.P. 1989: Design and Analysis of Computer Experiments. *Stat Sci* **4**, 409–435

Simpson, T.W.; Allen, J.K.; Mistree, F. 1998: Spatial Correlation Metamodels for Global Approximation in Structural Design Optimization. In *Advances in Design Automation,* DETC98/DAC-5613, Atlanta, GA, ASME

Simpson, T.W.; Mistree, F. 1998: Spatial Correlation Metamodels for Global Approximation in Computer-Based Engineering Design. In *INFORMS Seattle Fall 1998 Meeting,* Seattle, WA, INFORMS

Simpson, T.W.; Peplinski, J.; Koch, P.N.; Allen, J.K. 2001: Metamodels for Computer-Based Engineering Design: Survey and Recommendations. *Eng Comput* **17**(2), 129–150

Simpson, T.W.; Mauery, T.M.; Korte, J.J.; Mistree, F. 2001: Kriging Metamodels for Global Approximation in Simulation-Based Multidisciplinary Design Optimization. *AIAA J* **39**(12), 2233–2241.

Simpson, T.W.; Meckesheimer, M. 2004: Evaluation of a Graphical Design Interface for Design Space Visualization. *45th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics & Materials Conference,* April 19–22, AIAA-2004-1683

Toropov, V.; van Keulen, F.; Markine, V.; de Doer, H. 1996: Refinements in the Multi-Point Approximation Method to Reduce the Effects of Noisy Structural Responses. In *6th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization* **2**, 941–951. AIAA-96-4087, Bellevue, WA, AIAA

Wujek, B.; Renaud, J.E.; Batill, S.M.; Brockman, J.B. 1996: Concurrent Subspace Optimization Using Design Variable Sharing in a Distributed Computing Environment. *Concurrent Engineering: Research and Applications* **4**, 361–378, Technomic, USA