Design search and optimisation using radial basis functions with regression capabilities

A. J. Keane[1]

[1]School of Engineering Sciences, University of Southampton, Southampton, SO17 1BJ, U.K.
ajk@soton.ac.uk

Abstract

Modern design search and optimisation (DSO) processes that involve the use of expensive computer simulations commonly use surrogate modelling techniques, where data is collected from planned experiments on the expensive codes and then used to build meta-models. Such models (often termed response surface models or RSMs) can be built using many methods that have a variety of capabilities. For example, simple polynomial (often linear or quadratic) regression curves have been used in this way for many years. These lack the ability to model complex shapes and so are not very useful in constructing global RSM's for non-linear codes such as the Navier Stokes solvers used in CFD - they are, however, easy to build. By contrast Kriging and Gaussian Process models can be much more sophisticated but are often difficult and time consuming to set up and tune. At an intermediate level radial basis function (RBF) models using simple spline functions offer rapid modelling capabilities with some ability to fit complex data. However, as normally used such RBF RSM's strictly interpolate the available computational data and while acceptable in some cases, when used with codes that are iteratively converged, they find it difficult to deal with the numerical noise inevitably present. This paper describes a modification to the basic RBF scheme that allows a systematic variation of the degree of regression from a pure linear regression line to a fully interpolating cubic radial basis function model. The ideas presented are illustrated with data from the field of aerospace design.

1 Introduction

Design search an optimisation is concerned with improving designs by modifying parameters under the designer's control while at the same time meeting any constraints imposed by requirements such as structural integrity, safety, cost margins, etc. Commonly such improvements are based on the use of computer simulations. If used to justify important design changes these simulations must be accurate and they are thus often computationally expensive: runs times ranging from tens of minutes to days are not uncommon. When multiple design parameters are to be investigated, searching over design spaces directly with the full simulation code becomes unworkable. In such cases surrogate models are often used. These surrogate, response surface or meta-models seek to reproduce the behaviour of the full simulation code at much reduced computational cost and usually take the form of some kind of multi-dimensional curve fit through pre-computed results obtained from the expensive code. Figure 1 illustrates just such a design process (see Keane [1], for example).

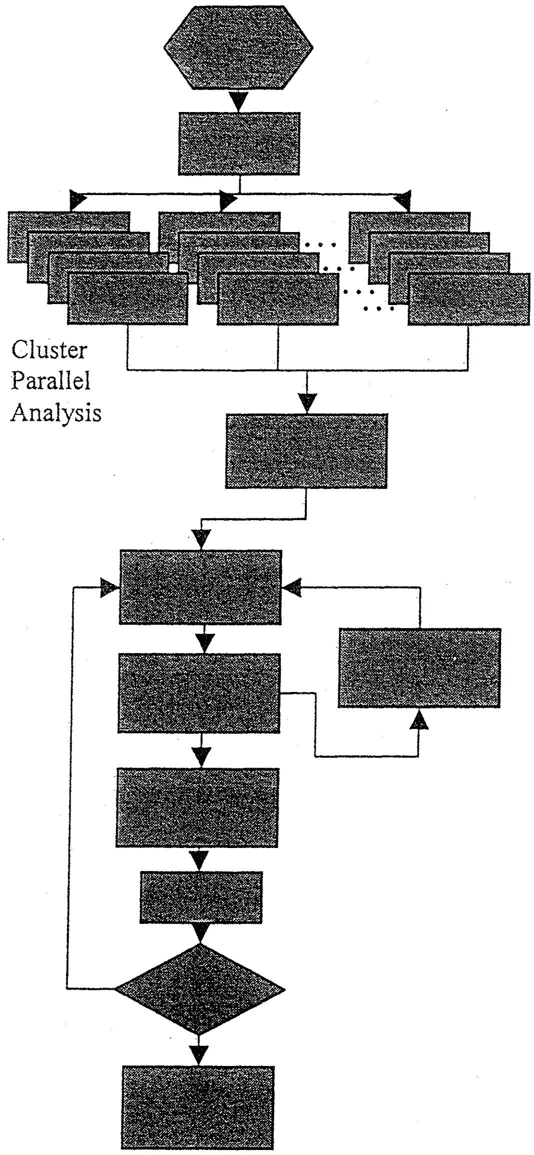


Figure 1 – a surrogate based search process for CFD.

used. These surrogate, response surface or meta-models seek to reproduce the behaviour of the full simulation code at much reduced computational cost and usually take the form of some kind of multi-dimensional curve fit through pre-computed results obtained from the expensive code. Figure 1 illustrates just such a design process (see Keane [1], for example).

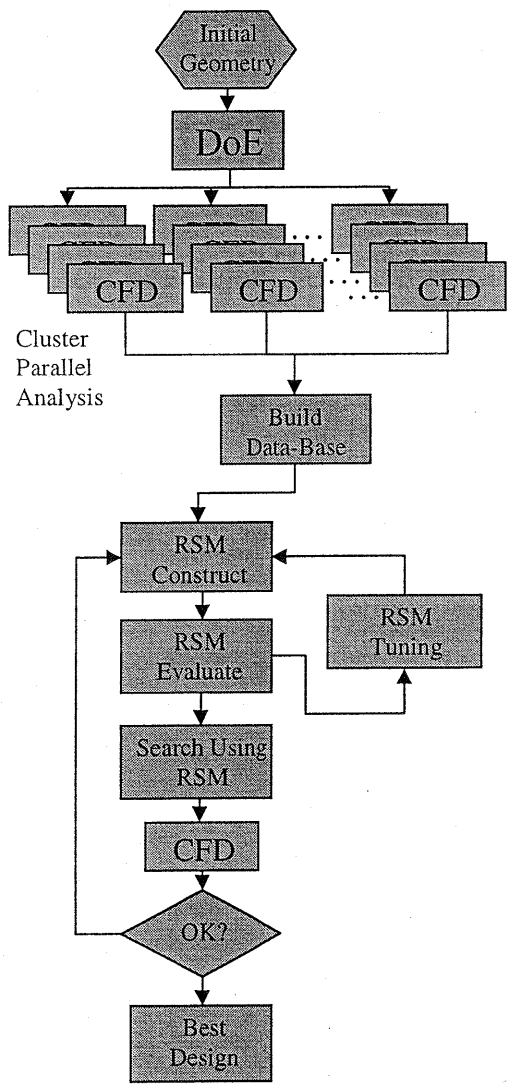


Figure 1 – a surrogate based search process for CFD.

In Figure 1 an initial geometric model is varied parametrically (e.g., by use of a parametric CAD system) according to some Design of Experiments (DoE) plan and then the resulting geometries are analysed (in parallel) to construct a data-base of results. A multi-dimensional RSM is then used to curve fit this data and this can then, in turn, be searched using optimisers to suggest improved designs for analysis. If the improved designs are not adequate the RSM can be updated and the search repeated until some form of convergence is achieved. There are a range of RSM types that can be used in this way (see Myers and Montgomery [2] or Jones [3]) and a range of update strategies (see Jones et al [4] for example).

One issue that arises in this process concerns any noise that may be present in the computed results. Such noise commonly arises when iterative solvers are used to carry out performance analysis, since the final convergence must usually be truncated before absolute precision is achieved if realistic run times are to be maintained. Also, most expensive PDE solvers use discretization schemes that must usually be changed with any geometric changes so that discontinuities can arise in the resulting data. In such circumstances the RSM curve fit must allow for this noise if over-fitting is to be avoided. It is also the case that response functions of many variables can have highly complex behaviour and any attempt to curve fit by interpolation with limited data can be misleading without some form of smoothing to establish overall trends, as opposed to a focus on local features.

In this paper a simple and robust technique is introduced to allow designers to control the degree of regression involved in radial basis function RSM construction in an easy to use fashion. While there are numerous more sophisticated approaches available in the statistics and neural networks literature, the suggested approach neatly fits between two existing methods widely used in the DSO RSM community, i.e., linear regression and interpolating cubic spline radial basis functions.

1.1 Theory

When carrying out optimisation a designer is seeking a solution vector in parameter space, x^* such that $y(x^*)$ is a minimum of the objective function $y(x)$, subject to k constraints $g_k(x)>0$. Here x is a vector of m design variables, $x=(x_1, \ldots, x_m)^T$. If a surrogate or response surface model $y'(x)$ is to be used, this model is set up using n training points $y_j(x_j)$, $j=1, \ldots, n$, $x_j=(x_{1j}, \ldots, x_{mj})^T$, taken from a pre-computed data-base of results (as per Figure 1). To begin with consider two such models for $y'(x)$: linear regression and cubic radial basis functions. In linear regression the RSM is given by

$$y'(\mathbf{x}) = a_0 + \sum_{i=1}^{m} a_i x_i \qquad (1)$$

where the coefficients a are chosen to minimize the sum of the squares of the errors between the surrogate and the training data, typically using an SVD least squares approach, see Figure 2.

By contrast an interpolating radial basis function model is constructed from a series of functions φ based at the sample points:

$$y'(\mathbf{x}) = \sum_{j=1}^{n} b_j \varphi(\mathbf{x} - \mathbf{x}_j). \tag{2}$$

Here the coefficients b are chosen so that the model passes through the training points, i.e., n equations are set up using the n training points to solve for the n unknown coefficients b. The functions, φ, can take many forms (see for example Jones[3]), here we use cubic splines, i.e.,

$$\varphi(\mathbf{x} - \mathbf{x}_j) = \| \mathbf{x} - \mathbf{x}_j \|^3 . \tag{3}$$

As Jones notes, however, this method can be numerically poorly conditioned and so it is better to use a modified approach which combines a regression line with the basis functions:

$$y'(\mathbf{x}) = a_0 + \sum_{i=1}^{m} a_i x_i + \sum_{j=1}^{n} b_j \varphi(\mathbf{x} - \mathbf{x}_j). \tag{4}$$

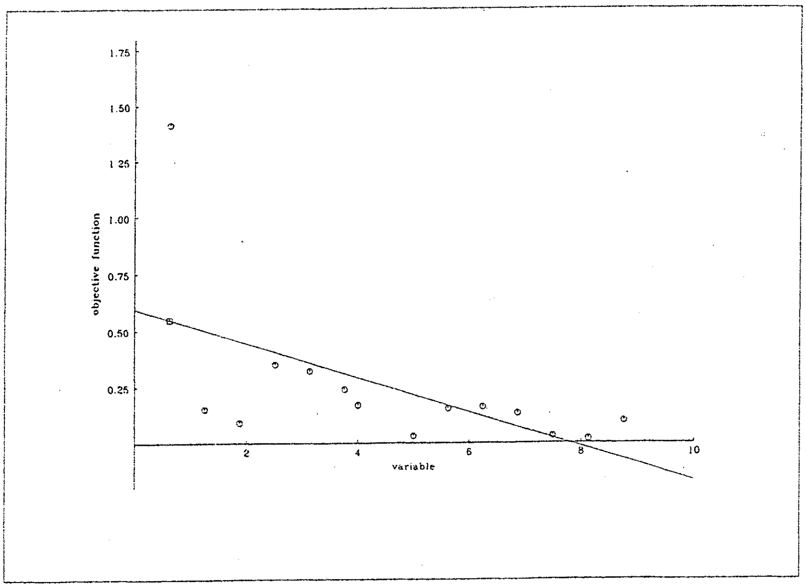
In this form the model has $m+n+1$ unknowns and so a set of $m+1$ additional constraints are added:

$$\sum_{j=1}^{n} b_j = 0 \tag{5}$$

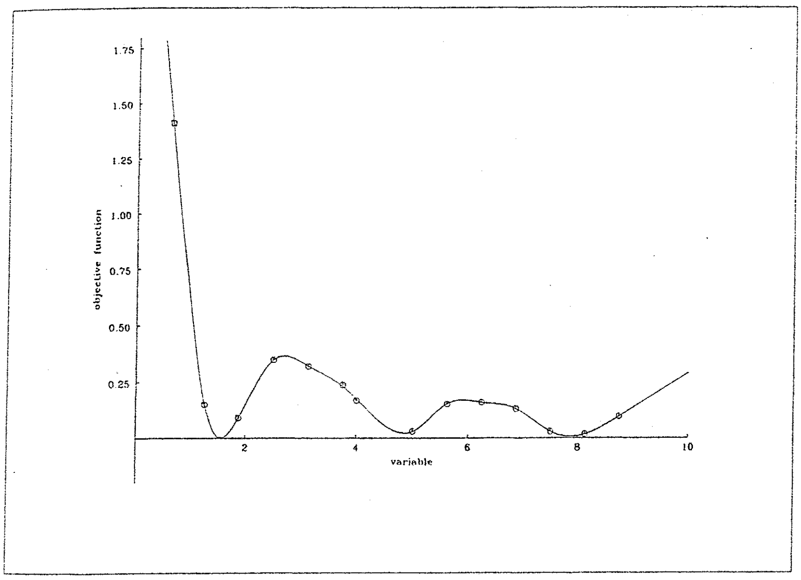
and

$$\sum_{j=1}^{n} b_j x_{ij} = 0 \tag{6}$$

for $x_{ij} = 1, \ldots, m$. In this form the basis functions model the differences between the regression line and the actual data points. The resulting model still strictly interpolates the data, but is now numerically better behaved (n.b., this approach can easily be extended to higher order regression models if desired). Figure 3 illustrates this approach applied to the same data as used in Figure 2.

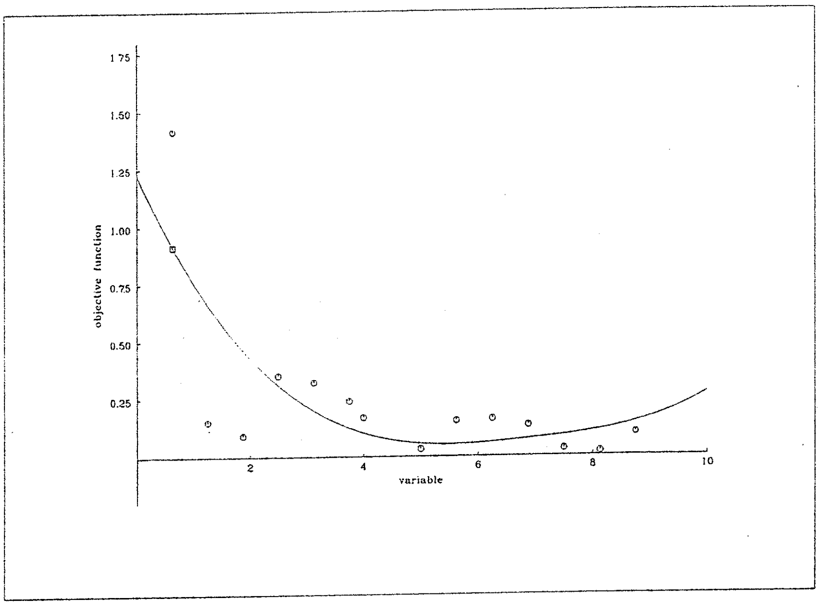


It is clear that these two approaches take a fundamentally different view of the training data: the linear regression approach assumes that there are errors in the data which cause it to deviate from a basically very simple trend; the RBF model assumes the data to be error free and the underlying trends to be much more complex. As has already been noted, however, engineering calculations usually lie somewhere between these two extremes. Certainly the results of engineering calculations are usually repeatable and often give complex trends, but they also contain noise. A number of sophisticated curve fitting schemes with regression abilities have been proposed to cater for this kind of issue, such as Kriging with regularisation, see for example Jones et al [4]. Such schemes often contain hyper-parameters which must be tuned and so it is natural to ask whether a simpler approach may be possible. That proposed here is based on the previous two models.



To see how to proceed we note that in equation (4) we may regard the summation over the n basis functions as a corrector to the underlying linear regression. If we include only a subset of these functions but still use all the training data to solve for the m coefficients a and reduced number of coefficients b we have an over-determined set of equations that can still be solved in a least squares sense using SVD. Similarly the summations in equations (5) and (6) are taken over the reduced numbers of basis functions. Therefore, as the number of basis functions increases from zero to n the resulting model moves from being one of simple linear regression up to full RBF interpolation. Figures 4 and 5 illustrate this process for one and seven basis functions, respectively. Figure 6 shows members of the family spanning the space between Figures 2 and 3.

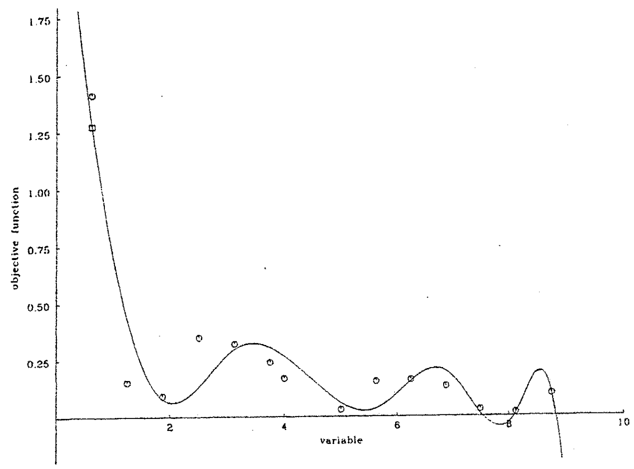
It is clear from figures 4 to 6 that a range of models can be built in this way. When setting them up the user must first decide how many basis functions to include and at which training data points to centre them. These decisions open up a degree of model tuning that can be exploited by the user according to the needs of the problem in hand. Choosing the number of functions to use, essentially addresses the problem referred to as "over-fitting" in the regression literature (see Ralston [5] for example): if too few functions are used the resulting model may fail to represent important trends in the data; conversely, if too many are used the resulting model may just be driven by the noise in the data.



One scheme for addressing this issue may be readily built from the Null hypothesis used in polynomial curve fitting (see Ralston [5] pp 254). When dealing with polynomials this hypothesis allows the order of the model to be tensioned against errors in the model, so that increasing the order must significantly reduce errors between the model and the data. Here we monitor the following expression:

$$\sigma^2 = \delta^2 / (n - n') \qquad (7)$$

where n' is the reduced number of basis functions being used and δ^2 is the sum of the squares of the errors between the model and the function values at the training points (note that both denominator and numerator become zero when all n basis functions are used, i.e., the model then interpolates the data). When this expression is nearly constant for several vales of n' we stop adding more basis function terms. Figure 7 shows a plot of this function for the data used so far and it can be seen that values around seven give a good fit., i.e., as per Figure 5. This might surprise the reader until it is realised that the data used for this test is, in fact, taken from the so called "bump" function (ref Keane [6]) which is oscillatory in just this sort of way.



The second issue that must be addressed in setting up the model concerns selection of the basis functions to include. Note that whichever functions are used, all the training data are still used in the SVD least squared process used to construct the model. It is just that some training points additionally become the centres for RBF construction, and thus have more impact on the resulting curve fit. In fact once fewer than n basis functions are used there is no need to place their centres at the training points at all. There are various schemes available in the literature for building models with arbitrary centres, such as those using clustering algorithms to select good locations, radial basis function neural networks (where similar models are used but complex training strategies are used to solve for the desired weights) or support vector machines. Such approaches lie outside the scope of this paper; suffice to say that they all involve a good deal of statistical analysis and tuning that rather defeats the desire for building simple models that can be rapidly set up – here the centres are selected randomly from the locations in the training set. If more complex models are required then more statistically soundly based approaches such as Kriging and stochastic process models should be considered. The approach adopted here simply relies on the fact that the model is bounded on one side by linear regression and on the other by an interpolating cubic spline RBF. Both of these schemes are well understood and simply to apply – by just tracking equation (7) for various degrees of model fit the user can decide if what is being proposed makes sense. A characteristic of the proposed approach is that the degree of fit, σ^2, tends to be oscillatory showing significant increases if the regression process is poor, because gradual addition of further basis functions then fails to consistently improve the model with diminishing returns at each stage.

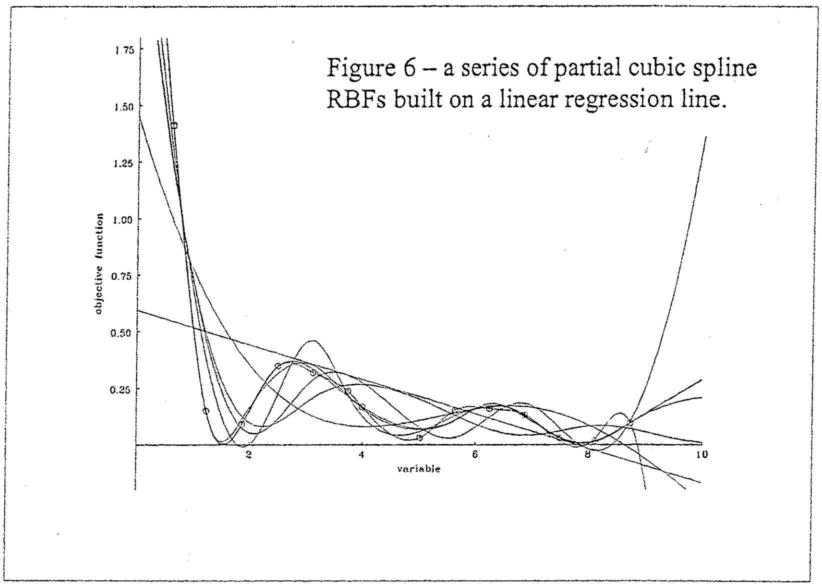


Figure 6 – a series of partial cubic spline RBFs built on a linear regression line.

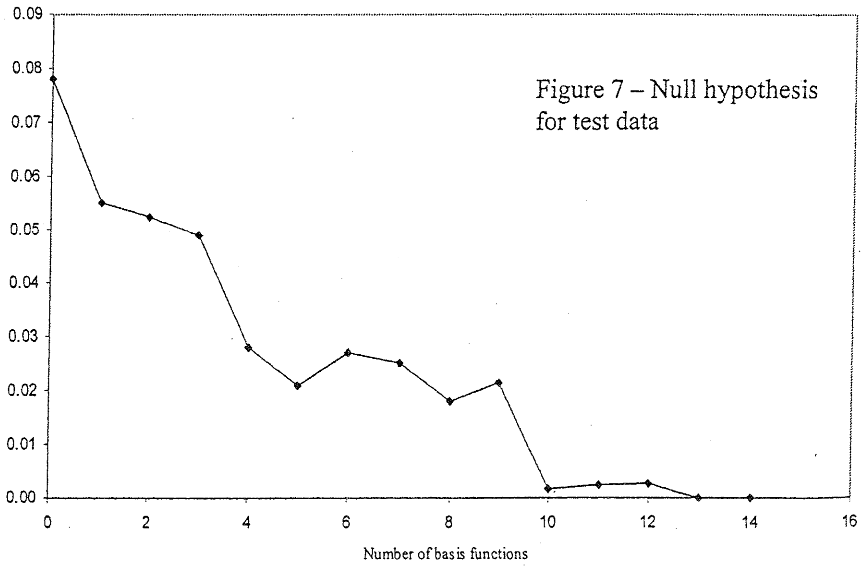


Figure 7 – Null hypothesis for test data

1.2 Example

To demonstrate this approach on a more practical example the wing design case presented by Keane[1] is re-analysed here. The problem concerns the design of a transonic civil transport aircraft wing defined by 11 variables, see Table 1. In the reference both empirical and CFD results were used; here only the empirical "Tadpole" model is tested for simplicity [7]. In the work by Keane a Krig was used to model the results coming from Tadpole and this was able, after careful tuning, to predict 250 unseen test data points with a correlation coefficient of 0.991. This test data is re-analysed here using various numbers of basis functions to regress the data. First the Null Hypothesis test of equation (7) is applied, see Figure 8. The figure shows that several levels of regression may be tried: there being plateaus in the curve at around 35, 100 and 150 basis functions. Figure 9 shows a correlation plot for a model with 150 basis functions and as can be seen this fits the unseen trial data quite well, yielding a correlation coefficient of 0.938. Adding more basis functions will, of course make the RSM fit the training data better but not necessarily improve its ability to predict unseen test data. Table 2 shows how the correlation coefficient for the test data varies with different numbers of basis functions. As can be seen the plateau in the Null Hypothesis at around 35 basis functions is matched by a good correlation, while that at 215 is slightly worse. Using a full set of basis function, which forces the model to interpolate the training data, massively over-fits the inputs and yields a very poor ability to predict unseen test data – a simple linear regression model is much better.

Conclusions

This paper has introduced a simple modification to the well known cubic radial basis function response surface model to enable it to have controllable regression capabilities. The proposed approach can be used to model noisy data without recourse to more sophisticated methods with little loss in accuracy. The only parameter tuning required is the selection of the degree of regression and this can set up either with a simple Null hypothesis approach or by using unseen test data to check for over-fitting. The advantage of the Null hypothesis is that it does not require additional unseen test data when setting up the model and is thus much cheaper to use. The approach taken lacks some statistical rigour but trades this for simplicity of definition and ease of use. It has been demonstrated on a simple one-dimensional test case and a realistic 11-dimensional aircraft wing design problem.

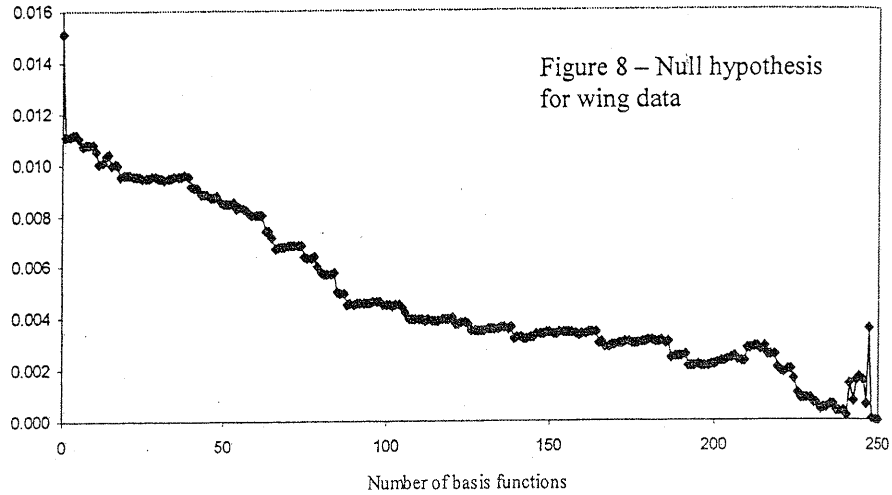


Figure 8 – Null hypothesis for wing data

Lower limit	Value	Upper limit	Quantity (units)
100	168	250	Wing area (m^2)
6	9.07	12	Aspect ratio
0.2	0.313	0.45	Kink position
25	27.1	45	Sweep angle (degrees)
0.4	0.598	0.7	Inboard taper ratio
0.2	0.506	0.6	Outboard taper ratio
0.1	0.150	0.18	Root t/c
0.06	0.122	0.14	Kink t/c
0.06	0.122	0.14	Tip t/c
4.0	4.5	5.0	Tip washout (degrees)
0.65	0.75	0.84	Kink washout fraction
	3.145		D/Q (m^2) – Tadpole

Table 1 – Design variables and objective function value

Number of basis functions	Correlation coefficient
0 (simple linear regression)	0.869
35	0.930
100	0.928
150	0.938
215	0.924
250	0.257

Table 2 – Correlation data

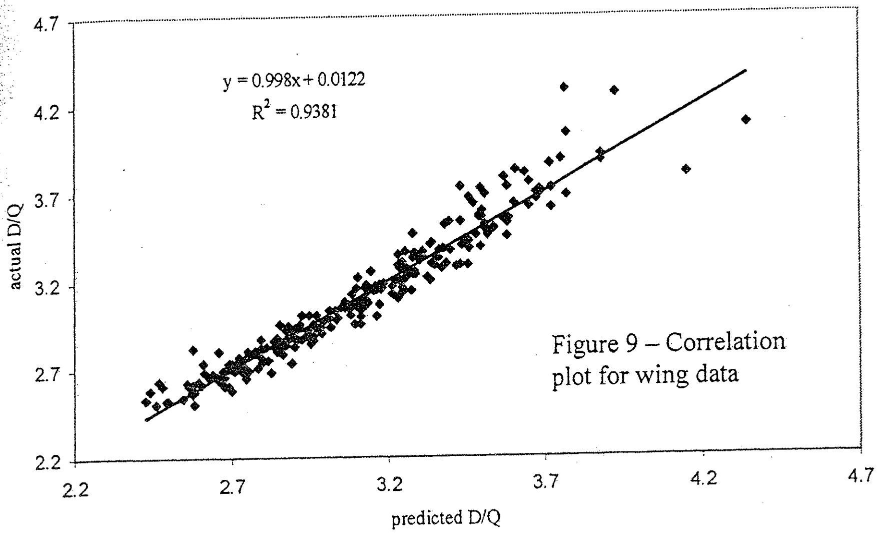


Figure 9 – Correlation plot for wing data

References

1 Keane, A.J., (2003) "The Use of Design of Experiment, Response Surface and Data Fusion Methods in Transonic Aircraft Wing Design Optimisation", *J. of Aircraft*, **40**(4) pp. 741-750.

2 Myers, R.H. and Montgomery, D.C., (1995) "*Response surface methodology: process and product optimization using design of experiments*", John Wiley and Sons inc.

3 Jones, D.R., (2001) "A taxonomy of global optimization methods based on response surfaces", *Journal of Global Optimization*, **21**, pp. 345-383.

4 Jones, D.R., Schonlau, M. and Welch, W.J., (1998) "Efficient global optimization of expensive black-box functions", *Journal of Global Optimization*, **13**, pp. 455-492.

5 Ralston, A. and Rabinowitz, P., (1978) "*A first course in numerical analysis*", McGraw-Hill, New York.

6 Keane, A. J., (1995) "Genetic algorithm optimization of multi-peak problems: studies in convergence and robustness", *Artificial Intelligence in Engineering* 9(2) pp. 75-83.

7 Cousin, J., and Metcalfe, M., (1990) "The BAE Ltd Transport Aircraft Synthesis and Optimization Program", AIAA-90-3295.