Cumulative global metamodels with uncertainty – a tool for aerospace integration

P. H. Reisenthel, J. F. Love, D. J. Lesieutre and R. E. Childs
Nielsen Engineering & Research
California, USA

ABSTRACT

The integration of multidisciplinary data is key to supporting decisions during the development of aerospace products. Multidimensional metamodels can be automatically constructed using limited experimental or numerical data, including data from heterogeneous sources. Recent progress in multidimensional response surface technology, for example, provides the ability to interpolate between sparse data points in a multidimensional parameter space. These analytical representations act as surrogates that are based on and complement higher fidelity models and/or experiments, and can include technical data from multiple fidelity levels and multiple disciplines. Most importantly, these representations can be constructed on-the-fly and are cumulatively enriched as more data become available. The purpose of the present paper is to highlight applications of these cumulative global metamodels (CGM), their ease of construction, and the role they can play in aerospace integration. A simple metamodel implementation based on a radial basis function network is presented. This model generalises multidimensional data while simultaneously furnishing an estimate of the uncertainty on the prediction. Four examples are discussed. The first two illustrate the efficiency of surrogate-based design/optimisation. The third applies CGM concepts to a data fusion application. The last example is used to visualise extrapolation uncertainty, based on computational fluid dynamics data.

NOMENCLATURE

b	RBF scale parameter
c_k	basis function coefficients, elements of [C]
C_D	drag force coefficient
C_m	pitching moment about centre of gravity
C_R	radial force coefficient
$[C]$	solution vector ($[c_1, c_2, \ldots, c_p]^T$)
$[\hat{C}]$	regression estimate for $[C]$
cov	covariance
$[e]$	regression error
$E[.]$	expected value
f	shape function
F	global interpolant (output of RBF network)
$[\hat{F}]$	regression estimate for F
$[I]$	identity matrix
M	Mach number
N	dimension (number of independent variables)
p	number of support vectors
$[Q_{1,2}]$	orthogonal matrices
r_c	canopy radius
RBF	radial basis function
ref	reference value
RS	response surface
SVM	support vector machine
var	variance
x_i	i-th independent variable (i-th co-ordinate of X vector)
X	independent variables vector
Y	dependent variable
α	angle-of-attack

Paper No. 3035. Manuscript received 26 October 2005, revised version received 22 February 2006 accepted 2 May 2006.
A shortened version of this paper was originally presented at the 1st International Conference on Innovation and Integration in Aerospace Sciences, CEIAT, Queen's University Belfast, Northern Ireland, 4-5 August 2005.

https://doi.org/10.1017/S0001924000001299 Published online by Cambridge University Press

δ_{ij} Kroenecker delta tensor
δY_k seed uncertainty of dependent variable at point k
φ roll angle
φ_k k-th radial basis function
χ_k radial basis function centre for φ_k
θ_c total canopy angle
θ_r canopy reference angle
θ_v canopy vent angle
σ^2 regression error variance
$[\Sigma]$ diagonal matrix of singular values
$[\Sigma^+]$ truncated inverse of $[\Sigma]$

1.0 INTRODUCTION

Global metamodels and response surface technology are increasingly used in a variety of fields, including structural reliability, instrument calibration, and aerodynamic and trajectory optimisation, to name a few[1-10]. Because of their analytical nature, these models can be used for automated searches in surrogate-based analysis and optimisation[11-13], and are naturally well-suited to the acceleration of optimisation tasks and rapid strategy evaluation. A central issue to constructing appropriate response surface models is the so-called curse of dimensionality, in which the number of data points required to characterise/support the surface increases exponentially with the number of independent variables. This difficulty is well-known and, in effect, precludes the use of conventional schemes, such as polynomial and piecewise-polynomial (finite-element) approximations. Neural networks[14,15], support vector machines[16], kernel methods[17,18] and multidimensional splines[19] all have proven capable of multidimensional data generalisation and of (partially) circumventing this difficulty. The particular approach used in this paper is based on self-training radial basis function networks which form the basis of the NEAR-RS (response surface) technology[20].

NEAR-RS is a software system consisting of two modules: a metamodel (response surface) identification module and a metamodel evaluation/interrogation module. A graphical user interface included in this second module serves as a multidimensional viewer, which facilitates the visualisation of trends in high-dimensional data[21]. A key aspect of the technology is the ability to estimate further sampling needs and model quality, based on automatic uncertainty estimation.

Four distinct applications illustrating the use of these methods are discussed in this paper. The first three illustrate the data adaptivity, data fusion, and feature extraction/data mining capabilities of the method. The last application presents an example of the uncertainty modeling capabilities of CGMs.

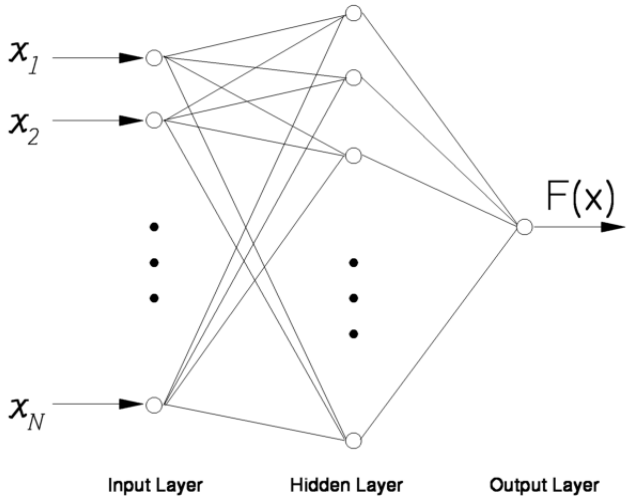


Figure 1. Radial basis function network.

2.0 TECHNICAL BACKGROUND

Response surface methods can be used to enhance the usefulness of limited experimental or computational data. The problem is akin to interpolating and extrapolating the data outside the range where these data were collected, a task which, without any regularising assumptions, constitutes a fundamentally ill-posed problem[22]. Regularising assumptions can come in various forms: physics-based models, mathematical equations (such as splines), implicit smoothness assumptions, or other empiricisms. The method used here employs a particular form of regularisation, in which a hyper-surface constructed using radial basis functions is made to approximate the supporting data. The basic theory behind the response surface identification and its uncertainty is presented next.

The task of formulating a response surface in N-dimensional space amounts to identifying a smooth mapping $F : R^N \rightarrow R$ on the basis of p available data points. If this response surface acts as an interpolant, then the function F must satisfy the constraints

$$F(X_i) = Y_i, \quad i = 1,\ldots,p \qquad \ldots (1)$$

where each X_1 represents a vector of independent variables (for example, spatial co-ordinates, flow conditions, and/or configuration parameters), and each Y_i is a dependent variable (for example, pressure). In the case where F represents, instead, a regression model fit to the data, then the response surface is required to minimise the distance $\|F(X_i) - Y_i\|$, typically in the least squares sense.

This goal can be achieved by a number of different means, for example Kriging[23,24] which is used in the popular DACE stochastic process model[25], multivariate adaptive splines[19], and support vector machine (SVM)[16] algorithms. The goal of this paper is not to compare these methods to each other, but to illustrate how this class of methods can be used to accomplish various tasks when analysing multidimensional data. Our approach uses a radial basis function (*RBF*) network to represent the function F. In this approach, F is expanded into basis functions φ_k which are radially symmetric about their control point, χ_k. By analogy with SVMs, we will refer to $(\chi_k; F(\chi_k))$ as the support vectors for the response surface. Thus;

$$F(X) = \sum_k c_k \varphi_k(X), \quad \varphi_k(X) = f(\|X - \chi_k\|; b) \qquad \ldots (2)$$

where f is a scalar shape function, b is an adjustable[26] scale or stiffness parameter, and $\|\cdot\|$ designates the Euclidean norm. The c_k are the basis function coefficients. They are parameters to be identified. Note that, if the basis functions (i.e., their shape, centres, and number) are known, then the determination of the nonlinear response surface boils down to an identification problem which is linear in the parameters. In other words, the coefficients c_k are the solution of a least-squares linear problem

$$[A]\,[C] = [Y] \qquad \ldots (3)$$

where each row of Equation (3) is an instantiation of the constraints expressed in Equation (1). Radial basis function models, such as Equations. (1) and (2), can be viewed[27] as a three-layer feedforward neural network with linear output mapping. This is shown schematically in Fig. 1, where the number of nodes in the hidden layer is equal to the number of basis functions, the inputs x_i are the co-ordinates of X, and the weights of the output layer are the coefficients c_k. Also, the shape function (f) is the activation function of the hidden layer nodes, which can take a number of forms, for example, Gaussian, thin plate spline, multiquadric, or reciprocal multiquadric[28,29].

There are, a priori, a number of different ways of selecting the control points χ_k. One possible approach is the use of sequential approximation and optimisation methods. This can be quite expensive, and a more efficient approach to control point selection consists of using a fixed subset of the existing training data. Algorithms such as generalised cross-validation (*GCV*) can be used

https://doi.org/10.1017/S0001924000001299 Published online by Cambridge University Press

for this purpose, resulting in parsimonious networks with good generalisation properties[30]. While the use of a small number of regressors φ_k is indeed desirable from the point of view of model robustness, we will confine the present discussion to simple networks where the training data are assumed deterministic and sparse. Thus, in this particular implementation, the number of basis functions is equal to the number of training data points, and the centres (control points χ_k) of the RBFs coincide with the data points. As a result of this simplification, there is no need for stepwise regression algorithms: the structure of the equivalent neural network (Fig. 1) is automatically determined by the data.

Multidimensional response surface identification in NEAR-RS is a three-step process:

(1) preconditioning/classification,

(2) formation of the [*A*] matrix in Equation (3), and

(3) solution method for [*C*].

The preconditioning step uses a classification algorithm to associate data points which have the same (or substantially similar) values of the independent variables. This step is a mechanism for mitigating problems associated with overfitting dense point clusters. At present, this operation is performed on the basis of a user-defined tolerance in the independent variables. If no tolerance is prescribed, then strict equality is required in order for two points to be associated. The purpose of preconditioning is to improve the solution characteristics by improving the condition number of the [*A*] matrix. The formation of the [*A*] matrix is relatively straightforward: it involves the calculation of distances between all training data points. If the response surface uncertainty is desired, then an additional step (weighted least squares, described below) is used. Finally, the solution method uses robust pseudoinversion techniques, the purpose of which is to take care of pathological situations, such as the handling of inconsistent data. Such data can occur, for example, as a result of improper or incomplete parameterisation, such as repeatability tests or the existence of data from various sources (different codes, different fidelity level, algorithms, etc.) at the same, or substantially similar, condition. These data 'inconsistencies' amount to an ill-posed problem in terms of interpolation, a difficulty which is circumvented using regularisation techniques.

One important addition to these ideas is the concept of response surface uncertainty. In the following, it is shown that, due to linearity in the parameters, it is possible to make use of well-established statistical results to propagate the uncertainty of the support data onto an uncertainty of the response surface itself.

Consider the original equation [*A*] [*C*] = [*Y*] as a regressor model for the data. Rewrite Equation (3) as

$$[A]\,[C] = [Y] + [e] \qquad \ldots (4)$$

where [*e*] is the modeling error. Let $[A]=[Q_1][\Sigma][Q_2^T]$ designate the singular value decomposition of the matrix [*A*]. The pseudoinverse solution is then given by[31] $[\hat{C}] = ([Q_2][\Sigma^+][Q_1^T])[Y]$. It can then be shown[32] under certain simplifying assumptions, that the covariance matrix of the solution vector is;

$$cov[C] \equiv E\left[(\hat{C}-C)(\hat{C}-C)^T\right] = \sigma^2 E\left[Q_2\Sigma^+\Sigma Q_2^T\right] \ldots (5)$$

where *E*[.] designates the expected value, and σ^2 is the variance, presumed uniform, of [*e*]. Thus, Equation (5) propagates the uncertainty in [*Y*] onto the solution vector [*C*]. Alternatively, the matrix $[Q_2\Sigma^+\Sigma Q_2^T]$ can be interpreted as a sensitivity matrix which redistributes the measurement noise onto the solution vector components. This uncertainty, in turn, ties into the uncertainty on the response surface itself. Let $F(X)=\sum c_k\varphi_k(X)$ where the φ_k are the members of a radial basis function set. It then follows that;

$$E\left[(\hat{F}-F)^2\right] = \sum_i\sum_j (cov[C])_{ij}\,\varphi_i(X)\,\varphi_j(X) \qquad \ldots (6)$$

Equation (6) represents the variance *var*(*F*) of the response surface. Assume a Gaussian probability distribution. The resulting uncertainty on *F*(*X*), defined as $\Delta F = \pm 3\sqrt{var(F)}$ ('three-sigma' uncertainty), is shown in the hypothetical example of Fig. 2. The dependent variable uncertainty of the training data is indicated in the form of vertical error bars. The resulting uncertainty on the response surface is indicated in the form of upper and lower bounds, $F \pm \Delta F$, using thin dashed lines.

Note that, in order to propagate uncertainty according to the method described above, the variable [*e*] must be a stochastic variable such that

$$cov[e] = \sigma^2[I] \qquad \ldots (7)$$

where [*I*] is the identity matrix. In other words, [*e*] must have zero cross-correlation, and must be of uniform variance σ^2 across all of its components. While the zero cross-correlation assumption is typically not justified if there is a deterministic bias between the regressor model and the data, it is still possible in practice to use the above equations to propagate uncertainty, if one assumes that [*e*] represents a vector of random measurement errors $\left[\delta Y_1, \delta Y_2, \ldots, \delta Y_p\right]^T$. When these seed uncertainty levels differ from support vector to support vector, such as in the example of Fig. 2, then both left- and right-hand sides of Equation (3) are multiplied by a weighting matrix [*W*], where the elements of [*W*] are defined as $W_{ij} = \delta_{ij} / \sqrt{var(Y_i)}$. This simple algorithm ensures that the variance of the transformed variables is uniform. At present, the uncertainties are assumed to be uncorrelated between data points. If this were not the case, then more sophisticated techniques, such as Markov estimators and/or instrumental variables[32] could be used.

3.0 REFUELLING DROGUE CANOPY DESIGN

Rapid optimisation through multidimensional surrogate modeling was used in a recent effort at Nielsen Engineering & Research involving the redesign of a refuelling drogue canopy. Figure 3 depicts a standard C-130 high-speed drogue during a wind-tunnel test (flow is from left to right). The picture shows the end of the refuelling hose, connected through a ball joint to the MA-3 coupling, to which the strut arms of the canopy are attached. The inflated canopy provides the appropriate drag force to both stabilise the hose-and-drogue assembly and maintain the hose at the proper angle for refuelling operations.

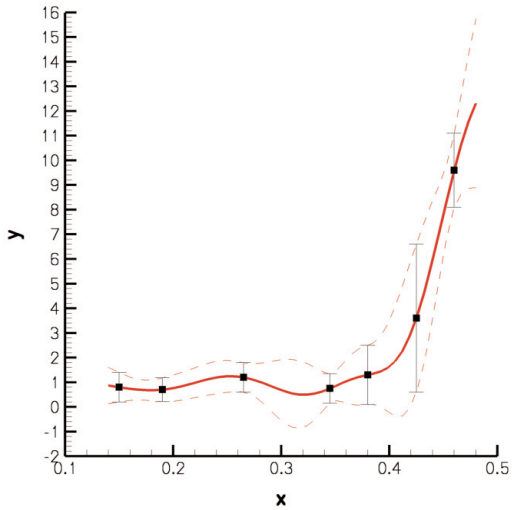


Figure 2. Response surface (solid line) plus/minus uncertainty (dashed lines).

https://doi.org/10.1017/S0001924000001299 Published online by Cambridge University Press

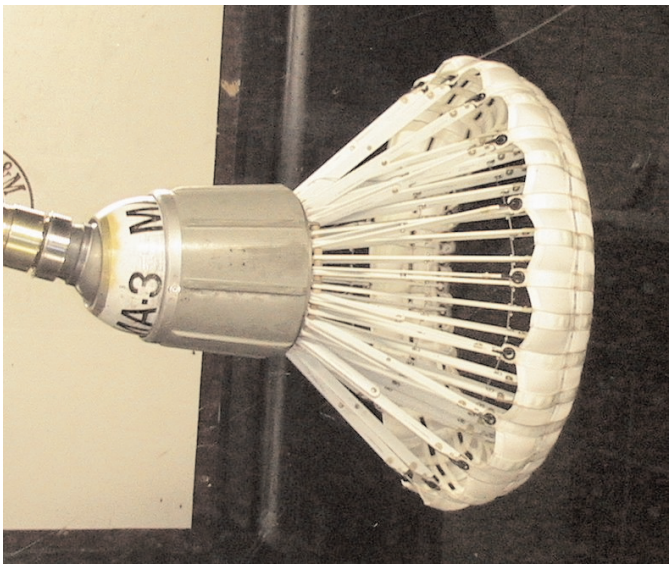


Figure 3. Standard C-130 refuelling drogue.

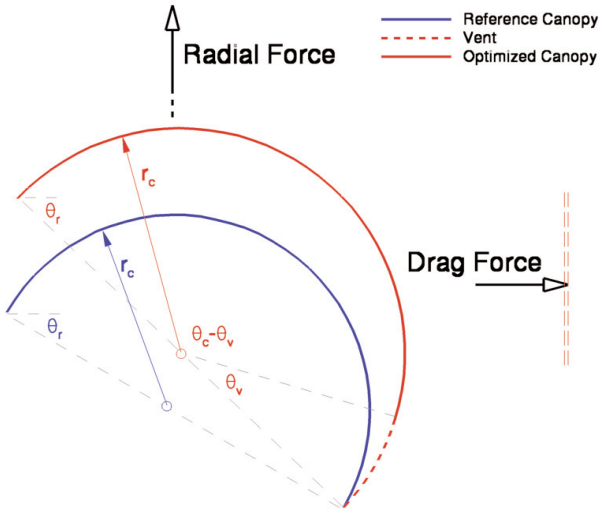


Figure 4. Canopy geometric parameter definition.

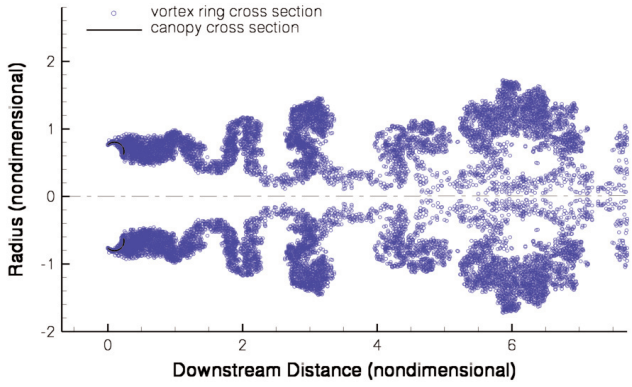


Figure 5. Canopy wake cross-section calculated using VPARA.

3.1 Objective and methods

For a given air speed, the design objective is to maximise the radial force on the canopy, while maintaining the drag within a narrow prescribed range. The canopy shape is modelled as a first approximation as being axisymmetric. The four geometric design parameters (see Fig. 4) are the canopy radius, r_c , the reference angle, θ_r, the vent angle, θ_v, and the total arc, θ_c. These four parameters define a family of canopy shapes with a common downstream (inner) strut arm attachment point.

Two examples of canopy geometry are superimposed in Fig. 4. Each geometry is an arc section representing a radial cut of the refuelling drogue canopy (cloth drag chute) in Fig. 3. The dashed portion of the arc (present in the optimised geometry) designates the porous or vented portion of the canopy. Not shown in Fig. 4 are the strut arms, which extend from each end of the arc towards the coupling.

The aerodynamic performance of the inflated canopies was estimated using Sandia National Laboratory's VPARA code.[33] VPARA predicts the axisymmetric flow field over rigid bluff bodies such as thin shells, and has been used to estimate the drag and radial force produced by canopies of a fixed shape. The solution method is based on the axisymmetric vorticity transport equation, where discrete ring vortices are generated at the body, satisfying the appropriate surface boundary conditions. These vortices are subsequently diffused and convected into the boundary layers and the wake of the body, while new vortices are generated at the body surface. VPARA was used to compute time histories of the pressure distributions and the forces on the modelled surfaces. Estimates of the mean drag and radial force coefficients were obtained by averaging the instantaneous loads over a portion of the total time history. As an example, Fig. 5 depicts the location of the shed vortices after a nondimensional convective time of 240.

3.2 Canopy optimisation

For this design problem, there are four independent variables (the four geometric parameters defined in Fig. 4), and two dependent variables, the drag C_D and radial force C_R. From a known reference condition (corresponding to an existing drogue), the design space was seeded initially with the results of 12 additional VPARA calculations. Using these 13 configurations as the initial sample size, a four-dimensional response surface was then constructed for both C_D and C_R. This small number of calculations represents an initial investment of less than 10% of the computational resources used. By comparison, Ref. (12) recommends 35% of the available computational budget as a safe choice of initial Latin Hypercube sample size when confronted with a novel application. The current 10% figure was based primarily on heuristics and the experience of the designer, taking into account the relative simplicity of the underlying physics and (correspondingly) the expected lack of complexity of the response surfaces for C_D and C_R when considered individually.

As previously mentioned, the objective was to maximise C_R, subject to C_D being confined to a narrow range of allowable values. Thus, the C_D and C_R response surfaces were combined into a single objective function, defining a new, highly nonlinear, hypersurface. A thorough, global search of this hypersurface was then conducted through repeated interrogations of the surrogate model. The result of the search is a hierarchy of maxima which can then be investigated through further VPARA calculations.

After collecting a small number of additional calculations, new response surfaces are formed for C_D and C_R. In this manner, the resulting, updated hypersurface is cumulatively enriched with the addition of each new data point. The designer used heuristics to determine where to acquire the next set of runs, choosing, at times, to follow the suggested optima on the current response surface. At other times, the corners of the space were systematically evaluated to help visualize the space. At other times still, curiosity and/or

https://doi.org/10.1017/S0001924000001299 Published online by Cambridge University Press

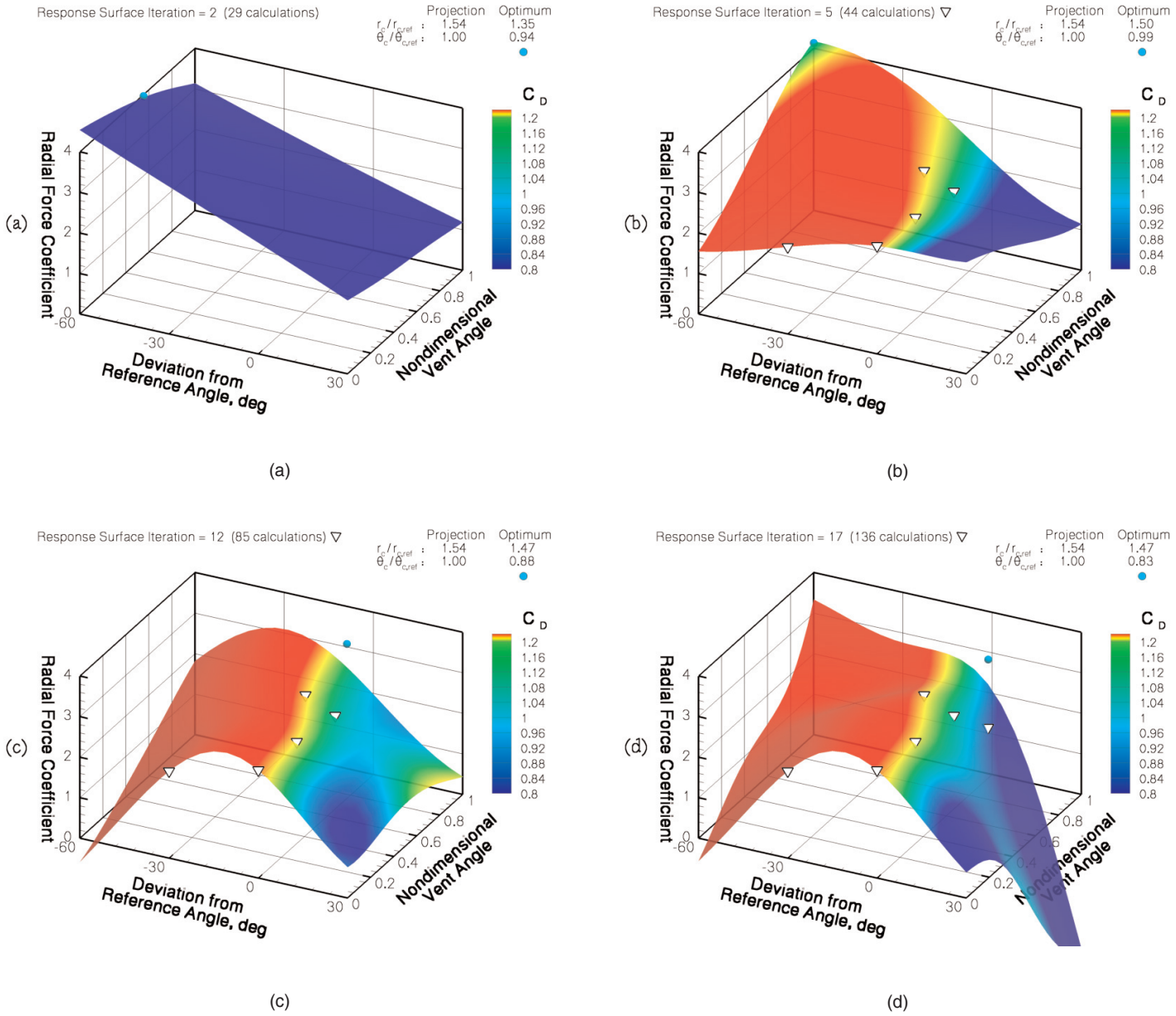


Figure 6. Radial force response surface evolution as a function of cumulative data set enrichment.

experience can be attributed as the biggest factors in selecting the infill samples. Thus, no algorithm was used to determine how many new points should be taken and where these calculations should be performed. Rather, this is a faithful documentation of how the response surface tool was used by an expert in the loop. Snapshots of the overall process steps are shown in Fig. 6. Figure 6 shows the response surface of the radial force coefficient at various stages of the design iteration process.

The surfaces are two-parameter projections of the hypersurface, i.e., the parameters r_c and θ_c are kept constant for illustration purposes. The results of actual VPARA calculations for this same projection are indicated using ∇ triangular symbols. Note that fewer symbols are shown than the total number of points because this is a two-dimensional projection of a four-dimensional surface. Thus, for example, Fig. 6(b) shows five points out of 44. This is because the other 39 points are out-of-plane, i.e., correspond to different values of the parameters r_c and θ_c. An indication of the corresponding drag coefficient response surface is also given through colouring. In this case, the goal was to keep C_D near 1·0. Figures 6(a), (b), (c) and (d) correspond to response surfaces formed after 29, 44, 95, and 136 calculations, respectively.

An assessment of the performance of the optimisation method is given in Fig. 7. Two curves are shown. Both represent the maximum radial force coefficient versus iteration number, subject to $C_D = 1{\cdot}0$. At a given iteration, the bottom curve represents the current best design based on the VPARA computations carried out to that point. By contrast, the top curve at the same iteration number represents the new suggested optimum, based on a search of the response surface computed at that iteration. Naturally, asymptotically, these two curves should converge. The purpose of the figure is not to show convergence, however, but to document the process actually used during the optimisation, as well as to furnish insight into how the method works. For this purpose, next to these performance metric curves (maximum C_R, conditional to $C_D = 1{\cdot}0$) is an indication of the

https://doi.org/10.1017/S0001924000001299 Published online by Cambridge University Press

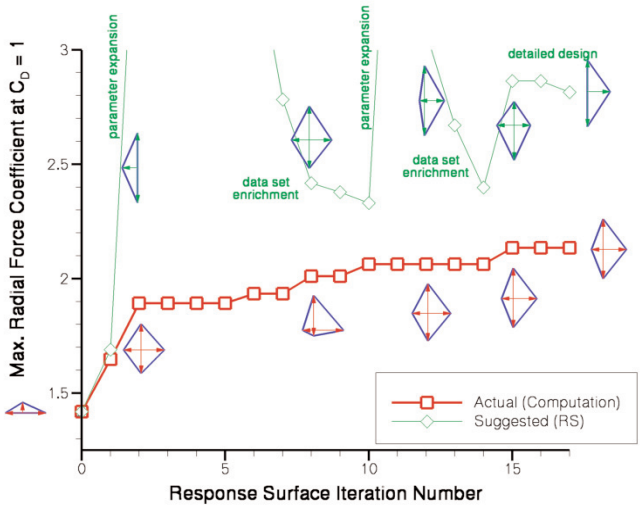


Figure 7. Design evolution as a function of response surface iteration.

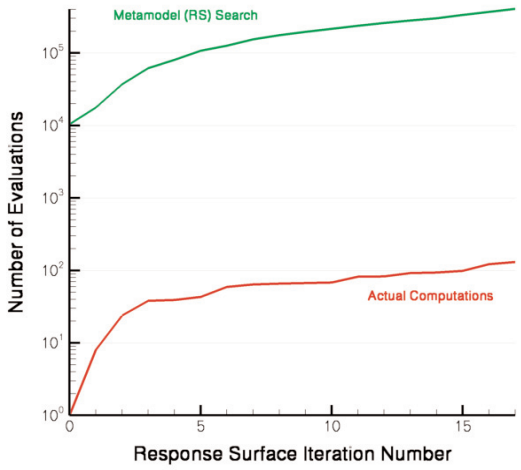


Figure 8. Efficiency of cumulative metamodel approach shown as a comparison of search count *versus* number of actual calculations.

independent variables for which this performance is achieved. The magnitudes of the independent (geometry definition) variables are shown in the form of radial histograms or 'radiograms,' for short. Each radiogram shows four vectors at 90-degree angles, corresponding (clockwise from top) to: r_c, θ_r, θ_v, and θ_c. In these graphs, the minimum and maximum length of each vector has been rescaled from 0 to 1 for convenience. Also, a line connecting the ends of the vectors is added to enhance the visual perception. Using these radiograms, one can visualise on the same plot as the performance curves (a) the evolution of the actual design in terms of the geometric parameters, and (b) the parameters corresponding to the current suggested maximum, based on the response surface.

In the early stages of design, the response surface is more erratic and tends to point the design in new directions which may or may not be fruitful ultimately, as more real data are collected. The addition of these new data points enriches the response surface, which becomes more realistic and more accurate in the process. Also, rather than a systematic convergence study, the graph of Fig. 7 is a faithful documentation of the design process used. During the course of this design, the parameter space underwent several expansions, i.e., redefinitions of the parameter boundaries by the designer. Each time this happened, the current response surface pointed to new, extrapolative extrema (as high as $C_R = 4{\cdot}1$). But, as the data set was subsequently enriched, this would be followed by a period of relaxation back to a more realistic maximum. Aside from parameter expansions, there were also changes in strategy throughout the process. For example, in the early phases, the designer chose to follow the suggestions made by the response surface optimiser (i.e., to investigate the response surface maximum). Later, the designer chose to perform calculations at some of the 'corners' of the parameter space, to get a better overall representation of the response surface. Later still, the designer was satisfied that the calculated optimum met the desired specifications, and was more interested in populating the design space near this optimum, in order to better understand the local shape of the response, such as the width of the allowable region, and whether there were any 'dangerous' sensitivities to stay away from. In the final stages, the designer chose to investigate suboptimal maxima, optimising in subranges of the canopy radius. This was done with the goal of fabricating and wind-tunnel testing several different canopies, in order to provide a smooth path from the existing, reference canopy to the (very different) optimised canopy. While optimisation methods have been developed[34] that can incorporate robustness from the onset, the current straightforward approach did yield a final design deemed to be robust, i.e., which was relatively insensitive to small perturbations in the design parameters. Most importantly, regardless of the optimisation strategy used, the response surface's accuracy continues to be cumulatively enhanced as more data become available.

Careful scrutiny of Fig. 7 also shows that the current approach did not amount to a mere gradient search, and that, indeed, significantly different directions and regions of the parameter space were investigated as a result of the surrogate model-based optimisation. Most importantly, the continuous and analytical description provided by the response surface is well-suited to rapid global search strategies, since interrogations of the surrogate model are many orders of magnitude faster than the individual data calculations, a fact that the authors took advantage of, giving a high degree of confidence that the parameter space was searched thoroughly.

The estimated gain in performance is illustrated in Fig. 8, which shows the cumulative number of surrogate (metamodel) evaluations, compared to the cumulative number of VPARA calculations needed to achieve the final design, a net three-and-a-half order of magnitude gain. For reference, each VPARA calculation took on the order of 20 to 30 minutes. This is in contrast to one evaluation of the RBF response surface, which takes less than a second. Thus, at the end of the design, approximately 50 CPU hours were spent on VPARA calculations, *versus* approximately two hours spent on evaluations resulting from the global search of the response surfaces. The quoted three-and-a-half orders of magnitude improvement implicitly assumes that, had the same optimisation strategy been used without the surrogate model, then it would have cost on the order of 10^5 VPARA calculations to achieve the same result.

Note that the criteria or strategies used to enrich the design space were based largely on heuristics and the designer's own experience, thoughtful reflection, and subsequent decisions at each iteration of the response surface. The example is a faithful depiction of a semi-empirical procedure used in a working environment with an expert in the loop, and no attempt was made to automate the population of the design space. This refuelling drogue canopy design example is but one illustration of surrogate-based optimisation, for which there are many examples in the literature. In particular, the reader is referred to the seminal work of Jones *et al*[11] concerning efficient global optimisation when dealing with limited quantities of expensive computational data, in which a closed-form expected improvement function is used to balance local and global searches of the design space. This balance between the need to continue exploring the space (global search) and the desire to focus in on a local optimum has also been addressed in recent studies[12], emphasising the possibility of automating these procedures for large problems.

https://doi.org/10.1017/S0001924000001299 Published online by Cambridge University Press

4.0 LARGE ASYMMETRIC LAUNCH VEHICLE PAYLOAD FAIRING DESIGN

A second application involving sequential optimisation concerns the aerodynamic and structural design of a payload fairing large enough to encompass a spacecraft with an optical mirror up to twice the diameter of an evolved expendable launch vehicle (EELV)[35].

4.1 Objective and methods

The principle design objectives at the early stages of this work were related to stability and control, as well as the mass of the fairing. Optimisation and computational fluid dynamics (CFD) methods were used to design an asymmetric exterior shape for the fairing that meets specific aerodynamic design goals. These aerodynamic goals include a balance of a low lateral force on the fairing and smooth variations in that force with respect to angle-of-attack, across a range of Mach numbers near M = 1.

The fairing surface was defined principally in terms of analytical functions. The cost to perform design optimisation increases significantly with the number of independent design variables; therefore, it is important to minimise the number of variables needed to parameterise the surface shape. This can be accomplished with an analytical surface shape definition, rather than a surface defined by discrete points. Also, the approach sought to impose the designer's understanding of aerodynamics on the optimisation process, and to focus the design on shapes that would trigger flow separation gradually, to alleviate excessive aerodynamic forces on the fairing.

Because the flow around the payload fairing involves flow separation, a Reynolds-averaged Navier-Stokes CFD methodology (Overflow[36]) was used to predict the steady-state forces and moments on the launch vehicle. Front and side views of the three-dimensional payload fairing geometry are shown in the insets of Fig. 9. The geometry resembles an asymmetrical ellipsoid with a flow separation trigger. This geometry was initially parameterised using nine free design variables (longitudinal minor axis and focus for the nose, depths of the minor and major axes, three variables controlling the height, location, and slope of the separation trigger, and two geometry transition parameters). Three of these design variables (transition variables, and nose focus) were found early on to have a minimal effect on aerodynamic performance and were, consequently, set to fixed values. The remaining six active independent variables were used to automatically remesh the surface and recompute a new volume grid using Hypgen[37]. Grid convergence studies were conducted to determine the optimal grid resolution needed. Using this grid resolution, each CFD calculation took between two and six hours, depending on the requested convergence level.

4.2 Payload fairing optimisation

The approach used was similar to the one described for the refuelling drogue canopy optimisation. A global optimisation scheme is applied to a continuous, analytical surrogate, rather than the CFD calculations themselves. This allows a thorough search of the solution space without the cost of performing actual CFD solutions. Naturally, the approach relies on the response surface being representative of the design space, which may not be the case, particularly at the early stages of the design. Nevertheless, and as demonstrated through the previous example and this one, the procedure often yields solutions satisfying the problem constraints within a relatively small number of CFD calculations.

Similar to the example of the refuelling drogue canopy, multiple dependent variables were combined into a single objective function. The dependent variables were the overall pitching moment, C_m, at multiple aerodynamic conditions (M,α,φ). The six-dimensional independent variables parameter space was initially seeded using latin hypercube sampling (LHS) design-of-experiments methods[38].

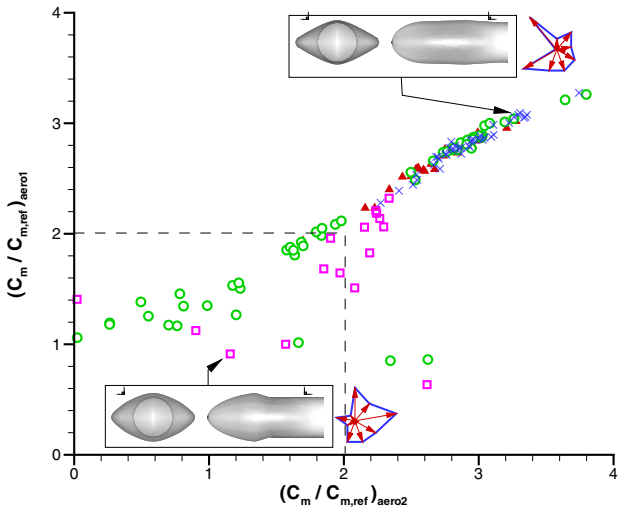


Figure 9. Rapid multiobjective optimisation using cumulatively enhanced metamodel.

The results of these initial calculations were used to form the first generation of response surfaces (one for each aerodynamic condition). A global search of the objective function over the response surfaces would then 'suggest' the next set of CFD runs. A mixed strategy was used to enrich the data set (i.e., compute further CFD solutions). As in the case of the refuelling drogue canopy, in the early stages of design the strategy involved successive waves of parameter expansion, followed by periods of populating the data set. Each generation of new CFD calculations used LHS randomisation for the independent variables, running 10 to 16 solutions at a time on a cluster of Linux workstations. By contrast, the intermediate and later stages of the design were characterised by a *focusing* of the parameter space on the most promising regions, at a given iteration of the design. Note that these regions of interest did not always remain the most promising ones, as further CFD calculations would either confirm or invalidate the trend predicted by the response surface at the previous iteration. Once again, the approach was an empirical one, rather than a systematic attempt at balancing exploration and exploitation (see Ref. 12).

Figure 9 shows the result of the optimisation procedure. In this example, the various symbols are associated with successive generations of CFD data points harvested in the process. Each data point generation is the result of an exhaustive 'scatter and poll' search strategy[39] on the global metamodel constructed as a result of the design's previous iteration. The particular implementation of the scatter and poll strategy used here involves parallel searches, each started from random locations on the surface (the 'scatter' phase) and, from each of these random locations, uses a local nongradient-based direct search[40] employing an interrogation stencil that is either translated or focused, depending on the results of the evaluations on the stencil (the 'poll' phase). Most importantly, the metamodel is enhanced at each generation by the new knowledge provided by the results of the latest CFD runs. In this case, the objective was to reduce the coefficient C_m to less than twice some reference coefficient $C_{m,ref}$ for two distinct aerodynamic conditions, *aero*1 and *aero*2, which are combinations of different (M,α) pairs near $M = 1$. Thus, any point such that $(C_m/C_{m,ref})_{aero1} < 2{\cdot}0$ and $(C_m/C_{m,ref})_{aero2} < 2{\cdot}0$ (i.e., inside the dashed line in Fig. 9) is an acceptable aerodynamic design. The main point of this figure is that, after just ten generations (less than 230 CFD runs), a large number of acceptable designs were found, guided by the cumulatively enhanced metamodel. This rather small number of CFD runs (for a six-dimensional, nonlinear parameter space) is to be contrasted with the total number of metamodel interrogations used to drive the optimisation, which is on the order of three million.

https://doi.org/10.1017/S0001924000001299 Published online by Cambridge University Press

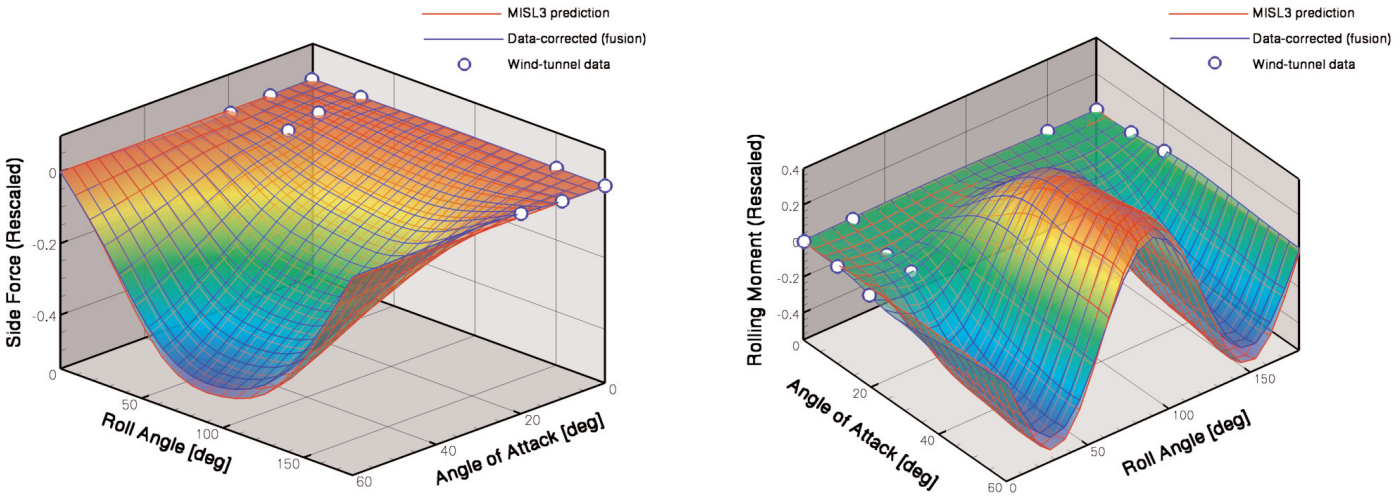


Figure 10. MISL3 and experimental side force (rolled co-ordinate system) and rolling moment predictions.

5.0 MUTUAL ENHANCEMENT OF DATA SETS

We now present a data processing technique that can help the engineer or scientist develop a better global understanding of multi-dimensional data, specifically limited experimental test data, with the aid of physics-based computational information. A data fusion technique is used to produce a response surface acting as a global interpolant of the data, both computational and experimental. The advantage of this technique, as opposed to conventional interpolation or data fitting techniques, is that the interpolation of the experimental data can be regarded as essentially computational- (model-) based. In other words, the physics of numerical simulations can be used to interpolate (and, possibly, extrapolate) the experimental data where the sampling is sparse or even absent, while still respecting the integrity of the experimental data. *Vice versa*, the resulting metamodel representation can also be regarded as a calibration of the computational model, based on experimental data. The method used here employs a particular form of regularisation, in which a hypersurface going through the experimental data is 'supported' by additional computational constraints. The calculation of this hypersurface uses the same technology as previously described for the refuelling drogue canopy and asymmetric payload fairing design applications.

5.1 Correction of aerodynamic databases using experimental data

The fusion of wind-tunnel data with aerodynamic databases is considered next. The goal is to assimilate wind-tunnel data for use in flight simulations. The example shown below corresponds to a generic body-tail missile configuration (not shown). This application merges two data sets: an experimental (wind-tunnel) aerodynamic data set and a computational data set. These data are used together to create a hypersurface metamodel. As before, the data point inputs to the hypersurface calculation are referred to as the support vectors of the surface, by analogy with SVMs. The 'computational' *support vectors* are supplied by the MISL3 code[41]. This MISL3 database consists of forces and moments predictions for a wide range of angles of attack, roll angles, and Mach numbers in the subsonic, transonic, and supersonic range. The experimental support vectors are supplied by a wind-tunnel test for a much smaller range of conditions consisting of three Mach numbers, four roll angles, and a subset of the angle-of-attack range. The goal was to produce an 'error database', defined as the difference between a fit to the MISL3 database and a fit to the experimental data. The fits were obtained by constructing a low-order analytic (smoothly varying) four-dimensional response surface generated using simultaneously the MISL3 support vectors (in one 'plane' of the parameter space), and the experimental support vectors (in a separate 'plane' of the parameter space). The error database can then be used to 'correct' the MISL3 database to take into account the experimental measurements. Conversely, this method is an example of how to construct smart interpolation and extrapolation schemes in cases where limited quantities of data are available. From a user perspective, this process is automatic and does not require the specification of any equations. Only support vectors from alternate sources of data are needed. In this case (see Fig. 10), there is a very specific shape to the extrapolation in angle-of-attack and roll angle, a shape which can be thought of as being 'inherited' from the MISL3 database.

In a different application, we have used this method to improve pressure drag predictions based on the fusion of high-resolution CFD calculations with limited pressure tap measurements in a wind-tunnel test, essentially using the experimental data to anchor the prediction, and the detailed spatial distribution provided by the CFD to accurately integrate the pressure. It is argued that the technique outlined above is quite general and can be used, either to 'fill-in' where limited quantities of experimental data are available, or to 'fine tune' the results of computational analyses using the limited data available.

6.0 UNCERTAINTY PREDICTION

As previously mentioned, a key aspect of this technology is the ability to estimate further sampling needs and estimate the response surface quality, based on uncertainty estimation. The uncertainty estimation in NEAR-RS is based on propagating statistical descriptions of uncertainty in measurements or input data to estimates of uncertainty in the response surface coefficients themselves (see Equation (5)). This approach makes use of the covariance of the output measurements and is based on the theory of best linear unbiased estimation[32].

Although the examples shown in the previous sections, pertaining to design/optimisation and data fusion, did not make explicit use of the uncertainty prediction, it is worth pointing out that the uncertainty prediction is an automatic byproduct of the response surface identification in NEAR-RS. Whether it can be used in optimisation in a manner similar or complementary to the

https://doi.org/10.1017/S0001924000001299 Published online by Cambridge University Press

expected improvement function one gets from Kriging[11] remains to be determined. Uncertainty estimates for the previously described data fusion of MISL3 and wind-tunnel data were generated, but were mainly used for the purpose of attributing confidence intervals to the overall prediction. In this section, we present a fourth application, in which the uncertainty produced by the model is explicitly visualised. By way of illustration, an example of response surface with uncertainty estimation is given in Fig. 11, for a model of the aerodynamic performance of an X-38-type forebody configuration as a function of Mach number and pitch angle upon atmospheric reentry. The X-38 data set consists of Euler CFD computations carried out by NASA Ames, and was used as a test case in the development of CFDExchange[20], a system for the automated archival, search, and results retrieval of CFD solutions, using a relational database. In this system, the retrieved CFD solutions are considered as points in a multidimensional space. In the present case, the solution space is parameterised by Mach number, pitch angle, grid resolution, and solution algorithm settings.

The response surface shown in Fig. 11 is for given resolution and algorithm settings. The CFD points for these conditions are the only inputs to the response surface calculation, and are indicated in the figure in the form of triangular symbols (∇). Although the data points are clustered only around the nominal trajectory parameters, the resulting surface is analytical and exists everywhere in space, including outside the intended X-38 reentry trajectory. The uncertainty prediction is indicated by the shading of the response surface. Because no seed uncertainty (±3σ) based on the CFD data was available in this case, the response surface uncertainty is shown in relative form, i.e., as the amplification factor $\sqrt{var(F)}/\sigma$ (denoted in the figure as $\sigma_{RS}/\sigma_{DATA}$). A large amplification of the prediction uncertainty is visible away from the data points. Note that these are not two separate models (one for the dependent variable, and one for the uncertainty), but a single metamodel (see Equations. (5 and 6)) that incorporates both the dependent variable and its uncertainty.

As computational technologies mature, there is a growing need to quantify prediction uncertainty. While the purpose of the uncertainty evaluation here is merely illustrative, giving a flavour what is possible, more quantitative uses of the prediction uncertainty have previously been suggested. For example[20], in cases where the goal is to characterise the overall response surface within an acceptable uncertainty bound, the uncertainty prediction can be used to guide the further population of the parameter space, for example, by 'placing' further runs/cases in areas of highest uncertainty. On the other hand, if the goal is purely to optimise (rather than globally characterise), then, as pointed out in Ref. 12, this type of strategy may require an excessively large number of evaluations, and needs to be balanced by the local exploitation of promising areas.

7.0 CONCLUSION

Several design optimisation and data fusion applications employing cumulatively enhanced global metamodels were presented. These metamodels play a key role in surrogate-based optimisation because global searches and design space exploration become practical as a result of the small computational cost of evaluating the surrogate. In the present applications, one involving the redesign of a refuelling drogue canopy, the other the design of an asymmetric payload fairing, the use of metamodel surrogates amounted to estimated computational savings of three to four orders of magnitude. This estimate is based on the total number of evaluations used to explore the response surfaces, and the ratio of computational times between a metamodel evaluation (less than a second) and the time required for a computational analysis (on the order of hours). We also showed how these global metamodels

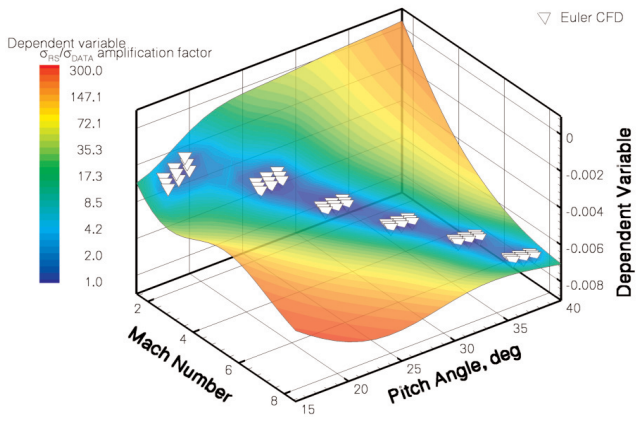


Figure 11. Uncertainty modeling for X-38 reentry.

can be used for data fusion. Specifically, computational data can be used to formulate a dimensionally augmented response surface based on both computational and experimental support vectors. Used in this manner, the computational data has the effect of regularising the response surface identification, which is highly beneficial when confronted with the problem of data generalisation based on sparse experimental data. Conversely, this can be viewed as a method for fine-tuning or calibrating computational codes based on limited experimental data. The NEAR-RS methodology provides simultaneously a metamodel and the uncertainty on this model. The response surface uncertainty is obtained by propagating statistical descriptions of dependent variable uncertainty in the input data to estimates of uncertainty in the response surface coefficients themselves. This approach makes use of the covariance of the output measurements and is based on the theory of best linear unbiased estimation. Both the response surface and its uncertainty are described by compact analytical equations whose coefficients are exported by NEAR-RS to a MATLAB® compatible file for subsequent use in other applications. Thus, cumulative global metamodels (1) allow data structure flexibility and the use of heterogeneous data sets, (2) provide a fully analytic, mathematical description that can be easily manipulated and shared between aerospace customers, and (3) have a rational basis for propagating uncertainty estimates, making these models suitable for risk assessment. A potential application of metamodel uncertainty is its use as a driver for decision making or in further populating data sets.

ACKNOWLEDGEMENTS

Financial support for this work was provided by the following sponsoring agencies: Air Force Research Laboratory (refuelling drogue/canopy design, and launch vehicle payload fairing design), Office of Naval Research, Program Officer Brian Almquist, Code 321OE (aerodynamic database enhancement), and NASA Ames Research Centre, Technical Monitor Terry Holst (X-38/response surface uncertainty). The authors would like to thank Daniel A. Pruzan, Marnix F. E. Dillenius, Michael R. Mendenhall, and Laura C. Rodman for supplying the application data that made possible the development of the methods presented in this paper. We are also grateful to the reviewers for their valuable suggestions.

https://doi.org/10.1017/S0001924000001299 Published online by Cambridge University Press

REFERENCES

1. BURMAN, J., PAPILA, N., SHYY, W. and GEBART, B.R. Assessment of response surface-based optimisation techniques for unsteady flow around bluff bodies, AIAA 2002-5596.
2. DELOACH, R. and ERICKSON, G.E. Low-order response surface modeling of wind-tunnel data over truncated inference subspaces, AIAA 2003-0456.
3. FORRESTER, A.I.J., BRESSLOFF, N.W. and KEANE, A.J. Response surface model evolution, AIAA 2003-4089.
4. HIROKAWA, N., FUJITA, K. and IWASE, T. Voronoi diagram based blending of quadratic response surfaces for cumulative global optimisation, AIAA 2002-5460.
5. KNILL, D.L., GIUNTA, A., BAKER, C.A., GROSSMAN, B., MASON, W.H., HAFTKA, R.T. and WATSON, L.T. Response surface models combining linear and Euler aerodynamics for supersonic transport design, *J Aircr*, January-February 1999, **36**, (1), pp 75-86.
6. KRISHNAMURPHY, T. Response surface approximation with augmented and compactly supported radial basis functions, AIAA 2003-1748.
7. PAPILA, N., SHYY, W., GRIFFIN, L. and DORNEY, D.J. Shape optimisation of supersonic turbines using response surface and neural network methods, AIAA 2001-1065.
8. WALKER, E.L. Statistical calibration and validation of a homogeneous ventilated wall-interference correction method for the National Transonic Facility, NASA TP-2005-213947.
9. RAIS-ROHANI, M. and SINGH, M.N. Efficient response surface approach for reliability estimation of composite structures, AIAA 2002-5604.
10. VITTAL, S. and HAJELA, P. Confidence intervals for reliability estimated using response surface methods, AIAA 2002 5475.
11. JONES, D.R., SCHONLAU, M. and WELCH, W.J. Efficient global optimisation of expensive black-box functions, *J Global Opt,* 1998, **13**, pp 455-492.
12. SOBESTER, A., LEARY, S.J. and KEANE, A.J. On the design of optimisation strategies based on global response surface approximation models, *J Global Opt*, 2005, **33**, (1), pp 31-59.
13. QUEIPO, N.V., HAFTKA, R.T., SHYY, W., GOEL, T., VAIDYANATHAN, R. and TUCKER, P.K. Surrogate-based analysis and optimisation, *Progress in Aerospace Sciences*, 2005, **41**, pp 1-28.
14. POGGIO, T. and GIROSI, F. Network for approximation and learning, *Proc IEEE*, September 1990, **78**, (9), pp 1481-1497.
15. RAETH, P.G., GUSTAFSON, S.C., LITTLE, G.R. and PUTERBAUGH, T.S. Stretch and hammer neural networks for n-dimensional data generalisation, Air Force Wright Laboratory Report WL-TR-91-1146, 1992.
16. VAPNIK, V.N. *The Nature of Statistical Learning Theory*. Springer, 1995.
17. GUNN, S.R. Structural modelling with sparse kernels, Proceedings of the 13th IFAC Symposium on System Identification, Rotterdam, Netherlands, 2003.
18. GRAMACY, R.B., LEE, H.K.H. and MACREADY, W. Parameter space exploration with Gaussian process trees, Proceedings of the International Conference on Machine Learning, Omnipress and ACM Digital Library, 2004, pp 353-360.
19. FRIEDMAN, J.H. Multivariate adaptive regression splines, *The Annals of Statistics*, 1991, **19**, (1), pp 1-67.
20. RODMAN, L.C., REISENTHEL, P.H. and CHILDS, R.E. An automated documentation and reporting system for CFD, AIAA 2002-0986.
21. http://www.nearinc.com/near/project/rs.htm
22. ZELDIN, B.A. and MEADE, A.J. Integrating experimental data and mathematical models in simulation of physical systems, *AIAA J*, 1998, **35**, (11), pp 1787-1790.
23. MATHERON, G. Principles of geostatistics, *Economic Geology*, 1963, **58**, pp 1246-1266.
24. CRESSIE, N. The origins of Kriging, *Mathematical Geology*, 1990, **22**, (3), pp 239-252.
25. SACKS, J., WELCH, W.J., MITCHELL, T.J. and WYNN, H.P. Design and analysis of computer experiments, *Statistical Science*, 1989, **4**, pp 409-435.
26. ORR, M. Optimising the widths of radial basis functions, Proceedings of the Fifth Brazilian Symposium on Neural Networks, Belo Horizonte, Brazil, 1998.
27. POTTMANN, M. and SEBORG, D.E. Identification of non-linear processes using reciprocal multiquadric functions, *J Process and Control*, 1992, **2**, (4), pp 189-203.
28. HARDY, R.L. Multiquadric equations of topography and other irregular surfaces, *J Geophys. Res*, 1971, **76**, (8), pp 1905-1915.
29. FRANKE, R. Scattered data interpolation: Test of some methods, *Math of Comp*, 1982, **38**, (157), pp 181-200.
30. ORR, M.J.L. Regularisation in the selection of radial basis function centres, *Neural Computation*, 1995, **7**, (3), pp 606-623.
31. STRANG, G. *Linear Algebra and Its Applications*, Academic Press, New York, USA, 1980, p 142.
32. NORTON, J.P. *An Introduction to Identification*, Academic Press, New York, USA, 1986, pp 87-119.
33. STRICKLAND, J.H. Axisymmetric bluff-body flow: a vortex solver for thin shells, Sandia Report SAND91-2760, 1992.
34. PEDIRODA, V., POLONI, C. and CLARICH, A. A fast and robust adaptive methodology for airfoil design under uncertainties based on game theory and self-organising-map theory, AIAA 2006-1472.
35. CHILDS, R.E., REISENTHEL, P.H., ROSE, J. and MALY, J. Large asymmetric launch vehicle payload fairing, NEAR TR 611, 2005.
36. BUNING, P.G., JESPERSEN, D.C., PULLIAM, T.H., KLOPFER, G.H., CHAN, W.M., SLOTNICK, J.P., KRIST, S.E. and RENZE, K.J. Overflow user's manual version 1.8r, NASA Ames Research Center, 2000.
37. CHAN, W.M. and STEGER, J.L. Enhancements of a three-dimensional hyperbolic grid generation scheme, *Appl Math and Comput,* 1992, **51**, pp 181-205.
38. GIUNTA, A.A., WOJTKIEWICZ, S.F. and ELDRED, M.S. Overview of modern design of experiments methods for computational simulations,' AIAA 2003-0649.
39. AUDET, C. and DENNIS, J.E. A pattern search filter method for nonlinear programming without derivatives, *SIAM J Optim*, 2004, **14**, (4), pp 980-1010.
40. KOLDA, T.G., LEWIS, R.M. and TORCZON, V. Optimisation by direct search: new perspectives on some classical and modern methods, *SIAM Review*, 2003, **45**, (3), pp 385-482.
41. LESIEUTRE, D.J., LOVE, J.F. and DILLENIUS, M.F.E. MISL3 aerodynamic analysis for finned vehicles with axisymmetric bodies, NEAR TR 561, 2004.

https://doi.org/10.1017/S0001924000001299 Published online by Cambridge University Press