# AN AUTOMATED DOCUMENTATION AND REPORTING SYSTEM FOR CFD

Laura C. Rodman,[*] Patrick H. Reisenthel,[**] and Robert E. Childs[†]
Nielsen Engineering & Research, Inc.
Mountain View, California

## ABSTRACT

Aerodynamics design requires a tremendous number of CFD calculations. Data management tools are needed which can improve the efficiency of these calculations, facilitate the understanding of errors and uncertainty in the data, and distill high-level knowledge and trends from the solution sets. This paper describes a knowledge system for documenting and reporting high-level CFD information. The documentation portion of the system gathers information from CFD calculations, such as the input values and solution quality measures, and stores this information in a database. The reporting mechanism allows users to query the information and to retrieve output, and includes a method for integrating data from a large number of runs so they can be viewed in a multidimensional design space. Metrics that quantify the uncertainty and inconsistency information are included in the design data for inspection by the user. Users can also learn how well the computational technology applies to a given problem, the limitations of the technology, the errors present in the solution, and the distribution of a flow feature throughout the design space.

## INTRODUCTION

Much of vehicle aerodynamics engineering is becoming increasingly dependent upon the use of CFD calculations for design and analysis tasks. The number of calculations required and the amount of data generated is predicted to increase tremendously in the next five years; for a major design effort the calculations might number in the thousands. With such a large number of calculations comes the need to efficiently manage and interpret the results as an aid to the human designer. Thus, data management tools must be developed which can improve the efficiency of these calculations, facilitate the understanding of errors and uncertainty in the data, and distill high-level knowledge and trends from the solution sets. This need has been recognized by personnel from both commercial and government organizations; therefore, successful data management systems will play a key role in future aerospace design environments.

The purpose of this project is to develop a knowledge management system for capturing, documenting, and reporting high-level CFD information. This knowledge can then be used to develop intelligent reasoning tools. High-level information for CFD calculations includes: (1) the objective of the calculation, (2) the algorithms, physical models, and fidelity level that compose the computational technology, (3) a description of the engineering problem being solved, (4) the input and grid parameters used during the calculation, (5) measures of the solution accuracy and convergence, (6) appropriateness of the turbulence and/or transition models used, (7) significant flow features present in the solution, (8) postprocessed data that are of interest to the end users of the calculation, and (9) the computational and manpower resources required to perform the calculation. This type of information can be used to analyze and interpret a set of results, and to assist in planning future calculations.

This project builds upon previous information management work at NEAR, called CFDExchange, in which a CFD database was developed that contains engineering problem descriptions, code usage information, project management information, and results from the calculations.[1] However, CFDExchange requires manual entry of much of the calculation description, and although the labor requirement for each individual run is small, for large design efforts the cumulative effort becomes prohibitive. The goal of the

---

* Senior Research Scientist, AIAA Senior Member
** Chief Scientist, AIAA Senior Member
† Exec. Vice President, AIAA Senior Member

current project is to develop a means of automating the information input into the CFD database and to compile the results from many runs to form a design space of interest to users.

The documentation and reporting system consists of components for data storage and management, information input, and information output. The CFD database design was extended during this work to manage a broader category of information. The documentation and reporting system design is modular so that different facilities for performing the various tasks can be used. To support this modular design, a data exchange standard based on XML (Extensible Markup Language) was developed.

The knowledge capture and documentation portion of the system provides a mechanism for automatically collecting and storing high-level information in a structured database. The reporting mechanism allows users to query the information and to retrieve output, and includes a method for integrating multifidelity data from a large number of runs so they can be viewed in a multidimensional design space. Uncertainty and inconsistency information is included in the design data for easy interpretation by the user. Users can also learn how well the computational technology applies to a given engineering problem, the limitations of the technology, the errors present in a solution set, and the distribution of a flow feature throughout the design space.

The information that can be input into the system during the course of a calculation includes the CFD code inputs, the grid parameterization, the convergence if applicable, grid and solution quality measures, and postprocessed quantities. In order to demonstrate this capability, a module that parses an input file, extracts calculation parameters, and inserts them into the database was developed. Also, an analysis that performs an error assessment on a CFD solution was used to determine the major sources of error, and this information was imported into the database.

The reporting mechanism retrieves relevant results from the database according to a user's request. For design studies, the user may request results from a family of calculations, each at a different point in a design space. The system compiles the results from the relevant runs and computes an analytic multidimensional representation of the data based on these results. This allows the user to create one-, two- or three-dimensional plots of the design space as desired for easy interpretation. Since data from a variety of sources are combined, and this data includes error measures, the analytical representation takes the inconsistency and uncertainty characterization of the data into account.

TECHNICAL OBJECTIVES

The specific objectives of this project are to:

(1) Develop a data model for the stored information and the characterization. Select an appropriate database scheme and management system for this model.
(2) Develop the system infrastructure and design the interface for the various file interrogation modules and the database access.
(3) Incorporate existing file interrogation schemes into the system.
(4) Develop a preliminary reporting mechanism including the multidimensional analytical representation.
(5) Develop the sample system with communication links between the various modules to test the approach.

SYSTEM DESIGN

An overview of the documentation and reporting system is given in this section, followed by a description of the enabling technologies behind this system. The various components of the system are described next, including the database design to model and store the high-level CFD information, the information input into the database from various sources, and the information output that extracts data from the database in response to user queries.

Functionality Overview

**Figure 1** shows a schematic for the system. The data sources for input into the system include a CFD code input file, grid file, solution file, and output file. User input can also be gathered. Modules are used to interrogate these input sources to extract the information that is of interest. These interrogation modules then output that information in a standardized format, which is then imported into the database. On the output side, data is exported from the database based on user queries. While the input data is typically associated with individual calculations, the output data compiles many calculations together to create a global metamodel of the data, hereafter referred to as the "design space." (The term "design space" is used interchangeably for either the design data itself or the response surface metamodel representing this data. This is distinct from the "parameter space," in which both the data and/or the metamodel are embedded.) The output data is provided in a standardized format. This output can then be used by various postprocessing modules, such as a design space interpolation routine, a plotting package, or a browser.
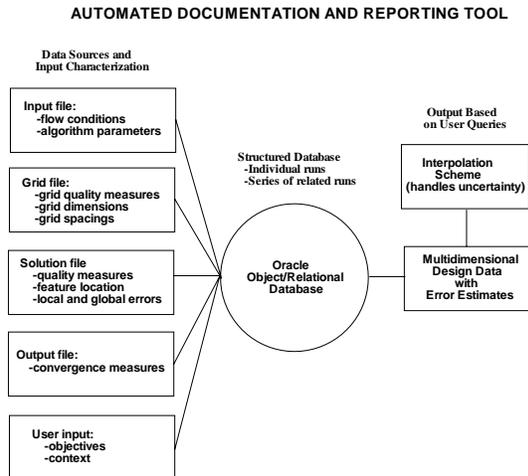
American Institute of Aeronautics and Astronautics

**AUTOMATED DOCUMENTATION AND REPORTING TOOL**

**Data Sources and Input Characterization**

Input file:
-flow conditions
-algorithm parameters

Grid file:
-grid quality measures
-grid dimensions
-grid spacings

Solution file:
-quality measures
-feature location
-local and global errors

Output file:
-convergence measures

User input:
-objectives
-context

**Structured Database**
-Individual runs
-Series of related runs

Oracle Object/Relational Database

**Output Based on User Queries**

Interpolation Scheme (handles uncertainty)

Multidimensional Design Data with Error Estimates

**Figure 1.-** Overview of the Automated Documentation and Reporting System.

### Enabling Technologies

The automated documentation and reporting system is feasible due to several recent enabling technology developments. These technologies are: (1) automated solution quality assessment, (2) multidimensional data representation using a response surface methodology, and (3) improved database and information management products for electronic data exchange. These technologies are described below.

### Automated Solution Quality Assessment

The automated solution quality assessment builds on the development of solution interrogation methods that can estimate the error due to several sources within a calculation, and can propagate that error throughout the flow field to obtain a measure of the global error in solution quantities.[2] The solution quality assessment (SQA) method makes quantitative estimates of spatial discretization errors within CFD solutions on structured grids. These errors result from flow features that are too fine to be resolved by the CFD spatial discretization scheme and the grid. Analyses are performed to determine the local error throughout the domain and the solution characteristics, such as flow separation. Ongoing work on the SQA method includes errors from sources other than spatial discretization, such as from turbulence modeling and boundary conditions. The goal of this work is to establish reliable uncertainty bars on data from CFD solutions. Both the local errors and the solution accuracy can be provided from this assessment, and this ability to postprocess CFD solutions and to extract high-level information is key to the automated documentation functionality of the system.

During the current project, the focus of the work was not on the development of new solution quality assessment modules, but rather on the ability to integrate existing modules for CFD in the current system.

### Multidimensional Data Representation

Each individual CFD calculation stored in the database represents a single point in a multidimensional parameter space that is of interest to a designer. A major portion of the automated reporting is the ability to compile results from a number of individual CFD calculations to create a global metamodel of data. Parameters such as Mach number, angle of attack, configuration, etc., define the dimensionality of the design space. Creating this design space from a number of individual runs is difficult, however, since the data may be sparse or irregular. Also, if the design space is of high dimension, it is difficult to visualize the data. In this project, a multidimensional response surface representation of the data is used to provide interpolation between the sparse data points, to handle the representation of uncertainty in the data, and to provide a method to generate one- or two-dimensional slices of the space for easy viewing. The analytical response surface provides a representation of the design space that is continuous and can be constructed "on the fly." Its analytical nature makes it amenable to automated searches, and low dimensional plots can be generated, based on suitable projections.

The response surface (sometimes referred to as a *metamodel*) is a nonlinear fit to scattered multidimensional data. A central issue to constructing an appropriate response surface model is the so-called *curse of dimensionality*, in which the number of data points required increases exponentially with the number of dimensions.[3] This difficulty precludes the use of conventional schemes, such as polynomial approximations.[4] The multidimensional aspects can be addressed using one of several methods. These include multidimensional splines,[5-6] neural networks,[7] and radial basis function approaches.[8-9] The framework adopted in this work is the radial basis function (RBF) approach. The RBF framework was selected for the following reasons:

1.  Artificial neural networks (ANN) and RBFs both are simpler in implementation and appear to have superior speed over multidimensional splines, which is an important consideration in terms of on-the-fly visualization of design data.
2.  Although RBF models can be viewed as three-layer feedforward neural networks, they have several key advantages over conventional ANNs, notably: linearity-in-the-parameters and guaranteed and rapid convergence, as opposed to the learning

algorithms (e.g., back propagation) used in conventional feedforward ANNs.[10]

The radial basis function model used is described in detail in Ref. 11. The solution method includes techniques that allow the response surface model to be a nonlinear fit to data rather than an interpolant that must go through each data point. These techniques are *classification*, in which nearby points are associated, and *singular value decomposition*, which is a method of handling ill-posed data redundancy.

One important addition to the multidimensional response surface technology that was explored in this project was the concept of response surface uncertainty. The presentation of uncertainty information in CFD data to engineers is a goal of this work. It was found that due to the linearity in parameters, it is possible to make use of well-established statistical least squares results. In particular, the data uncertainty can be correctly propagated onto an uncertainty on the multidimensional response surface itself. The derivation of this extension is given in Ref. 11.

Results of the response surface derivation are shown in **Figures 2** and **3. Figure 2** shows how uncertainty information would be presented for a response surface. 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, using a thin dashed line. **Figure 3** shows an example of uncertainty estimates using both classification and singular value decomposition. Note that the overall response surface uncertainty takes into account not only the individual variances of the training data, but also the bias between these data and the nominal solution. Note also the effect of the individual variances on the location of the resulting multiquadric: the points with the least uncertainty are those that tend to "pull" the response surface towards them, as one would intuitively expect. Results such as those presented in **Figure 3** also tell us that more data are required, in order to reduce/resolve the uncertainty in the neighborhood of $x = 0.4$. This is, of course, obvious in the one-dimensional example shown here, but provides the basis for automation in a high-dimensional space (see Results section).

One of the issues with which one is typically confronted when attempting to perform an interpolation is whether the interpolant is *local* or *global* . Although the goal is to obtain a global analytical representation of the data, global interpolants may not be the best solution to this problem. This has to do with the lack of independence of certain regions of the parameter space with respect to others. For example, artificial "sympathetic curvature" in one area of the design

space may be generated as a consequence of gradients in a completely different area of that space. This unfortunate consequence of global interpolation schemes is one of the reasons that most interpolation schemes are local in nature (take, for example, the multitude of spline methods). It is, therefore, often preferable to focus on the task of producing high-quality local interpolants which are smoothly blended together.
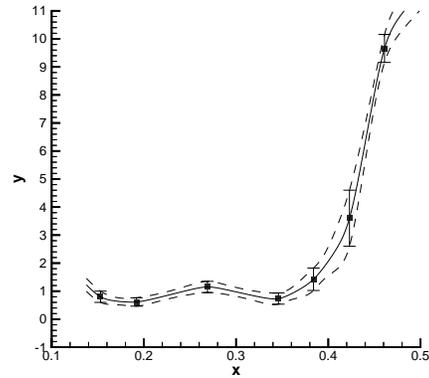


**Figure 2**.- Response Surface Uncertainty.

This problem can be handled by subdividing the data into subgroups referred to as "clusters." (There are no particular constraints on the clusters: a given data point can belong to several clusters, and the cluster regions can overlap, whether or not they have any points in common). A membership function can be defined in standard fuzzy-logic fashion, as a probability-like measure that the dependent variable at a given point in the parameter space "belongs" to a cluster. The framework of fuzzy logic offers several practical advantages: (1) it is the mechanism for generating smooth transients between patches, (2) it
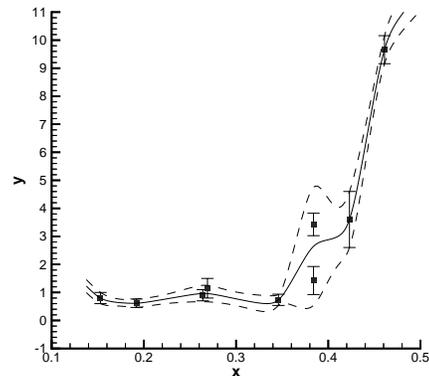


**Figure 3.-** Uncertainty, Classification, and SVD Regularization Combined.

circumvents convoluted and difficult to implement IF-logic constructs in multiple dimensions, and (3) can be either automated with suitable defaults or customized (through the membership functions) for greater control. This concept is explained in more detail in Ref. 11, and an example of this customization is provided in the Results section.

## Database and XML Technologies

Part of the task for this project is to determine a computational infrastructure that will facilitate the structuring, importing, and exporting of data into and out of a database. This infrastructure includes the type of database to use to store the information and the communication scheme by which various code modules share information.

Various choices for a database management system were considered. The previous work on CFDExchange[1] used a relational database to store knowledge about CFD runs. It was decided that the continued use of an RDBMS was most appropriate for this task. Relational databases are common for high volume transactions. Relational database management systems (RDBMS) are well-understood and provide excellent performance in search and retrieval tasks. Queries are placed against a relational database using the SQL language. Modern RDBMS have a provision for storing and managing graphics, numerical data, and large text through the use of binary or text Large Objects (LOBs).

A decision was made early on to base part of this infrastructure on XML (Extensible Markup Language). XML provides a standardized means of creating a hierarchical markup language to describe a given application. The markup can be used to create XML documents with content that is self-describing, and the standard ensures that multiple applications can use the common format for data exchange. XML is rapidly becoming widespread for this purpose, and off-the-shelf software toolkits exist to parse XML documents and create an application programming interface (API) for accessing the document's internal data. XML can play several roles in an information system. XML markup format was chosen for the data transfer tasks, but not for database storage and management. There are two data transfer tasks. First, the data input into the database comes from multiple sources. This data comes from the CFD code, the code environment, the associated files, and the user. XML is used to take this input, translate it to a common format, and then use one tool to convert this common format to the database format. The other data transfer task is the output for the automatic reporting. This output may need to be sent to a variety of output devices, such as a postprocessor, a browser, a word processor, or a printer. Thus, the output from the database will be in the form of XML, and various conversions or style sheets can be used to translate the common output to the appropriate format for the different output devices.

An XML-based markup language for CFD was written, and this markup was based on the existing CFDExchange database structure. A second XML format was developed for the database output. While the input XML describes information about a single run, the output format must represent data from multiple runs that can be viewed together in a design space.

## System Development

As mentioned above, the automated documentation and reporting system consists of three components: (1) the database design, (2) information input modules, and (3) information output modules. The development of these components is discussed below.

## Database Design

The database design in this work was based on the CFD database used in CFDExchange. A detailed description of this baseline design can be found in Reference 1, and an overview is given here. This database is intended to document CFD calculations and their use within an organization. Four main categories of information are stored: (1) CFD Technologies, (2) Engineering Problems, (3) Project Management and Resources, and (4) Results. CFD technologies include CFD codes, and the algorithms and physical models incorporated within that code (many codes provide multiple algorithm or model options to be selected by the user). A goal of CFDExchange is to provide a platform for documenting these technologies along with the results that were obtained by using them. It is well-known, however, that an evaluation of CFD codes or technologies cannot be made without considering the specific engineering problem to which they are applied. CFDExchange documents both CFD technologies and engineering problems. A combination of the two is required in order to understand the results and to use those results to determine the applicability, limitations, accuracy, and proper usage of a CFD code.

In addition to the CFD technologies and engineering problems, resource information is important for evaluating the cost and risk of a calculation. Project management information should also be included, such as the projects or tasks that a calculation supports, the organization for which the work was performed, and the personnel responsible for the work.

A wide variety of results can be included in this system. Formal results, such as those found in technical papers, can be included, and these results consist of both text and graphics. Informal results, which typically are not described in technical reports,

should also be documented within CFDExchange. Comments by the users could describe useful information such as special steps required to get a code to run, and instances where a code did not perform adequately. CFD calculations create a suite of data files (input, output, grid, and solution files, for example). CFDExchange maintains a record of the data files associated with each calculation and their locations.

The database design also stores the postprocessed output quantities that are of most interest to designers. These output quantities will vary depending upon the engineering problem being solved. For example, if a CFD calculation supports the external aerodynamics design of a vehicle, then quantities of interest will be the force and moment coefficients and the center of pressure. Along with the values, the error and variance in each quantity can be stored, along with a text object that contains information used to generate a plot of the output. The text object includes the grid zone, location, plane of orientation, length scale, and priority of any errors in the solution.

## Information Input

Information that is used to populate this system originates from several sources: the user, the CFD code input file(s), the solution files, and the grid file. Input that comes from the user is acquired through Web-based forms which connect to a database.

As mentioned in a previous section, information extracted from CFD files is written in XML format. To demonstrate, an OVERFLOW input file parsing module and a solution file interrogation module were developed.

The OVERFLOW input file module was written so that all input quantities for every grid zone are read in and parsed. In addition, the default value for each input parameter is used if the input file does not specify a value. Thus, all possible parameters are populated and output in XML format.

An existing solution interrogation module was modified to output XML markup that is used to characterize error information. The regions in which error occurs in a solution are identified, and the top ten regions are prioritized for output. The output for each error includes the grid zone, location (in both grid and physical space), the orientation, an appropriate length scale (used for plotting purposes), and the components of the solution (e.g., streamwise momentum) that display the greatest error (used to determine which quantities show the error when plotted).

Software was developed that parses the XML provided and inputs it into the relevant tables in the relational database. These modules were used to extract the data from the documents and to insert them into the appropriate database fields. Most of the XML content has a one-to-one correspondence with a database field. The exception is the error plotting information that is described above. This entire section of markup is inserted into the database as a text large object, within one field. The information is not required to be reflected in the database design for querying or reporting tasks, and will only be extracted as input for a plotting package.

## Information Output

Information that is stored in the database is exported into an XML format, which is used as the standard data exchange format for input into the multidimensional design space module. The data required to create a multidimensional design space is stored in tables in the relational database. This data is retrieved via user query, and it is exported in a format that can be read by the response surface module.

A software module was written that takes a series of CFD runs as input and selects the design data from the database for those series. The data are then exported as XML format. This format provides for data from multiple runs to be combined, and provides for the integrated quantities of most interest to designers. This format also accounts for null entries or irregular data in the design space matrix. The XML format is converted to tabular data for the response surface program or for a plotting program to read.

## RESULTS

The above technologies and system components were tested using a sample data set. As mentioned above, software modules were developed to create the input XML format from CFD files and insert from this format into the database. Other modules query the database and output the results in XML.

To test the system functionality, a data set was created within the relational database for use with the output module and with the response surface module. The data set used was the Kistler Aerospace K-1 Orbiter Vehicle database, which consists of a series of CFD runs designed to estimate the vehicle's aerodynamic characteristics during the design stage. These calculations were performed by Nielsen Engineering & Research under contract with Kistler Aerospace between January 1997 and March 1999. The CFD runs performed cover a wide range of atmospheric reentry flow conditions as well as a limited number of parametric variations of the vehicle shape.

For the present purpose, only a subset of the Kistler CFD database was used. This subset consists of Euler and Navier-Stokes runs carried out using the OVERFLOW code. In these calculations, the

freestream Mach number was varied between 1.1 and 25 and the angle of attack (AOA) ranged from 0 to 25 degrees. (No yaw angle variations were considered because of the vehicle's axial symmetry). Variations in the geometric configuration were considered in terms of three parameters: the flare diameter (D), the overall length of the vehicle (L), and the nose shape. The baseline configuration has a nose shape which is referred to as "flat." A separate configuration was investigated using a rounded nose.

This subset of CFD runs (referred to hereafter as the data set) consisted of 37 Navier-Stokes runs and 237 Euler runs. The Navier-Stokes runs covered the transonic/low supersonic flight regime ($1.1 \leq$ Mach $\leq 2.5$) at angles of attack $0° \leq$ AOA $\leq 25°$, with the exception of one row of data at AOA $= 8°$ and $2.0 \leq$ Mach $\leq 25$. The range of flow conditions calculated using the Euler code was $2.0 \leq$ Mach $\leq 20$, and $0° \leq$ AOA $\leq 20°$. The geometric configurations included two nose radii, two flare diameters, and three vehicle lengths.

The Kistler data set consists of processed results, i.e., integrated quantities such as the axial force coefficient, $C_A$, normal force coefficient, $C_N$, moment coefficient, $C_m$, lift coefficient, $C_L$, drag coefficient, $C_D$, and center of pressure location, $X/L_{cp}$ These integrated quantities constitute the dependent variables of the interpolation. They are generically denoted $y$. The five independent variables are: Mach number, AOA, flare diameter, D, vehicle length, L, and nose roundness, %round. The goal is to construct a five-dimensional response surface $y = F(Mach, AOA, D, L, \%round)$ which is continuous and analytic.

The present data set comprised both Navier-Stokes and Euler CFD runs. In some regions of the parameter space (namely, $2.0 \leq$ Mach $\leq 2.5$), these runs overlapped. Two data point clusters were defined accordingly (see **Figure 4**). In **Figure 4**, Cluster #1 is made of Navier-Stokes results exclusively, while Cluster #2 comprises a majority of Euler runs. Since the Navier-Stokes runs were deemed more accurate than the corresponding Euler runs at the same conditions, the peak membership associated with the Navier-Stokes cluster was chosen in the present example to be twice that of the Euler cluster.

All of the results presented below are the result of a five-dimensional interpolation. All visual representations, either one- or two-dimensional, are, therefore, *projections* of a five-dimensional space. Whenever a two-dimensional plot (either carpet plot or contour plot) is shown, three of the five parameters are held constant. Similarly, whenever a one-dimensional plot (line graph) is shown, four out of five parameters are held constant.

**Figures 5** and **6** depict projections of the local response surfaces for clusters 1 ("Navier-Stokes")

and 2 ("Euler"), respectively, for the normal force $C_N$ as a function of Mach and AOA. Each of these response surfaces (RS) are obtained independently of one another, using the radial basis function (RBF) network methodology mentioned previously. The resulting blended, fuzzy-logic-based, global RS is shown in **Figure 7**. Note that, because each RBF network operates on data which are scaled to the unit hypercube, the global RS is *de facto* associated with a multiresolution anisotropic network of RBFs in physical space.

The two-dimensional representations of **Figures 5-7** are useful to formulate an intuitive idea of overall trends in the parameter/design space. Quantitative detailed information is most often obtained from one-dimensional cuts.

The usefulness of having a continuous, analytic description of the design space can be further illustrated using less conventional plots, such as **Figure 8**. **Figure 8** depicts the normal force coefficient as a function of vehicle length and flare diameter at a hypothetical Mach number of 5.5, hypothetical angle of attack of 19 degrees, and for a hypothetical nose roundness factor of 33%. Such information can be useful to designers for the examination of global trends. Note that the absence of any data symbols in **Figure 8** simply reflects the fact that no data points exist for these design conditions.





**Figure 5.-** 2-D Projected Local Response Surface of $C_N$ versus Angle of Attack and Mach Number ($1.1 \leq M \leq 2.5$, Navier-Stokes) for a Vehicle Length of 611″, Flare Diameter of 22', and Nose Roundness Factor of 0%.

**Figure 4.-** Schematic Illustrating the Location of Available Computational Data and the Definition of Clusters Based on the Methodology Used. The Points Shown in this Figure are for a Vehicle Length of 611″, Flare Diameter of 22', and Nose Roundness Factor of 0%.

## Use of Multidimensional Response Surface Uncertainty

As mentioned earlier in this paper, the use of covariance analysis, combined with the desirable property of linearity-in-the-parameters that is afforded by the RFB network formulation, provides us with the ability to propagate uncertainty from the data points to the response surface itself. The ability to estimate RS uncertainty is a powerful tool. First, it provides a confidence level on any evaluation of the response surface, whether in interpolation or extrapolation mode, a distinction which is not always easy to make in multiple dimensions. Second, the RS uncertainty can be used as a driver for further populating the database (database augmentation).
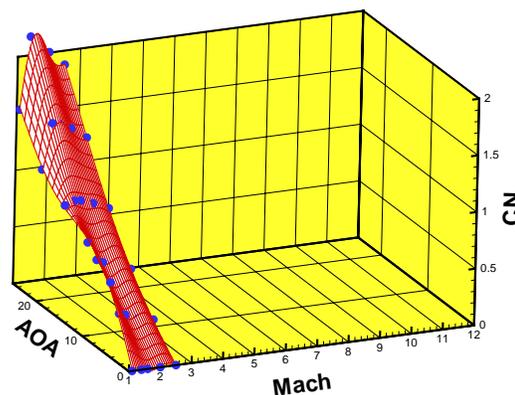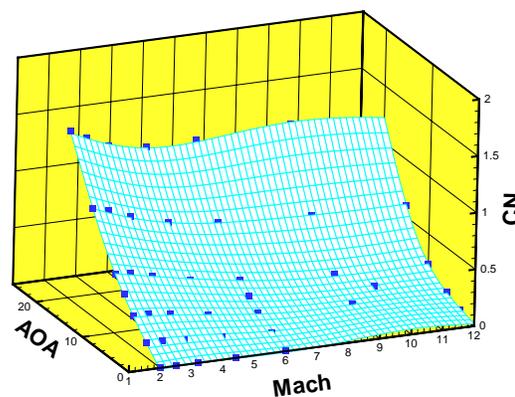


**Figure 6.-** 2-D Projected Local Response Surface of $C_N$ versus Angle of Attack and Mach Number ($2.0 \leq M \leq 12.0$, Euler) for Vehicle Length 611″, Flare Diameter 22', and Nose Roundness Factor 0%.

It is interesting to note that uncertainty remains low and commensurate with the data point uncertainty in the more densely populated, interpolative areas of the parameter space. By contrast, the (five-dimensional) uncertainty blooms in the areas of the parameter space that are largely extrapolative. This is a very useful property and is consistent with the earlier results obtained in one dimension.
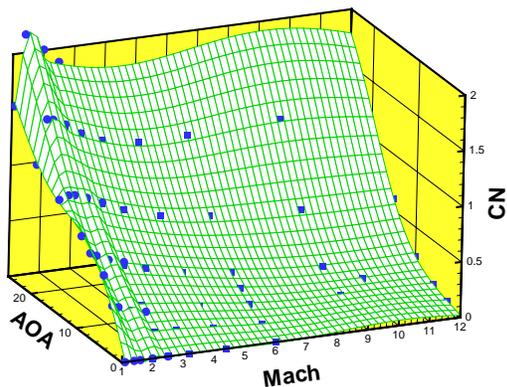
American Institute of Aeronautics and Astronautics

**Figure 7.-** 2-D Projected Global Response Surface of $C_N$ versus Angle of Attack and Mach Number $(1.1 \leq M \leq 12.0)$ for Vehicle Length 611″, Flare Diameter 22', and Nose Roundness Factor 0%.
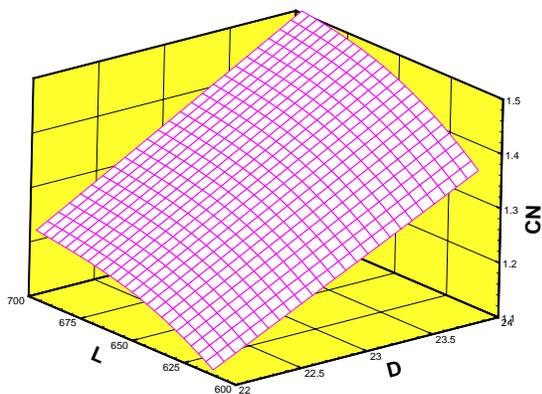


**Figure 8.-** 2-D Projected Response Surface of $C_N$ versus Vehicle Length and Flare Diameter at Mach 5.5, Angle of Attack 19˚, and nose roundness factor of 33%.

These observations suggest that RS uncertainty could be used as a metric for driving the automatic population of a high-dimensional data base. Suppose eight parametric CFD runs are carried out simultaneously on a parallel computer. The RS uncertainty metric can help answer the question: what parameters should be used for the next eight runs? In other words, RS uncertainty can help maximize the amount of useful information and, therefore, accelerate the overall process by populating the database more efficiently. In reality, one would expect uncertainty to be one of several drivers, which could include

objectives, as well as confidence memberships in various codes/fidelity levels to further optimize the process.

To test the RS uncertainty component of this idea, one would ideally like to carry out additional CFD runs until the uncertainty is reduced below a certain level. For the purpose of this study, it is easier, instead, to decimate the data set and track the evolution of the RS uncertainty. The data set decimation experiments were performed using a random number generator to "throw away" certain points in the data base. As previously mentioned, the full data set for Cluster #2 consisted of 244 CFD runs. Various five-dimensional response surfaces (along with their five-dimensional uncertainties) were constructed on the basis of 124, 68, and 26 randomly chosen CFD data points.

A comparison of the two-dimensionally projected RS uncertainties over the range $2.0 \leq \text{Mach} \leq 12.0$, $0° \leq \text{AOA} \leq 20°$ for various decimation levels is shown in **Figures 9** through **12** in the form of uncertainty contours. The data points at the length, flare diameter, and nose roundness factor corresponding to the correct projection are indicated, when available, using white squares. It is clear, from these results, that not only does the uncertainty diminish as the space is increasingly populated, but also that this is a convergent process.

## CONCLUSIONS AND POTENTIAL APPLICATIONS

The future of aerodynamics design points to the need to manage a large amount of CFD solution data, far more than can be analyzed and interpreted by human effort alone. Thus, there is a need for data management tools to automatically gather information about the calculations, assess the accuracy of the solutions, detect significant features, and report to a user an aggregate view of a large number of solutions integrated into a multidimensional space that portrays performance and uncertainty in that performance.

The goal of this work was to demonstrate how the recent technologies of solution quality assessment, multidimensional data representation, recent database management tools, and XML can be used to lay the foundation for this automated documentation and reporting system for high-level CFD data. To show feasibility of the system, this project accomplished the following objectives:

(1) Extended an existing data model for the high-level information and adapted it to a new database management system platform,
(2) Developed the system infrastructure to import and export XML data into the database,

9

(3) Demonstrated that independent solution and file interrogation modules can be used to extract the information required for input into the database and can produce this data in the common format required,

(4) Investigated several options for plot generation,

(5) Developed a multidimensional analytical representation methodology which is capable of interpolating data from disparate sources, blending them together, and providing a global uncertainty model of the data.

This work demonstrates that the key technologies are available to create a full-scale CFD knowledge system in future work. This system can be used to assist engineers in a variety of tasks, such as archiving the CFD runs that support a given project, tracking resource usage, determining the limitations of a particular computational technology, or assessing uncertainty and risk on the basis of computational error information, data sparseness, and a comprehensive representation of parameter space variations. The system also provides an infrastructure for new solution interrogation and feature extraction modules to be easily plugged in. These new modules can increase the feature detection capabilities and lead to data mining within large numbers of solutions.

With the addition of intelligent feedback in the system, information from the database can be used to assist in the task of defining where in a design space, and at what fidelity level, future calculations should be run. Minimizing the number of calculations required to adequately blanket a design space, will lead to greater efficiency in data generation.

To meet these goals, future research and development efforts will expand the CFD knowledge management system around four interacting functions: (1) automated knowledge capture and feature extraction, (2) reporting function and metamodel synthesis, (3) expanded database functionality, and (4) high-level control function.

## ACKNOWLEDGMENT

## REFERENCES

1. Rodman, L. C. and Tzou, K. T. S., "An Online System to Support CFD Technology Exchange," 1st Symposium on Marine Applications of CFD, Hydrodynamics/Hydroacoustics Technology Center, NSWC, May 1998.

2. Childs, R. E., Faller, W. E., and Bower, W. W., "Automated System for CFD Solution Quality Assessment," NEAR TR 549, Jun. 1999.

3. LeClair, S., "Innovations in Materials Design - Computational Approach to the Design/Discovery of New Materials," *AFRL Technology Horizons*, Mar. 2000, pp. 20-21.

4. 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," *Journal of Aircraft*, Vol. 36, No. 1., Jan.-Feb. 1999, pp. 75-86.

5. Klein, V. and Batterson, J. G., "Determination of Airplane Model Structure From Flight Data Using Splines and Stepwise Regression," NASA TP 2126, Mar. 1983.

6. Riemenschneider, S. D. and Shen, Z., "Multidimensional Interpolatory Subdivision Schemes," *SIAM Journal of Numerical Analysis*, Vol. 34, No. 6, Dec. 1997, pp. 2357-2381.

7. Raeth, P. G., Gustafson, S. C., Little, G. R., and Puterbaugh, T. S., "Stretch and Hammer Neural Networks For N-Dimensional Data Generalization," Air Force Wright Laboratory Report WL-TR-91-1146, Jan. 1992.

8. Poggio, T. and Girosi, F., "Network for Approximation and Learning," *Proc. IEEE*, Vol. 78, No. 9, 1990, pp. 1481-1497.

9. Zheng, G. L. and Billings, S. A., "Radial Basis Function Network Configuration Using Mutual Information and the Orthogonal Least Squares Algorithm," *Neural Networks*, Vol. 9, No. 9, pp. 1619-1637.

10. Broomhead, D. S. and Lowe, D., *Complex Syst.,* Vol. 2, 1988, p. 321.

11. Rodman, L. C., Reisenthel, P. H., and Childs, R. E., "An Automated Documentation and Reporting Tool for Aerospace Design Data," NEAR TR 555, June 2000.
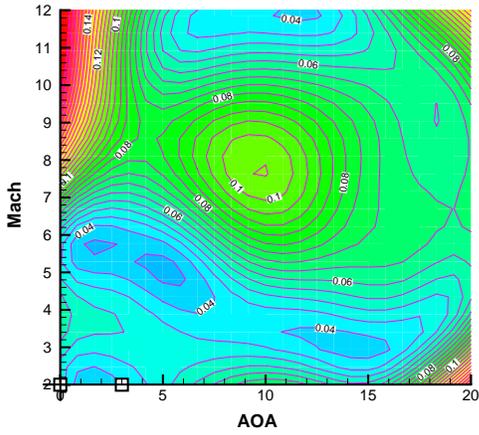
**Figure 9.-** Contour Plot of Response Surface Uncertainty When Using 26 Data Points out of 244.
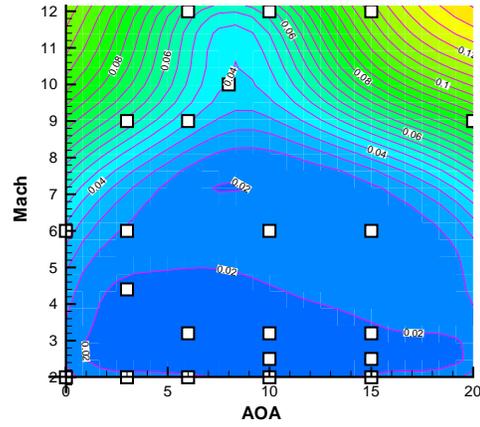


**Figure 10.-** Contour Plot of Response Surface Uncertainty When Using 68 Data Points out of 244.



**Figure 11.-** Contour Plot of Response Surface Uncertainty When Using 124 Data Points out of 244.
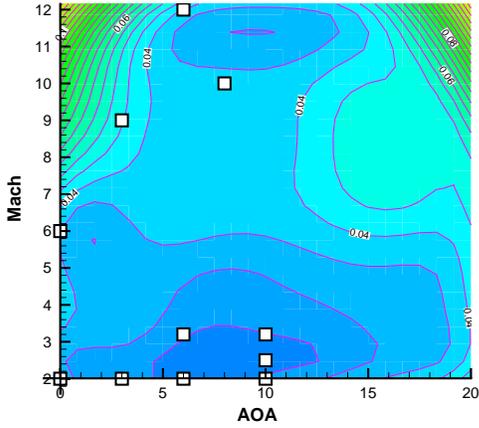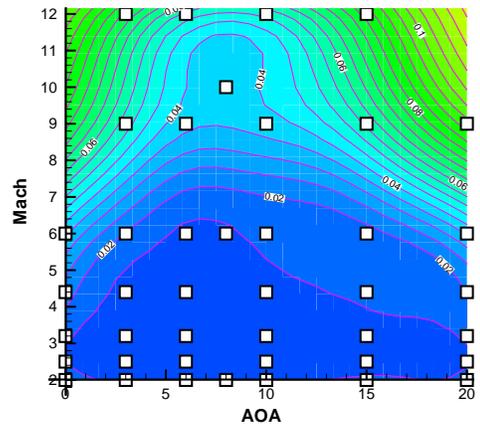


**Figure 12.-** Contour Plot of Response Surface Uncertainty When Using 244 Data Points out of 244.

American Institute of Aeronautics and Astronautics