

Development of a Prototype Model-Form Uncertainty Knowledge Base

Lawrence L. Green

NASA Langley Research Center, Hampton, Virginia 23681

Uncertainties are generally classified as either aleatory or epistemic. Aleatory uncertainties are those attributed to random variation, either naturally or through manufacturing processes. Epistemic uncertainties are generally attributed to a lack of knowledge. One type of epistemic uncertainty is called model-form uncertainty. The term model-form means that among the choices to be made during a design process within an analysis, there are different forms of the analysis process, which each give different results for the same configuration at the same flight conditions. Examples of model-form uncertainties include the grid density, grid type, and solver type used within a computational fluid dynamics code, or the choice of the number and type of model elements within a structures analysis. The objectives of this work are to identify and quantify a representative set of model-form uncertainties and to make this information available to designers through an interactive knowledge base (KB). The KB can then be used during probabilistic design sessions, so as to enable the possible reduction of uncertainties in the design process through resource investment. An extensive literature search has been conducted to identify and quantify typical model-form uncertainties present within aerospace design. An initial attempt has been made to assemble the results of this literature search into a searchable KB, usable in real time during probabilistic design sessions. A concept of operations and the basic structure of a model-form uncertainty KB are described. Key operations within the KB are illustrated. Current limitations in the KB, and possible workarounds are explained.

Nomenclature

<i>ARMD</i>	=	Aeronautics Research Mission Directorate
<i>CDF</i>	=	Cumulative Distribution Function
<i>CFD</i>	=	Computational Fluid Dynamics
<i>DB</i>	=	Database
<i>DPW</i>	=	Drag Prediction Workshop
<i>EDF</i>	=	Empirical Distribution Function
<i>FY</i>	=	Fiscal Year
<i>KB</i>	=	Knowledge Base
<i>JPEG</i>	=	Joint Photographic Experts Group, also JPG
<i>NASA</i>	=	National Aeronautics and Space Administration
<i>MDAO</i>	=	Multi-Disciplinary Analysis / Optimization
<i>PDF</i>	=	Portable Document Format
<i>SA</i>	=	Systems Analysis
<i>SQL</i>	=	Structured Query Language

I. Introduction

DURING fiscal year (FY) 2014, an effort was initiated within the National Aeronautics and Space Administration (NASA) Aeronautics Research Mission Directorate (ARMD), Fundamental Aeronautics Program, Aeronautical Sciences Project, Multi-Disciplinary Analysis / Optimization and Systems Analysis (MDAO & SA) Research Theme to investigate and quantify the effects of a particular class of uncertainties, known as model-form uncertainties, within aerospace design. The effort was continued during FY 2015 within the reorganized NASA ARMD, under the Transformative Aeronautics Concepts Program Transformational Tools & Technologies Project¹, Aeronautical Sciences Subproject, MDAO & SA Research Theme.

There are two general types of uncertainties: aleatory and epistemic². Aleatory uncertainties are usually due to random variations, such as the naturally occurring variation in material properties. Aleatory uncertainties might also be introduced through imperfect and variable manufacturing processes. These uncertainties are considered to be irreducible, meaning that expending resources here may better quantify the uncertainty, but will not eliminate the uncertainty. Epistemic uncertainties generally result from a lack of knowledge, or from approximations to the true physics behaviors, whether intended or not. Epistemic uncertainties are generally considered to be reducible, meaning that expending resources here will better quantify the uncertainty and may help to reduce the uncertainty.

One type of epistemic uncertainty is called model-form uncertainty. The term model-form uncertainty means that among the choices to be made within an analysis, there are different forms of the analysis process, which each give different results, for example, for the same configuration at the same flight conditions. Thus, uncertainty in the design results exists simply due to choices within the analysis process. Examples of model-form uncertainties include computational grid density, grid type, and solver type used within a computational fluid dynamics code, or the number and type of model elements within a finite element analysis. There may be better and worse disciplinary choices from the stand point of technical accuracy, but the best choice for a particular application may instead be constrained by the project schedule and/or budget. That is to say, that the best way to get a system-level answer in some situations may not be the best combinations of disciplinary choices for the problem. Thus, some level of disciplinary approximation and uncertainty may be deemed acceptable so that the system analysis can be accomplished within a timely and cost-effective means. The impact of the various model-form uncertainties must be propagated through the design process to the system figures of merit. The issues, in this scenario, then become: 1) how best to quantify the uncertainty as associated with the modeling choices that are made during the design process, 2) how best to allocate any resources that may be available for uncertainty reduction, and 3) determining how much input uncertainty may be tolerated within the scope of imposed response requirements.

The objectives of this work are to identify and quantify a representative set of model-form uncertainties and to make this information available to designers through an interactive knowledge base (KB) that can be used during probabilistic design sessions, so as to enable the reduction of uncertainties in the design process through resource investment. An extensive literature search has been conducted to identify and quantify typical model-form uncertainties present within aerospace design; this literature is briefly described herein, and discussed more fully in a companion document to this paper³, the KB User Guide. An initial attempt has been made to assemble the results of this literature search into a searchable knowledge base using Microsoft Access⁴. In this context, the term “knowledge base” follows the usage of Chudoba⁵ in which the elements of the knowledge base include a mixture of experience, values, contextual information and expert insight that provides a setting for evaluating and incorporating new experiences and information. Knowledge derives from information as information derives from data. Knowledge develops over time and assembles via certain key components, such as experience, truth, complexity, judgment, rules of thumb and intuition, values and beliefs⁵. Thus, some of the key elements of model-form uncertainty knowledge base are:

1. Data sources (publications) and data items (text blocks, figures and tables from specific publications)
2. Contextual information
3. Referential information
4. A means to enable qualitative/quantitative comparisons, judgments, values and beliefs about the data and information⁵⁻⁷
5. Rules of thumb derived from the data, information and experience
6. A means to date and update the data, information, experience and rules of thumb
7. Ability of users to rate the credibility^{6,7} of the content and to provide comments about out of date or incorrect information that can be changed or access removed.

The remainder of this paper illustrates how these elements have been brought together. First, a brief concept of operations is presented. Then, a description of a typical data item is presented. Then, the structure of the current database is described. Finally, some typical uses and limitations of the current knowledge base within an aerospace design session are illustrated. Workarounds to the limitations are suggested. Appendix A provides a brief example to show the importance of correctly categorizing epistemic uncertainties.

II. Concept of Operations

A brief concept of operations is now presented for the model-form uncertainty knowledge base; more information on this topic is presented in the accompanying User Guide³. At the highest level, a Microsoft Access database (DB) file consist of tables, queries, forms and reports. The tables are similar to those within Microsoft Excel⁸.

First, one can imagine different classes or levels of users that might interact with the knowledge base in different ways. Microsoft Access offers a natural way to classify users, by what permissions the users are granted: Read Only, Add New, or Edit. The Read Only permission simply allows users to utilize existing tables. The Add New permission also gives users the ability to create new records within existing tables. The Edit permission allows users to edit existing records in tables. All of these permissions can be enforced by macros, however, the permissions can be easily over-ridden. Any user could create their own new queries, forms, reports and macros. An additional means to restrict user permissions may be desirable; this can be accomplished through custom coding via Structured Query Language (SQL⁹) or Visual Basic¹⁰, but such options were deemed to be beyond the scope of this effort.

If the KB is to be a community resource, a user should be required to request access to the KB through an administrator. The user would then be assigned certain permissions within the KB, depending upon their database (DB) skills and disciplinary needs. One might expect that a community of active, contributing KB users develops with common interests related to a wide variety model-form uncertainties, and that each KB user has an interest in continually improving the KB. Each KB user has specific needs or desires related to finding qualitative or quantitative information about particular model-form uncertainties; these needs and interests likely change over time.

A KB user has several ways to go about obtaining the qualitative or quantitative information about specific model-form uncertainties. They might search the KB for particular authors, publications or even author affiliations with which they are familiar that address the uncertainty of interest. More commonly, the user might simply search for any KB entries related to specific combinations of keywords. Having found some information on the topic of interest, they might refine their search to restrict the authors, publication dates or types of interest.

Additionally, the user might wish to consider only KB entries that have a certain average or minimum credibility rating, as assigned by other KB users. The KB user might review specific data items or data sources and find items upon which they would like to comment or rate the credibility. User supplied comments might take the form of suggested corrections that would require some disposition and possible arbitration by others, or they might simply provide additional insights about the applicability of the data items or data sources in certain situations. If an insufficient amount of information is found related to a particular search topic, the user may wish to add new data sources and/or data items to the KB for future use, or they may simply want to add additional keywords to items already in the KB to make them more readily accessible.

Having identified a group of data sources and data items that are applicable to a particular search topic, the KB user may wish to create a reference list suitable for incorporation into their own publications to document the uncertainty sources used for a specific design study. The KB user might also wish to perform additional processing of the data items found in the KB; for example, several data items might be combined through additional processing to develop rules of thumb for cost/benefit tradeoffs. The KB user may want to extract the uncertainty estimates from selected searches into Microsoft Excel or to make them available for other applications, including web pages. The two limitations just previously noted also affect these desirable and expected operations.

A user may have certain uncertainty requirements imposed externally on the system or disciplinary responses. For example, the program administration in a certain context might require that there be no more than 10% uncertainty in a particular response. The KB user then can investigate what uncertainty demands this requirement places upon the contributing disciplinary tools by iteratively employing information from the KB. For example, a user might discover that 20% uncertainty in a given set of discipline inputs results in 15% system level response uncertainty; if they then reduce the input uncertainty to 10%, they might find that the response uncertainty now meets the program-imposed 10% margin of uncertainty in the response. But that improvement in response uncertainty (from 20% for method A to 10% for method B) could only be achieved at a certain cost and with certain disciplinary tools and methods, which may or may not be acceptable to the program. The user is able to determine: 1) if any methods documented within the KB meet the programmatic requirement, and 2) the cost / benefit for achieving the desired response uncertainty.

The KB administrators might wish to collect and report user comments that require disposition or arbitration. Once action has been completed related to specific user comments, they may want a means to report back to user community the comment dispositions, and to make new public KB resources available to the user community.

III. Overview of the Knowledge Base Structure

As noted previously, a Microsoft Access DB file consist of tables, queries, forms and reports. The tables, as in Microsoft Excel, consist of columns and rows. The columns are the data fields and rows are the data elements. The data elements include data items, described in terms of referential, contextual and judgmental information. Referential information includes, for example, the author(s), publisher and publication date. Contextual information includes, for example, the data applicability and the process used to obtain the data. Judgmental information includes, for example, the correctness and credibility of the data. As described in the companion document to this paper³, multiple data items can be combined to form rules of thumb about the cost benefit relationships for various possible resource investment choices that might be exercised to reduce the impact of model-form uncertainty within the design process (e.g., increased grid resolution in aerodynamics or higher fidelity propulsion models).

The queries, forms and reports provide an interactive means to use the KB. If properly constructed, these enable novices to use the database without interacting directly with the data tables. However, these queries, forms and reports must be developed and provided by database (and subject matter) experts ahead of time for novice users. Anticipating all the ways in which a person might wish to use the knowledge base may also be difficult. Macros and other custom coding programming techniques such as SQL⁹ or Visual Basic¹⁰ are also available to package a series of operations that might be handled by individual queries, forms and reports. In Access, the tables can be linked together to create relationships. Most relational databases employ this structure and also follow certain normalization rules about how the tables are constructed¹¹; these rules help to minimize the redundancy of information in the database and improve the maintainability of the database.

The current KB includes two types of tables: 1) those which include data items and their associated referential, contextual and judgmental information, and 2) those which describe the relationships between the various tables and the included information. Typically, the latter type of tables only include item numbers which enumerate the associations between rows in two or more specific tables. The table and field names within the tables are intended to be self-explanatory, but brief additional comments are also generally provided within the table descriptions³. The first field name for each table is generally an auto-number, primary key field, used to ensure that no two records in a table are identical; these fields are also used in the type 2 tables noted above to construct relationships between the various type 1 tables. The reader should note that if a record has been deleted from a table for any reason, the auto-number, primary key field for that record is forever lost, resulting in an out of sequence numbering for the remaining records. That is to say, if the table originally contained nine records and Record Number 4 was then deleted, the remaining records will be numbered as 1, 2, 3, 5, 6, 7, 8 and 9; if a new record is later added, it will be auto-numbered as 10 and Record Number 4 will never be reused. This situation occurs in some of the tables.

The basic building block of the KB are data items and data sources. Currently, each data item is a hyperlink to an image (usually of the Joint Photographic Experts Group, JPEG, or more commonly, JPG, variety) of a text block, figure or table from a particular data source (publication). Alternatively, the data item could also be a hyperlink to a Microsoft Word or Excel file that contains the actual text, figure or table of data. Each data item is intended to provide qualitative or quantitative model-form uncertainty estimates for a particular dependent variable (response), such as total drag coefficient of an aircraft, or the lift coefficient of aircraft. The model-form uncertainty estimate may be a function of one or more factors (independent variables), such as Mach number, or grid size. Also, each data item may be relevant only in a certain context, such as subsonic, transonic or supersonic flight and for a particular type of aerospace vehicle (e.g., transport aircraft or rocket). Each of these nuances should be available to the KB user to specify as part of their search criterion. Each data item is linked through the KB to a specific data source. Currently, each data source is a hyperlink to a document, a Portable Document Format (PDF) file, containing the data item.

Each data source has a primary author (and perhaps secondary authors), a title, publisher, publication type, a publication date and a specific location in the source publication. The location of a given data item within a publication might be associated with a page, column, paragraph, figure number or table number. Each author might be associated with certain organizations, which may change over time as they move through their careers. The same

kind of material might be published at various times over numerous years and by different authors, each with specific data sets considered and perhaps different conclusions drawn as the body of knowledge about the subject matter increases.

Figure 1 attempts to capture the complex nature of the KB as it currently exists. The figure illustrates a network of nodes and links. Each of the nodes roughly corresponds to a table or set of tables in the KB and each of the links (arrows) correspond to the implied or actual relationships (associations) in the KB information, many of which should really be bi-directional. Consider the figure starting with the purple node labeled KB Users at the middle bottom of the figure. The KB Users have certain credentials (password) that can be authenticated enabling them certain predefined permissions (blue nodes). The KB Users interact directly with the data items and data sources (shown in the green nodes). The data items are associated with a particular response and perhaps one or more factors that may take on specific values (shown in the yellow nodes). Each data item is also associated with a unique data source that has certain information associated with it (all shown in orange nodes), such as the authors and their affiliations, the author order, the publication type and date and the applicability in terms of speed range and aerospace configurations. Also, each data item has a specific location in the data source.

The user may also be interested in the keywords, credibility assessments and comments about specific data items that have been provided by other users (shown in cyan nodes). As seen in the figure, many factors can influence the credibility assessments of the data items and data sources. And, having proper keywords to identify the data items and data sources to other users is critical to making the KB useful. Comments might be provided when users have either found something in error within the database that needs to be fixed, or if they believe additional qualifications are needed for particular data items or data sources.

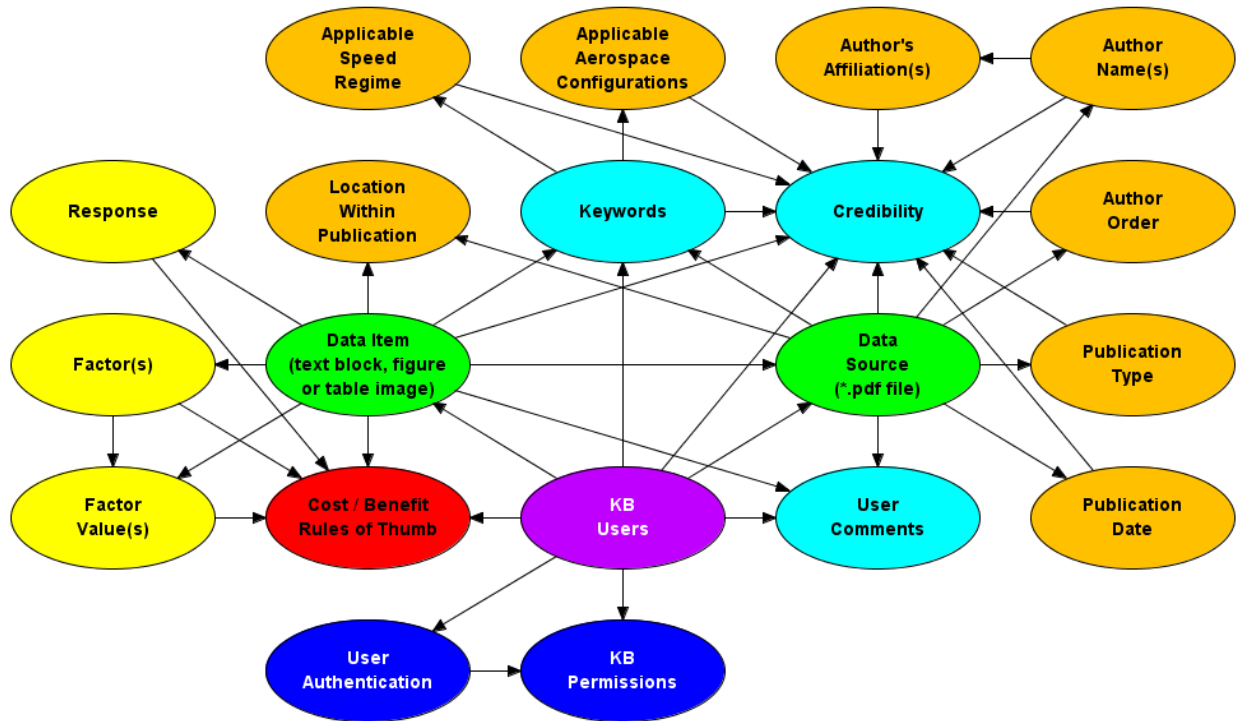


Figure 1. Illustration of the current KB information content and relationships.

IV. Data Item Description

As noted above, the basic building block of the KB are data items and data sources. Each data item is intended to provide qualitative or quantitative model form uncertainty estimates for a response. The model form uncertainty estimate may be a function of one or more factors and each data item may be relevant only in a certain context.

The following table and figure are used to illustrate some of the assumptions and operations embedded in the development of the model-form uncertainty knowledge base. Table 1 provides numeric values for two comparable discrete data sets, shown in Figure 2 as solid black square symbols and open red circle symbols. The table and figure represent two possible forms of a typical data item. The KB is built up for many such images (copies of figures or tables from the data sources). The factor value is plotted along the x-axis (abscissa), while the response value is plotted along the y-axis (ordinate) in the figure. The model-form uncertainty between the two data sets can be computed by differencing the response values at same factor value. For example, at a factor value of 5, the difference in the response values is $51.28 - 26.77 = 24.50$. However, this number may have specific dimensional units associated with it, whereas the uncertainty might be better represented in a non-dimensional form. Here, two different percentage uncertainty estimates are constructed as the ratio of the difference and either of its contributors, multiplied by 100. For example, the first uncertainty estimate is computed as $24.50 / 51.28 * 100 = 47.79\%$; the second uncertainty estimate is computed as $24.50 / 26.77 * 100 = 91.52\%$. For response values that may fall between Data Set 1 and Data Set 2, the two percentage estimates provide robust bounds on the data uncertainty, for this factor value, compared to either of these data sets.

The reader should note that no response value is provided at Factor Value = 4 for Data Set 2; furthermore, no response value is provided for either data set at intermediary values, such as Factor Value = 3.5. For this reason, a desirable option may be to fit smooth curves, via regression techniques¹², to the existing data provided in Table 1. In Figure 2, quadratic fits to both data sets, for Data Set 1, as a solid black curve, and for Data Set 2, as a dashed red curve. If equations for the response surface fits are available, an interpolation of each curve can be performed to enable computing uncertainty estimate at any factor value; this is illustrated with the blue dashed construction lines wherein a factor value of 4 is chosen and the response value for Data Set 1 has been interpolated to have a value of about 33.4.

Table 1. Sample discrete data values.

Factor Value	Data Set	Response Value
0	1	0.92
1	1	2.90
2	1	8.67
3	1	18.63
5	1	51.28
0	2	6.74
1	2	6.91
2	2	9.02
3	2	13.07
4	2	19.03
5	2	26.77

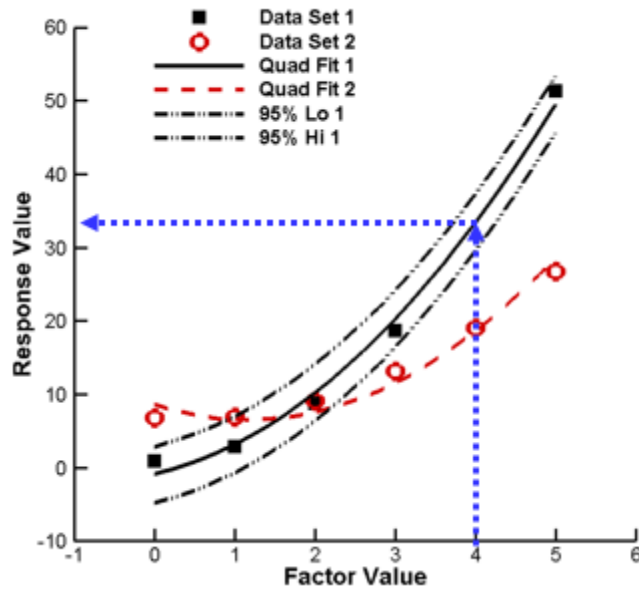


Figure 2. Illustration of the sample data set, associated curves and construction lines.

Carrying this idea one step further, the reader can see that the response surface curve (least squares fits) of Fig 2 do not actually pass through all the data points; thus, there is some additional uncertainty that can be considered simply associated with quadratic fits to the raw data. This leads one to construct confidence bounds via an appropriate standard deviation and statistically determined k factor¹³, as shown in the figure for Data Set 1 using black dash-dot-dot line. Again, construction lines and/or analytic interpolation can be used to extract relevant values from any of the curves in the figure, leading to a wide range of uncertainty estimates for a given factor value that can be extracted. These interpolation techniques will be leveraged for data items within the model-form uncertainty database. The same kinds of techniques can be applied to groups of data items, over a range of factor values to develop cost / benefit models for a particular response. Furthermore, cost / benefit models can be developed for multiple responses (i.e., aerodynamics and structures) that may be competing for resource investment within a design application, to enable cost-effective allocation of the resources (i.e., additional computational time), toward the reduction of system level uncertainties. Additional discussion of the development of cost / benefit rules of thumb for resource investment is provided in the companion document³.

An extensive literature search was conducted to identify data sources and data items within each data source that quantify typical model form uncertainties present within aerospace design. The literature search has, so far, identified almost 900 documents that could be included in the model-form uncertainty knowledge base. However, most of these publications have yet to be incorporated into the knowledge base because the current manual process for entry of data is time consuming and error prone. In fact, with just one person working part time for about six months (between ¼ to ½ of a work year total) to develop both the knowledge base structure and its content, less than 20 publications have been included so far in the current knowledge base; this does include more than 200 data items and more than 500 associations of data items to keywords. Every data item has at least one keyword association, some data items have several keyword associations. A significant need for automation employing the techniques known as “datamining¹⁴⁻¹⁵” and “big data analytics¹⁶⁻²⁰” has been identified, yet many of the techniques needed for this project are still being developed or proven to be reliable²¹⁻²². Still, the infrastructure and necessary operations for a model-form uncertainty knowledge base can be described and prototyped.

One class of model-form uncertainty documentation that was found and used in this KB arises from workshops in which common problems are solved under common guidelines by independent contributors, perhaps by using different methods. These workshops are intended to be a form of independent verification / validation, but they frequently illuminate differences (uncertainties) among the methods or contributor’s best practices. The workshop results are compared and may be used to construct bounds of the inherent uncertainties.

One such series of workshops are the five instances of the Drag Prediction Workshop (DPW) sponsored by the American Institute of Aeronautics and Astronautics (AIAA) (2001, 2003, 2006, 2009 and 2012)²³⁻²⁸. From the onset, the DPW Organizing Committee defined a set of primary objectives for the DPW Series that make the workshops quite useful and exemplary for quantifying model-form uncertainties²³.

The series of workshops involved numerous researcher using various CFD codes and methods to make predictions for common geometries of: the total drag coefficient, the pressure drag coefficient, the skin friction drag coefficient, the angle of attack to achieve a specified lift coefficient at a given Mach number, the total pitching moment coefficient, drag rise curves over a series of Mach numbers, drag polar curves, grid convergence metrics and separation bubble metrics. Statistical summaries of the computational results have been generated²⁴⁻²⁸. The DPW workshop result summaries are an excellent source of model uncertainty estimates over time, with learning and method improvements incorporated; these references are used in this abstract subsequently to illustrate some key database operations. Other similar workshops have also been identified and used to extract results for the model-form uncertainty knowledge base. For example, a workshop from 2011 examined the uncertainties associated with various ablation models for re-entry flow physics²⁹. Another workshop from 2011 examined the differences of various CFD code in the realm of high lift predictions³⁰⁻³¹. Yet another workshop of interest from 2013 documented the uncertainties among various aero-elastic codes and solutions³².

Another class of comparisons that have been employed in the knowledge base arises when different test facilities or methods are applied to the same problem³³⁻⁴¹. Examples of such have been identified for the common research aerodynamic test model³⁶ and for radiated fan noise³⁷. Still other data sources presented comparisons between test and calculations⁴².

V. Illustration of Key Features of the Database

This section of the paper illustrates some of the key features of the KB through a series of screen shots. The user opens the KB by either double clicking on the desktop Microsoft Access icon, or by double clicking the correct (*.accdb) file name within a given directory structure. As shown in Figure 3, the KB is also currently set to open the form “UserSignIn2F”. This form displays the current KB user names from a predefined list. The user would select their own user name from the list provided, enter their password and click on the button labeled “Run Macro”. This should enable the KB to compare the password entered by the user in the form to that previously stored with the user name, to enable a customized view of the available tables, queries, forms, reports and macros within the KB. This action requires retrieving the predefined user permission level value for the user from an existing table, which (unbelievably) has not yet been successfully accomplished with the available macro commands, despite numerous attempts to do so. If the user level value can be successfully retrieved, a customized view of the available tables, queries, forms, reports and macros within the KB has been demonstrated.

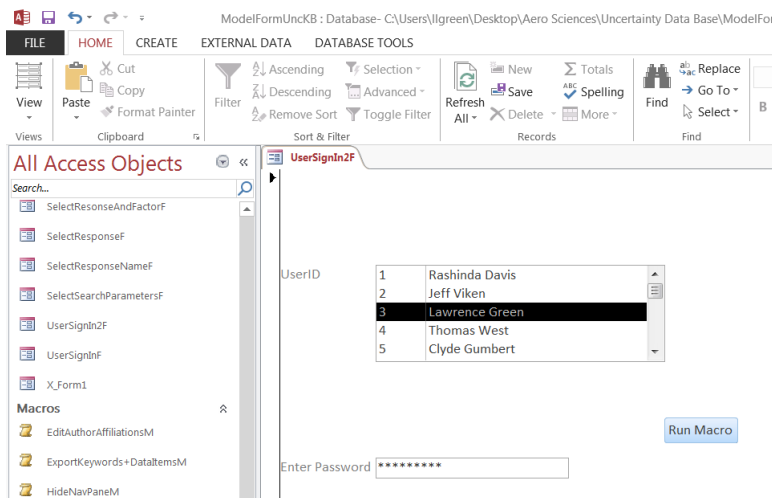


Figure 3. Illustration of user sign in to the KB.

In general, macros can be easily defined that open a form. The form will typically build a list of possible user choices from one or more of the existing tables. The form accepts input from the user. The user input is then inserted into a query that has been paired with the form. The query selects records from another table matching the user-specified choices. For example, assuming the user is interested in searching for data items by keyword(s), the user might first want to know what keywords have already been defined. As shown in Figure 4, by double clicking on the macro, highlighted in pink on the left side of the screen, named “ShowKeywordsM”, the table “KeywordChoicesT” is opened and displayed. At present 21 keywords have been defined.

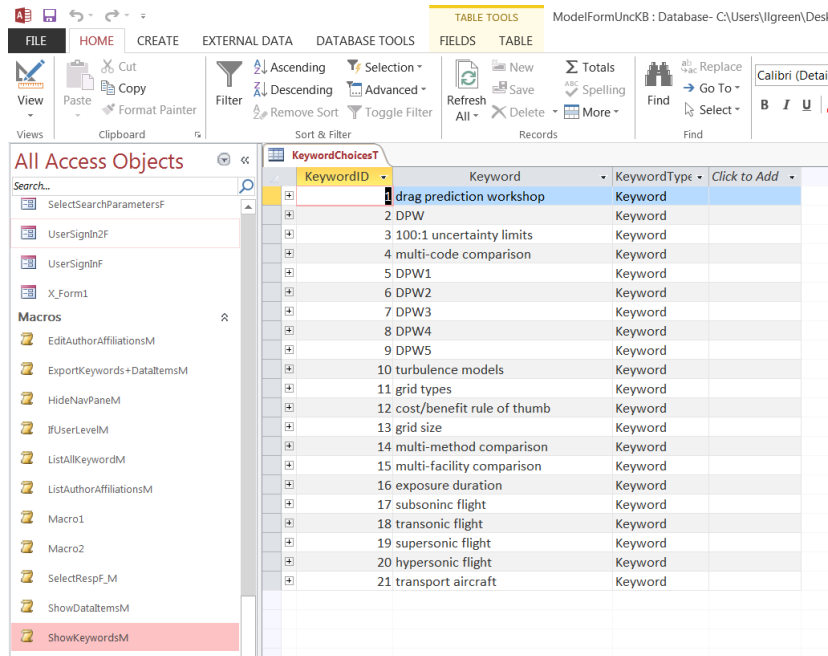


Figure 4. Illustration of keyword display in the KB.

The user can search for data items associated with a given keyword by double clicking on the macro named “SelectKeywordM”. As shown in Figure 5, this macro, highlighted in pink on the left side of the figure, opens a form (SelectKeyword2F) that displays the current keywords from the table named “KeywordChoicesT”. The user can then select any one keyword from the list and click on the button labeled “RunMacro”. This action runs a query that creates a new temporary table that associates the keyword with the data items and selects only those data record that match the keyword selected by the user. The new table is exported to Microsoft Excel; as shown in Figure 6, the user is prompted to provide a name and directory for the new table.

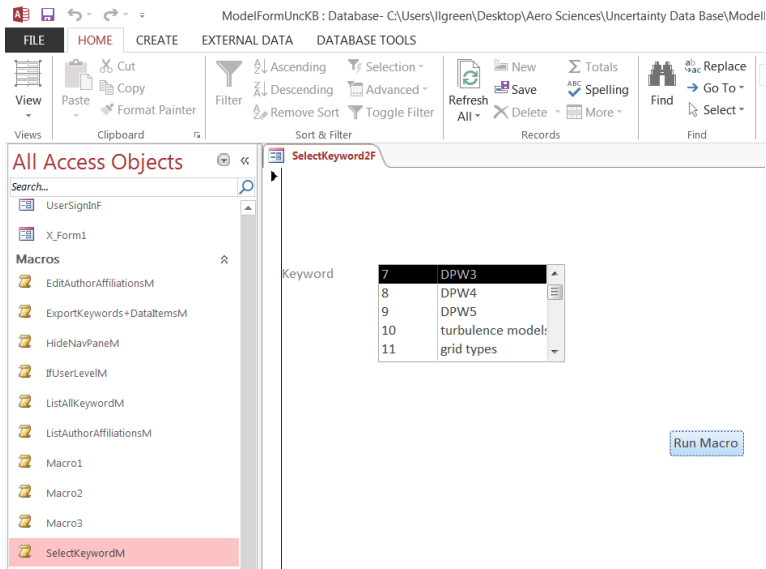


Figure 5. Illustration of keyword selection in the KB.

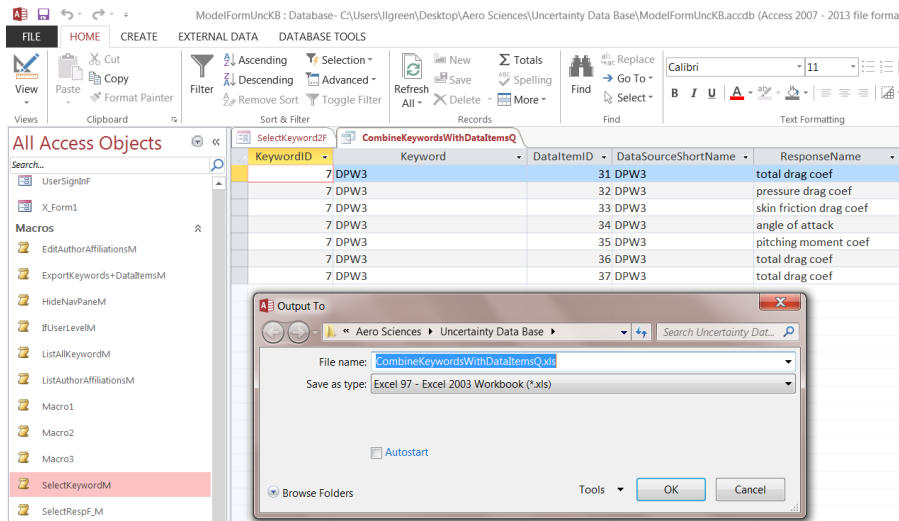


Figure 6. Illustration of keyword query export to Microsoft Excel.

Ideally, there would be no need to export the query results to Microsoft Excel, but this action is provided as a workaround to limitations that were encountered when attempting to do this operation strictly in Microsoft Access. First, if the query attempts to create a new permanent table within the KB, Microsoft Access prohibits this action from competing when computed values (the uncertainty bounds described in section above) are found in the query. In order to preserve the computed value fields, the query result has been exported to Microsoft Excel. Second, ideally, the user might like to select multiple keywords simultaneously. No means has been identified in the available Form Design Wizard or in the available macro commands to enable multiple simultaneous selections. Additional processing of the query results in Microsoft Excel would allow for down-selecting to a particular set data items associated with a single response and factor, and one or more keywords.

Finally, if the user double clicks on the highlighted macro “ShowDataItems2M”, this executes a simple query that yields a subset of the fields (for illustration purposes only) from the table named “DataItemsT”. As shown in Figure 7, the user has opened a hyperlink to the file called “Images\DPW3-Chart9b.JPG”. The image file opens on the screen while the KB is in use. In this case, it has opened the query results table with Edit permissions, so the user could add new records to the table or edit existing data records. Uncertainty bounds for this data item have been provided by the data source author (dotted lines). In many cases, the KB user will need to determine suitable bounds for the data presented in a data item image.

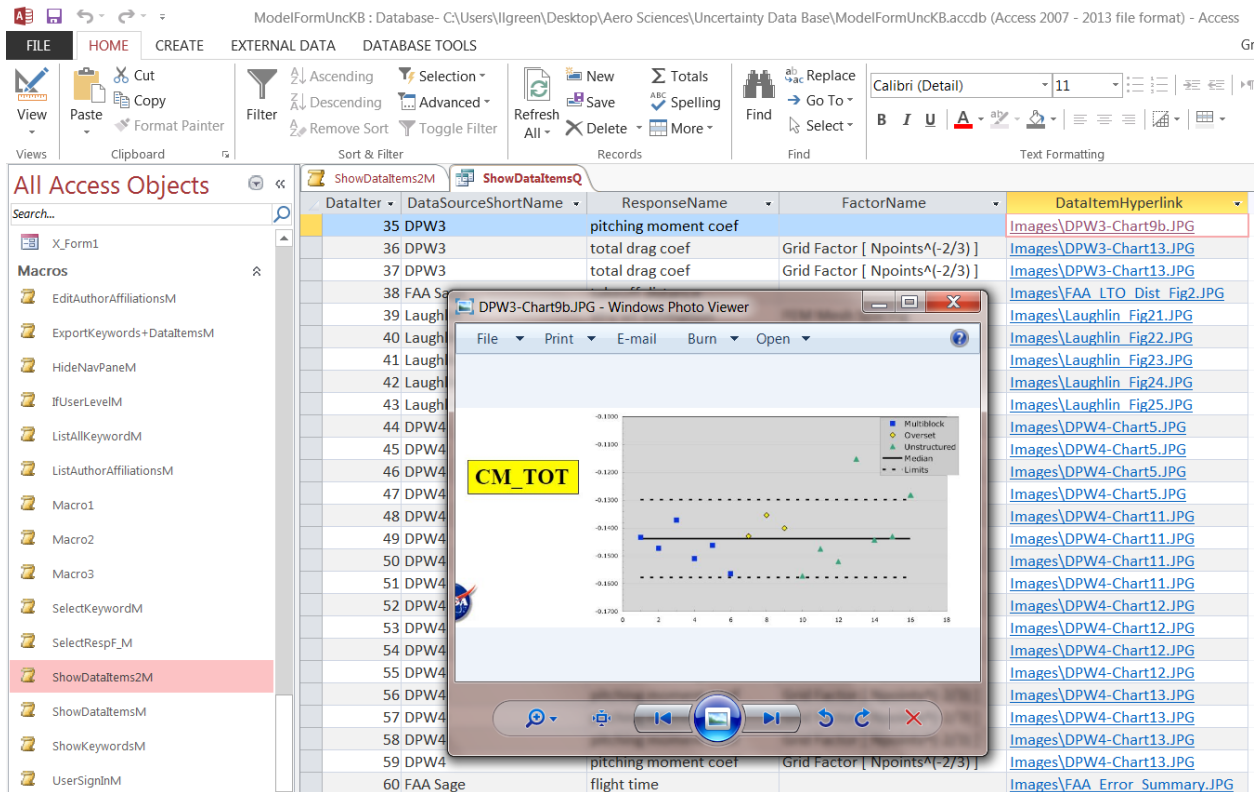


Figure 7. Illustration of hyperlink access from the KB.

VI. Concluding Remarks

The need for a model-form uncertainty knowledge base was discussed. The basic building blocks of the knowledge base, the data item and the data source, have been described. The basic structure and features of the model-form uncertainty database, including some important limitations in the current implementation, have been described. The literature search and the data items and data sources included in the current knowledge base were discussed. Several key operations of the database were also illustrated. Several important limitations in the current knowledge base implementation were identified and some workarounds were suggested. An example was also discussed which illustrates the importance of correctly categorizing and quantifying epistemic uncertainties.

The work described represents a starting point for an important effort. The need for automation in the data entry portion of the knowledge base (KB) operations was discussed. A community of interested and contributing users should be cultivated through training about the benefits of using such a KB. The ability to directly and automatically process collections of data items into cost/benefit rules of thumb needs to be developed. And, the ability to directly and automatically insert knowledge from the KB into design efforts needs to be developed and demonstrated.

Appendix A: Epistemic Uncertainty Example

To illustrate the importance of correctly categorizing epistemic uncertainties, consider the example problem $A^2 * B = C$, over the range $A = [1, 3]$ and $B = [1, 3]$. The behavior is illustrated in Figure A1 with 100 randomly chosen pairs of (A, B) values, shown as red or pink circles in the figure, overlaid on the response surface color-coded by value from blue (low) to red (high). In this domain, $C = [1, 27]$. If attention is restricted to the center of the domain, it is correct to say that if $A = 2$ and $B = 2$, then $C = 8$. However, if there is some uncertainty in the values of A and B , it is correct to say that if $A \sim 2$ and $B \sim 2$, then $C \sim 8$. The uncertainty in A and B might be aleatory, in which a normal distribution might be used to describe the parameters, such as, $A = \text{Norm}(2.0, 0.2)$, in which the

mean value is 2.0 and the standard deviation is 0.2, and $B = \text{Norm}(2.0, 0.2)$. Using a 10,000 sample Monte Carlo approach to propagate the uncertainty in A and B through to the response C, 99% of the values for response C are found in the interval [4.22, 13.51]. If however, the uncertainty in A and B is epistemic and described as a uniform distribution given by $A = B = [1.48, 2.52]$ (the 0.005 and 0.995 points on the normal cumulative distribution function (CDF) of the normal distributions for A and B noted above), then 99% of the values for response C are found in the interval [3.53, 15.08]. This represents more than a 24% increase in the response uncertainty interval for the epistemic treatment compared to the aleatory treatment.

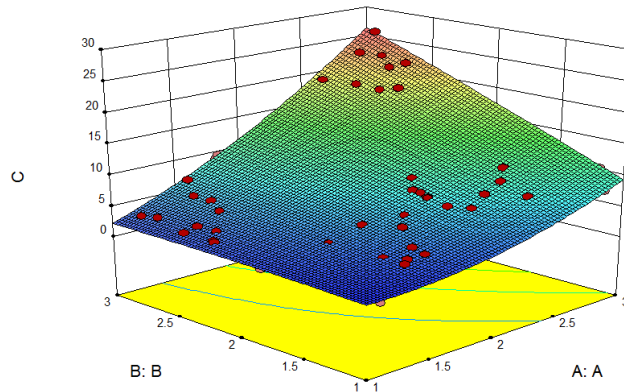


Figure A1. Illustration of sample problem.

The CDF and empirical distribution function (EDF, a discrete approximation to the CDF based upon the data available) for the response uncertainties shown in Figure A2 for both the epistemic and aleatory treatments of the input variables A and B. Comparing the blue curves (solid for CDF and dash dot dot for EDF) to the red curves, one can see that treating the uncertainties in A and B as normally distributed variations results in a more normally distributed response, with a smaller dispersion, compared to those for interval uncertainties. The degree of normality of the response is indicated by the difference between same color curves, and the amount of dispersion is indicated by the difference between different color curves of the same type (solid or dash dot dot). Clearly, it is important to categorize and model epistemic uncertainties differently than aleatory uncertainties.

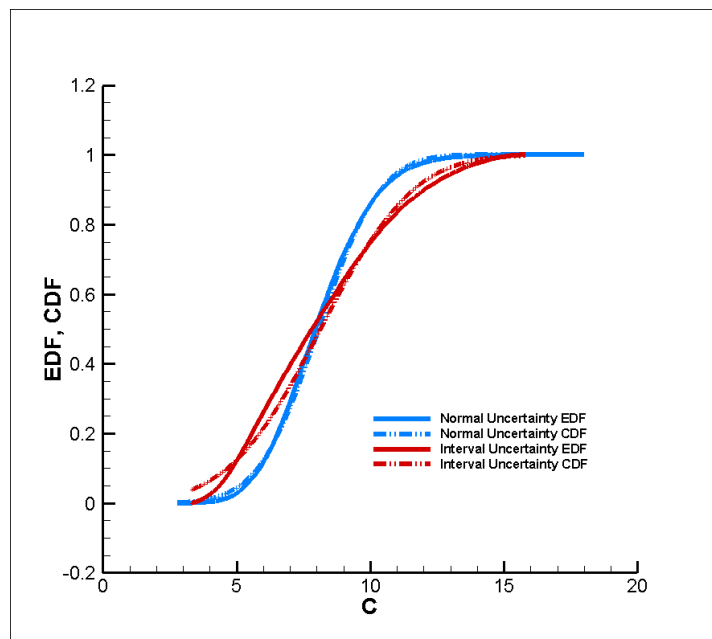


Figure A2. Illustration of the sample problem CDF and EDF.

Acknowledgments

The author would like to acknowledge NASA, NASA Langley Research Center (LaRC), the Vehicle Analysis Branch of NASA LaRC, Dr. Dennis M. Bushnell, Jeffery K. Viken, and Glenn J. Bobskill for their patience and support of the Distinguished Research Program that allowed this work to continue after the author retired as a civil servant from NASA LaRC while this work was in progress.

References

1. The NASA / ARMD Transformative Aeronautics Concepts Program, Transformational Tools & Technologies Project website: <http://www.aeronautics.nasa.gov/tacp/ttt/index.html>, accessed 08/30/2015.
2. Oberkampf, W. L., Helton, J. C. and Sentz, K., "Mathematical Representation of Uncertainty", AIAA Paper 2001-1645, 2001.
3. Green, Lawrence L.: Model Form Uncertainty Knowledge Base User Guide, Version 1.0, NASA TM-2015-XXXXXX, 2015.
4. The Microsoft Access website: <https://products.office.com/en-us/access>, accessed 05/07/2015.
5. Chudoba, Bernd and Huang, Xiao: "Development of a Dedicated Aerospace Vehicle Conceptual Design Knowledge-Based System. AIAA 2006-0225, Presented at the 44th AIAA Aerospace Sciences Meeting and Exhibit, Reno, NV, January 9-12, 2006.
6. Standard for Models and Simulations, NASA-STD-7009, 2008.
7. Green, L. L., Blattmig, S. R., Luckring, J. M. and Tripathi, R.: "An Uncertainty Structure Matrix for Models and Simulations." AIAA 2008-2154, 2008.
8. The Microsoft Excel website: <https://products.office.com/en-us/excel>, accessed 05/07/2015.
9. The SQL wikipedia website: <https://en.wikipedia.org/wiki/SQL>, accessed August 30, 2015.
10. The Visual Basic wikipedia website: https://en.wikipedia.org/wiki/Visual_Basic, accessed August 30, 2015.
11. Kent, William, "A Simple Guide to Five Normal Forms in Relational Database Theory", website: <http://www.bkent.net/Doc/simple5.htm>, 1982, (accessed 05/08/2015).
12. The wikipedia regression analysis webpage: http://en.wikipedia.org/wiki/Regression_analysis (accessed 05/19/2015).
13. Department of Defense, "COMPOSITE MATERIALS HANDBOOK, VOLUME 1. POLYMER MATRIX COMPOSITES GUIDELINES FOR CHARACTERIZATION OF STRUCTURAL MATERIALS", MIL-HDBK-17-1F, 17 JUNE 2002, Superseding MIL-HDBK-17-1E, 23 January 1997.
14. The SAS Data mining from A to Z website: http://www.sas.com/en_us/offers/sem/data-mining-2273479/register.html?gclid=CIKft4nMwMcCFcsXHwodPyOKsO&keyword=data%20mining&matchtype=e, accessed September 1, 2015.
15. The wikipedia data mining website: https://en.wikipedia.org/wiki/Data_mining, accessed September 1, 2015.
16. Snijders, C., Matzat, U., and Reips, U.D. 2012. Big data: big gaps of knowledge in the field of Internet science, International Journal of Internet Science, Vol. 7, pp. 1-5.

17. The TechTarget Big Data Analytics Definition website:
<http://searchbusinessanalytics.techtarget.com/definition/big-data-analytics>, accessed September 1, 2015.
18. The SAS Big Data Analytics website: http://www.sas.com/en_us/insights/analytics/big-data-analytics.html, accessed September 1, 2015.
19. The IBM What is Big Data Analytics? Website: <http://www-01.ibm.com/software/data/infosphere/hadoop/what-is-big-data-analytics.html>, accessed September 1, 2015.
20. The wikipedia Big Data website: https://en.wikipedia.org/wiki/Big_data, accessed September 1, 2015.
21. The Marketing Land 7 Limitations Of Big Data In Marketing website: <http://marketingland.com/7-limitations-big-data-marketing-analytics-117998>
22. The ParAccel Hadoop's Limitations for Big Data Analytics website:
http://www.enterprisemanagement360.com/wp-content/files_mf/1360922634PARACCEL1.pdf,
accessed September 1, 2015.
23. Drag Prediction Workshops IV and V website: <http://arc.aiaa.org/doi/pdf/10.2514/1.C033081>, accessed 05/11/2015.
24. Hemsch, Michael J.: "Statistical Comparison of Computational Results", Presented at the American Institute of Aeronautics and Astronautics (AIAA) Applied Aerodynamics (APA) Technical Committee (TC) CFD Drag Prediction Workshop, Anaheim, CA, June 9-10, 2001, http://aaac.larc.nasa.gov/tsab/cfdlarc/aiaa-dpw/Workshop1/pdf/Statistics_Hemsch.pdf, accessed 1/2/2015.
25. Hemsch, Michael J.: "Statistical Analysis of CFD Solutions", Presented at the 2nd AIAA CFD Drag Prediction Workshop, Orlando, FL, June 21-22, 2003, http://aaac.larc.nasa.gov/tsab/cfdlarc/aiaa-dpw/Workshop2/pdf/82Hemsch_Orlando_v.4.pdf, accessed 1/2/2015.
26. Morrison, J. H. and Hemsch, Michael J.: "Statistical Analysis of CFD Solutions from the 3rd AIAA Drag Prediction Workshop", Presented at the 3rd AIAA CFD Drag Prediction Workshop, San Francisco, CA, June 3-4, 2006, <http://aaac.larc.nasa.gov/tsab/cfdlarc/aiaa-dpw/Workshop3/presentations/S9-Morrison-Statistical-Summary.pdf>, accessed 1/2/2015.
27. Tinoco, Edward N., Levy, David, and Brodersen, Olaf: "DPW-IV Summary of Participants Data", Presented at the 4th AIAA CFD Drag Prediction Workshop, San Antonio, TX, June 20-21, 2009, http://aaac.larc.nasa.gov/tsab/cfdlarc/aiaa-dpw/Workshop4/presentations/DPW4_Presentations_files/D2-8_DPW4Summary_2009-07-20.pdf, accessed 1/2/2015.
28. Tinoco, Edward N., Levy, David, and Brodersen, Olaf: "DPW 5 Summary of Participant Data", Presented at the 5th AIAA CFD Drag Prediction Workshop, New Orleans, LA, 23-24 June 2012, http://aaac.larc.nasa.gov/tsab/cfdlarc/aiaa-dpw/Workshop5/presentations/DPW5_Presentation_Files/14_DPW5%20Summary-Draft_V7.pdf, accessed 1/2/2015.
29. Thermal Performance Database Team, Overview of Intercalibration Results, presented at 4th Ablator Modeling Workshop, Albuquerque, NM, March 211.
30. Rumsey, C. L., Long, M., Stuever, R. A. and Wayman, T. R., "Summary of the First AIAA CFD High Lift Prediction Workshop (invited)", AIAA Paper 2011-0939, 2011.
31. Pandya, M. J., Abdol-Hamid, K. S. and Parlette, E. B., "CFD Computations for a Generic High-Lift Configuration Using TetrUSS", AIAA Paper 2011-3008, 2011.

32. Heeg, J., Chwalowski, P. and Wieseman, C. D., “Lessons Learned in the Selection and Development of Test Cases for the Aeroelastic Prediction Workshop: Rectangular Supercritical Wing”, AIAA Paper 2013-0784, 2013.
33. Malwitz, A., Kim, B., Fleming, G., Lee, J., Balasubramanian, S., Waitz, I., Klima, K., Locke, M., Holsclaw, C., Morales, A., McQueen, E., and Gillette, W., “System for assessing Aviation’s Global Emissions (SAGE) Version 1.5 Validation Assessment, Model Assumptions and Uncertainties”, FAA-EE-2005-03, September 2005.
34. Trevor W. Laughlin, “A Parametric and Physics-Based Approach to Structural Weight Estimation of the Hybrid Wing Body Aircraft”, Master of Science Thesis in the School of Aerospace Engineering, Georgia Institute of Technology, December 2012.
35. The International Federation of Air Line Pilots’ Associations, “Certified versus advisory landing data on Boeing aircraft”, 12ADOBL03, January 2012.
36. Rivers, M. B. and Dittberner, A., “Experimental Investigations of the NASA Common Research Model in the NASA Langley National Transonic Facility and NASA Ames 11-Ft Transonic Wind Tunnel (Invited)”, AIAA Paper 2011-1126, 2011.
37. Morin, B. L., “Broadband Fan Noise Prediction System for Turbofan Engines Volume 3: Validation and Test Cases”, NASA/CR-2010-216898/VOL3, November 2010.
38. Hensch, M. J. and Walker, E. L., “The Crucial Role of Error Correlation for Uncertainty Modeling of CFD-Based Aerodynamics Increments”, AIAA Paper 2011-0173, 2011.
39. Addy, H. E. Jr. and Lee, S., “Icing Encounter Duration Sensitivity Study”, NASA/TM—2011-216367, 2011
40. Bell, J., “Pressure-Sensitive Paint Measurements on the NASA Common Research Model in the NASA 11-ft Transonic Wind Tunnel”, ARC-E-DAA-TN2759, 2011.
41. Nark, D., “Assessment of Radiated Fan Noise Prediction Capabilities using Static Engine Test Data”, NF1676L-11679, 2011.
42. Hanke, J., “Assessment of CFD-based Response Surface Model for Ares I Supersonic Ascent Aerodynamics”, AIAA Paper 2011-3648, 2011.