NASA Technical Memorandum 100609

A Description of an Automated Database Comparison Program

John D. McMinn
John D. Shaughnessy
P. Douglas Arbuckle

May 1988

SUMMARY

An interactive FORTRAN computer comparison program designed to automatically locate regions of incongruity between two databases, in order to reduce the amount of required analysis, is described. The software, guided by user input parameters, incrementally compares the databases and generates plots of those regions in the databases which do not compare within a specified tolerance. Additionally, to help the user further reduce the number of plots, tools are provided within the software which enable the user to statistically reduce the number of data points in the databases compared. To facilitate the description of these tools, the procedures used to compare two aerodynamic databases for an F-18A fighter aircraft are detailed.

INTRODUCTION

In scientific work, the situation often arises where it is necessary to compare data from different sources. This process can range from being relatively simple, as is the case when comparing two functions of the same independent variable, to a process involving the comparison of databases with dissimilar data points and a number of independent variables which may or may not be identical. A typical method of comparing two databases representative of the latter case is the tedious and time consuming process of plotting certain data from each and making comparisons manually.

A computer program could greatly reduce the magnitude of this effort. This software could conduct the comparison and show the user where large discrepancies exist between the data, allowing the user to look only at those regions in the databases which do not compare within a specified tolerance.

For example, a situation where such software would have been helpful arose during the comparison of the aerodynamic data from a real-time simulation (mostly wind tunnel data) to data generated by the USAF Digital Datcom (ref. 1). After a limited search of the NASA Langley computer software library, no software was found that performed the desired task; therefore, it was decided to develop such a computer program.

The purpose of this report is to describe the computer program that was developed to automate the task of comparing two databases. The capabilities of the software are described and an example of the procedures used to make a comparison is given. A comparison of two aerodynamic databases for an F-18A aircraft is made where one database was generated by the USAF Digital Datcom computer program (ref. 1) and the other was obtained from wind tunnel tests. Also included in this report are examples of the types of data generated.

SYMBOLS AND ABBREVIATIONS

ADC   Automated Database Comparison
CL   lift coefficient
CD   drag coefficient
CM   pitching moment coefficient
CY   sideforce coefficient

C1	rolling moment coefficient
CN	yawing moment coefficient
CLA	lift coefficient due to vertical acceleration
CMA	pitching moment coefficient due to vertical acceleration
CNR	yawing moment coefficient due to positive yaw rate
C1R	rolling moment coefficient due to positive yaw rate
CNP	yawing moment coefficent due to positive roll rate
C1P	rolling moment coefficient due to positive roll rate
CMQ	pitching moment coefficient due to positive pitch rate
CLQ	lift coefficient due to positive pitch rate
NR-	yawing moment coefficient due to negative yaw rate
1R-	rolling moment coefficient due to negative yaw rate
NP-	yawing moment coefficent due to negative roll rate
1P-	rolling moment coefficient due to negative roll rate
MQ-	pitching moment coefficient due to negative pitch rate
LQ-	lift coefficient due to negative pitch rate
RTS	real-time simulation

THE COMPARISON PROGRAM

The fundamental objective of the software is to automate the comparison of two databases to free the user from analyzing the large number of plots that could be generated during a manual comparison (for example, in a comparison of 10 terms such as: CMQ, CLQ, CNP, CL, CD, CM, CLA, CY, C1, and CMA generated from two databases where each of five independent variables such as: angle of attack, angle of sideslip, Mach number, leading and trailing edge flap angles range over 10 values; 100,000 plots could be generated). Therefore, the software should be able to compare two tabular, numerical databases and indicate where the databases do not compare within a certain specified tolerance. By writing appropriate data to an output file for later plotting, the number of plots to be viewed is greatly reduced. Additionally, the program should be able to examine databases which do not necessarily contain like terms by bringing the databases to a common reference point through total coefficient reconstruction equations like those discussed in reference 2 (i.e., summing items in an aerodynamic database to form total force or moment coefficients and perform the comparison on that basis).

With those guidelines in mind, the Automated Database Comparison (ADC) program was developed. This FORTRAN program is interactive and runs on systems such as Control Data Corporation's Network Operating System. The inputs for ADC are: (1) parameters, from a file or the keyboard; (2) names of the independent variables; (3) the independent values; and (4) acceptable comparison tolerance for each term.

Once the input parameters have been provided, the ADC begins searching the databases for dissimilarities that exceed the specified tolerance. These dissimilarities and tolerances are measured in the form of absolute errors between the data. Originally, the software was designed to compare databases on the basis of percentage error (dividing the absolute error by the largest value from one of the databases; called the reference database), but the resulting information proved to be misleading. In cases where the magnitude of the largest value was small and the curves were fairly similar, the software would indicate errors on the order of 100 percent. For that reason, the decision was made to compare on the basis of absolute error and display on the output

2

file the largest value from the reference database. Absolute error comparisons can also be misleading; however, this was determined to be the preferred means of comparison.

The ultimate result from using ADC is plots of the data which do not compare within the specified tolerance. Two tools within the software that help isolate these regions in the databases are the error summary file and the file splitting function which breaks large data files into smaller data files for plotting.

The error summary file shown in figure 1 displays on the first line the names of the databases being compared and identification information from the title line of the second database. The first 50 characters of the second line are reserved for displaying any information supplied by the user in response to a prompt for supplementary information. The other piece of information on the second line is the suffix for the data files that correspond to the information in this file (the data file names are formed by the concatenation of a term name and a suffix which is supplied by the user in response to a prompt).

Below the supplementary information is the section of the file labeled "TABLE OF SCHEDULES." Here all the independent variables are shown along with their first 10 values. To the left of the names of the independent variables are two columns of numbers. The first column is the ordinal number associated with the independent variable (i.e., in fig. 1, rudder is the seventh independent variable). The second column of numbers under the heading "CONTROL LOOP" show the program iteration loop in which the value of each variable is changed. If two or more independent variables have the same numbers in the "CONTROL LOOP" column, then they are being parametrically varied at the same time (this option allows the user to eliminate combinations of independent variables that are not desired, which further reduces the amount of data generated).

At the bottom of the error summary file is the Error Table. The left-most column of the table shows the terms being compared and the next column shows the maximum value encountered for a term from the first database (given as a reference from which to gauge the relative magnitude of the absolute errors). In the remaining columns of the table, absolute errors of similar magnitude are grouped into ranges (which can be altered by the user). For example, in figure 1 it can be seen that for drag coefficient, an absolute error of a magnitude between 0.05 and 0.1 units of CD occurred 18 times. By referring to this table the engineer can set the tolerance for the terms either to better partition the error groups or to limit the size of the resulting data files; the higher the tolerance, the lower the occurrence of discrepancies and therefore, less information is written to the data files.

The second tool in ADC is the file splitting function that reduces a data file into one of specific error magnitudes. This option allows the user to make plots of only the data which he considers to be in error. For example, if the user is working with the data in figure 1 and desires to know the conditions causing the instances of sideforce error between 0.01 and 0.02 units of sideforce, then this function can be used to create a file containing only the data for the desired plots. The alternative is to plot all the data and locate the desired 47 plots

from the 144 that result; obviously this function is most useful when there are a large number of plots.

Despite the flexibility evidenced by the aforementioned features, ADC does have some limitations. The databases must be in a specific tabular function-based format (ref. 2). A maximum of 250 errors between the databases for a particular term can be plotted. The variable on the horizontal axis of all plots is the first independent variable given in the input parameter list. The plots can display the values of up to 11 independent variables (one across the horizontal axis and 10 in a legend on the plot). A limited amount of user reconfiguration of the ADC source code is required to allow for the diverse variety of databases the software is capable of comparing.

A SAMPLE APPLICATION

In order to further illustrate the capabilities of the ADC, the following example is given. In this sample comparison the two databases to be compared contain aerodynamic data for the F-18A aircraft and were derived from the USAF Digital Datcom computer program (ref. 1) and wind tunnel data, respectively. Additionally, the databases have a large number of different independent variables and do not contain the same terms; therefore, direct comparison is not possible. Given this situation and the fact that the software allows for each independent variable to iterate through a range of 50 values during the comparison, it is possible that on the order of 10^{50} plots could be generated. Using more realistic numbers (see fig. 1), on the order of 100 plots would be required. Therefore, the following procedures are used to compare the databases.

First, the input parameters are specified. In order to make a thorough comparison of the databases, 11 independent variables are used to represent the major control surfaces and flight conditions of the F-18A. Those variables are: angle of attack, angle of sideslip, Mach number, left and right horizontal tail deflections, left and right aileron deflections, rudder deflection, leading and trailing edge flap deflections, and altitude. The values of these independent variables are assigned to completely cover the ranges of the databases but limited to a reasonable number whenever possible without compromising the quality of the comparison (e.g., using five angle-of-attack values as opposed to 50). To complete the input parameter list, the tolerances for the terms are entered and initially set to large values (e.g., the CL tolerance is set to 10 units of CL) to quickly generate the error summary file.

The second step is to use the error summary tool to perform a sensitivity analysis with respect to independent variables that, as determined by engineering judgment, might not have much influence on the error distribution within the databases. Mach number is the variable in this comparison that is tested. This is done by allowing all of the independent variables to vary, with the exception of Mach number. The output from the error summary file is then recorded and compared to the error summaries from subsequent runs at different Mach numbers. By analyzing the distribution pattern of the errors and realizing that they are nearly the same, it is shown that Mach number has little effect on the error distribution. With that confirmed, it is possible to select

one Mach number that is fairly representative of the rest, thus reducing the number of comparison plots to be viewed.

The third step in the strategy is to continue use of the error summary file to discover and take advantage of any symmetries in the databases that might exist. Due to the symmetric properties of the aircraft, these databases should have symmetry. By verifying these symmetries it is possible to remove those values that cause a duplication of errors already encountered from the schedules of the independent variables, e.g., eliminating -5° sideslip since it can be shown that it produces the same errors as +5° sideslip. The potential for duplication exists not only with sideslip but also with rudder, ailerons, and differential horizontal tails. Due to the number of independent variables that this stage applies to, there is a great potential for significantly reducing the amount of data to be compared.

Once the schedules of the independent variables have been selected, the accumulation of results begins. ADC is run multiple times, each time with some independent variables allowed to range over their list of input values while the other independent variables are not varied. Engineering judgment is used to decide which independent variables will be allowed to vary for a given run. The resulting error summary files, such as the one shown in figure 1, are then analyzed to determine the values for the independent variables which cause the databases not to compare well. The tolerances for each term are then lowered sufficiently to have the software create data files for those areas when ADC is rerun.

Next, the file splitting tool is applied. This is useful to quickly find the conditions causing a particular group of errors. For example, the error table in figure 1 shows 32 occurrences of CM errors with magnitudes from 0.1 to 0.2. The table also shows the largest magnitude of CM from the first database to be 0.194 units of CM. If it were desired to view only those plots exhibiting data in the error range of the same magnitude as the largest value of CM, the file splitting function could be used. The result is a data file of only those conditions causing CM errors between 0.1 and 0.2.

Finally, the data files are turned into plots similar to those shown in figure 2. On each plot, the variable across the horizontal axis is the first independent variable from the input parameter list. The solid curve represents data from the first database listed in the key on the upper right. The dashed curve represents data from the second database listed. The items listed below the key are the remaining independent variables with the values used to generate the curves shown. The bottom most item in the key, ERROR, is not an independent variable, but rather the maximum absolute error between data points on the plot.

CONCLUDING REMARKS

The software described herein (ADC) performs an automated database comparison to reduce the number of plots that would have to be analyzed in a manual comparison. ADC can automatically compare terms generated from two databases by varying independent variables over a range of values. To reduce the number of plots, the software contains two special tools which isolate the dissimilar areas within the databases.

The first is an error summary file which is in tabular form and provides a concise, comprehensive review of the distribution of the absolute errors between the two database values for each term of the comparison. The other tool provides a capability of plotting specific error ranges in the data files. To facilitate the description of these tools, the procedures used to compare two aerodynamic databases for an F-18A fighter aircraft are detailed.

REFERENCES

1. Vukelich, S. R.; and Williams, J. E.: The USAF Stability and Control Digital Datcom. AFFDL-TR-79-3032. April 1979.

2. Arbuckle, P. Douglas; Buttrill, Carey S.; and Zeiler, Thomas: A New Simulation Model Building Process for Use in Dynamic Systems Integration Research. AIAA Paper No. 87-2498-CP. August 17-19, 1987.

RTS /DATCOM F18 AERO DATA MAR-87 (ALIH -10,0,10)   ** SUFFIX FOR TERMS IS run1

supplementary information appears here

TABLE OF SCHEDULES

CONTROL NUM	LOOP	IND. VARIABLE	VALUES
1	1	ALPHA	0.00 5.00 10.00 15.00
2	2	BETA	0.00 5.00
3	3	MACH	0.80
4	4	HTLEFT	0.00 -10.00 -10.00
5	5	HTRGHT	0.00 10.00 10.00
6	6	TEFLAP	0.00 10.00
7	7	RUDDER	0.00 10.00
8	8	AILERR	0.00 -10.00 -10.00
9	9	AILERL	0.00
10	10	LEFLAP	0.00
11	3	ALT'S	20000.00

ERROR TABLE

TERMS	MAX	.0	0.0001	0.001	0.005	0.01	0.02	0.05	0.1	0.2	0.4	0.8	1.6	3.2	6.4
CL	1.31210														
CD	0.38991														
CM	0.19437														
CY	0.15240	2			4	27	47	44	18	144					
Cl	0.04058	2		2	22	45	73	64	68	126					
CN	0.02603	2		6	70	56	10			32					
CLA	3.62000											72	72		
CMA	2.28000													144	
CLQ	0.23842							144	144						
CMQ	0.23025						72	72	144						
CLP	0.10875														
CLR	0.48000								144						
CNP	5.79003								144						
CNR	5.16011												72	72	
LQ-	0.23842							144	144						
MQ-	0.23025						72	72	72						
1P-	0.10875														
1R-	0.48000								144						
NP-	5.79000								144						
NR-	5.16011												72	72	144

THE NUMBERS IN EACH COLUMN REFLECT THE NUMBER OF TIMES THE ERROR FOR A TERM FELL BETWEEN THE ERROR SHOWN AT THE TOP OF THE COLUMN TO THE LEFT AND THE ERROR AT THE TOP OF THE COLUMN TO THE RIGHT. IN THE CASE OF THE LAST COLUMN, THE NUMBERS REPRESENT ERRORS GREATER THAN 6.40

Figure 1.- Error summary file for the comparison of a database generated by the USAF digital datcom versus the Langley/McDonnell Douglas aerodynamic database for the F-18A.
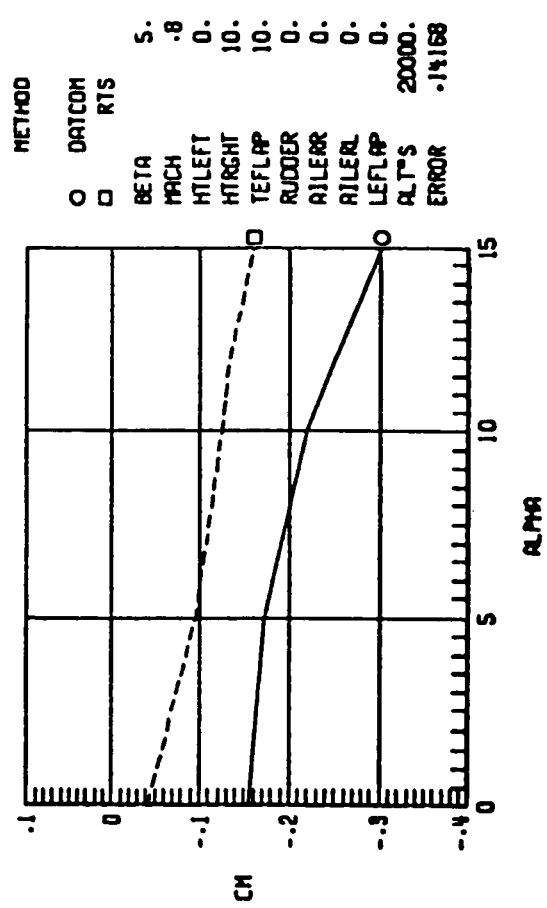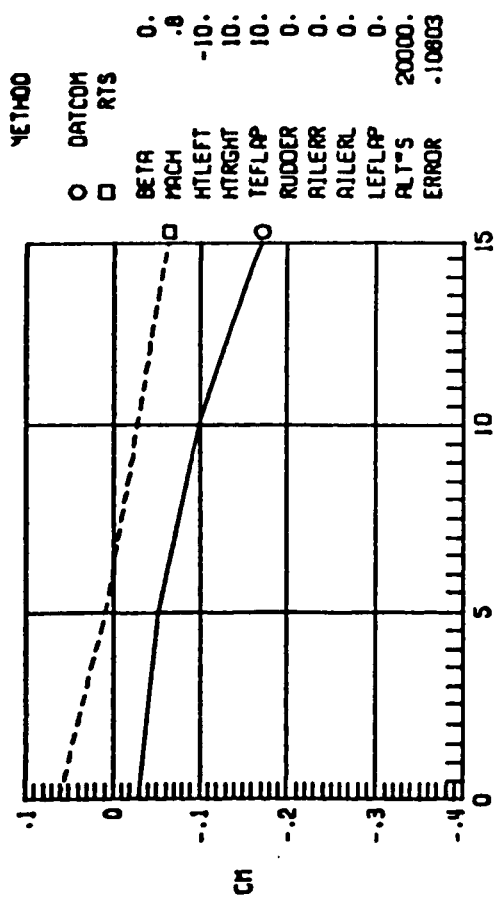
Figure 2.- Sample plots of F-18A aerodynamic data that were generated by ADC because they exhibit differences greater than the allowable tolerance for CM.

1. Report No. NASA TM-100609	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle A Description of an Automated Database Comparison Program		5. Report Date May 1988
		6. Performing Organization Code
7. Author(s) John D. McMinn, John D. Shaughnessy, and P. Douglas Arbuckle		8. Performing Organization Report No.
9. Performing Organization Name and Address NASA Langley Research Center Hampton, VA 23665-5225		10. Work Unit No. 505-66-01-02
		11. Contract or Grant No.
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, DC 20546-0001		13. Type of Report and Period Covered Technical Memorandum
		14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

An interactive FORTRAN computer comparison program designed to automatically locate regions of incongruity between two databases is described. The software, guided by user input parameters, incrementally compares the databases and generates plots of those regions in the databases which do not compare within a specified tolerance. Additionally, tools are provided within the software which enable the user to statistically reduce the number of data points in the databases compared. To facilitate the description of these tools, the procedures used to compare two aerodynamic databases for an F-18A fighter aircraft are detailed.

17. Key Words (Suggested by Author(s)) interactive database comparison FORTRAN program	18. Distribution Statement Unclassified – Unlimited Subject Category – 61, 05		
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 9	22. Price A02