

ARMY RESEARCH LABORATORY



Statistical Analysis of Surface Pressure Measurements vs. Computational Predictions

by Malcolm S. Taylor
and Walter B. Sturek

ARL-TR-1318

March 1997

Approved for public release; distribution is unlimited.

19970326 026

DUO QUALITY INSPECTED 1

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

Army Research Laboratory

Aberdeen Proving Ground, MD 21005-5067

ARL-TR-1318

March 1997

Statistical Analysis of Surface Pressure Measurements vs. Computational Predictions

Malcolm S. Taylor, Walter B. Sturek
Information Science and Technology Directorate, ARL

Abstract

This report details a statistical approach to the comparison of results from an international study in which Navier-Stokes computational techniques were applied to a complex flow field with highly separated flow about a fixed geometry. The traditional method for evaluating flow-field computations is for the engineer to visually inspect the results, make qualitative comparisons, and draw conclusions—a process that can be effective for a limited number of data sets. In this study, the number of data sets was so large as to make the evaluation and comparison of the results a formidable task. In response, a statistical approach with emphasis on data visualization was developed. The procedure extends directly to higher dimensions, and is of particular value when a large number of data sets are to be compared and an impartial quantitative assessment is sought.

ACKNOWLEDGMENTS

The authors wish to acknowledge the contributions of Duane Frist, formerly at the U.S. Army Research Laboratory, and Hugh Thornburg, formerly at the Mississippi State University National Science Foundation Engineering Center, for computational results and graphics, which appear in this report as Figures 2 and 6, respectively.

INTENTIONALLY LEFT BLANK.

TABLE OF CONTENTS

	<u>Page</u>
ACKNOWLEDGMENTS	iii
LIST OF FIGURES	vii
LIST OF TABLES	vii
1. INTRODUCTION	1
2. STATISTICAL APPROACH	3
2.1 Data Structure	3
2.2 Error Analysis	3
2.3 Summary	5
3. APPLICATION TO TTCP DATA	6
4. CONCLUSION	8
5. REFERENCES	9
DISTRIBUTION LIST	11
REPORT DOCUMENTATION PAGE	15

INTENTIONALLY LEFT BLANK.

LIST OF FIGURES

<u>Figure</u>		<u>Page</u>
1.	Schlieren photograph of model in wind tunnel.	1
2.	Surface pressure C_p vs. circumferential position ϕ	2
3.	Contour surface of error for data set $k=1$	4
4.	Box and whisker plot of errors; $k=1, 2, \dots, 9$	5
5.	Scatterplot of mean and standard deviation for error sets $k=1, 2, \dots, 9$	6
6.	Surface pressure C_p vs. circumferential position ϕ	7
7.	Box and whisker plot of errors for BB, BL, ke, and kw turbulence models.	7
8.	Scatterplot of errors for BB, BL, ke, and kw data sets.	8

LIST OF TABLES

<u>Table</u>		<u>Page</u>
1.	Distance Measurements for (\bar{x}, s)	8

INTENTIONALLY LEFT BLANK.

1. INTRODUCTION

This report discusses the development and application of a statistical analysis approach to the evaluation of some results of a collaborative study carried out under the auspices of The Technical Cooperation Program (TTCP) with participants from Canada, the United Kingdom, and the United States. The purpose of the study was to apply Navier-Stokes computational techniques to a complex flow field with highly separated flow about a fixed missile geometry to evaluate the predictive technology.

The computational study encompassed 8 computer codes, 9 turbulent viscous models, several laminar viscous results, as many as 15 grid resolutions, and 6 test cases. This produced more than 90 test-case results for evaluation with respect to each other and to effects of grid resolution, choice of turbulence model, and computational technique.

A Schlieren photograph showing the test model mounted in the wind tunnel for a Mach 3.5 flow condition and 8-deg angle of incidence is shown in Figure 1. The bow shock and lee-side separated vortical flow are clearly visible. In Figure 2, experimental measurements of surface pressure taken at eight axial locations, x/d , are plotted along with the computed values from laminar and turbulent viscous models. The computed results show generally close agreement with the experiment on the windward side ($0 \leq \phi \leq 90$ deg) and disagreement on the leeward side ($90 \leq \phi \leq 180$ deg). The lee-side separated flow is the critical region for evaluating the predictive capability of the computational procedures.

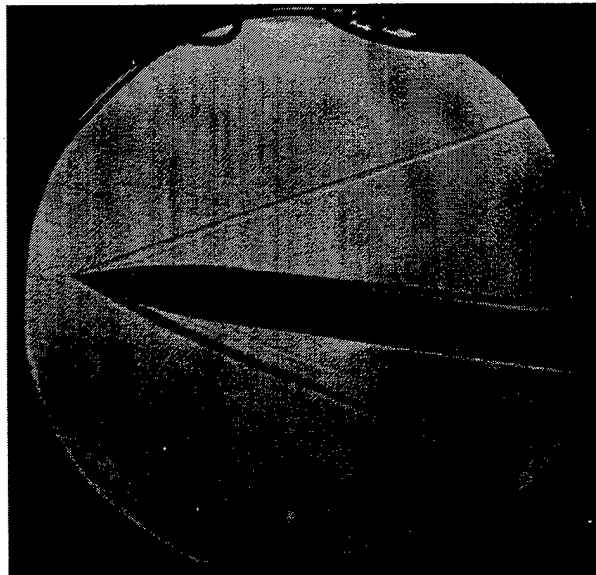


FIGURE 1. *Schlieren photograph of model mounted in wind tunnel; Mach 3.5, angle of incidence 8 deg.*

The traditional method for evaluating flow-field computations is for the engineer to visually inspect the results, make qualitative comparisons, and draw conclusions—a process that can be quite satisfactory for a limited number of data sets. However, for this study, the number of data sets was so large as to make the evaluation and comparison of the results a formidable task, and it was decided to explore the contribution that statistical data analysis techniques might be able to make in the evaluation process.

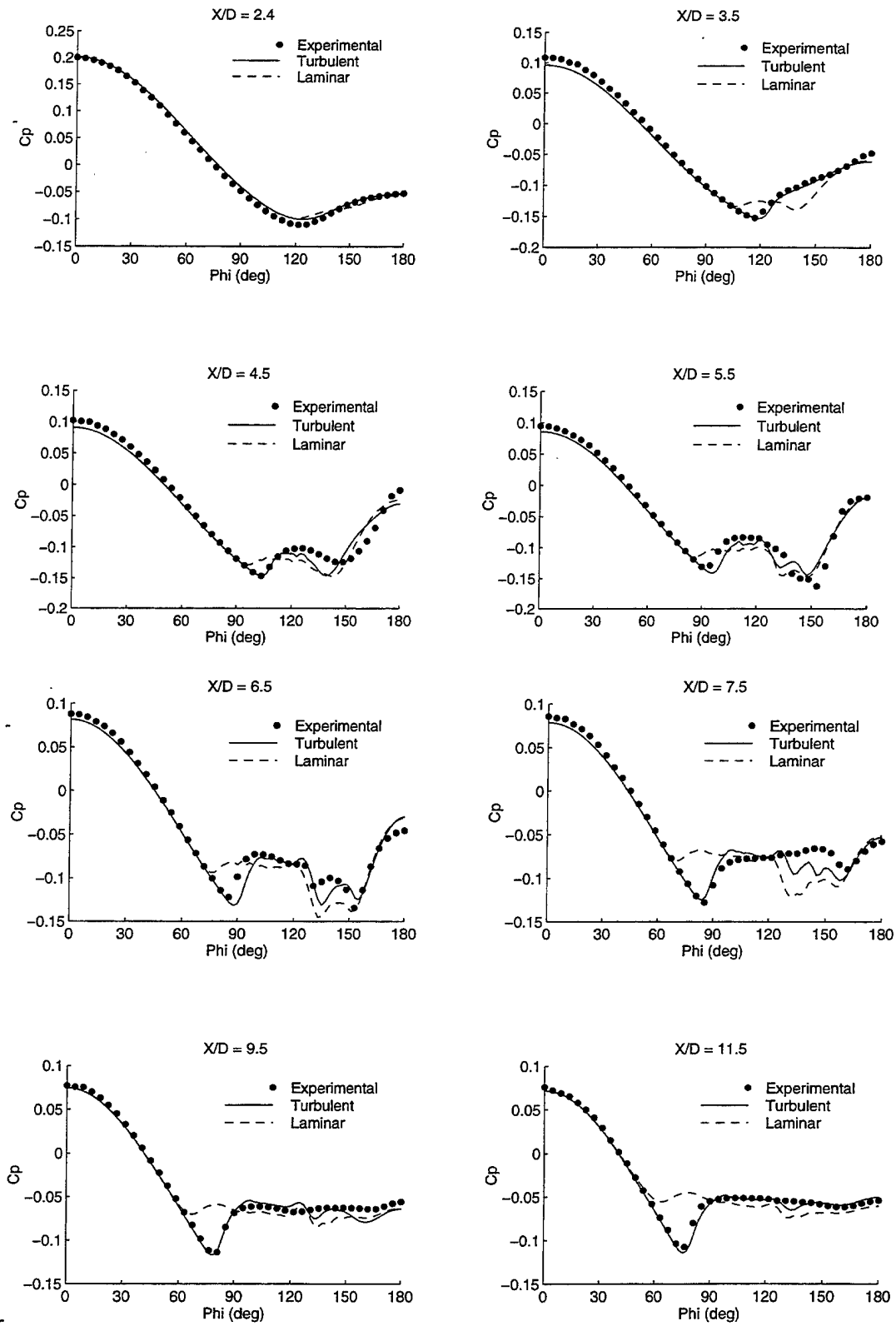


FIGURE 2. Surface pressure C_p vs. circumferential position ϕ ; Mach 2.5, angle of incidence 14 deg. Comparison between computation and experiment.

2. STATISTICAL APPROACH

The wide availability of powerful and affordable computing resources has impacted the way in which data analysis is conducted at a fundamental level, and expressions such as “exploratory data analysis” and “data visualization” have made their way into the statistical lexicon. Visualization of quantitative data is an active area of research, with professional journals and books devoted exclusively to the topic (Tufté 1983, 1990; Cleveland 1993; DuToit, Steyn, and Stumpf 1986). The basic idea is that data, if properly collected and effectively portrayed, can often speak directly to the subject-matter expert without resorting to a plethora of statistical jargon and assumptions. This is the approach undertaken in this section; a subset of TTCP test-case results is used for illustrative purposes.

2.1 Data Structure

A set of experimental data was chosen to serve as a baseline against which all the flowfield predictors would be compared. The experimental data pertinent to this investigation consist of measurements of surface pressure taken at 41 circumferential positions, ϕ , about each of 9 axial locations, x/d , for a test model such as that shown in Figure 1.

The technique involves the comparison of two data sets—the experimental (or baseline) data and a set of computed results. Both data sets are three-dimensional, consisting of the triples: (axial location, circumferential position, surface pressure). To facilitate discussion, the experimental data are denoted as

$$Y_{ij} \quad i = 3.5(1)11.5, \quad j = 0(4.5)180, \quad (1)$$

where the subscript i indexes axial location, x/d ; the subscript j indexes circumferential position, ϕ ; and Y_{ij} is the measured surface pressure. The computed value is denoted as

$$X_{ij}^k \quad k = 1(1)9, \quad (2)$$

where subscripts i and j are the same as for the experimental data; the superscript k indexes the data sets chosen for analysis.

2.2 Error Analysis

If we define the difference between a computed value of surface pressure and the corresponding experimental value as error, then

$$E_{ij}^k = X_{ij}^k - Y_{ij} \quad (3)$$

is the error at location (i, j) for data set k . A three-dimensional contour plot provides an informative display of error over the entire axial-circumferential grid; such a plot for the data set corresponding to $k = 1$ is shown in Figure 3.

The approximately planar surface in Figure 3 represents the windward region where experimental and computed values are in reasonably good agreement. The projection of the surface onto the horizontal plane further aids in the definition of this region; for data set $k = 1$, the substantial irregularities occur on the leeward side, corresponding to values of ϕ greater than 90 deg.

The highly irregular contour surface in Figure 3 provides convincing evidence of the need for further data reduction before the simultaneous comparison of several flowfield calculations can be

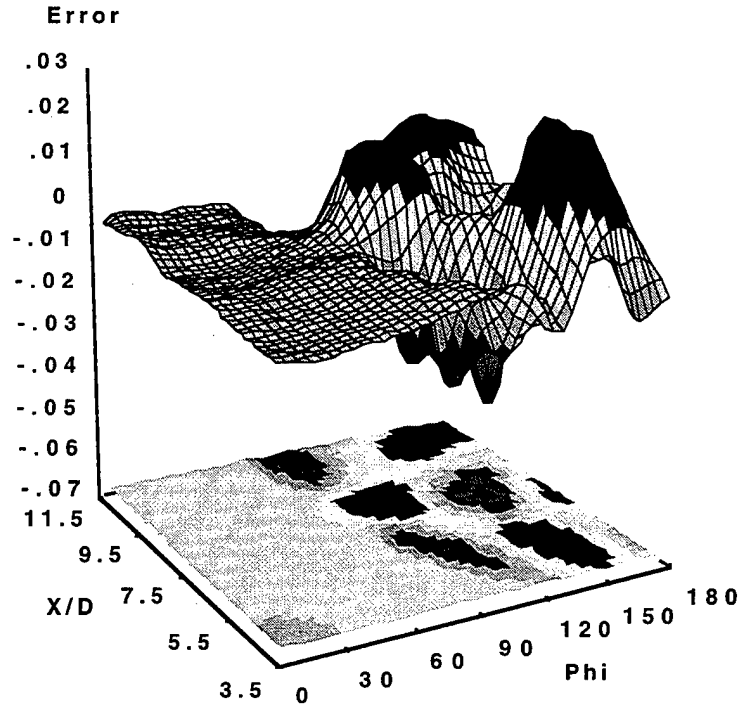


FIGURE 3. Contour surface of error for data set $k=1$.

effectively undertaken. A further data summary using box and whisker plots is shown in Figure 4. To construct a box and whisker plot, the errors over the entire grid are combined and ranked in value from smallest to largest. The median, or middle, value is visually represented by the dot (\bullet) inside the boxes in Figure 4; the top and bottom of the box mark the 75th and 25th percentiles, respectively, of the ranked errors; the whiskers are vertical lines that terminate at the extreme values of the data set. The width of the box holds no significance. These five-number data summaries provide a compact description of how the errors are distributed over the entire grid, and, as can be seen, facilitate comparison across several data sets.

The perfect case, where computed and experimental values coincide at every location (i,j) , would cause the box and whisker plot to collapse to a single value—zero. Failing that, a thin box with short whiskers indicates good overall agreement. A cursory glance at Figure 4 suggests that data sets 3 and 4 may be closest to this ideal, while data set 8 appears in some difficulty. Since the whiskers terminate at the maximum and minimum error values, a spurious, or outlying, observation could distort the whisker length and mislead the viewer. The data sets in Figure 4 were inspected and found to be free from outliers—a precautionary procedure that should be routinely performed.

The statistical graphics in Figures 3–4, while enlightening, suggestive, and highly appropriate for an initial screening, are not entirely adequate for an impartial assessment of which error sets represent “closest” agreement between experiment and model. Clearly, we would like the errors to be tightly clustered about zero. This would be reflected in a location parameter (a mean or median) and a dispersion parameter (a standard deviation or interquartile range) both close to zero. This suggests that an attempt to formally rank the effectiveness of the computation procedures as revealed through the error sets $\{E_{ij}^k\}$, $k = 1, 2, \dots, 9$, should involve a statistic that includes a measure of both location and dispersion. And this is exactly what was done.

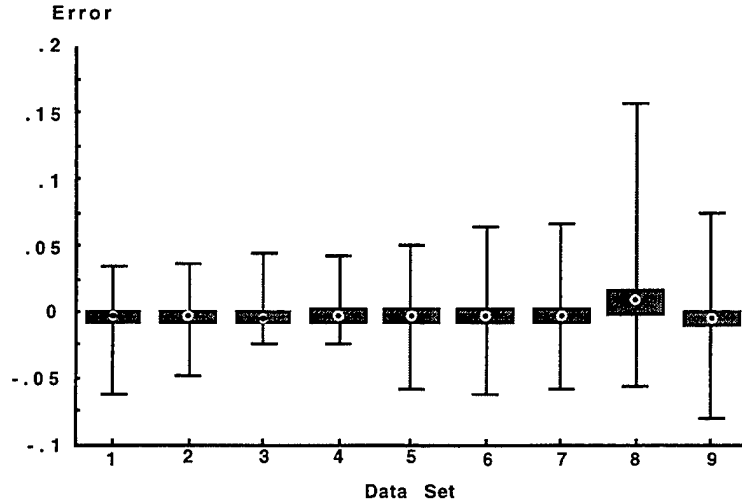


FIGURE 4. Box and whisker plot of errors; $k=1, 2, \dots, 9$.

We chose as a statistic the distance between the point determined by the sample mean and standard deviation (\bar{x}_k, s_k) and the origin. Recall that the origin $(0,0)$ corresponds to perfect agreement between calculation and experiment. The distances

$$\sqrt{\bar{x}_k^2 + s_k^2} \quad k = 1, 2, \dots, 9 \quad (4)$$

were determined, and the ranking induced under this procedure in terms of the index k is as follows:

$$4 < 3 < 2 < 7 < 5 < 6 < 1 < 9 < 8. \quad (5)$$

Expression (5) asserts that the set of computed data corresponding to $k=4$ is in closest agreement with the experimental measurements; data set $k=3$ is second closest, ..., data set $k=8$ is furthest away.

In Figure 5, the sample mean and standard deviation (\bar{x}_k, s_k) for each of the 9 data sets are plotted by case number k . More complicated ranking procedures are, of course, possible. An obvious modification would be to form a weighted combination of mean and standard deviation, but to do so without a compelling reason serves no practical purpose, and was not undertaken for these data.

2.3 Summary

The results of the direct ranking procedure are altogether consistent with the preliminary graphics. This is reassuring. A surprise at this stage usually means that an error in logic or calculation is still at large. As mentioned earlier, and as seen in Figure 2, the flowfield predictors seem to experience the most difficulty for larger circumferential values, so focusing on that subregion rather than the entire flowfield may be more appropriate.

For the statistician, failure to account for error in the experimental measurements Y_{ij} is troublesome. The statistician's model for experimental data is usually of the form: *observed* $Y_{ij} = \text{actual } Y_{ij} + \text{error } e_{ij}$. Engineering experience leaves the fluid dynamicist convinced that concern

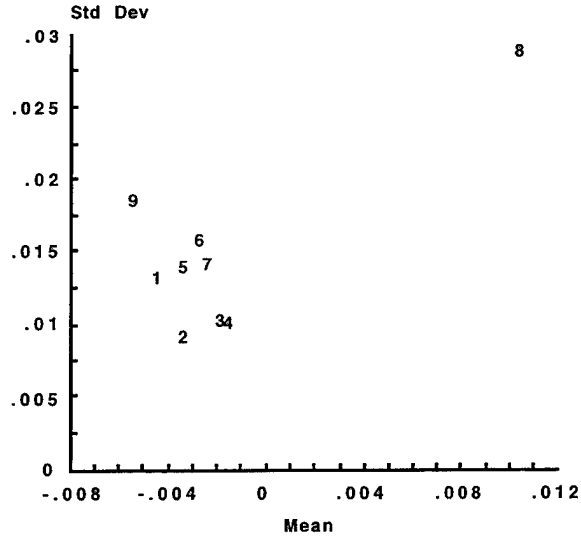


FIGURE 5. Scatterplot of mean and standard deviation for error sets $k=1, 2, \dots, 9$.

over this point is unwarranted. Moreover, in this instance, only a single set of experimental data is available and all the flowfield predictors are deterministic, so the error, whatever its magnitude, remains forever confounded with the recorded observations.

3. APPLICATION TO TTCP DATA

The approach developed in section 2 was applied to output from four turbulence models: the Baldwin-Lomax(BL), Baldwin-Barth(BB), k-omega(kw), and k-epsilon(ke), all using the NPARC computational code (Cooper and Sirbaugh 1989). The theory and implementation of these models is well documented and is apart from our considerations here.

The difficulty of visual comparison between experiment and computation over the four models is convincingly illustrated in Figure 6. Only four axial stations are shown, but, even from that small number, the difficulty facing the engineer attempting to rank the procedures is readily apparent.

The statistical approach was applied to the entire data sets: 41 circumferential locations at each of 14 axial stations for the 4 turbulence models. The box and whisker plot visualization is shown in Figure 7. The BL result has the shortest whiskers, indicating the smallest excursion from the experimental data. The mean error for all the data sets is close to zero, with little apparent advantage for one turbulence model over another. The results are shown in a scatterplot format in Figure 8. Again, the BL model seems to have the advantage; however, as shown in Table 1, the level of disagreement between the four turbulence models is very small.

The formal ranking is $BL < kw < ke < BB$, but, importantly, the small numerical differences listed in Table 1 hold substantial practical significance for the engineer. The four turbulence models vary widely in complexity and computer resource requirements. The most computer efficient model is the algebraic model BL, followed by the BB model. The two-equation models, kw and ke, are more complex and require significantly greater computer resources than either the BB or BL model.

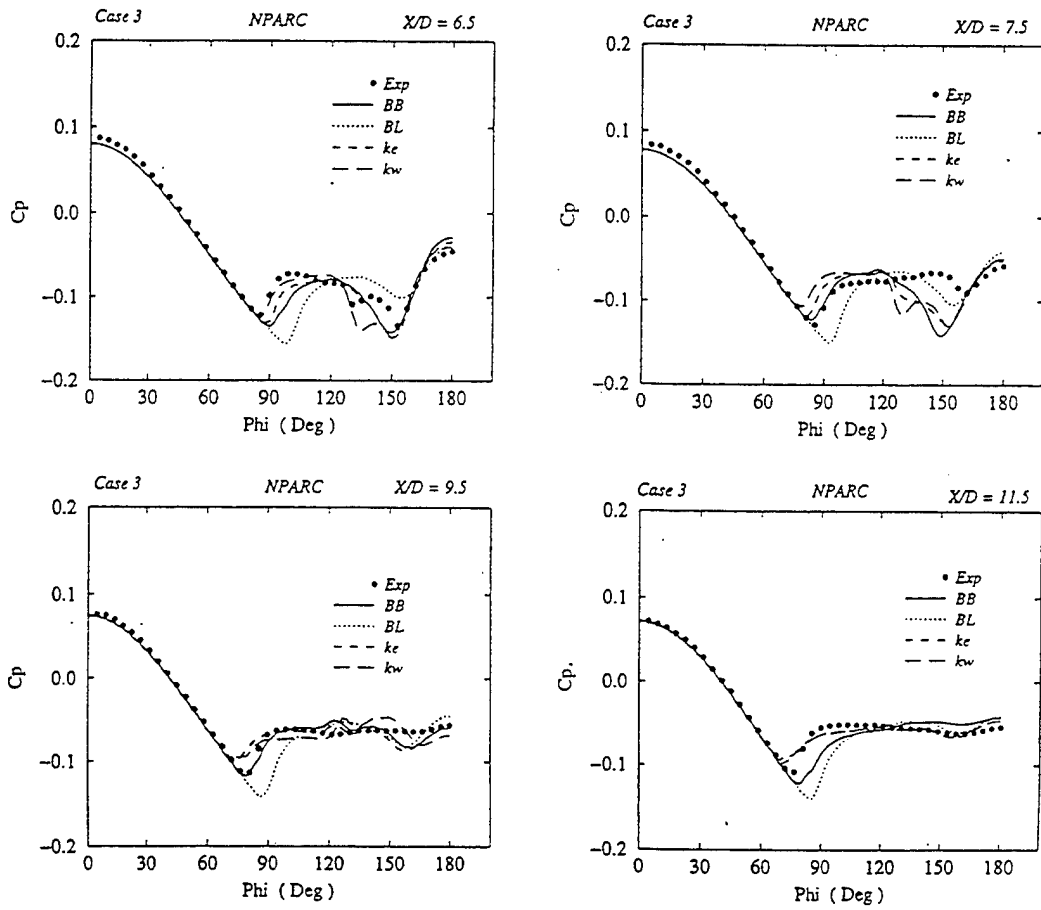


FIGURE 6. Surface pressure C_p vs. circumferential position ϕ ; Mach 2.5, angle of incidence 14-deg. Comparison between computation and experiment for both laminar and turbulent viscous modeling.

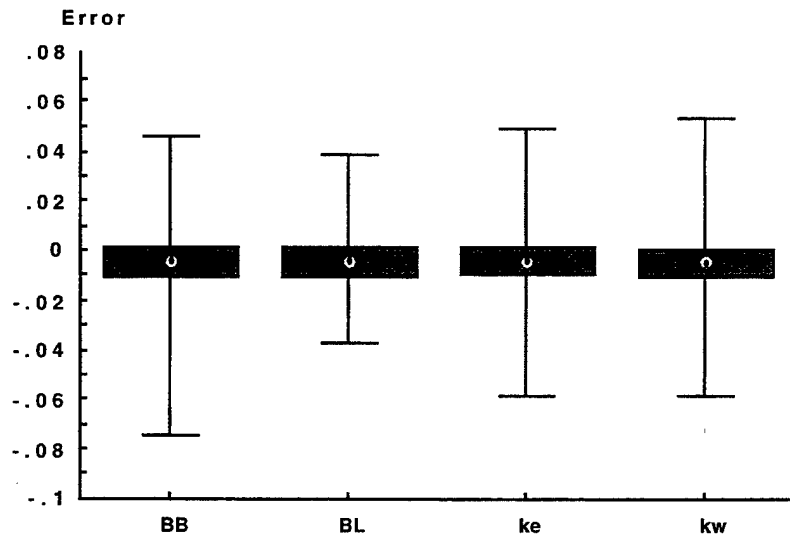


FIGURE 7. Box and whisker plot of errors for BB, BL, ke, and kw turbulence models; Mach 3.5, angle of incidence 14 deg.

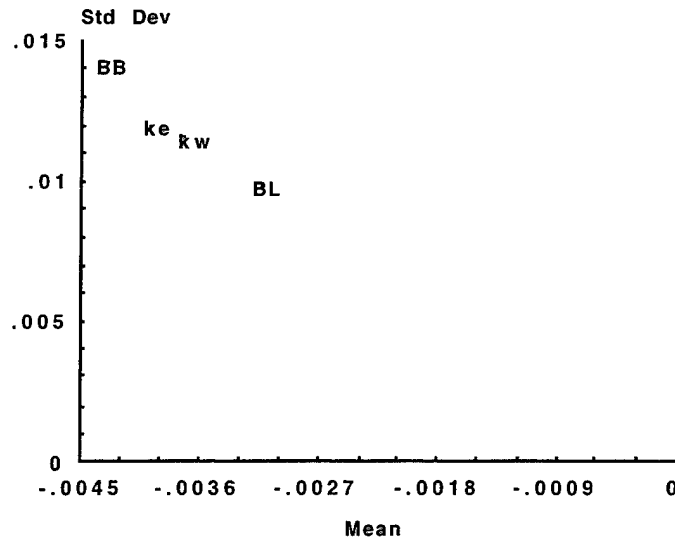


FIGURE 8. Scatterplot of errors for *BB*, *BL*, *ke*, and *kw* data sets; Mach 3.5, angle of incidence 14 deg.

TABLE 1. Distance Measurements for (\bar{x}, s)

case:	BB	BL	ke	kw
distance:	0.01257	0.01125	0.01149	0.01131

4. CONCLUSION

A statistical approach, with emphasis on data visualization, has been developed to assist the engineer in the comparison of computer models intended to predict surface pressure about a physical model in a complex flow field. The approach was applied to the evaluation of computational results obtained from four distinct turbulence models and showed that the use of more sophisticated computer-intensive turbulence models did not provide correspondingly superior results. The statistical analysis procedure extends directly to higher dimensions and should be particularly valuable when a large number of data sets are to be compared, and when an impartial quantitative assessment is sought.

5. REFERENCES

Cleveland, W. S. *Visualizing Data*. New York: Hobart Press, 1993.

Cooper, G. K., and J. R. Sirbaugh. "PARC Code: Theory and Usage." AEDC-TR-89-15, Arnold Engineering Development Center, Tullahoma, TN. December 1989.

DuToit, S. H. C., A. G. W. Steyn, and R. H. Stumpf. *Graphical Exploratory Data Analysis*. New York: Springer-Verlag, 1986.

Tufte, E. R. *The Visual Display of Quantitative Information*. Cheshire: Graphics Press, 1983.

Tufte, E. R. *Envisioning Information*. Cheshire: Graphics Press, 1990.

INTENTIONALLY LEFT BLANK.

<u>NO. OF COPIES</u>	<u>ORGANIZATION</u>
2	DEFENSE TECHNICAL INFO CTR ATTN DTIC DDA 8725 JOHN J KINGMAN RD STE 0944 FT BELVOIR VA 22060-6218
1	HQDA DAMO FDQ ATTN DENNIS SCHMIDT 400 ARMY PENTAGON WASHINGTON DC 20310-0460
1	US MILITARY ACADEMY MATH SCI CTR OF EXCELLENCE DEPT OF MATHEMATICAL SCI ATTN MDN A MAJ DON ENGEN THAYER HALL WEST POINT NY 10996-1786
1	DIRECTOR US ARMY RESEARCH LAB ATTN AMSRL CS AL TP 2800 POWDER MILL RD ADELPHI MD 20783-1145
1	DIRECTOR US ARMY RESEARCH LAB ATTN AMSRL CS AL TA 2800 POWDER MILL RD ADELPHI MD 20783-1145
3	DIRECTOR US ARMY RESEARCH LAB ATTN AMSRL CI LL 2800 POWDER MILL RD ADELPHI MD 20783-1145
	<u>ABERDEEN PROVING GROUND</u>
2	DIR USARL ATTN AMSRL CI LP (305)

NO. OF
COPIES ORGANIZATION

- 1 COMMANDER
 NAVAL WEAPONS CENTER
 ATTN C HOUSH CODE C2892
 1 ADMINISTRATION CIRCLE
 CHINA LAKE CA 93555
- 1 COMMANDER
 NAWCAD
 ATTN D FRIST CODE 4321 MS 3
 BLDG 2187 SUITE 1340 D3
 PATUXENT RIVER MD 20670-5304
- 3 WL/FIMC
 ATTN J MANTER
 E JOSYULA
 D KINSEY
 BLDG 450 SUITE 7
 2645 FIFTH ST
 WPAFB OH 45433-7913
- 1 MSU ENGINEERING RESEARCH CENTER
 ATTN B SONI
 PO BOX 6176
 MISSISSIPPI STATE MS 39762

ABERDEEN PROVING GROUND

- 20 DIR USARL
 ATTN: AMSRL-WM-PB
 P. PLOSTINS
 P. WEINACHT
 J. SAHU
 AMSRL-WM-PA, S. RAY
 AMSRL-CI, W. MERMAGEN
 AMSRL-CI-H,
 C. NIETUBICZ
 W. STUREK (5 cps)
 AMSRL-IS, J. GANTT
 AMSRL-IS-C,
 M. TAYLOR (5 cps)
 M. KINDL
 AMSRL-IS-CI,
 A. BRODEEN
 B. BROOME

NO. OF
COPIES ORGANIZATION

- 1 DEFENSE RSRCH ESTAB-VALCARTIER
ATTN MARC LAUZON
PO BOX 8800
COURCELETTE, QUEBEC
CANADA GOA1RO

- 1 DEFENSE RESEARCH AGENCY
ATTN TREVOR BIRCH
HIGH SPEED AND WEAPON AERODYNAM
BEDFORD MK41 6AE
UNITED KINGDOM

INTENTIONALLY LEFT BLANK.

REPORT DOCUMENTATION PAGE			<i>Form Approved OMB No. 0704-0188</i>	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project(0704-0188), Washington, DC 20503.				
1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE March 1997	3. REPORT TYPE AND DATES COVERED Final, Jun 96 - Jan 97		
4. TITLE AND SUBTITLE Statistical Analysis of Surface Pressure Measurements vs. Computational Predictions			5. FUNDING NUMBERS JONO: 7TE3CO	
6. AUTHOR(S) Malcolm S. Taylor and Walter B. Sturek				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Research Laboratory ATTN: AMSRL-IS-C Aberdeen Proving Ground, MD 21005-5067			8. PERFORMING ORGANIZATION REPORT NUMBER ARL-TR-1318	
9. SPONSORING/MONITORING AGENCY NAMES(S) AND ADDRESS(ES)			10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) This report details a statistical approach to the comparison of results from an international study in which Navier-Stokes computational techniques were applied to a complex flow field with highly separated flow about a fixed geometry. The traditional method for evaluating flow-field computations is for the engineer to visually inspect the results, make qualitative comparisons, and draw conclusions--a process that can be effective for a limited number of data sets. In this study, the number of data sets was so large as to make the evaluation and comparison of the results a formidable task. In response, a statistical approach with emphasis on data visualization was developed. The procedure extends directly to higher dimensions, and is of particular value when a large number of data sets are to be compared and an impartial quantitative assessment is sought.				
14. SUBJECT TERMS computational fluid dynamics, data visualization, exploratory data analysis, Navier-Stokes procedures			15. NUMBER OF PAGES 18	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UL	

INTENTIONALLY LEFT BLANK.

USER EVALUATION SHEET/CHANGE OF ADDRESS

This Laboratory undertakes a continuing effort to improve the quality of the reports it publishes. Your comments/answers to the items/questions below will aid us in our efforts.

1. ARL Report Number/Author ARL-TR-1318 (Taylor) Date of Report March 1997

2. Date Report Received _____

3. Does this report satisfy a need? (Comment on purpose, related project, or other area of interest for which the report will be used.) _____

4. Specifically, how is the report being used? (Information source, design data, procedure, source of ideas, etc.) _____

5. Has the information in this report led to any quantitative savings as far as man-hours or dollars saved, operating costs avoided, or efficiencies achieved, etc? If so, please elaborate. _____

6. General Comments. What do you think should be changed to improve future reports? (Indicate changes to organization, technical content, format, etc.) _____

**CURRENT
ADDRESS**

Organization

Name E-mail Name

Street or P.O. Box No.

City, State, Zip Code

7. If indicating a Change of Address or Address Correction, please provide the Current or Correct address above and the Old or Incorrect address below.

**OLD
ADDRESS**

Organization

Name

Street or P.O. Box No.

City, State, Zip Code

(Remove this sheet, fold as indicated, tape closed, and mail.)
(DO NOT STAPLE)

DEPARTMENT OF THE ARMY

OFFICIAL BUSINESS

BUSINESS REPLY MAIL
FIRST CLASS PERMIT NO 0001, APG, MD

POSTAGE WILL BE PAID BY ADDRESSEE

DIRECTOR
US ARMY RESEARCH LABORATORY
ATTN AMSRL IS C.
ABERDEEN PROVING GROUND MD 21005-5067



NO POSTAGE
NECESSARY
IF MAILED
IN THE
UNITED STATES

