

7000

INTERIM REPORT

7-23-79

Accession No. _____

Contract Program or Project Title: LOCA Analysis Assessment

Subject of this Document: "Preliminary Application of Quantitative Assessment Procedure"

Type of Document: Interim Report

Author(s): J. A. Dearien, C. B. Davis

Date of Document: April 1979

Responsible NRC Individual and NRC Office or Division: S. Fabric, NRC/RSR

This document was prepared primarily for preliminary or internal use. It has not received full review and approval. Since there may be substantive changes, this document should not be considered final.

Ray Rose
H. P. Pearson, Supervisor
Information Processing
EG&G Idaho

Prepared for
U.S. Nuclear Regulatory Commission
Washington, D.C. 20555

NRC File #A6047

INTERIM REPORT

NRC Research and Technical Assistance Report

520 265
79d8d884527

CAAP-TR-047

Date: April 1979

CODE ASSESSMENT AND APPLICATIONS PROGRAM

PRELIMINARY APPLICATION
OF
QUANTITATIVE ASSESSMENT PROCEDURE

NRC Research and Technical
Assistance Report



EG&G Idaho, Inc.



IDAHO NATIONAL ENGINEERING LABORATORY

DEPARTMENT OF ENERGY

IDAHO OPERATIONS OFFICE UNDER CONTRACT EY-76-C-07-1570

520 266



FORM EG&G-398
(Rev. 12-78)

INTERIM REPORT

Accession No. _____

Report No. CAAP-TR-047

Contract Program or Project Title: LOCA Analysis Assessment

Subject of this Document: Preliminary Application of Quantitative Assessment Procedure

Type of Document: Interim Report

Author(s): J. A. Dearien
C. B. Davis

Date of Document: April 1979

Responsible NRC Individual and NRC Office or Division: S. Fabric, NRC/RSR

This document was prepared primarily for preliminary or internal use. It has not received full review and approval. Since there may be substantive changes, this document should not be considered final.

EG&G Idaho, Inc.
Idaho Falls, Idaho 83401

Prepared for the
U.S. Nuclear Regulatory Commission
and the U.S. Department of Energy
Idaho Operations Office
Under contract No. EY-76-C-07-1570
NRC FIN No. 6047

INTERIM REPORT

520 267

ABSTRACT

A procedure for the quantitative assessment of computer codes is described and applied to two Semiscale tests. Results of the application of this procedure are described and conclusions drawn regarding the use of this procedure in the assessment of computer codes. A method has been developed as part of the procedure to indicate the percentile of knowledgeable persons who would deem the code as doing an acceptable job.

SUMMARY

A procedure has been developed to quantify the results of code assessment and relate this quantified result to the percentage of knowledgeable analysts who would be expected to find the code acceptable.

The quantification procedure was applied to two Semiscale tests and the results of the assessment process indicate that 50% to 70% of knowledgeable analysts would find the code results an acceptable representation of the experimental data.

The results are preliminary due to the limited amount of data at present but are very encouraging with respect to future application of the procedure.

CONTENTS

ABSTRACT i

SUMMARY. ii

I. INTRODUCTION 1

II. DESCRIPTION OF PROCEDURE 2

III. DESCRIPTION OF EXAMPLE PROBLEM 4

IV. DISCUSSION OF RESULTS. 7

V. CONCLUSIONS. 13

VI. FURTHER WORK NEEDED. 15

APPENDIX A

APPENDIX B

520 270

I. INTRODUCTION

The assessment of complex computer codes is a procedure by which the capabilities and limitations of these codes are determined by comparison of predicted results with experimental data. Code Assessment has been carried out at the INEL for over four years with the fuel codes (FRAP-S, FRAPCON, and FRAP-T) and the first assessment of a large thermal hydraulic code (RELAP4/MOD6) was completed in December 1978.

Assessment of these codes has, to date, been primarily a graphic approach with data/prediction comparisons presented in various graphical form with the readers left to their own judgement as to the quality of the comparisons. In those cases where assessors' comments are made on the quality of the comparisons, those comments are of the form "good", "not so good", "compare quite well", etc. The lack of a quantitative assessment criteria has precluded anything but qualitative statements as assessment results.

A procedure has been developed at INEL to quantify the type of data/prediction comparisons encountered in the assessment of thermal hydraulic codes. This report describes the procedure (Section II) and presents an example of its application to two Semiscale tests, (Section III) using the RELAP4/MOD6 code. Results are described in Section IV and conclusions regarding the use of the procedure are covered in Section V. Section VI describes future work planned and needed in the area of code assessment.

Appendix A describes the quantitative assessment procedure in detail. Appendix B describes research done on the human perception of code acceptability and how the results of this research are used in the assessment process.

520 271

II. DESCRIPTION OF PROCEDURE

A detailed description of the quantitative assessment procedure has been previously documented and is included in this report as Appendix A. The procedure will be discussed briefly with a description of a modification made to the appended procedure.

In the comparison of predicted results with experimental data, a problem arises in that the experimental data has a certain amount of uncertainty associated with it. This uncertainty arises from several sources and must be considered when comparing it to a "single valued" prediction from a computer code. In Figure 1, the reader can see the five different areas considered as error indicators, A) deviation of prediction from the data mean (absolute), B) deviation of prediction above the data mean, C) deviation of prediction below the data mean, D) deviation of prediction outside data error bands and E) deviation of prediction trend (slope) from data trend.

The procedure for quantifying the five error indicators is described in Appendix A with the exception of indicator E. The error indicators on prediction trend vs. data trend was recognized as a significant indicator and has since been added to the four transient error indicators of Appendix A. The equation for quantifying trend error is

$$E(t) = 1 - \frac{1}{t} * \int_{0.0}^t \frac{(P(t) - P(t-T) - (X(t) - X(t-T)))}{X_{MAX}(t) - X_{MIN}(t)} dt$$

where E measures the trend error, t is time, T is constant time offset, P and X are the calculated and measured parameters respectively, and XMAX and XMIN are the error bands of the data. The results described in Section V were obtained with T=1 s.

Weighting factors are applied to each of the error indicators as well as to the individual parameters being assessed for comparison (clad temperature, flows, etc). The combined effects of the weighted parameters and weighted error indicators are accumulated by the summing procedure described in Appendix A. This accumulation results in a quantification of the code results between 0 and 100 which can then be used as an indicator of the code capability or a relative indicator when scores of two codes are known.

The score calculated for the code is used to determine a percentile acceptance (PA) of the code to perform its desired function. The PA of a code is the percentage of people who would deem the code as doing an acceptable job. Appendix B describes the procedure and research used to arrive at the PA relations.

520 273

III. DESCRIPTION OF EXAMPLE PROBLEM

The objectives of this initial test of the quantification procedure were to determine 1) if it could be applied to a real problem, 2) the effort involved in applying the procedure and 3) if the results obtained are of a nature that can be realistically used in the assessment process. Items 1 and 2 have been resolved satisfactorily in that the procedure can be applied to a real problem with little effort (given that the uncertainty bands are available). The remainder of the report will address the applicability of the results in the assessment process.

1. SEMISCALE TEST PROBLEMS

Two Semiscale tests (S-04-6, S-06-6) were used in the example evaluation. These tests were selected because they were from a series of tests for which the data uncertainty bands were derivable. Furthermore, RELAP4/MOD6 system calculations had been made for these tests and were available for comparison with the data.

The Semiscale experimental program is part of the investigation of the thermal and hydraulic phenomena accompanying a hypothesized loss-of-coolant accident in a water cooled nuclear reactor system. Semiscale Tests S-04-6 and S-06-6 simulated the response of a pressurized water reactor during a loss-of-coolant experiment initiated by a 200% double-ended offset shear in the cold leg piping. Test S-04-6 was conducted from initial conditions of 15.6 MPa pressure, 557 K cold leg fluid temperature, 1.44 MW core power, and 38 K fluid temperature rise across the core. Test S-06-6 was conducted from initial conditions of 15.8 MPa pressure, 563 K cold leg fluid temperature, 1.00 MW core power, and 36 K fluid temperature rise across the core.

Nine key parameters were selected for comparison and quantification of error. These nine key parameters and the key parameter weighting functions (Section IV, Appendix A) are listed in Table I. The weighting functions were selected by a group of thermal hydraulic assessment engineers.

TABLE I

KEY PARAMETER WEIGHTING FACTORS

<u>Key Parameter</u>	<u>Weighting Function</u>
Clad Temperature - Middle of the Core	20
Volumetric Flow - Vessel Side Break	15
Fluid Density - Core Inlet	13
Fluid Density - Vessel Side Break	11
Volumetric Flow - Core Inlet	8
Mass Flow - Pump Side Break	7
Clad Temperature - Lower Core	5
Clad Temperature - Upper Core	5
Fluid Pressure - Upper Plenum	4

The next selections by the assessment engineers were the weighting factors for the five individual error indicators. Table II lists these weighting factors.

520 275

TABLE II

ERROR INDICATOR WEIGHTING FACTORS

<u>Error Indicators</u>	<u>Weighting Factors</u>	
	<u>Fluid Parameters</u>	<u>Clad Temperatures</u>
A	0.05	0.05
B	0.30	0.10
C	0.30	0.50
D	0.05	0.05
E	0.30	0.30

The weighting factors shown in Table II were determined in the following manner. Several hypothetical curves which compared a calculation with data (including error bands) were shown to a sample of 150 engineers (see Appendix B). The engineers subjectively graded the "goodness" of each calculation by assigning it a score between 0.0 and 1.0 (1.0 corresponding to a perfect calculation). A set of weighting factors was determined so that the scoring procedure, when applied to each hypothetical case, yielded a score approximately equal to the mean of scores subjectively estimated by the sample of engineers. This set of weighting factors was applied to the fluid parameters as shown in Table II. Similar weighting factors were used to score the clad temperature predictions except that modifications were made to penalize underpredictions.

2. APPLICATION OF PROCEDURE

The procedure as defined in Appendix A was applied to the RELAP4/MOD6 predictions of the two Semiscale tests with the weighting factors described in the preceding section. The results of the

application were individual time dependent scores for each of the nine key parameters and a combined score which included the weighting contribution of each of the key parameter scores.

The only "specific" that bears noting here is the procedure for handling negative scores. In the evaluation of the individual error indicators, it is possible to obtain a negative value if the prediction is of sufficient error. The adopted procedure for handling negatives was to allow the individual error indicators to be negative but not allow the weighted sum of the error indicators for any one key parameter to be negative. The logic being that the worst comparison one can obtain for a key parameter is zero.

IV. DISCUSSION OF RESULTS

The results are presented by key parameters in which the data/prediction plots for each of the two Semiscale tests are shown and then followed by the key parameter score plot. A discussion is given on each key parameter and the resulting score for that key parameter. Only the behavior of the data/prediction comparison relative to the computed parameter score will be addressed; the behavior of the experiment and/or the calculation will not be discussed. The total code score and percentile acceptance (PA) are presented at the end of the key parameter series and discussed.

1. CLAD TEMPERATURE NEAR THE CORE CENTER

Figures 2 and 3 compare predictions with clad temperatures measured near the axial center of the core, where the highest clad temperatures occurred for Tests S-04-6 and S-06-6. The figures show the prediction, the mean of the measurements, and the range of measured values versus time after rupture. Figure 4 compares the computed parameter scores, which were based on the results shown in Figures 2 and 3, for the two predictions. The parameter score for the S-06-4 calculation was high (0.85) while the score for the S-06-6 calculation was generally low (0.40). The parameter score for the S-04-6 calculation was higher because 1) its prediction was nearer the mean and 2) the range of the data was much larger in S-04-6 than S-06-6.

2. VOLUMETRIC FLOW NEAR THE VESSEL-SIDE BREAK

Figures 5 and 6 show predicted and measured flow near the vessel-side break for both Semiscale tests. Figure 7 shows the

parameter scores for the two predictions. The data/prediction comparisons are quite similar for the two tests resulting in similar scores for both predictions.

3. FLUID DENSITY AT THE CORE INLET

Figures 8 and 9 show predicted and measured fluid density at the core inlet. Figure 10 shows the parameter score for each prediction. In both tests, the sudden density decrease at about 0.5 s is predicted to occur later than actually measured causing the predictions to be far outside of the uncertainty limits during the first second of blowdown. Consequently, the scores for both predictions are 0.0 early in the test. When the predictions return inside the error bands, both parameter scores improve.

4. FLUID DENSITY NEAR THE VESSEL-SIDE BREAK

Figures 11 and 12 show predicted and measured fluid density near the vessel side break. Figure 13 shows the parameter score for each prediction. The predictions are generally within the data error bands. However, because the calculations are smooth and the measurements noisy, the trends are somewhat different resulting in lower scores than would have occurred if the measurements were smooth.

5. VOLUMETRIC FLOW AT THE CORE INLET

Figures 14 and 15 show predicted and measured volumetric flow near the core inlet. Figure 16 shows the parameter score for each prediction. Both predictions are frequently outside the error bands resulting in the relatively low parameter scores shown in Figure 16.

520 219

6. MASS FLOW NEAR THE PUMP SIDE BREAK

Figures 17 and 18 show predicted and measured mass flow near the pump side break. Figure 19 shows the parameter score for each prediction. Since the predictions are near the measurements, the parameter scores are relatively high. Because the measurements are noisy and the predictions smooth, the parameter scores are lower than would have occurred if the measurements were smooth.

7. CLAD TEMPERATURES NEAR THE BOTTOM OF THE CORE

Figures 20 and 21 show predicted and measured clad temperatures near the bottom of the core. Figure 22 shows the parameter score for each prediction. Both predictions are generally about 50 K higher than the data mean. Because the range of data is much smaller in S-06-6, the prediction is relatively further away from the mean and the procedure produces a lower score for the S-06-6 prediction.

8. CLAD TEMPERATURE NEAR THE TOP OF THE CORE

Figures 23 and 24 show predicted and measured clad temperatures near the top of the core. Figure 25 shows the parameter score for each prediction. The calculated temperatures are far higher than the measurements for both tests. The overprediction is large enough that both scores are 0.0 by 20 s after rupture.

9. FLUID PRESSURE IN THE UPPER PLENUM

Figures 26 and 27 show predicted and measured fluid pressure in the upper plenum. Figure 28 shows the parameter score for each prediction. The pressure is generally underpredicted for both tests. Since the S-04-6 prediction is considerably closer to the data, the S-04-6 prediction received a better score than the S-06-6 prediction.

The comparison in Figure 26 and the resultant score (50%) is an example of where the known data uncertainty may not be realistic for use in calculating a score. Over much of the range the prediction is within 5-10% of the data mean. The low score illustrated in Figure 28 is due to the code's inability to predict within the tight error bands. If this prediction were considered "good", the error bands might be broadened to reflect "allowable" error bands and thus give a score more compatible with the subjective evaluation of the comparison.

10. TOTAL CODE SCORE AND PERCENTILE ACCEPTANCE

The parameter weighting functions shown in Table I were applied to the parameters shown previously in this section to obtain the total calculation scores shown in Figure 29. Both total calculation scores are quite constant with respect to time. The two total calculation scores behaved similarly since they are based on similar calculations. The score for the S-04-6 calculation is higher than the S-06-6 calculation primarily because of the results shown in Figure 4 where the S-04-6 calculation received a higher score for the most heavily weighted parameter.

The important thing about the scores is that they are very close to each other and are relatively constant over the duration of the analyses. This indicates that the procedure is measuring something about the code that is relatively constant since the individual parametric scores vary a great deal.

The total code scores for these two test cases (55 ± 5) were used to obtain a PA of the code of $60\% \pm 10\%$. This PA was obtained from Figure B-9 (Appendix B) and indicates that from 50% to 70% of knowledgeable persons would find these analyses an acceptable representation of the data. No statement can be made, at present, if this is an acceptable PA since a formal acceptance criteria has not been agreed upon.

11. TOTAL CODE SCORE SENSITIVITY

The relatively constant nature of the total code score is an important factor in the quantitative assessment procedure since a large variance in this quantity could render it unusable as an evaluation criteria. Because of this importance, a sensitivity analysis was performed on the total code score to determine its dependence on the relative weighting factors of the key parameters. One can see that for the case where one parameter is given all the weight, the total code score would be equal to the key parameter score but for practical applications, there would be a reasonable distribution of weight based on the importance of the key parameters.

The sensitivity analysis was performed by modifying the individual key parameter weights of Table I as shown in Table III.

520 282

TABLE III

KEY PARAMETER SENSITIVITY MODIFICATIONS

<u>Key Parameter</u>	<u>Orig. Wt.</u>	<u>MOD1</u>	<u>MOD2</u>	<u>MOD3</u>
Clad temperature - middle of the core	20	10	20	20
Volumetric flow - vessel side break	15	10	15	15
Fluid density - core inlet	13	10	13	13
Fluid density - vessel side break	11	10	11	50
Volumetric flow - core inlet	5	10	50	8
Mass flow - pump side break	7	10	7	7
Clad temperature - lower core	5	10	5	5
Clad temperature - upper core	5	10	5	5
Fluid pressure - upper plenum	4	10	4	4

Results of the sensitivity study are shown in Figure 30 and indicate that 1) the relatively constant behavior of the total code score is maintained and 2) the variations are within the range of the two base case runs (55 \pm 5). Thus, the results indicate that for reasonable variations in the key parameter weighting functions, the constant behavior of the total code score is maintained within a small range. Also, this near constant behavior is not a function of the natural smoothing obtained in integral functions as the final score is only an addition of the key parameter scores (which do benefit from integral smoothing and yet still exhibit significant variations with respect to time). Thus, this near constant behavior gives further credence to the assumption that a basic, inherent capability of the code is being reflected through this total code score.

520 283

V. CONCLUSIONS

The conclusions reached here are based on a small sample but the results obtained are encouraging enough to pursue the procedure further.

1. Code Assessment results can be quantified. The procedure has been tried and results of an assessment comparison were quantified. The procedure is such that it can be automated and thus used in the assessment process with very little cost.
2. Individual parametric scores give quick identification of problem areas in code. It is possible to look at the individual parametric scores and get a quick idea of the relative calculational capabilities of the code in different aspects of the analysis (temperature, flow, etc) since the procedure is a normalizing process.
3. Individual parametric scores can be used as an indication of experimental data needs. If a parametric score is high and the analyst feels (or knows) that the calculation was really in substantial absolute error, an examination of the data-scoring plots will most likely indicate that the error bands on the data are large. Better instrumentation or experimental procedures may be called for.
4. Total code score is stable and appears to be measuring an inherent capability of the code. This is the most important finding to date with the procedure. The relatively constant values of the code scores for the two example tests were obtained from two sets of highly variable and individually differing sub scores. The consistency and near single-value output of the procedure appear to indicate the measurement of something inherent to the overall code itself rather than any

particular model. While the numeric score does not have sufficient meaning at the present to indicate, by itself, the capability of the code, it appears to be something which can be used in the future for establishing an acceptance criteria.

5. The percentile acceptance value of the code is a better indicator of code acceptability than code score. This conclusion is drawn primarily from the research findings discussed in Appendix B. The variability in human perception is such that there will always be a range of value judgement (scores) even when the decision of acceptable or nonacceptable is the important final outcome. Also from the standpoint of a person(s) having to make a decision on the acceptance or nonacceptance of a code, there will (probably) never be 100% consensus on whether or not something is acceptable. Therefore, having a process which indicates that 70%, 80%, or 90% of the knowledgeable analysts find a code acceptable gives the decision making person (or body) better idea of the worth of the code.

VI. FURTHER WORK NEEDED

The results presented in this report are preliminary in nature and are based on the limited amount of data available at this time. Although the initial results are very encouraging, there are several areas in which additional work and data are needed.

1. ADDITIONAL APPLICATIONS OF PROCEDURE

This preliminary application of the quantitative assessment procedure was made with two Semiscale tests taken from two similar Semiscale test series. Before concrete conclusions can be drawn on the applicability of the procedure, applications should be made on 1) other Semiscale tests series and 2) other facilities. Application of the procedure for these other tests must wait until error bands are obtained for the other tests. These error bands are to be a part of future experimental data reports and tests of the procedure will be made on these tests as the data and error bands are available.

2. ADDITIONAL DATA FOR PERCENTILE ACCEPTANCE ANALYSIS

Data on acceptability of results have been taken from 118 analysts and these data have been used to correlate code scores with acceptability of the code. While the number of engineers surveyed is large, there are other factions of the engineering and scientific field that could provide additional input and perceptions into the data base.

3. DEVELOPMENT OF AN ACCEPTANCE CRITERIA

Given the documented research on human perception and the fact that different percentages of people are willing to accept different results, how do we set an acceptance level for a code or particular results? The acceptance level must be set recognizing that there may never be 100% of the reviewers that will agree on a code being acceptable. The task at hand is to decide what percentage acceptance is reasonable.

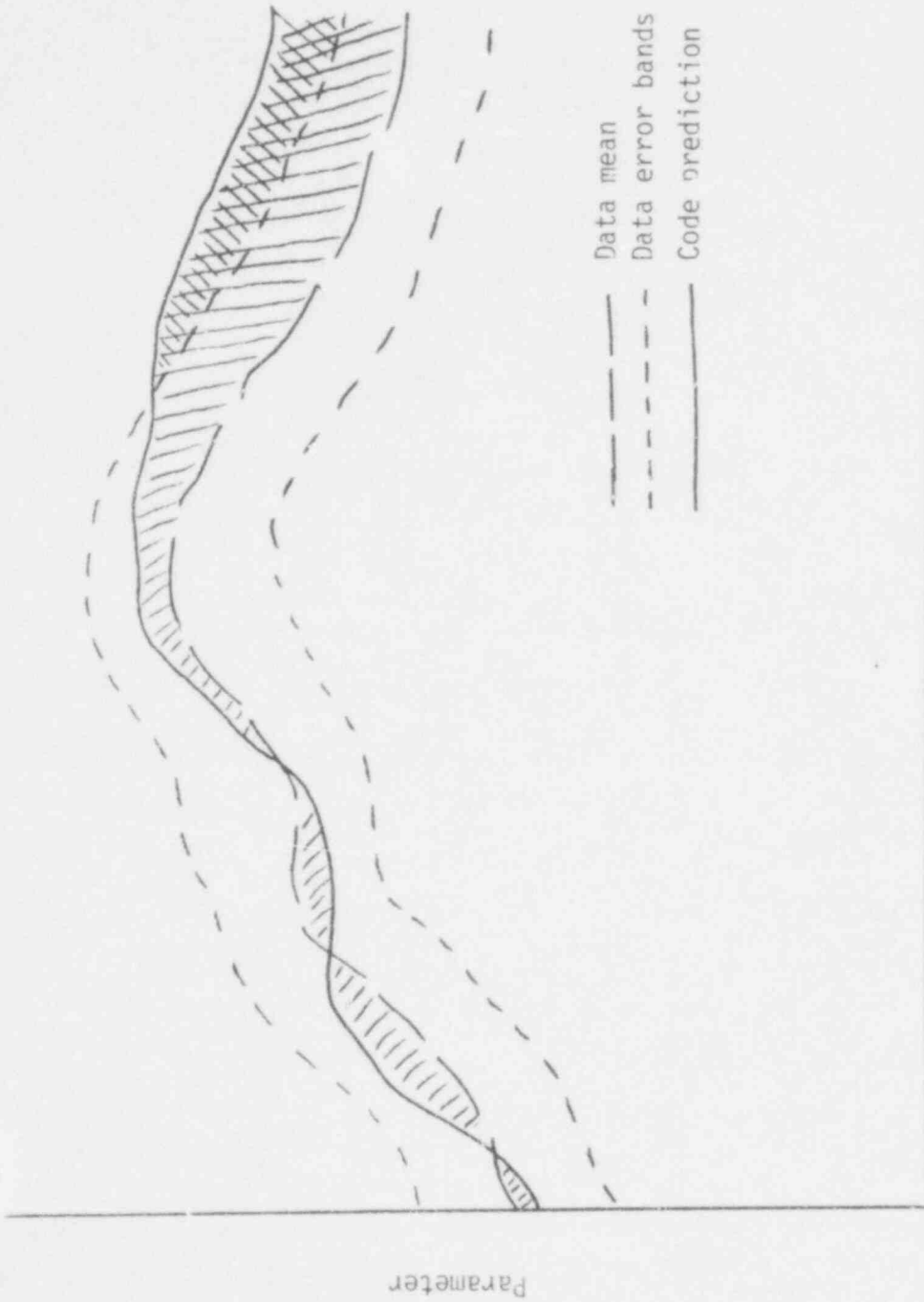


Figure 1

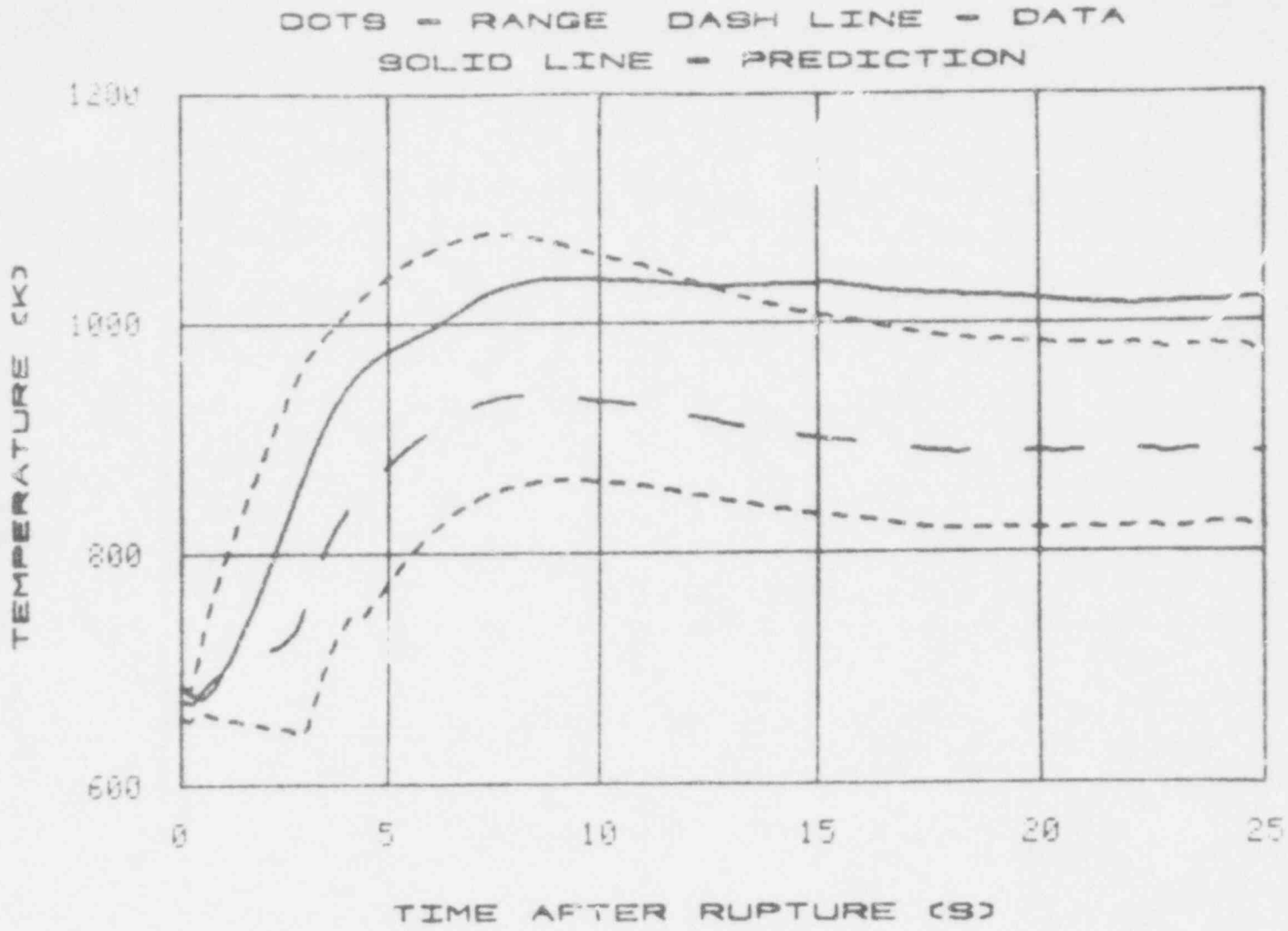


Figure 2

20

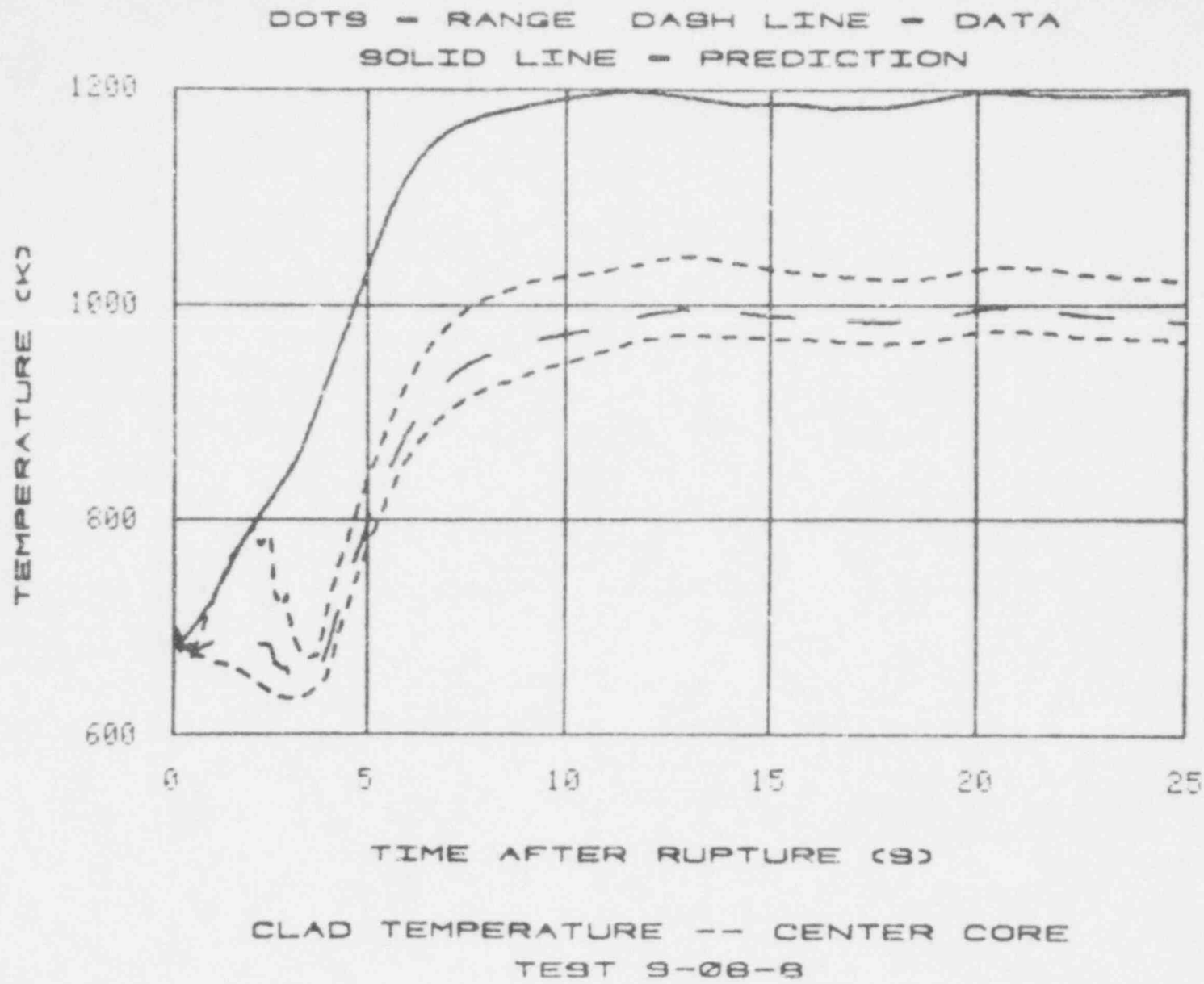
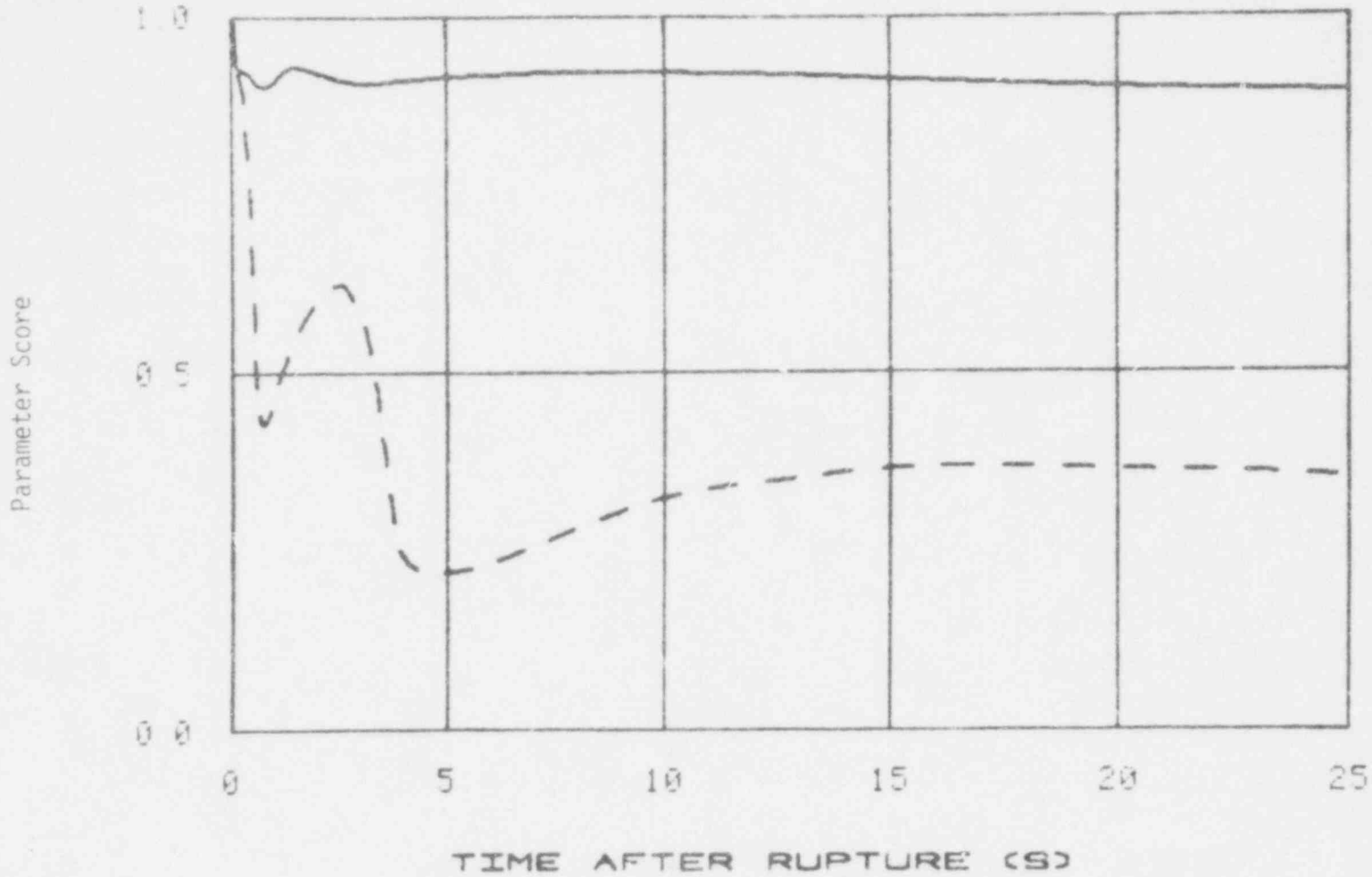


Figure 3

520 290

SOLID LINE - S-04-B DASH LINE - S-08-B



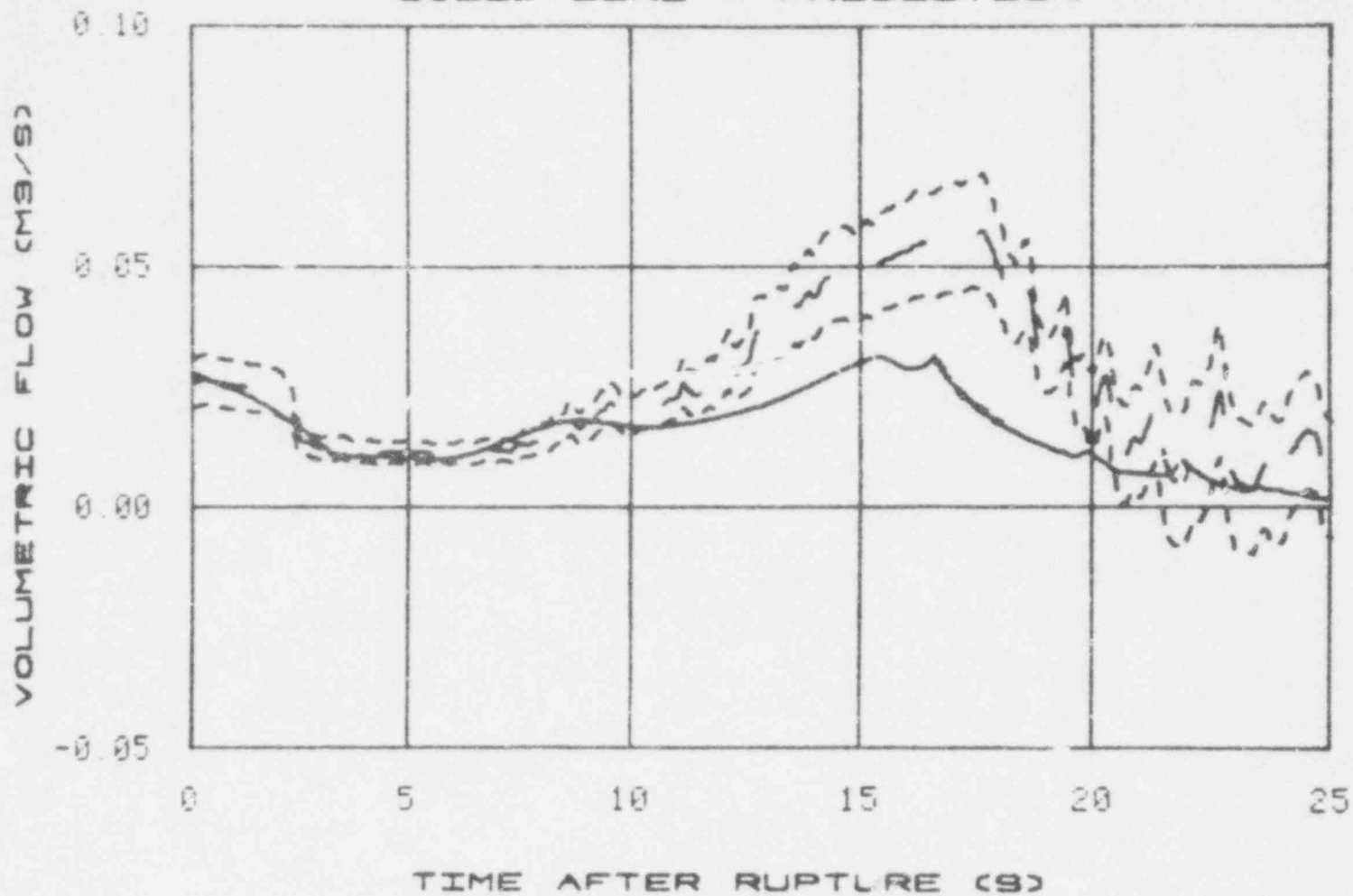
CLAD TEMPERATURE -- CENTER CORE

Figure 4

21

520 291

DOTS - RANGE DASH LINE - DATA
SOLID LINE - PREDICTION



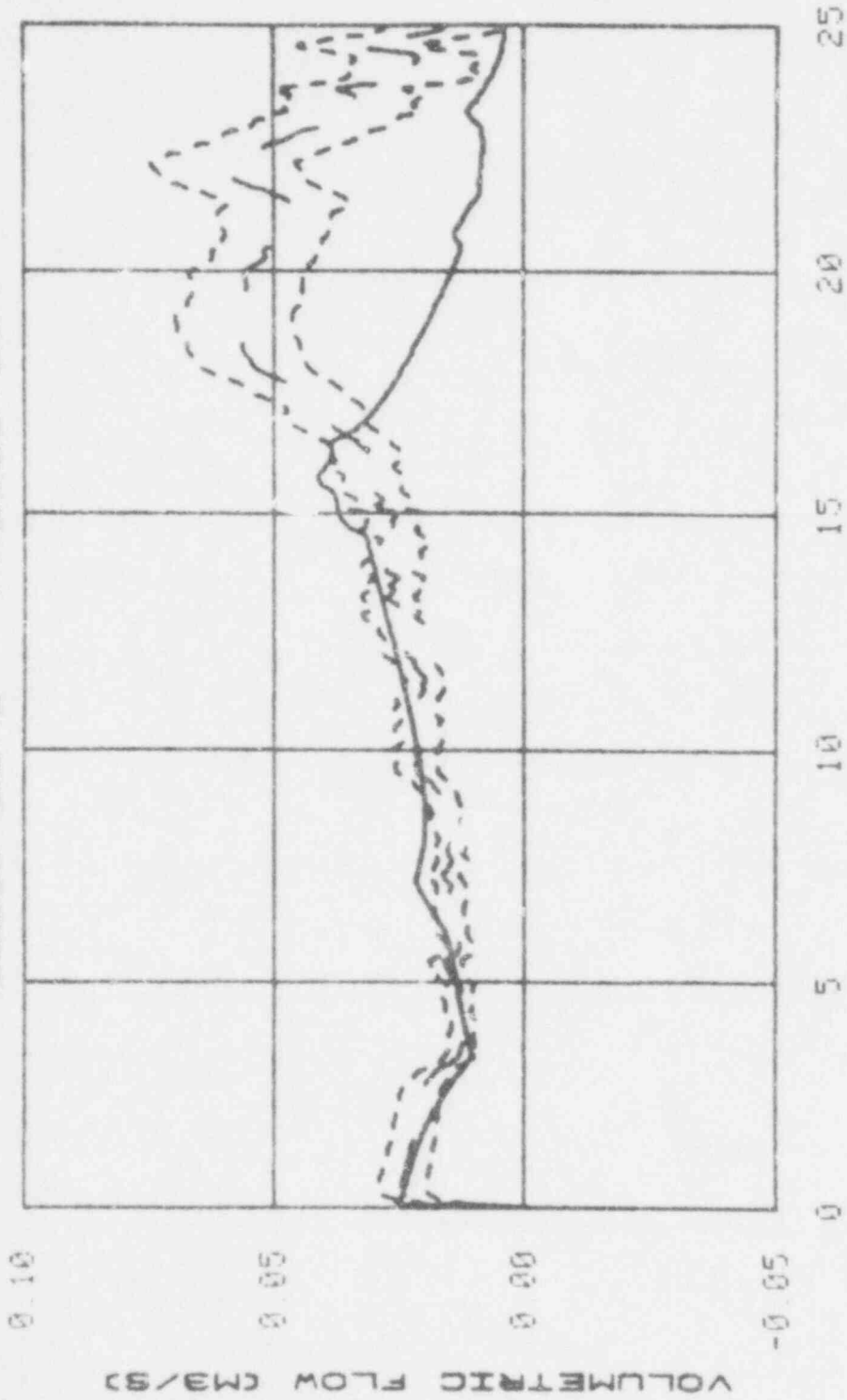
VOLUMETRIC FLOW -- VESSEL SIDE BREAK
TEST 9-04-8

Figure 5

22

520
292

DOTS - RANGE DASH LINE - DATA
SOLID LINE - PREDICTION

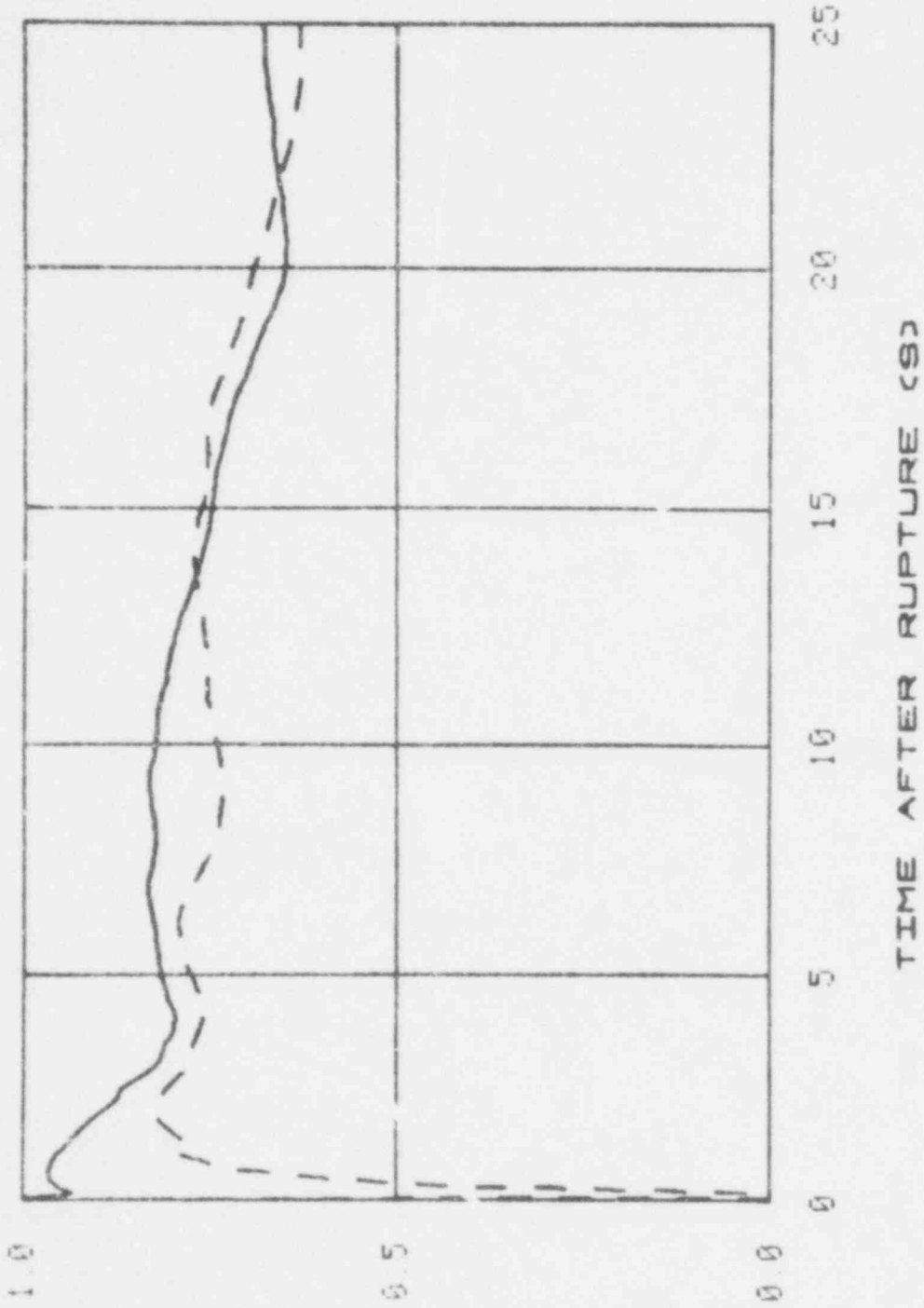


TIME AFTER RUPTURE (S)

VOLUMETRIC FLOW -- VESSEL SIDE BREAK
TEST 9-08-8

Figure 6

SOLID LINE - 9-04-8 DASH LINE - 9-08-8

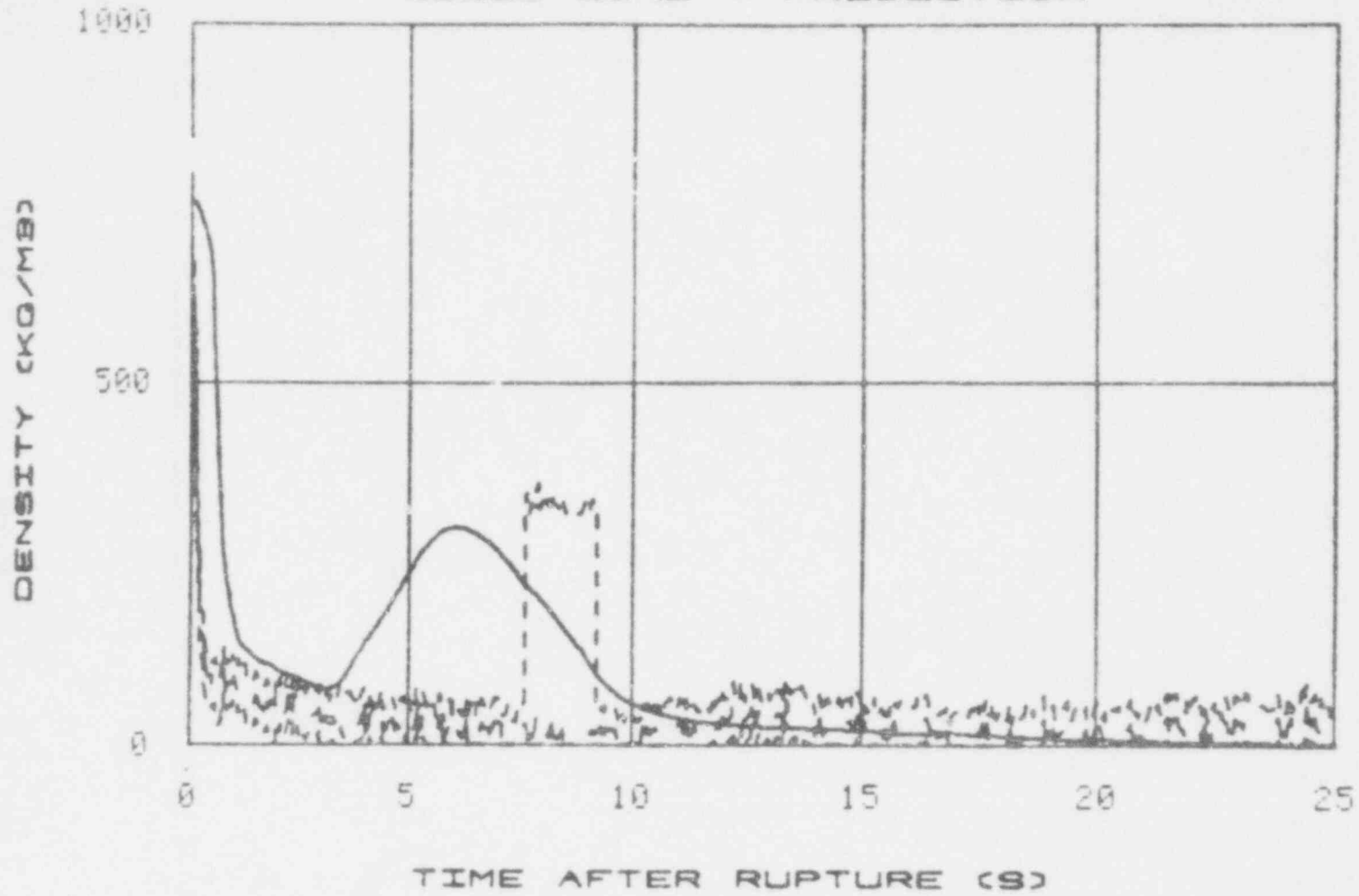


VOLUMETRIC FLOW -- VESSEL SIDE BREAK

Figure 7

Parameter Score

DOTS - RANGE DASH LINE - DATA
SOLID LINE - PREDICTION

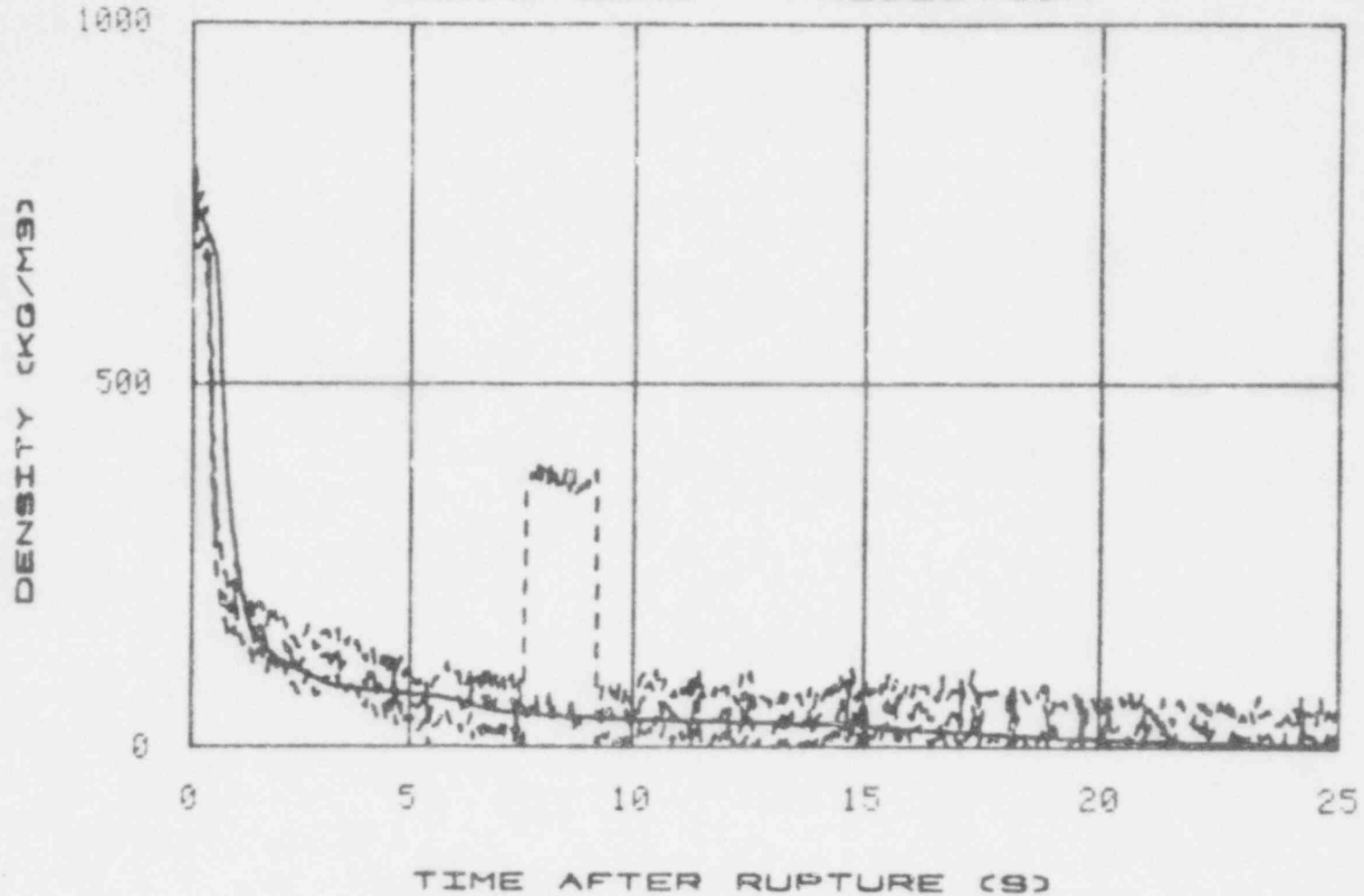


DENSITY -- CORE INLET
TEST 9-04-B

Figure 8

520 295

DOTS - RANGE DASH LINE - DATA
SOLID LINE - PREDICTION



DENSITY -- CORE INLET
TEST 9-08-8

Figure 9

26

520
296

SOLID LINE - 9-04-B DASH LINE - 9-08-B

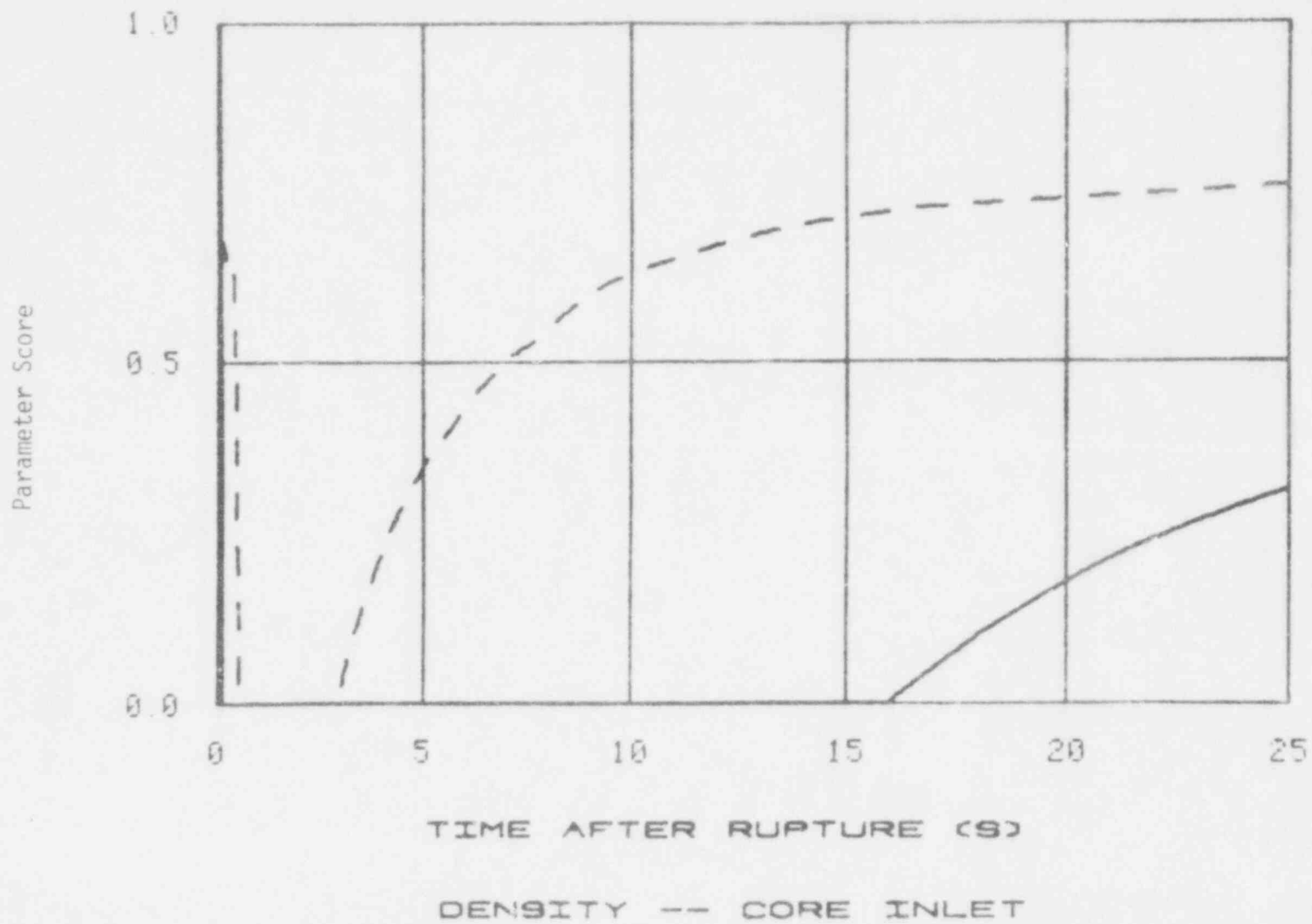
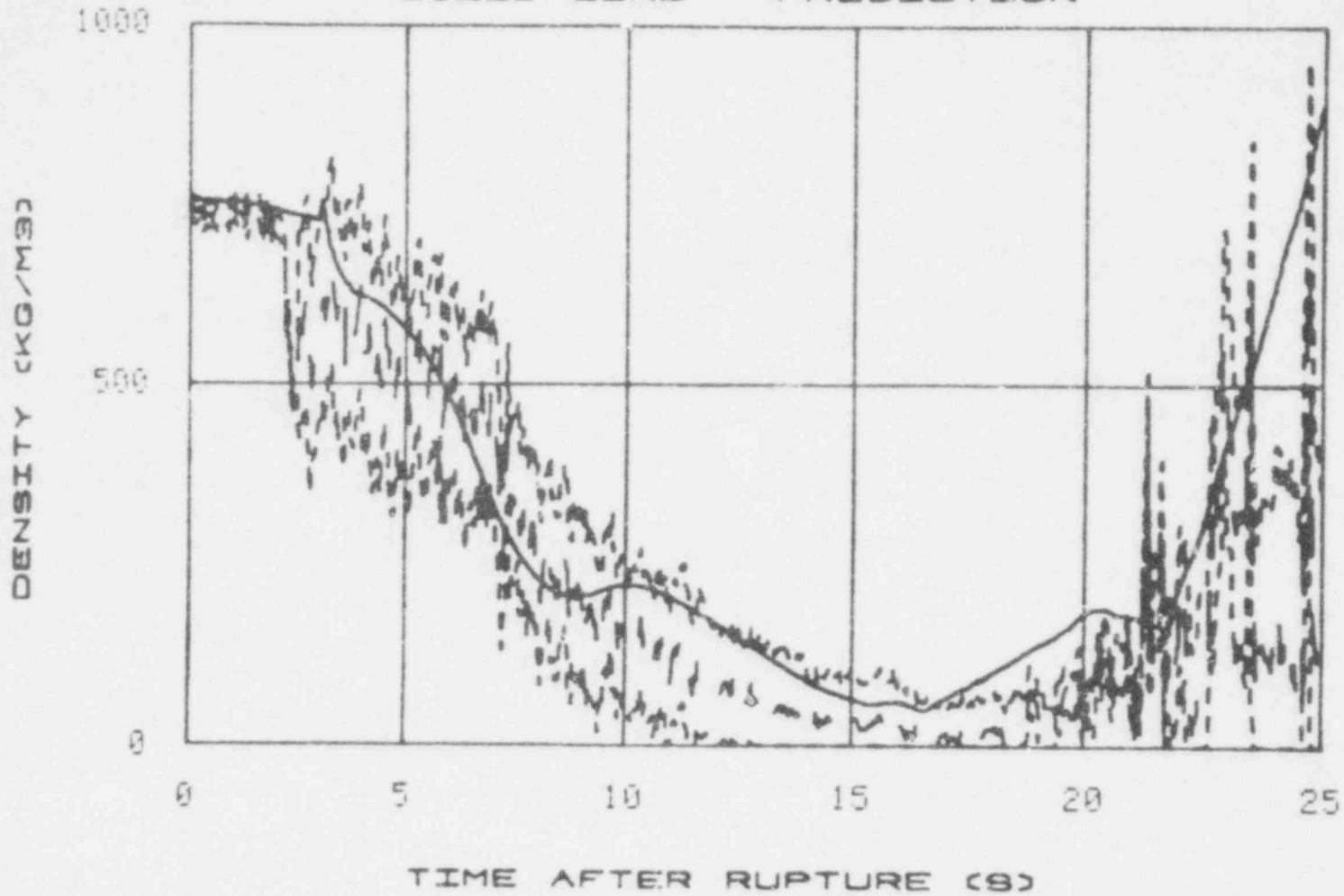


Figure 10

27
520
297

DOTS - RANGE DASH LINE - DATA
SOLID LINE - PREDICTION



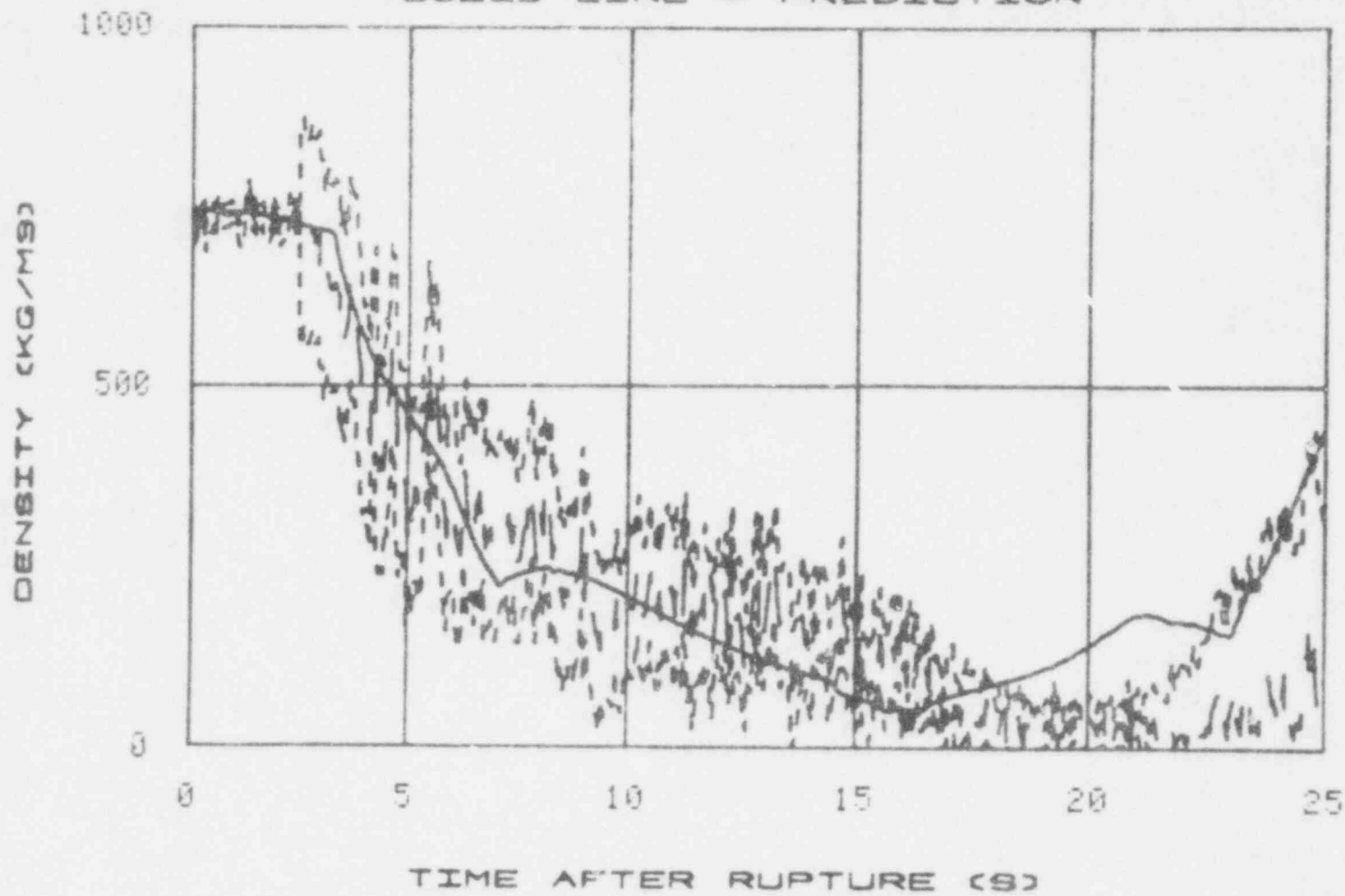
DENSITY -- VESSEL SIDE BREAK
TEST 9-04-8

Figure 11

28

520 298

DOTS - RANGE DASH LINE - DATA
SOLID LINE - PREDICTION



DENSITY -- VESSEL SIDE BREAK
TEST 9-08-8

Figure 12

29

520 299

SOLID LINE - 9-04-B DASH LINE - 9-08-B

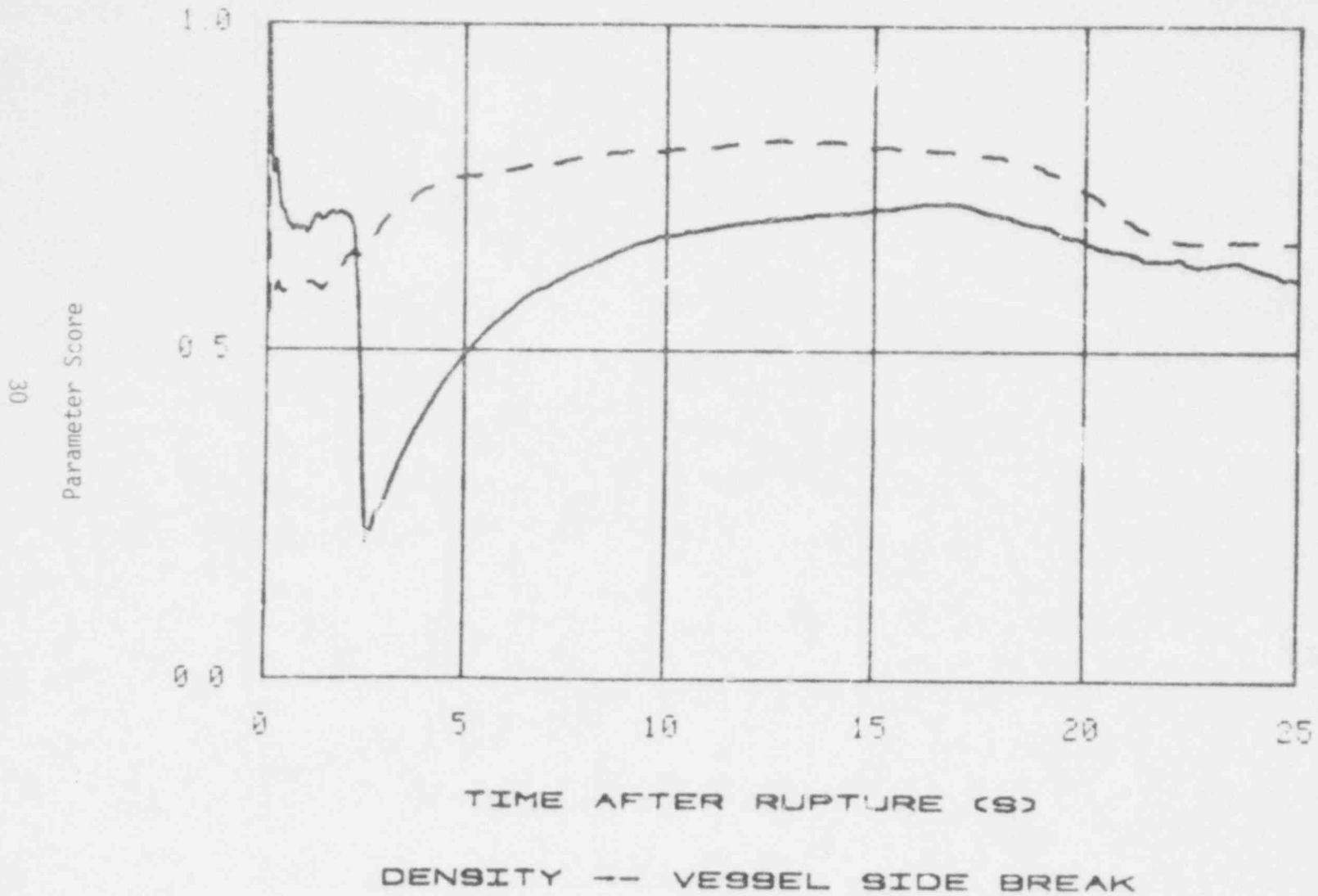
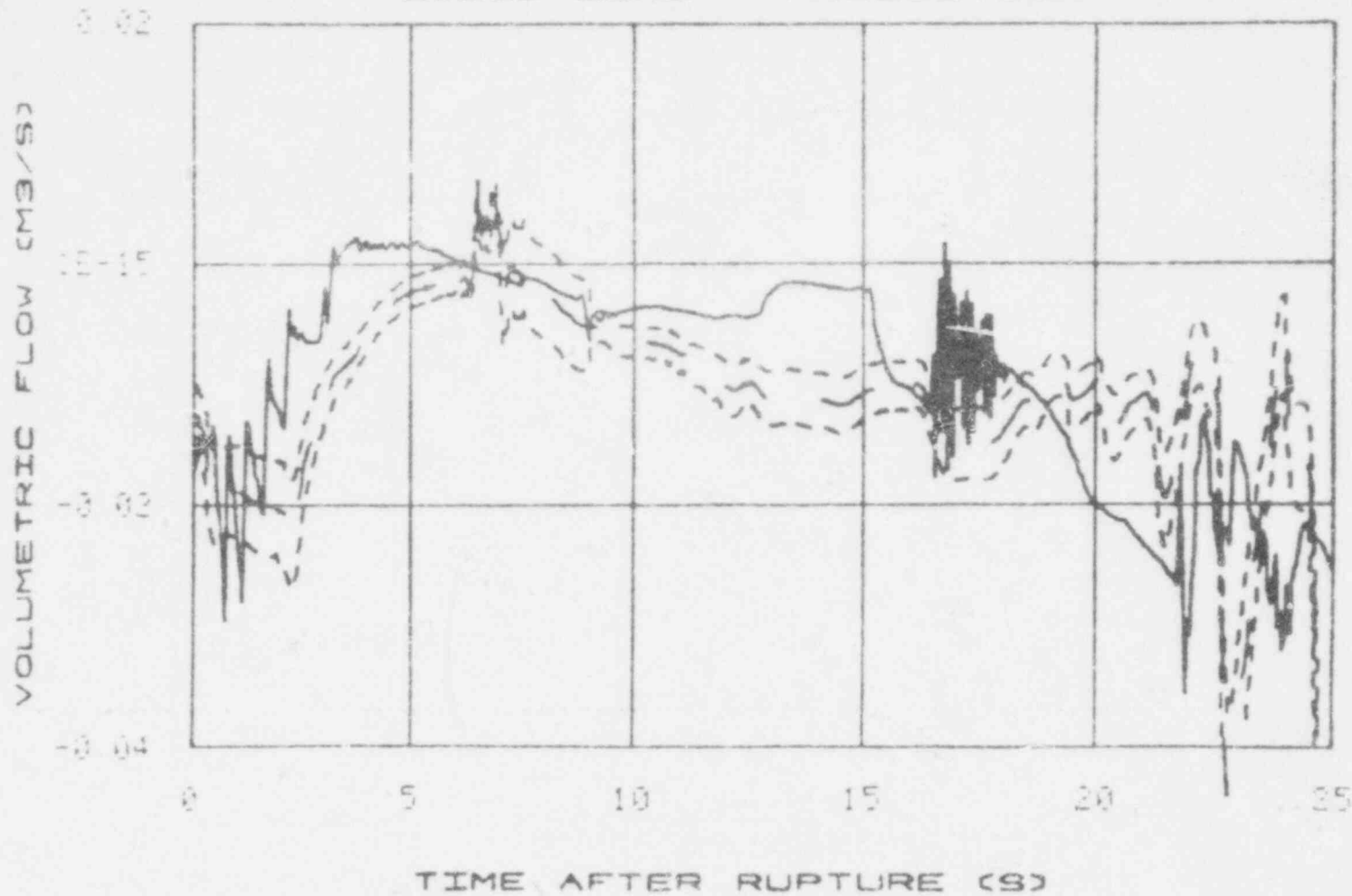


Figure 13

520 300

DOTS - RANGE DASH LINE - DATA
SOLID LINE - PREDICTION

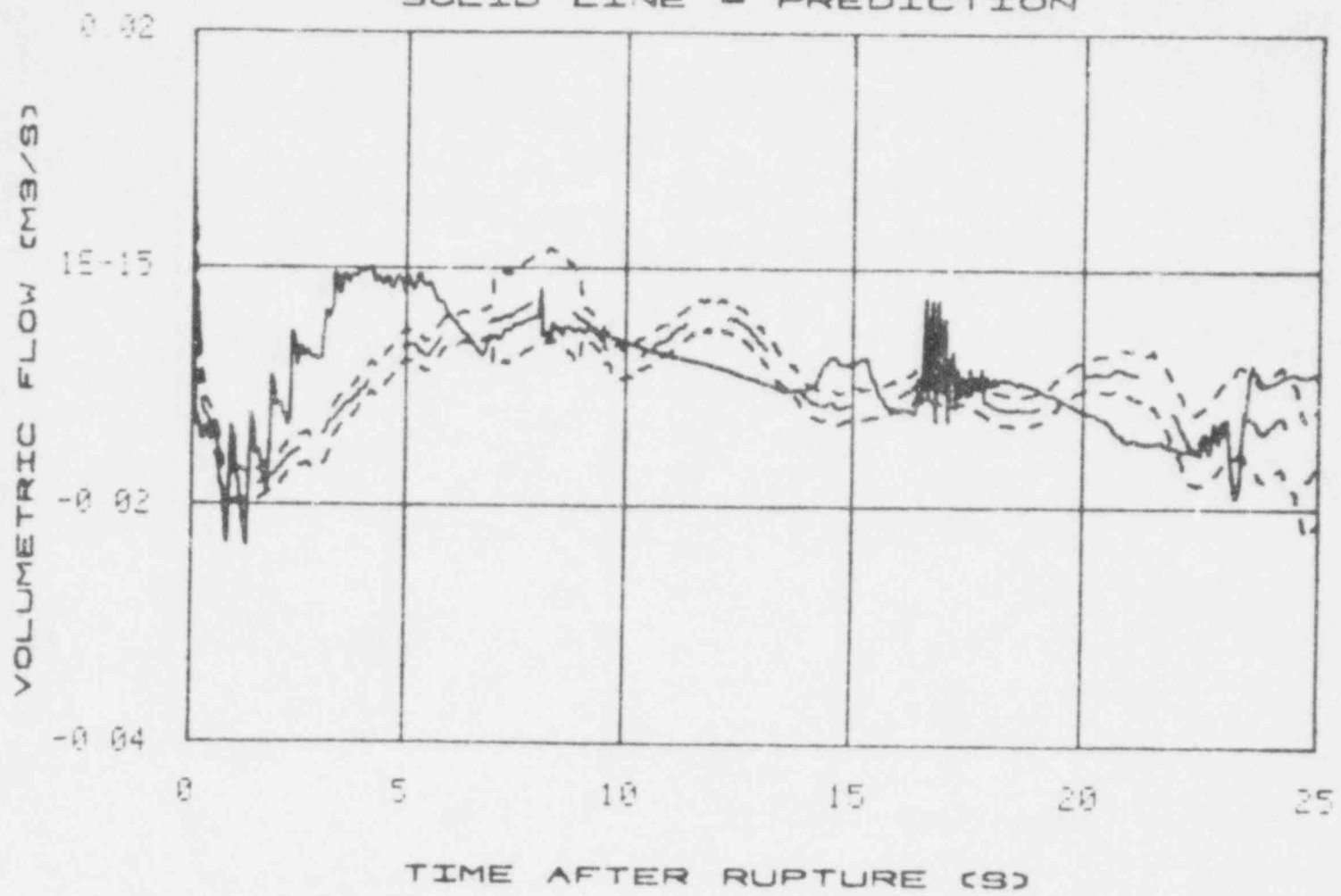


VOLUMETRIC FLOW -- CORE INLET
TEST 9-04-B

Figure 14

31
520 301

DOTS - RANGE DASH LINE - DATA
SOLID LINE - PREDICTION



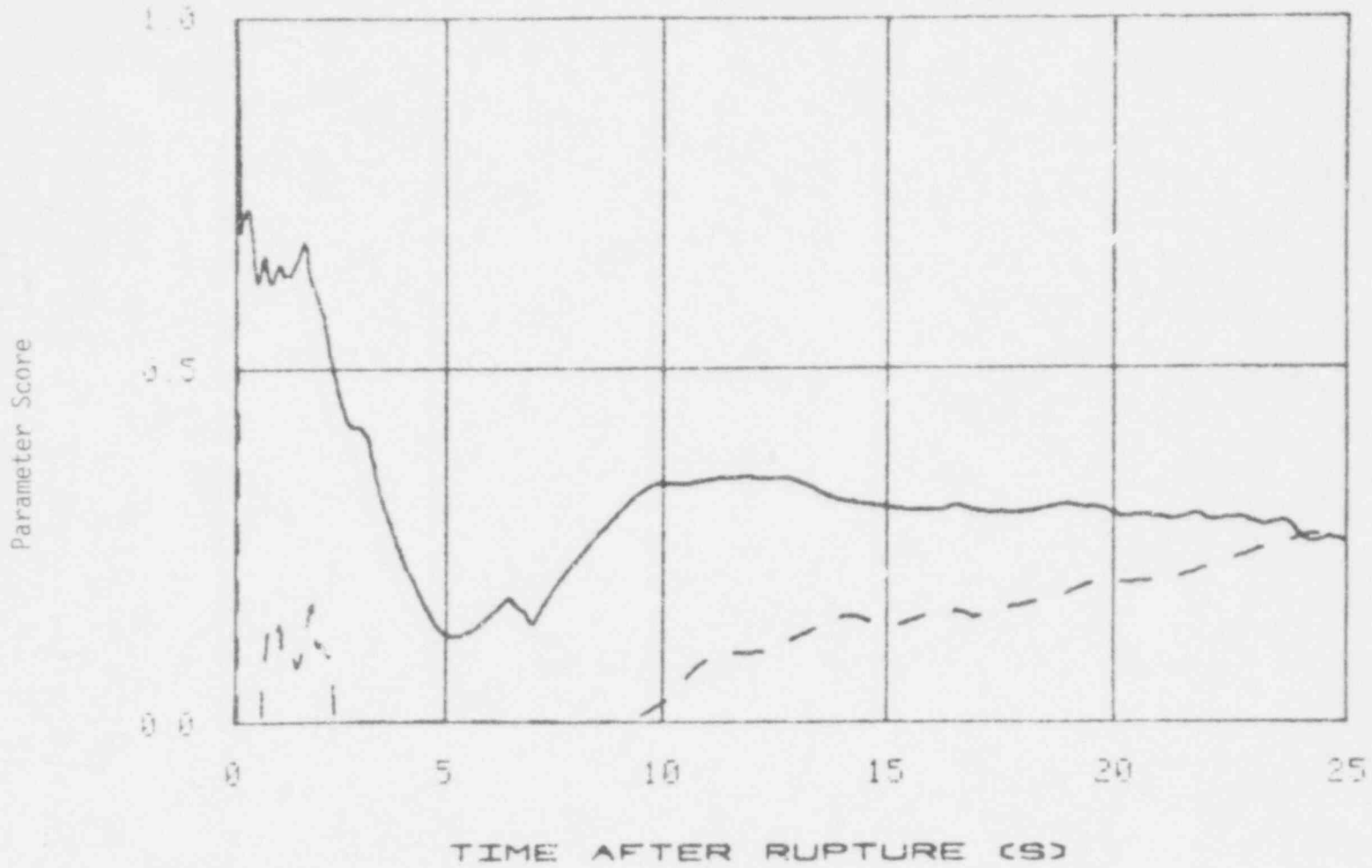
VOLUMETRIC FLOW -- CORE INLET
TEST S-08-8

Figure 15

32

520 302

SOLID LINE - S-04-B DASH LINE - S-08-B



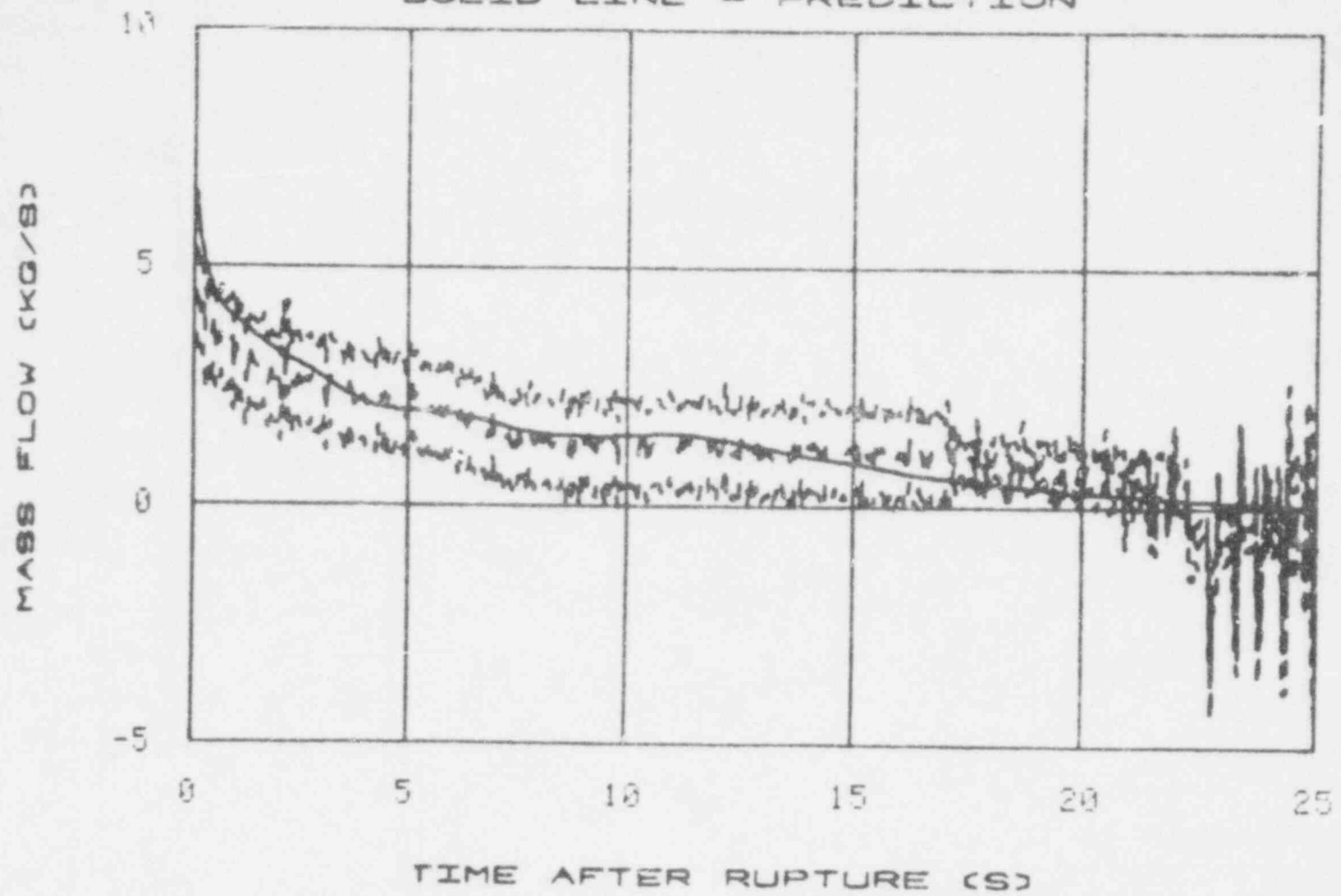
VOLUMETRIC FLOW -- CORE INLET

Figure 16

520 303

Parameter Score

DOTS - RANGE DASH LINE - DATA
SOLID LINE - PREDICTION



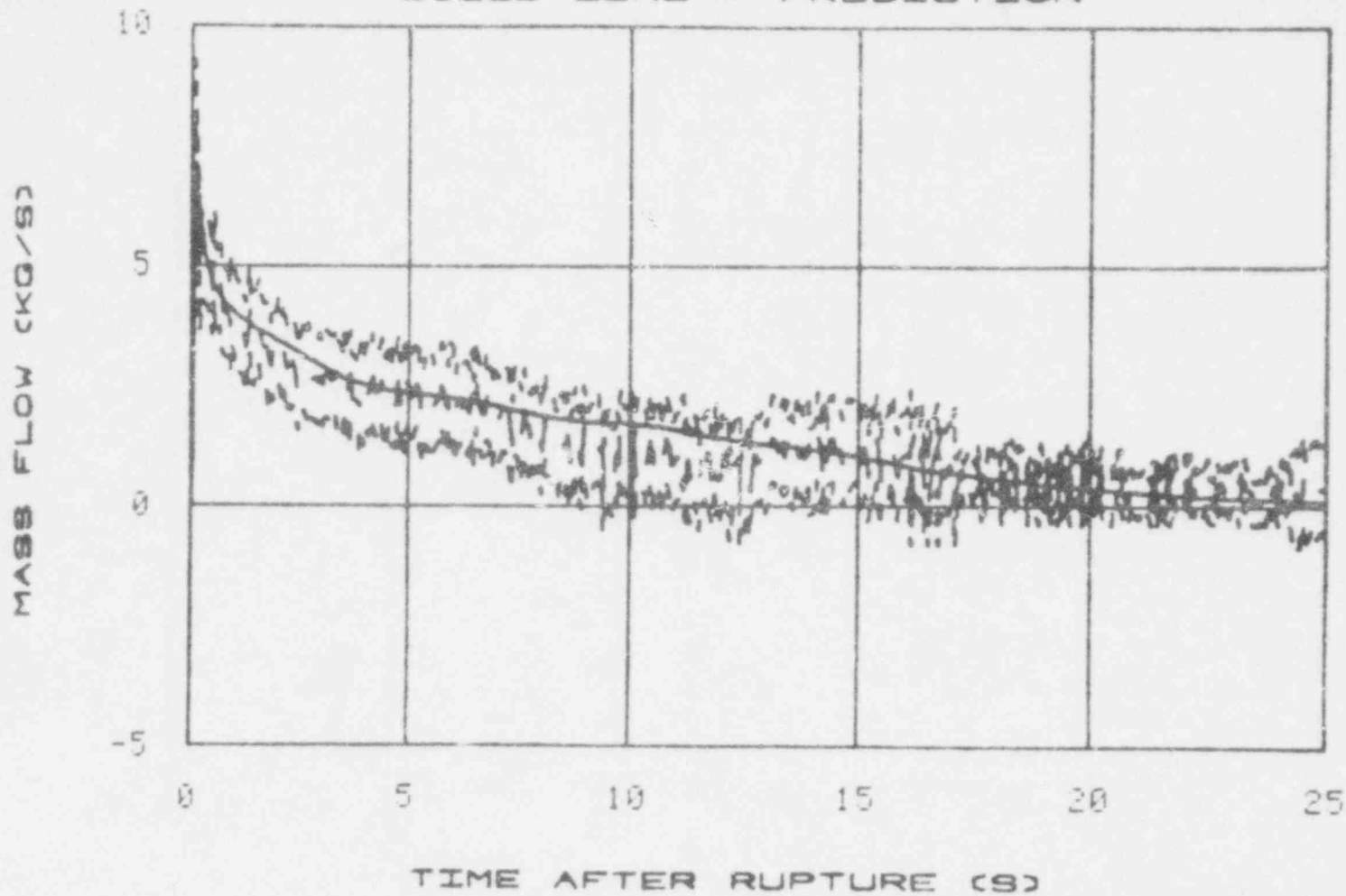
MASS FLOW -- PUMP SIDE BREAK
TEST 9-04-8

Figure 17

34

520 30

DOTS - RANGE DASH LINE - DATA
SOLID LINE - PREDICTION



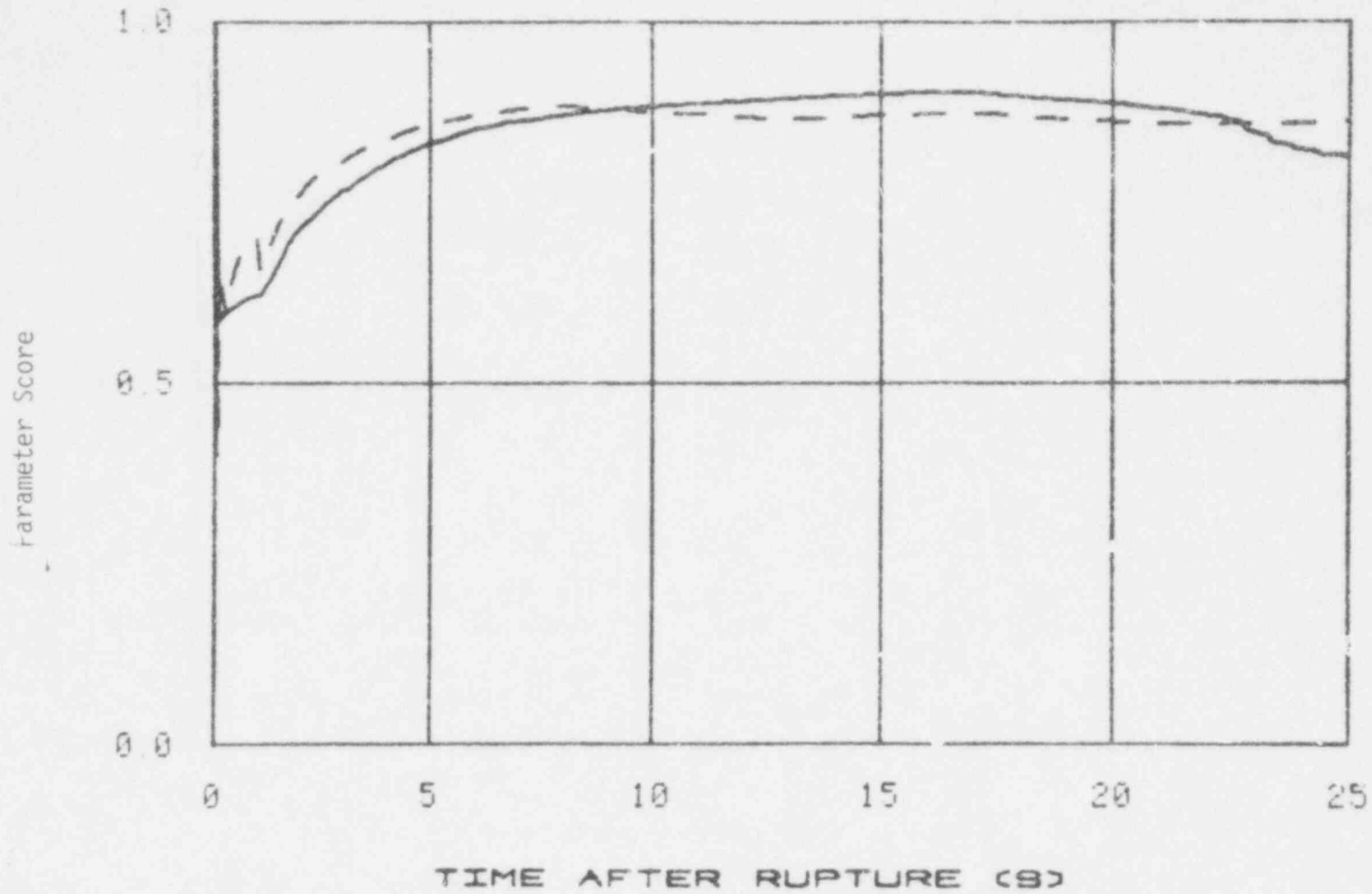
MASS FLOW -- PUMP SIDE BREAK
TEST 9-08-8

Figure 18

35

520 305

SOLID LINE - S-04-B DASH LINE - S-08-B



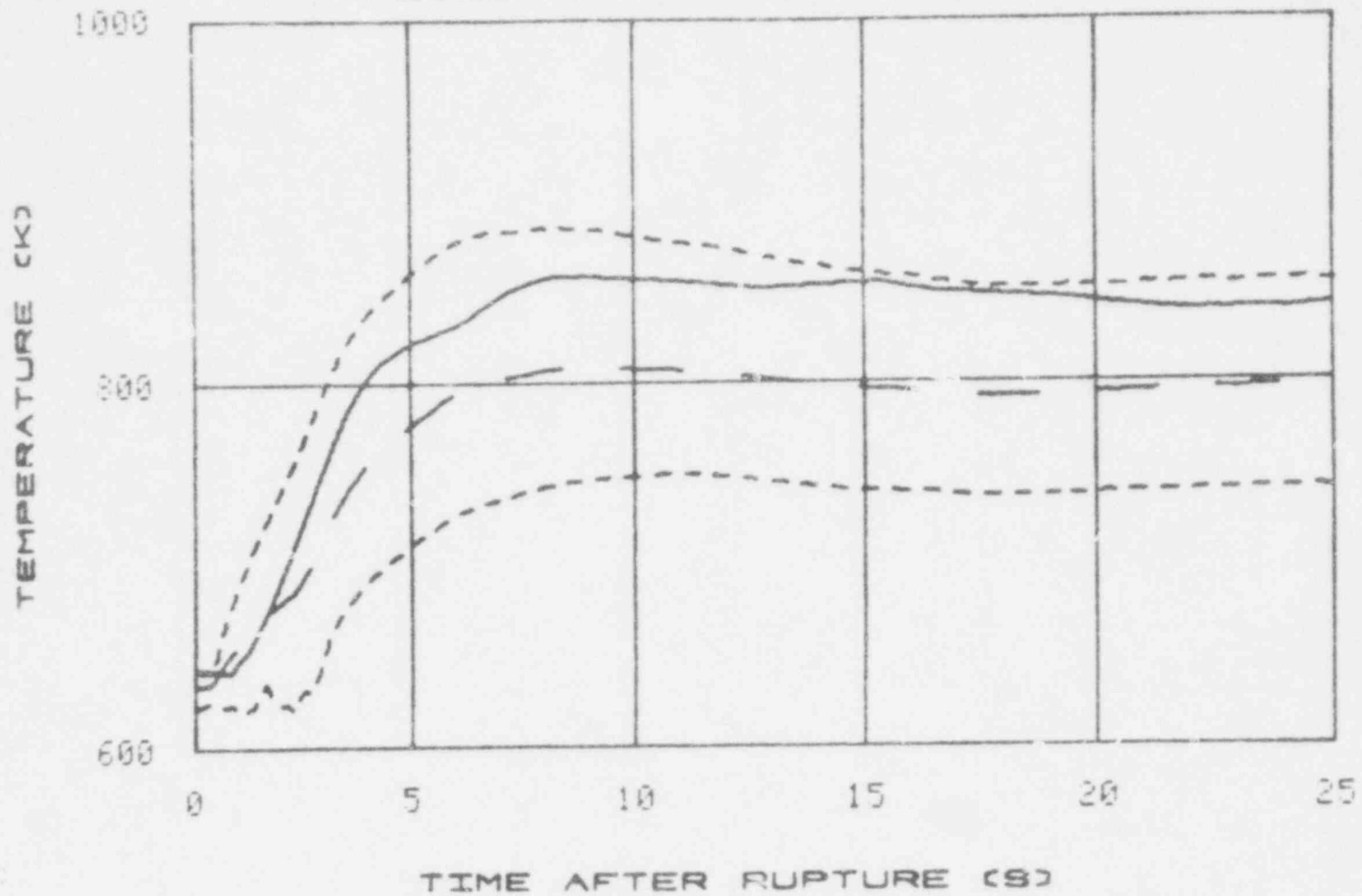
MASS FLOW -- PUMP SIDE BREAK

Figure 19

96

520 306

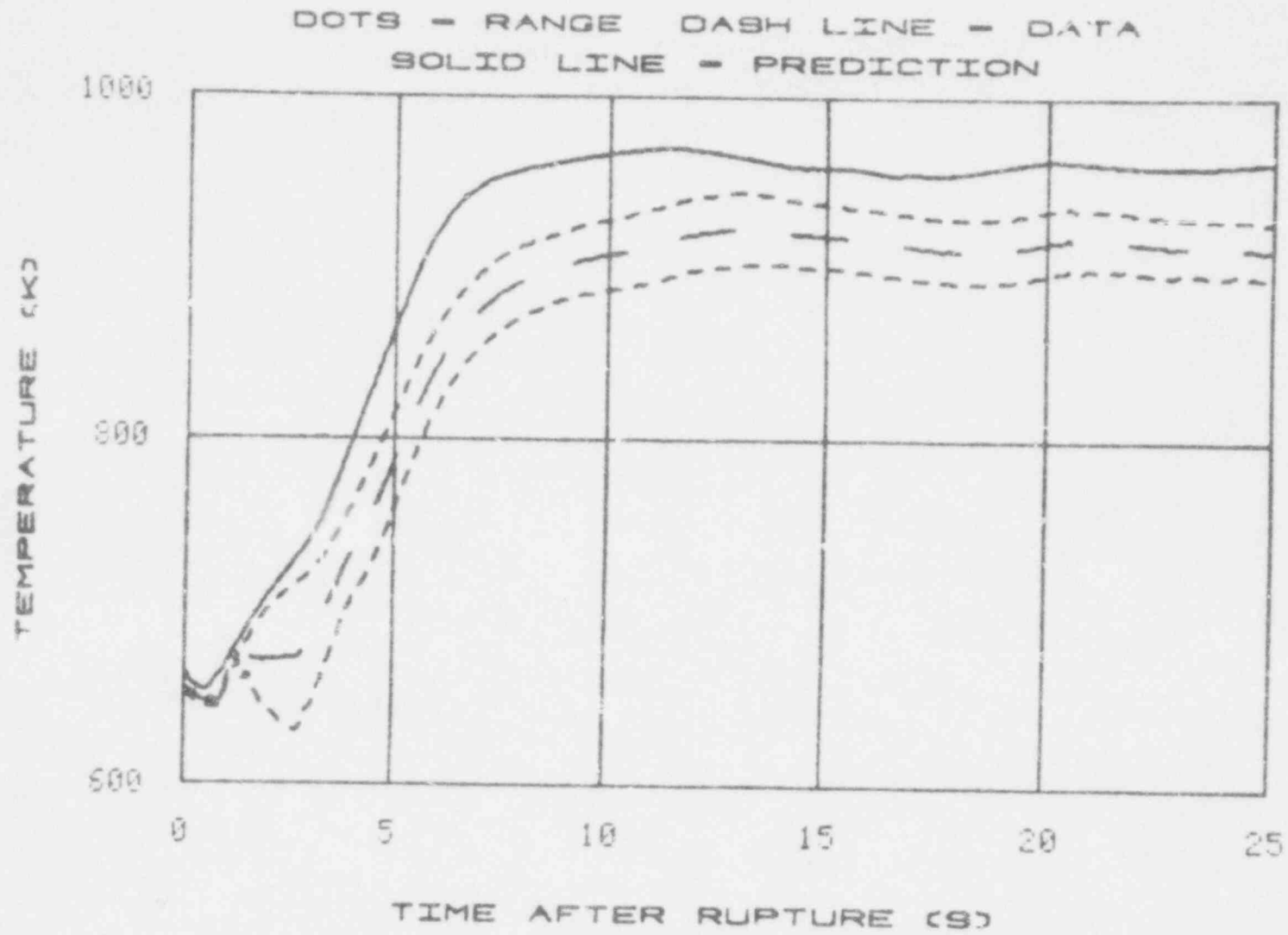
DOTS - RANGE DASH LINE - DATA
SOLID LINE - PREDICTION



CLAD TEMPERATURE -- LOWER CORE
TEST S-04-B

Figure 20

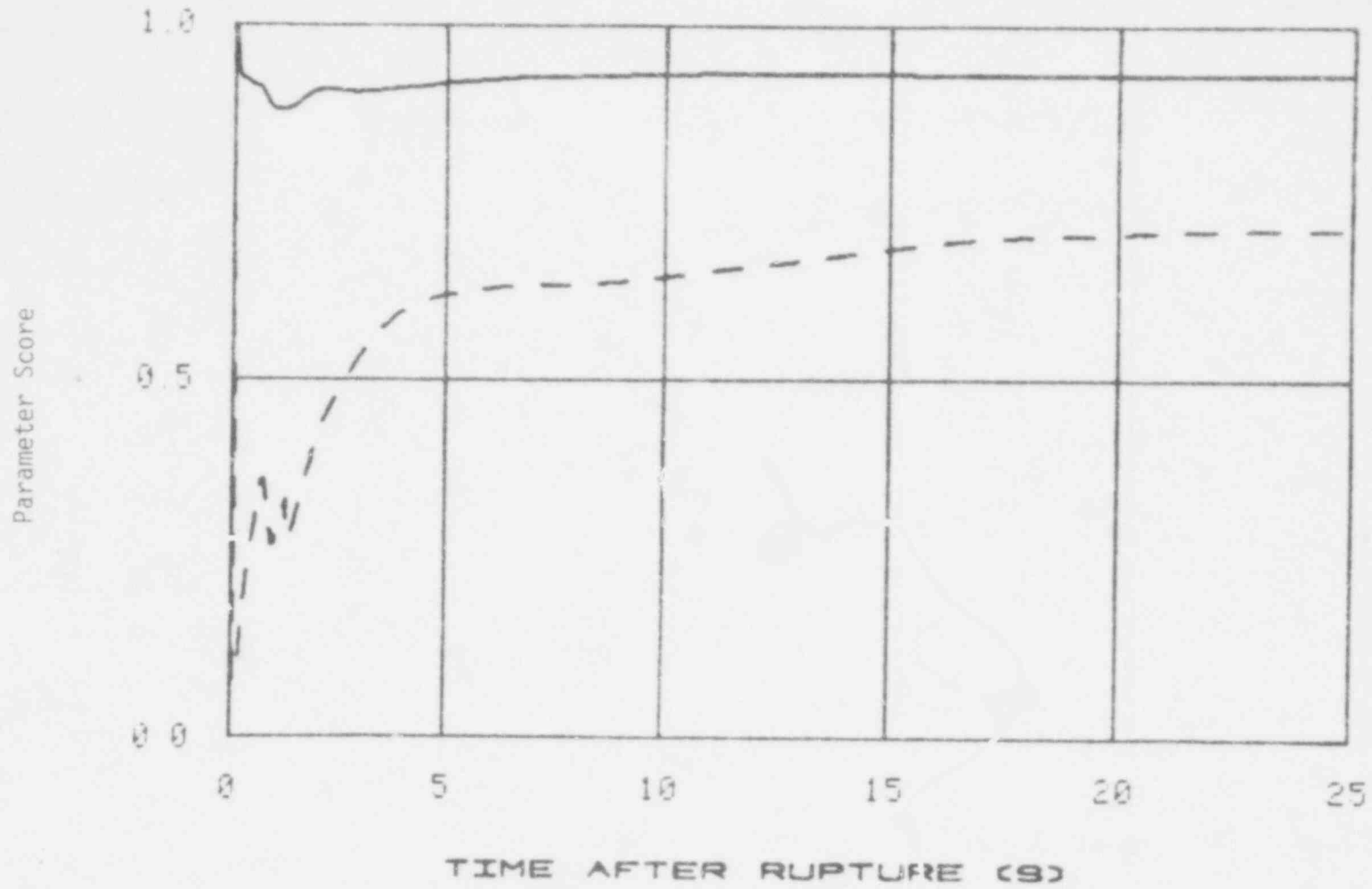
37
520 307



38

520 508

SOLID LINE - 9-04-B DASH LINE - 9-06-B



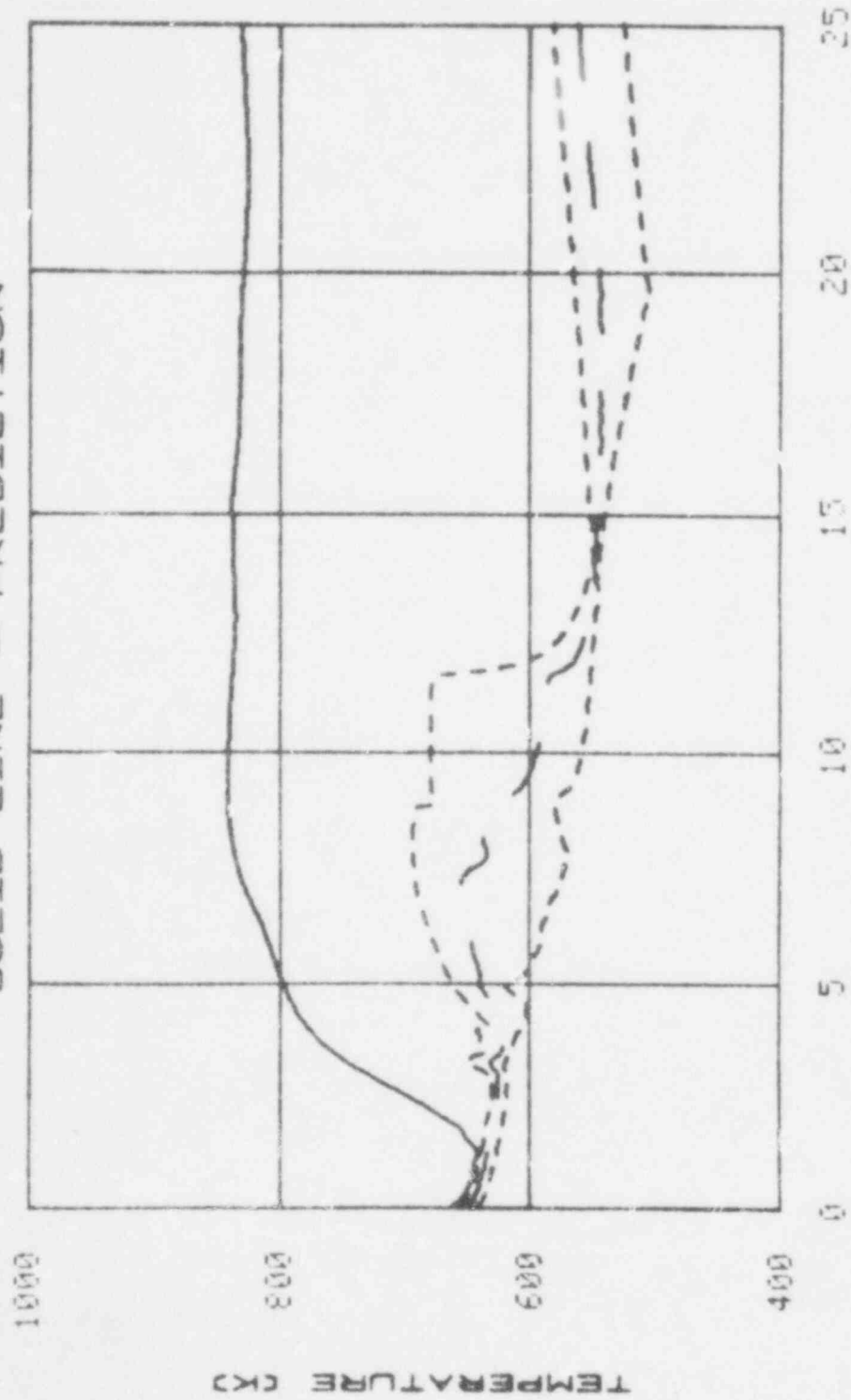
CLAD TEMPERATURE -- LOWER CORE

Figure 22

39

520 309

DOTS - RANGE DASH LINE - DATA
SOLID LINE - PREDICTION



CLAD TEMPERATURE -- UPPER CORE

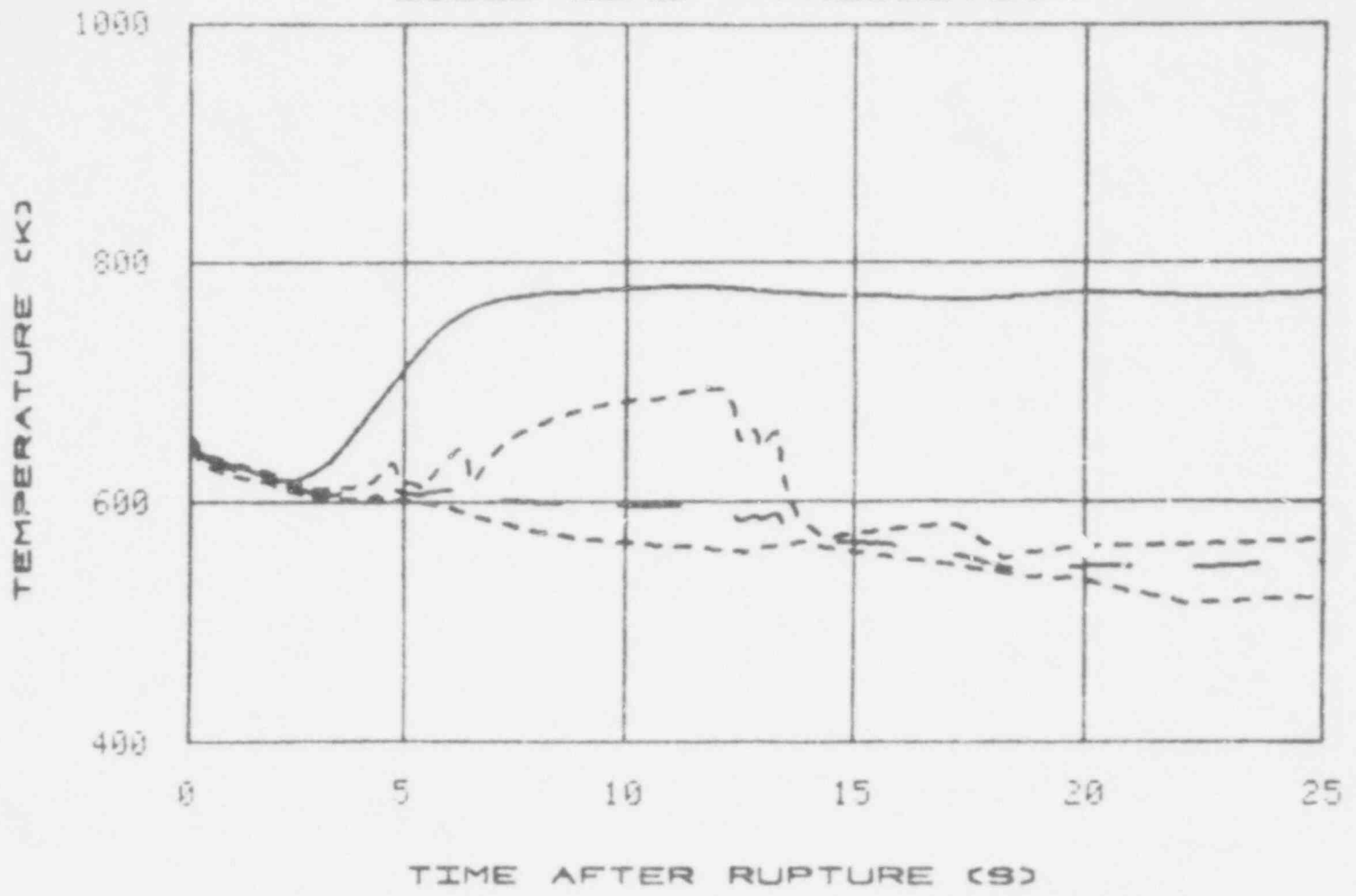
TEST 9-04-B

Figure 23

40

520 310

DOTS - RANGE DASH LINE - DATA
SOLID LINE - PREDICTION

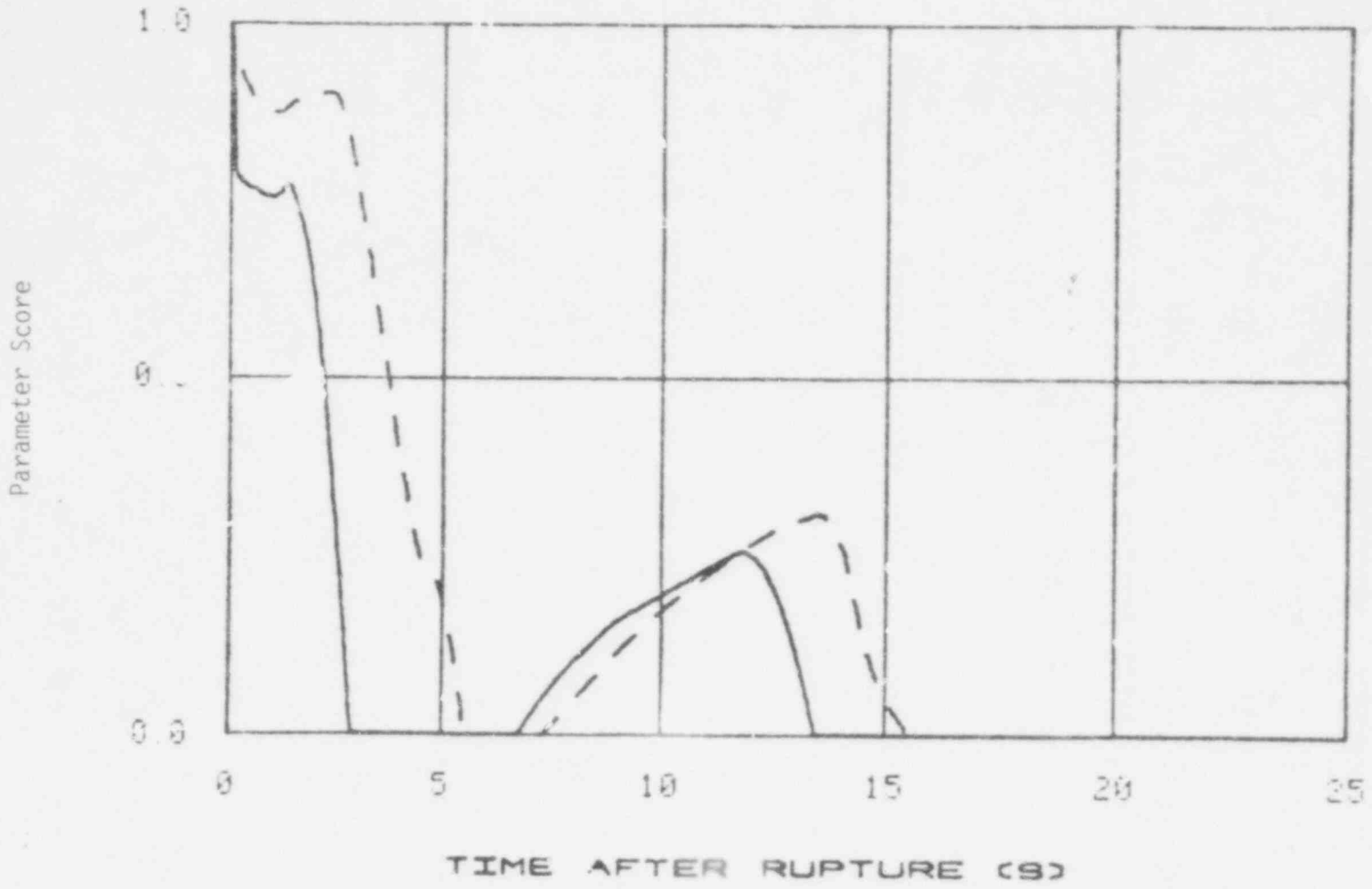


CLAD TEMPERATURE -- UPPER CORE
TEST 9-08-8

Figure 24

41
520 311

SOLID LINE - 9-04-8 DASH LINE - 9-08-8



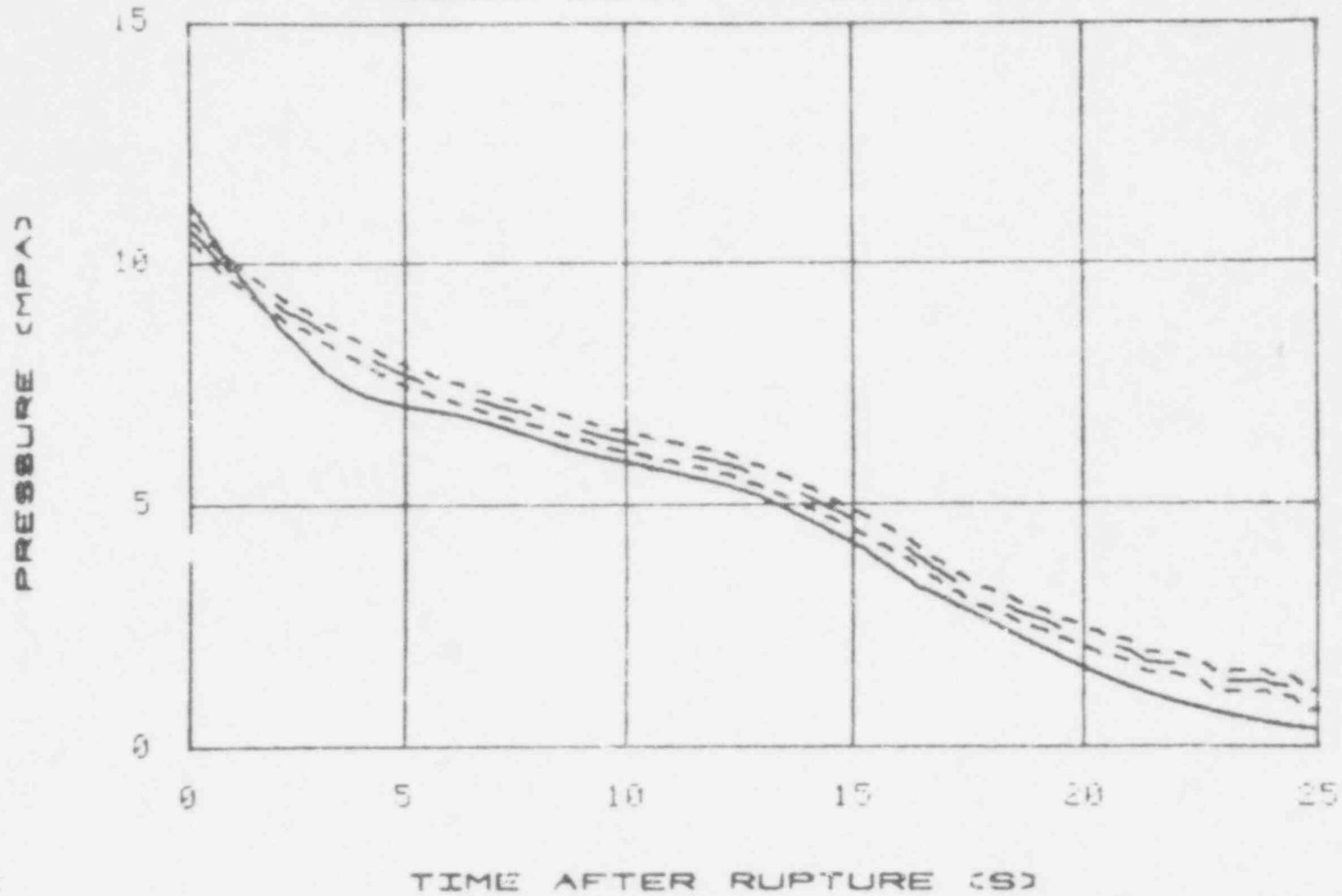
CLAD TEMPERATURE -- UPPER CORE

Figure 25

42

520 312

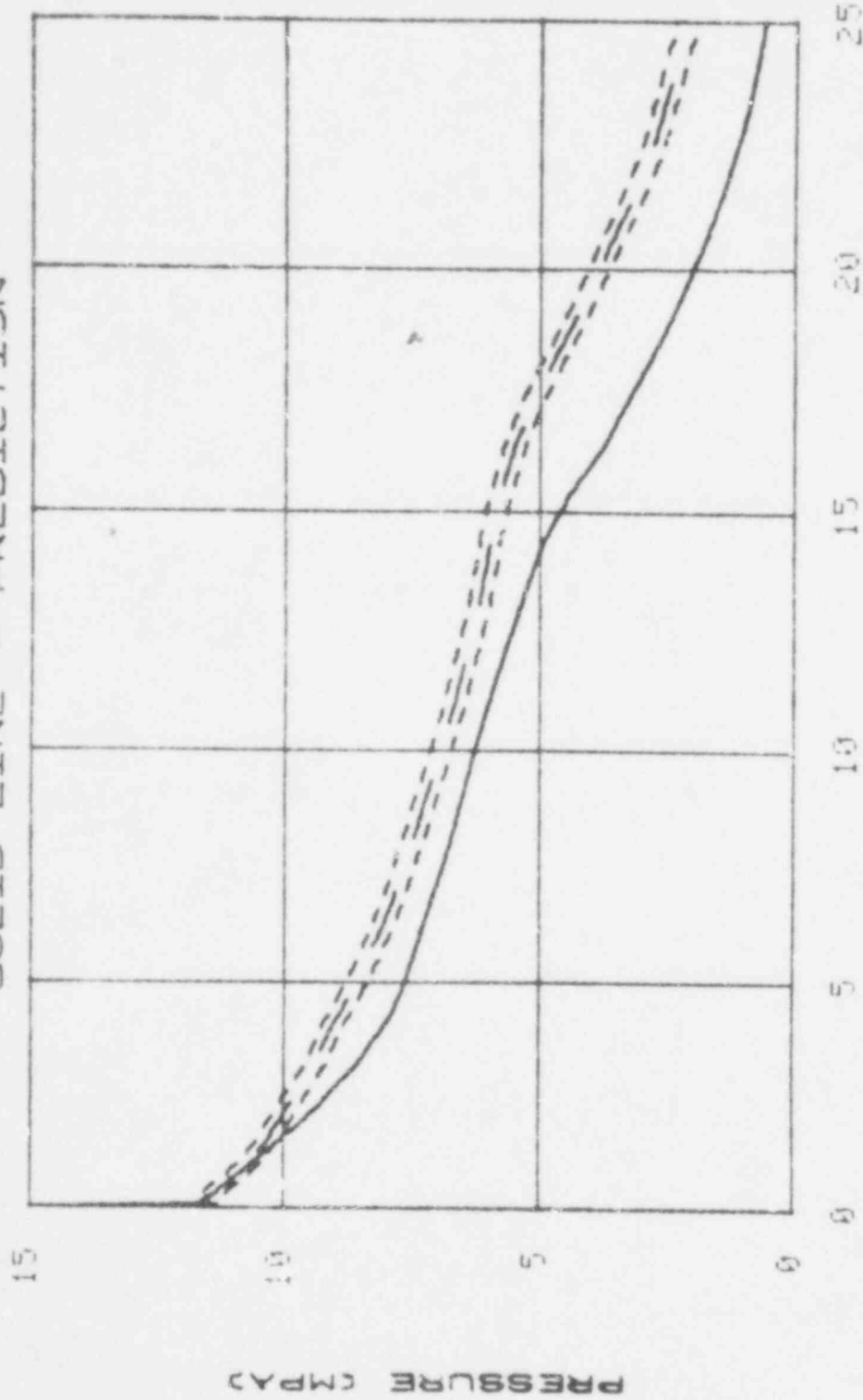
DOTS - RANGE DASH LINE - DATA
SOLID LINE - PREDICTION



PRESSURE -- UPPER PLENUM
TEST 9-04-8

Figure 26

DOTS - RANGE DASH LINE - DATA
SOLID LINE - PREDICTION

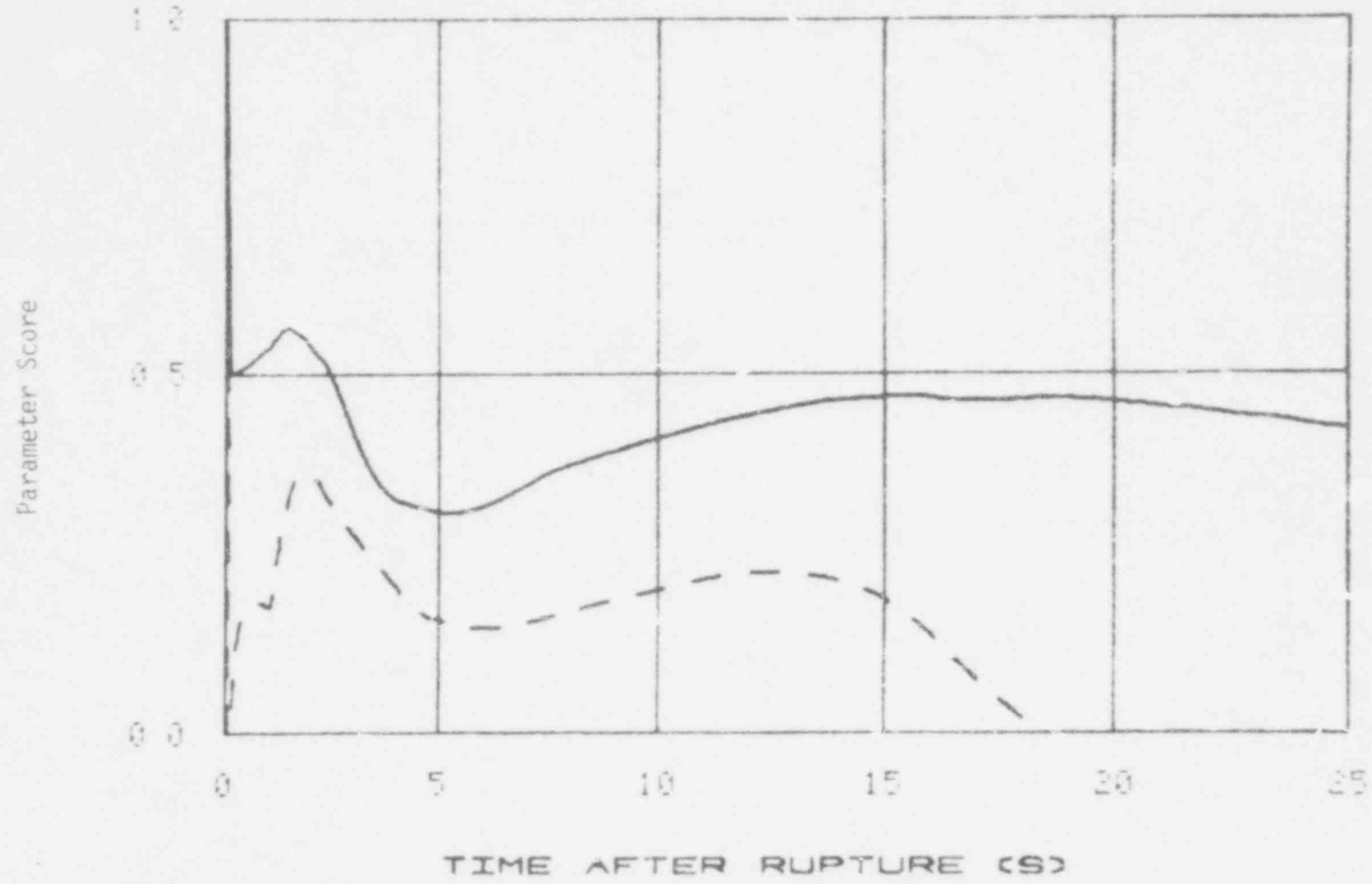


TIME AFTER RUPTURE (S)

PRESSURE -- UPPER PLENUM
TEST 9-08-8

Figure 27

SOLID LINE - 9-04-B DASH LINE - 9-08-B

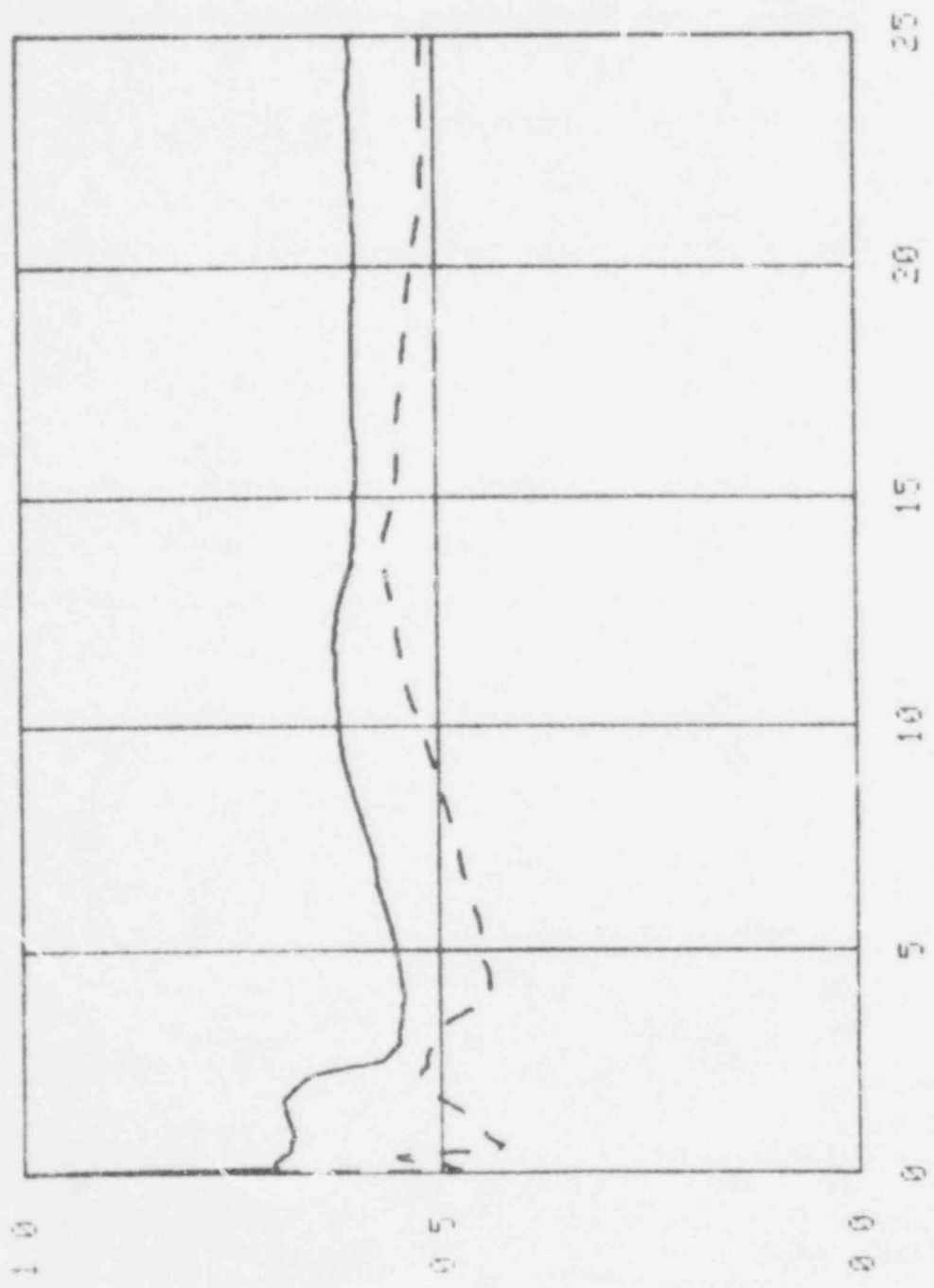


PRESSURE -- UPPER PLENUM

Figure 28

45
520 315

SOLID LINE - 8-04-0 DASH LINE - 9-08-0



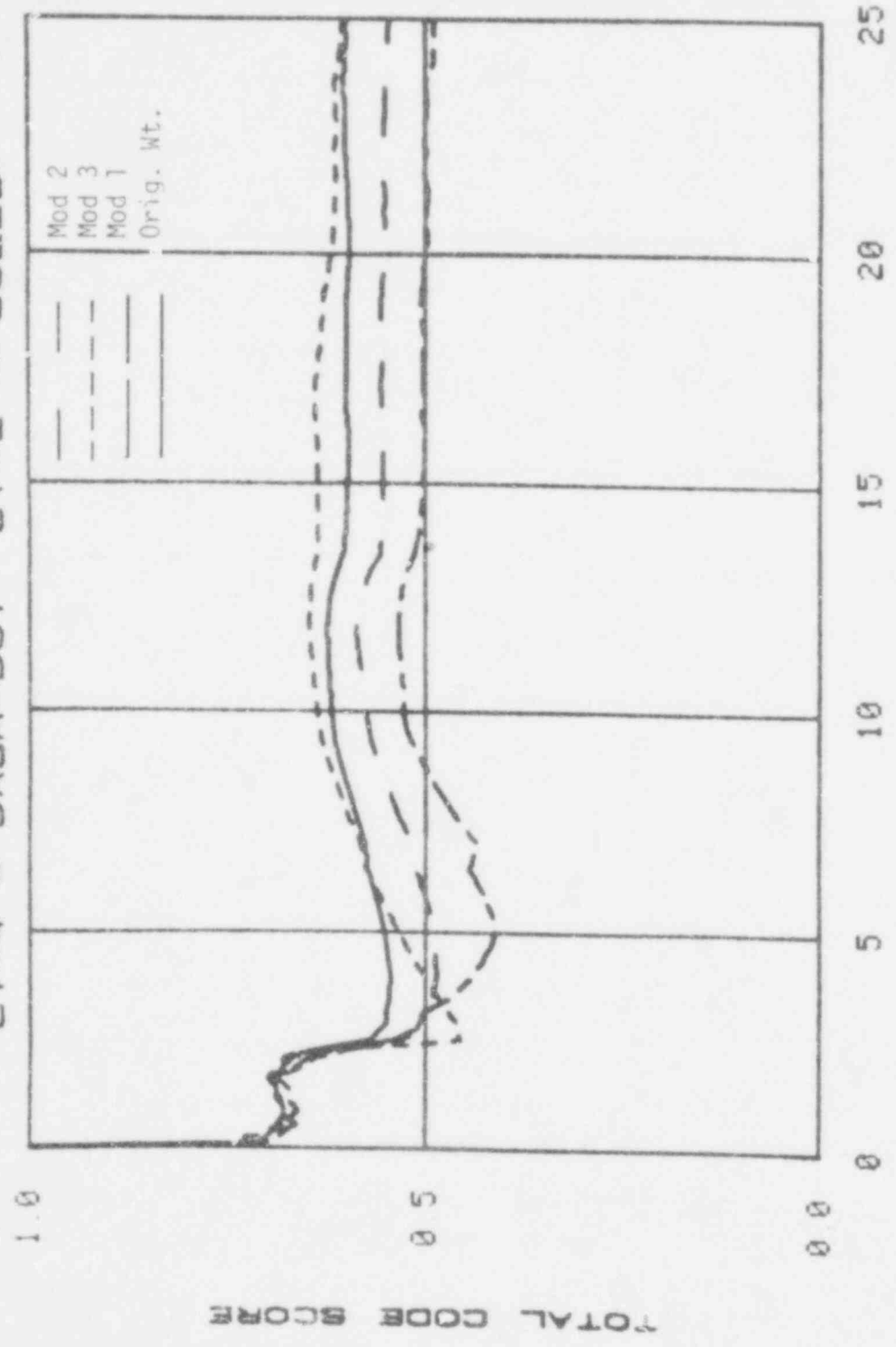
TIME AFTER RUPTURE (S)

Figure 29

Total Calculation Score

520 316

CY-6 - DOTS CY-2 - DASH
 CY-4 - DASH-DOT CY-1 - SOLID



TIME AFTER RUPTURE (S)

Figure 30

TOTAL CODE SCORE

APPENDIX A

520 318

QUANTITATIVE ASSESSMENT OF CODES
AND
ACCEPTANCE CRITERIA

by

J. A. Dearien

November 7, 1978

EG&G Idaho

520 319

CODE ACCEPTANCE CRITERIA

i. INTRODUCTION

The purpose of this document is to present a procedure for the quantitative assessment (QA) of code accuracy and development of an acceptance criteria (AC) based on the quantitative assessment.

The procedure for both QA and the AC of the code is based on the fact that data used in the evaluation process has an inherent scatter and an exact comparison between data and prediction will never be possible. Therefore, any procedure used to evaluate the code and any criteria used in judging the acceptability of the code must be in the form of a numeric/subjective evaluation. This document presents some representative data from the fuel area and the thermal-hydraulic area (Section 2), a discussion on how data exhibiting inherent scatter can be used to quantify errors in the code calculation (Section 3) and a proposed method for overall quantification of code accuracy (Section 4). Section 5 is a summary which describes certain fine points of the procedure and what can be gained from the procedure. Attachment 1 to this document is a draft standard practice directed toward obtaining the experimental data needed for this procedure. Attachment 2 to this document is a brief discussion of a similar procedure in which a combination of numerical and subjective approaches are used to evaluate options for a problem solution.

2. SCATTER IN REACTOR BEHAVIOR DATA

The data on which we base the development and evaluation of our reactor safety codes exhibit a great deal of scatter. This scatter is due to many things (see attachment 1) and is more than just instrument error. It is not the purpose of this document to perform an in-depth analysis of the error, only recognize that it exists and deal with it as part of the procedure.

The data we deal with have quite a range of variance, depending on the phenomena and its pertinent parameters such as time, temperature, flow regime, etc. Figure 1 shows the measured centerline temperature of a large number of fuel rods as a function of power level. The reader will note that the logical trend of increased centerline temperature with increased power is evident but that the range of temperatures for a given power level is on the order of 500-600 K. Since a fuel code analyzing a rod at one power would be expected to produce only one value for centerline temperature, it is obvious that some account must be taken for the real world in which the code is trying to compute. The $\pm 2\sigma$ lines on Figure 1 define the 2σ , 95% confidence limits of the data and will be addressed in the following section.

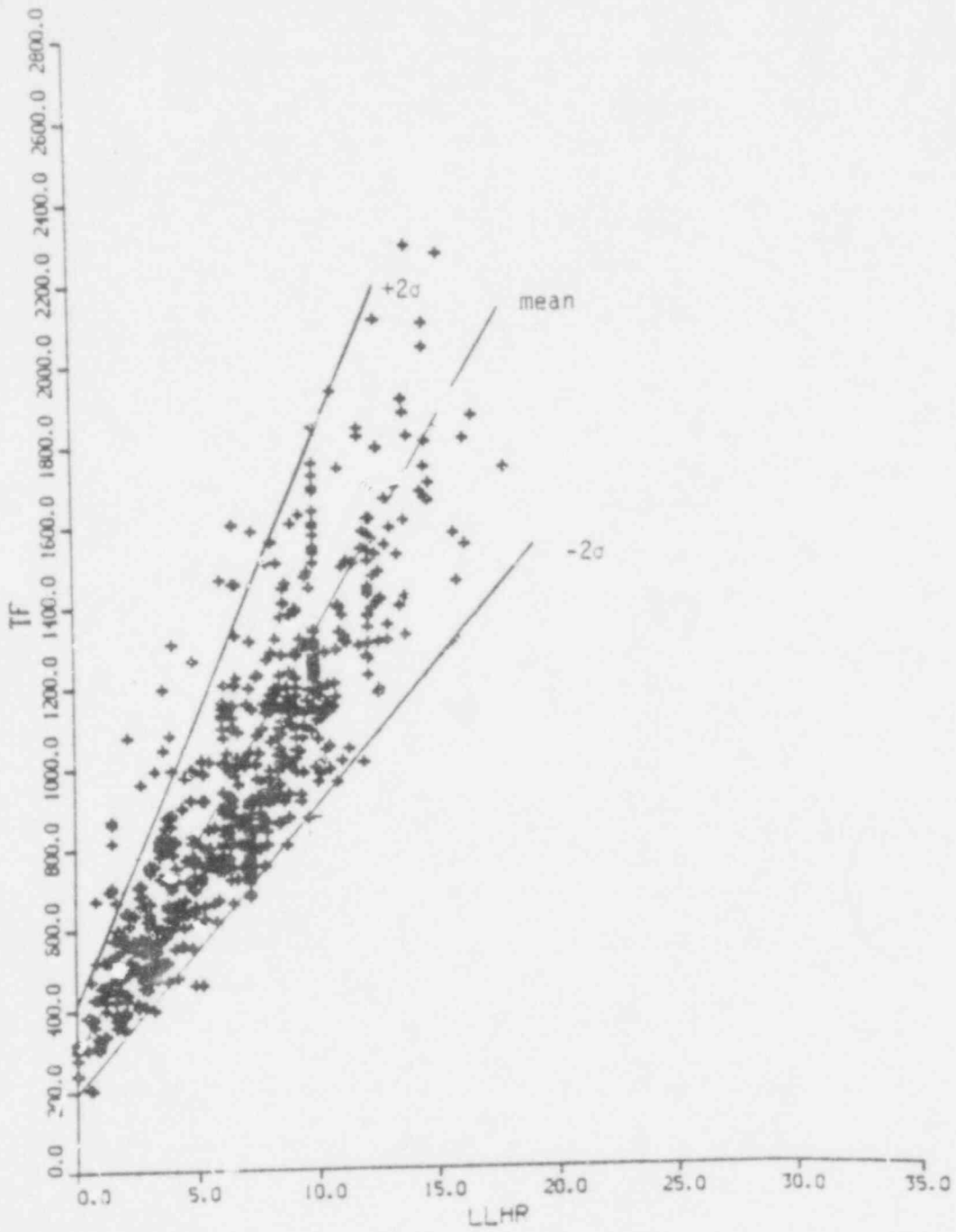
Figure 2 is a similar plot showing a typical set of thermocouple traces from THTF Test . . . approximate 2σ lines have been added to the data along with the code prediction of this test. Further comment on this plot and its use in a QA and AC of the code will be discussed in the following section.

Figures 1 and 2 are only two examples of typical data used in code assessment work and the scatter associated with that data. One has only to survey the literature to see that there are data with tighter bands (pressure decay during blowdown) and far broader bands (zircaloy burst strain and burst at high temperature). A procedure for dealing with this data behavior and arriving at a quantitative assessment of code accuracy is discussed in the following section.

3. QUANTITATIVE ASSESSMENT OF CODE ACCURACY

The objective of this section is to identify several different methods for evaluation of data/calculation comparisons when one calculation exists and a number of "equally applicable" data points exist. Section 3.2

TF VS LLHR



11-28-14 TUES 31 OCT. 1978 08- IL DISPLA REP 6.8

Fig. 1 Fuel centerline temperature at a 3 mil gap

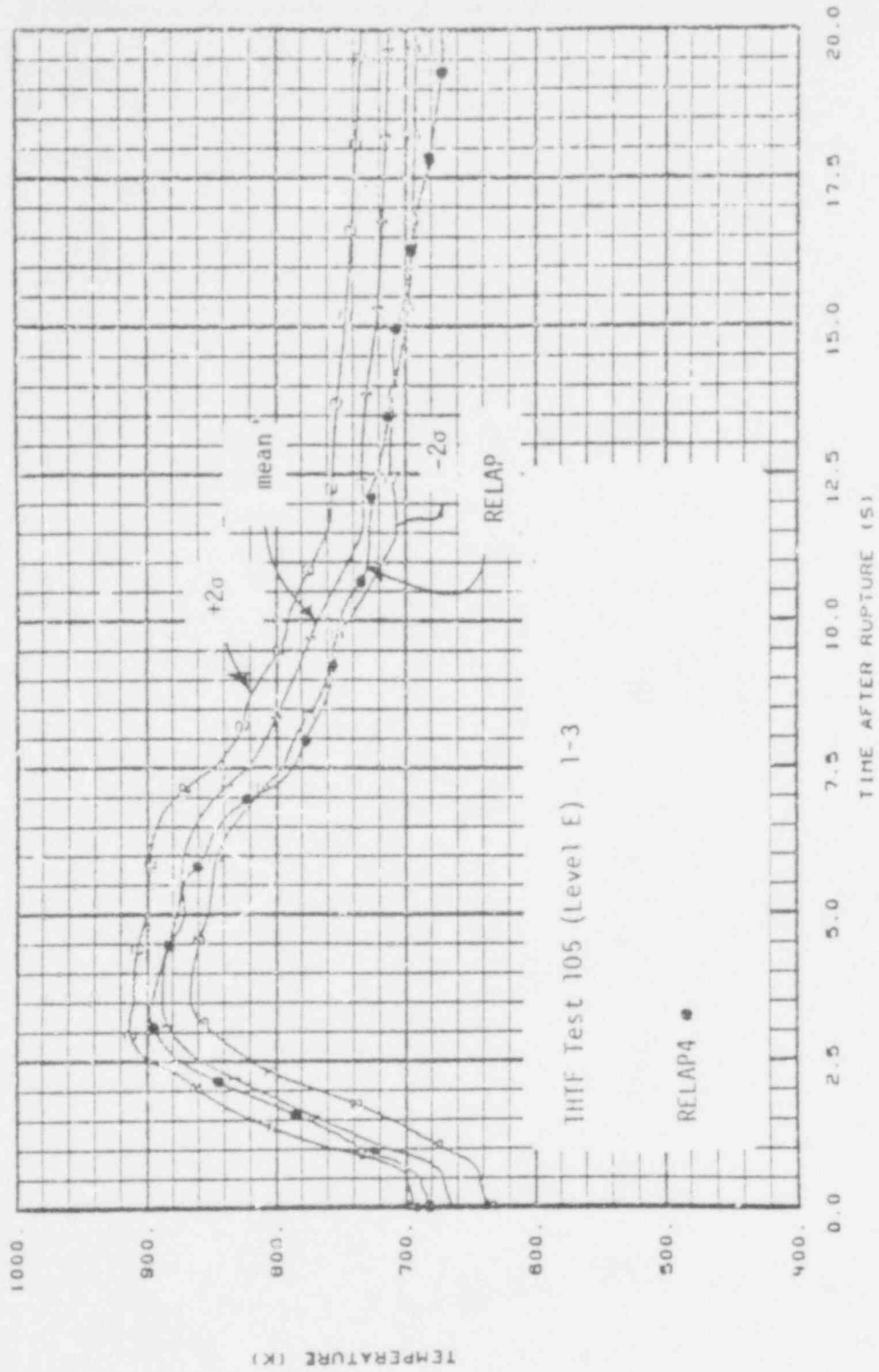


Fig 2 Clad temperatures at 1.09 m in the THF core.

discusses the situation where the number of "equally applicable" points is too small for a good statistical evaluation. Quantitative evaluation of the code using these evaluation criteria is covered in Section 4.

The procedure described in this document deals with a single prediction or single prediction "history" and a distribution of data. The basic procedure is equally applicable to situations where a distribution of predictions are available (some results of this nature are now available from the fuel codes and are being developed at Sandia for RELAP). Modifications to the assessment relations of Section 3.1 can be easily made to quantify the comparisons between data spread and prediction spread for this situation.

3.1 "Twenty to One You're Right"

The basic premise of this title, and this section, is that if the code is calculating answers that fall within the 2σ bands of the data, you are getting answers that have less than a 5% probability of being wrong. This in itself is not totally sufficient for acceptance of a code for two reasons, 1) if the code consistently predicts a response near a 2σ band of the data, a bias in the calculation is evident and further development is probably required and 2) if the code occasionally goes beyond the 2σ limits, it may still be producing overall integral response which is acceptable.

Use of the 2σ bands on the data as an evaluation criteria has the logical basis of recognizing the inherent behavior of the phenomena through the data, and not requiring the code to be significantly more precise than nature. The mechanics of this evaluation procedure is, therefore, to calculate the 2σ limits on the experimental data and check the calculation for behavior within these limits. The quantitative assessment of the calculation is not made on whether or not the calculation remains within the $\pm 2\sigma$ data bands but the manner in which the calculation behaves in this regime. This behavior of concern is relative to the two items described above and is discussed below in more detail.

3.1.1 Calculational Bias in the 2σ Region. The ideal behavior for a best estimate calculation is for it to split the data band and if not overlay the mean of the data, at least remain close to it. The degree to which this is obtained can be measured quite readily by the following relations;

$$A = 1 - \frac{1}{a_2 - a_1} \int_{a_1}^{a_2} \frac{(P-M)}{(d_{+2\sigma} - d_{-2\sigma})} \quad 1$$

$$B = 1 - \frac{1}{a_2 - a_1} \int_{a_1}^{a_2} \frac{(P-M)}{(d_{+2\sigma} - d_{-2\sigma})} \quad P > M \quad 2$$

$$C = 1 - \frac{1}{a_2 - a_1} \int_{a_1}^{a_2} \frac{(M-P)}{(d_{+2\sigma} - d_{-2\sigma})} \quad M > P \quad 3$$

where M is the mean of the data, P is the predicted parameter and $d_{+2\sigma}$, $d_{-2\sigma}$ are the $\pm 2\sigma$ limits of the data. The integration limits a_1 , a_2 represent the parametric range over which the evaluation is to be made. This range could be, for example, time (beginning to end of blowdown), temperature (normal operating to cladding burst) or length (bottom of core to hot spot). Equation 1 gives the integral difference between mean data and predictions and equation 2 and 3 reflect any oscillatory behavior in the comparisons. These three factors give the assessor an indication of the bias and the timewise behavior of the calculation and can be factored into an overall evaluation. The values of these integrals for Figure 2 are 0.77, 0.94, and 0.71 respectively.

The assessment relations of equations 1-4 are examples of linear additive models. These are but one example of functions which can be used to quantify the behavior of a prediction relative to applicable

data. Other relations, such as those involving root-mean-square calculations, 1σ limits vs. 2σ limits are equally applicable in the quantifying of data/prediction comparisons.

3.2.1 Excursion Beyond the 2σ Region. Many types of data and code calculations can be expected at some time to show spikes or deviations from the expected or desired path. The importance of these deviations depends on both magnitude and duration. Quantification of these excursions beyond the 2σ region is obtained with the following relation

$$D = 1 - \frac{1}{a_2 - a_1} \int_{a_1}^{a_2} \frac{(P - d_{+2\sigma})}{(d_{+2\sigma} - d_{-2\sigma})} [P > d_{+2\sigma}] \, da - \frac{1}{a_2 - a_1} \int_{a_1}^{a_2} \frac{(d_{-2\sigma} - P)}{(d_{+2\sigma} - d_{-2\sigma})} [d_{-2\sigma} > P] \, da \quad 4$$

This relation includes both biased deviation from the $\pm 2\sigma$ region and oscillatory behavior. It is not deemed necessary to separate the two since any significant deviation from the $\pm 2\sigma$ range is not desirable.

3.1.3 Single Value Comparisons. Section 3.1.1 and 3.1.2 dealt with the procedure for quantifying data/prediction comparisons of a transient nature or over a range of another variable. It is often required that an assessment be made on how well a point phenomena is calculated, such as peak clad temperature, time to DNB, rod burst, etc. Two relations are defined to indicate this degree of fit,

$$E = 1 - \left| \frac{X - Y}{Z_{+2\sigma} - Z_{-2\sigma}} \right| \quad 5$$

$$F = 1 - \left| \frac{Z_{2\sigma} - Y}{Z_{+2\sigma} - Z_{-2\sigma}} \right| \quad 6$$

where

- Y = calculated parameter
X = mean of data parameters
 $Z_{2\sigma}$ = 95% confidence limits on data parameters.

The reader will note that it is possible to obtain values for all the relations from Figure 2. By selective weighting of the individual assessment relations, it is possible to emphasize the importance of, say, calculating a conservative peak clad temperature as opposed to calculating a less than conservative value.

3.2 "What to do Until the Data Comes"

The data in Figures 1 and 2 are typical in that when large amounts of data are available on a particular parameter, the data show a considerable scatter. The figures are atypical in that many of the parameters for which we are concerned have little and sometimes no data. It is just not possible to have as many flow meters or gamma densitometers as thermocouples. Yet it is necessary to make some sort of evaluation of the accuracy of the code in predicting these parameters.

The point to be made and recognized in this section is that you cannot realistically evaluate the code against a criteria that is more restrictive than your understanding of the data. For example, if only two measurements are taken on flow (actual case where orag disk reads 7 kg/s and turbine meter reads 10 kg/s) the 95% confidence limits can be exceptionally large (-10.5 to 27.5 kg/s for the above case). A procedure that can produce seemingly ridiculous results, like the above, can be used to an advantage, however, in highlighting areas that 1) should have more data taken and 2) should have special attention paid when selecting acceptance criteria.

4. QUANTIFICATION OF CODE ACCURACY

The procedure for quantification of code accuracy (and therefore acceptance) is based on the selection of a number of key parameters, deriving a value(s) by the methods of Section 3 for how well the parameter is calculated and combining the calculated parametric values (with weighting factors) into a total value for the code (TCV). Subjective analysis of the TCV for various problem types then gives a level of acceptance for the code. This section discusses key parameter selection, acceptance matrix formulation, weighting factors and subjective analysis procedure.

4.1 Key Parameters

The code acceptance procedure requires that a set of key parameters be selected and weighted for use in obtaining the TCV. The following list is selected as an example only.

Parameter#	Parameter	Weight
1	Peak Clad Temp	10
2	Temperature History	9
3	Critical Flow	8
4	Core Flow	7
5	Time to DNB	7
6	Rod Burst	6
7	Time to Turnover	2
8	Time to Quench	3
9	Stored Energy	1
10	Fission Gas Inventory	1
	Total Weight	54

The reader will note that the weighting does not have to be sequential, but is meant to be a relative weighting.

520 328

4.2 Acceptance Matrix

The purpose of the acceptance matrix is to sum the weighted and normalized values of the key parameters and arrive at a TCV that can be used as an indication of the code accuracy. Figure 3 illustrates a code acceptance matrix showing key parameters, assessment relations from Section 3 and the weighting factors applied to each assessment relation. The various matrix weighting factors are discussed below.

4.2.1 Acceptance Matrix Weighting Factors. There are a number of different weighting and normalizing factors in the acceptance matrix. This large number of factors is for the purpose of removing subjectiveness from the total evaluation and placing it in areas where only relative evaluations are required. The first set of factors are the Key Parameter Weighting Factors (KPWF) discussed in Section 4.1. The KPWF's are summed to give a total weight and this total weight is used as a normalizing factor to obtain a "perfect score" distribution for the key parameters. A value for the sum of the assessment relation weighting factors is obtained by normalizing the "perfect score" to the total weight of the KPWF's ($100/54 = 1.85$). This total is then subjectively apportioned to the individual assessment relations. These numbers (≤ 1.85) are shown in the right side of each matrix location. It is by choice of these weightings that the assessor emphasizes the relative importance of each of the individual response calculations of the code.

4.2.2 Calculational Procedure for Matrix Evaluation. The philosophy of the acceptance matrix is that if the code is perfect in the calculation of all key parameters, all the assessment relation entries would be 1.0 and the following calculational procedure would produce a score of 100% for the code.

The calculational procedure is to generate a value for each of the key parameter assessment relations and place this value in the left side of each matrix position. These values are then multiplied by their associated assessment relation weighting factors (right side of each matrix location) and summed horizontally to give a total key parameter

$$\text{Key Parameter Score} = W_{KP} \times \sum W_{AR} \times \text{Value}_{AR}$$

$$\text{Total Code Value (TCV)} = \sum \text{Key Parameter Scores}$$

Assessment Relations (AR) Key Parameter	A		B		C		D		E		F		1	2	3	4
	A Weight	B Weight	C Weight	D Weight	E Weight	F Weight	100 Total Weight	Key Parameter Weight	Key Parameter Score	Perfect Score						
1. Peak Clad Temperature	.8	.4	.7	.3	.85	.3	.92	.6	.5	.25			1.562	10	15.62	18.5
2. Temperature History	.9	.5	.6	.2	.65	.2	.7	.95					1.365	9	12.28	16.6
3. Critical Flow	.65	.3	.82	.6	.93	.6	.87	.35					1.549	8	12.4	14.8
4. Core Flow	.77	.65	.65	.4	.5	.4	.84	.4					1.28	7	8.86	13.0
5. Time to INB									.91	1.85			1.68	7	11.76	13.0
6. Rod Burst	.87	.6	.74	.6			.67	.65					1.43	6	8.4	11.1
7. Time to Turnover	.91	.4	.8	.3	.73	.3	.86	.85					1.55	3	4.65	5.55
8. Time to Quench	.85	.5	.92	.5			.75	.85					1.53	2	3.06	3.7
9. Stored Energy	.87	.45					.78	.4	.95	.5	.94	.5	1.65	1	1.65	1.8
10. Fission Gas Investigation									.88	1.85			1.63	1	1.63	1.8
															TCV	
													54		80.41	100

Example for Key Parameter #1 $.8 \times .4 + .7 \times .3 + .85 \times .3 + .92 \times .6 + .9 \times .25 = 1.562$

$$\frac{\text{Perfect Key Parameter Weight}}{\text{Key Parameter Weight}} = \frac{100}{54} = 1.85$$

Fig. 3 Acceptance Matrix

POOR ORIGINAL

520 330

assessment weight (column 1). The key parameter assessment weight is then multiplied by the key parameter weighting factor (column 2) to give a key parameter score (column 3). The key parameter scores are then summed to give the total code value (TCV).

The sample acceptance matrix in Figure 3 has been filled out to show the procedure with numbers and results in a TCV of 80.41. The significance of this number and how it can be used is discussed in the following section.

4.3 Application of TCV for Code Acceptance

The TCV derived from the Acceptance Matrix is still a subjective value until there are known gauge points to judge the TCV against. We have two gauge points, 100% and 0%, but the significance of TCV's between these limits requires a disciplined evaluation. A procedure for establishing a more complete set of acceptance gauge points is described below.

4.3.1 Selection of Acceptance Gauge Points. There are data/prediction comparisons developed as part of the Independent Assessment of RELAP4/MOD6 that everyone would agree on as being "OK", "pretty good" or "not bad". These favorable comparisons are generally in the blowdown phase of the LOCA. There are also comparisons developed which elicit comments such as "that's terrible", "we've got to do better than that", or "the code stinks". Comparisons of this nature are generally found in certain reflood situations and temperature comparisons in the upper core.

The process of selecting significant TCV gauge points is to select a number of comparisons on which experts can formulate a subjective opinion and calculate the TCV for these comparisons. Figure 4 illustrates some hypothetical results from such analyses. Since the key parameters can be changed and their weighting factors varied, the flexibility of the evaluation procedure can be utilized to obtain a quantitative assessment of what has, to date, been a purely subjective evaluation. Once a

TCV

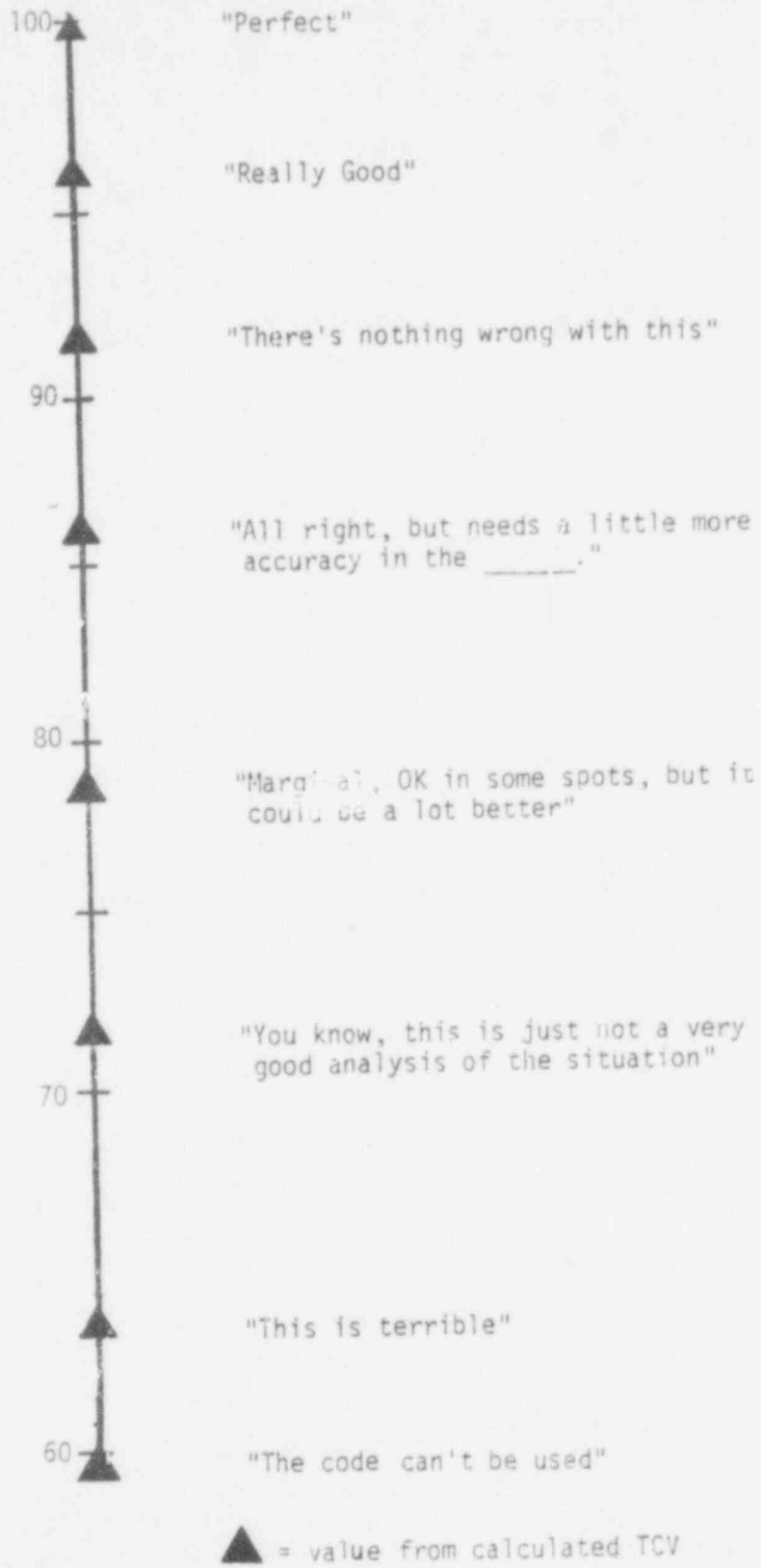


Fig. 4 Subjective Assessment

significant number of TCV's have been calculated and compared with the experts' subjective opinion, a break point or transition zone should emerge which separates the acceptable from the unacceptable.

4.3.2 Use of TCV for Application Regime Acceptance. The preceding section dealt with overall code acceptance. A function of the Independent Assessment process is also to define those areas where the code does well or areas where it should be used with caution. The variable integration limits on the assessment relations allow the assessor to evaluate a TCV for various regimes of a accident sequence. Thus, a TCV for only the blowdown phase can be calculated and compared with the TCV calculated from the reflood portion of the accident. The same process can be used for component evaluation. The relative as well as the absolute values of the TCV's are an indicator of the capability of the code in various regime and component analyses.

5. SUMMARY

A procedure has been described which can be used for the quantification of code accuracy.

The procedure described allows the code assessor to quantify code accuracy while taking into consideration the inherent scatter in experimental data. The procedure has sufficient flexibility to allow quantification of bias in the calculation, deviations from specified bounds and quantification of both transient and single value comparisons.

The subjective nature of code evaluation, while not completely removed, has been placed in the evaluation process at points where it can be used most effectively.

As long as there is some deviation between data and the predicted response, subjective analysis will be required by experts close to the problem. The assessment procedure described in Section 4 channels this subjective analysis into a decision on relative worths and then lets the

fundamental procedure quantify the outcome. Expertise in an area can be applied quite rigorously and with more consensus on the relative worth of a parameter than on the numeric value of worth.

The developed assessment procedure can be used to define application areas where the code does well and where the code should be used with caution.

By analyzing different problems and different response regimes within specific accidents (blowdown, refill, reflood) relative magnitudes of the total code value (TCV) can be used to quantify the applicability of the code in these various response regimes. After sufficient evaluation of the TCV numbers with subjective analyses of different problem solutions, the adequacy of the code to perform a particular calculation can be assessed directly from the indicated TCV for that problem type (see Section 4.2).

The developed assessment procedure can be used to evaluate the need (or lack of need) for further model development and experimental data.

In the analysis of TCV for component behavior (steam generators, pressurizers, etc.), if an evaluation is indicated that is contrary to a subjective analysis of the problem (i.e., the procedure indicates low performance and observations indicate acceptable performance or vice-versa) the problem may be traceable to a lack of suitable data for comparison (see Section 3.2). If the data required to perform a realistic comparison is truly lacking, more data should be obtained before requiring the code to meet an ill defined criteria. If adequate data is available, then further modeling is indicated.

The evaluation procedure is adaptable to automatic data processing and can thus be done as part of the independent assessment process.


In the independent assessment process, both data and predictions are manipulated and plotted within the computer from remote terminals. Since the assessment relations described in Section 3 are fixed procedures

520 334

for calculating various data/prediction relations, values for these various relations can be calculated at the same time the data/predictions are manipulated for plotting. These calculated assessment relation values can then be output and used with the subjective weighting factors for calculation of the TCV.

The TCV's calculated by one code can be compared with the TCV's calculated by another code to allow comparison between codes.

This application of the evaluation procedure can be a very important aid in the comparison of codes or comparison of various models or options within one code. Use of the procedure for this application has an added advantage of not requiring an exact association of TCV and code accuracy since only a relative evaluation is required.

 EG&G Idaho, Inc. WRRD STANDARD PRACTICE	Title: REQUIREMENTS FOR QUANTIFYING MEASUREMENT UNCERTAINTIES OF WRRD AND LOFT EXPERIMENTAL DATA	No.: WRRD #11 Page 1 of 21 Date: February 1979
	Approved: <u>L. J. Gossards</u>	LEGEND * REVISION # ADDITION
	Reviewed By: <u>[Signatures]</u>	

1.0 Purpose

The purpose of this Standard Practice is to establish requirements for the presentation of data reported by WRRD experimental programs and the LOFT project. It is the intent of the practice that reports of experimental results

- a- describe the source, recording, and processing of the data
- b- quantify the accuracy of the data

so that the users of the data, who may be separate or remote from the experimental program, can fully understand and interpret the accuracy of the data and any uncertainties in it.

Section 6 gives detailed practices for reporting both the data and data uncertainty. Uncertainty information is essential to both assessment and code development activities in comparing calculations to data and in determining the significance of differences.

References

- 2.1 ASME, "Guideline on Uncertainty Prediction and Presentation Techniques" distributed by

Executive Secretary,
 Journal of Fluids Engineering,
 Box 69,
 Hanover, New Hampshire 03755.

- 2.2 S. J. Kline and F. A. McClintock, "Describing Uncertainty in Single-Sample Experiments," Mechanical Engineering, Vol. 75, No. 1, pp. 3-8, January 1953.

WRRD STANDARD PRACTICE	Title: REQUIREMENTS FOR QUANTIFYING MEASUREMENT UNCERTAINTIES OF WRRD AND LOFT EXPERIMENTAL DATA	No.: WRRD #11
		Page 2 of 21
		Date: February 1979

- 2.3 C. Eisenhart, "Expression of the Uncertainties of Final Results" and H. H. Ku "Expressions of Imprecision, Systematic Error, and Uncertainty Associated with a Reported Value" Precision Measurement and Calibration, National Bureau of Standards special publication 300, Volume 1, pp 69-80, February 1969, Washington, D. C., U. S. Government Printing Office.
- 2.4 Yardley Beers, Introduction to the Theory of Error, Reading Mass., Addison-Wesley, 1957.
- 2.5 N. E. Dooney and C. Eisenhart, "On Absolute Measurement," Precision Measurement and Calibration, National Bureau of Standards Special Publication 300, Volume 1, p. 50, February 1969, Washington, D. C., U. S. Government Printing Office.
- 2.6 W. Mendenhall and R. Scheaffer, Mathematical Statistics with Applications, North Scituate, Mass., Duxbury Press, 1973.
- 2.7 E. O. Doebelin, Measurement Systems Application and Design, New York, McGraw Hill, 1975.
- 2.8 F. R. Hampel, "Robust Estimation: A Condensed Partial Survey," Z. Wahrscheinlichkeitstheorie Verw. Gebiete 27, (1973) pp 87-104.

3.0 Policy

- 3.1 Estimates of uncertainty must be reported for all experimental results produced under WRRD and LOFT management. This policy applies to all reported quantities. It should be noted that the policy requires only that uncertainty be estimated and therefore, it applies to Quick Look Reports as well as Experimental Data Reports, journal publications, and presentations at meetings.

Where valid statistical procedures for estimating uncertainty exist, they should be used. Where formal statistical procedures are impractical the experimental group's best subjective estimate of the uncertainty should be used. As the data presentation progresses from the Quick Look Report to more formal levels of presentation the uncertainty estimates may become correspondingly more formal.

WRRD STANDARD PRACTICE	Title: REQUIREMENTS FOR QUANTIFYING MEASUREMENT UNCERTAINTIES OF WRRD AND LOFT EXPERIMENTAL DATA	No.: WRRD #11
		Page 3 of 21
		Date: February 1979

- 3.2 Uncertainty estimates must be individually reviewed in a manner determined by the manager of the experimental program. As determined by the manager, review may be by one person, by several persons acting in rotation, or by a committee.
- 3.2.1 Reviews of uncertainty should be thorough. If uncertainty estimates appear larger than necessary, reviewers should ask: is there a valid basis for narrowing our uncertainty estimates? On the other hand, if in-situ calibrations of some measurement frequently show calibration errors larger than its stated uncertainties, reviewers should ask: should the uncertainty estimates be increased?
- 3.2.2 When data is transmitted by a letter or report, the reviewer's signature must appear in an approval block which accompanies the data. When data is transmitted by computer tape, the reviewer's signature must appear in an approval block contained in the letter of transmission for the tape. In cases where review is by a committee, the committee chairman signs for the committee.
- 3.3 The uncertainty estimates generated pursuant to this Standard Practice are to be reported along with the data, but when uncertainty estimates remain relatively fixed throughout a series of experiments, the uncertainty may be documented in a TREE (Technical report E.G.&G External) report with any discrepancies described in the data reports.
- 3.4 Reported uncertainties must include the effects of all types of error - specifically the effects of random errors, systematic error and possible mistakes.*
- 3.4.1 In calculating or estimating bounds to systematic error, deliberately "optimistic" and deliberately "pessimistic" assumptions should both be avoided. Instead, "best estimate" assumptions should be used in order to calculate reasonable, credible bounds.

* See Section 5 and Appendix B

WRRD STANDARD PRACTICE	Title: REQUIREMENTS FOR QUANTIFYING MEASUREMENT UNCERTAINTIES OF WRRD AND LOFT EXPERIMENTAL DATA	No.: WRRD #11
		Page 4 of 21
		Date: February 1979

- 3.5 Experimental groups must recognize an obligation to supply the users of their data with uncertainty estimates for all reported quantities. This obligation exists because the experimental results produced under WRRD and LOFT management are extensively used to check calculations and to support decisions affecting nuclear system safety, and because many users are separate or remote from the experimental programs.

The ASME (American Society of Mechanical Engineers) requires such estimates for all papers accepted by its Journal of Fluids Engineering. This requirement is stated in Reference 2.1. Further, Kline and McClintock (in Reference 2.2, cited by Reference 2.1) express the point of view that a subjective estimate is better than no estimate at all:

"Determination of the actual value of the uncertainty interval . . . is one of the jobs of the experimenter. As already noted, at least some of these intervals will have to be based on estimates. . . . Despite this the experimenter owes it to himself and to his readers to go ahead and do the best he can; no one else is in an equally good position to make the required estimates which are essential to . . . interpretation of the results. Such estimates are, of course, not pure guesses. Factors such as instrument backlash*, sensitivity, and fluctuation, as well as the accuracy of the basic theory of operation of the instrument, sometimes can be accounted for. Calibration of the instrument against some type of standard is sometimes available, and experience based on prior experiments or auxiliary experiments can be used."

4.0 Responsibilities

- 4.1 The manager of the experimental program is responsible for developing the information required by this standard practice, for keeping it current, and for assuring its accuracy.
- 4.2 This Standard Practice gives general guidelines for the quantification of measurement uncertainties, and should be regarded as a minimum standard. Where the members of experimental programs are aware of additional information which is pertinent to measurement uncertainty, this information should be reported. Deficiencies in this Standard Practice should be reported to the Director, WRRD or director, LOFT.

* "backlash" is used here as a metonym for all hysteresis effects.

WRRD STANDARD PRACTICE	Title: REQUIREMENTS FOR QUANTIFYING MEASUREMENT UNCERTAINTIES OF WRRD AND LOFT EXPERIMENTAL DATA	No.: WRRD #11
		Page 5 of 21
		Date: February 1979

4.3 The director, WRRD, and Director, LOFT are responsible for assuring consistent application of this Standard Practice among experimental programs and for correction of deficiencies reported according to 4.2, above.

5.0 Exposition of Assumptions, Definitions and Basic Terms

5.1 "True Values"

In this Standard Practice, a basic assumption is that "true values" of measured quantities exist. This assumes for example, that effects of the Heisenberg uncertainty principle are negligible and that philosophical questions such as "what is truth?" may be bypassed.

5.2 Error and Uncertainty

In practice, the measured values differ from the true values because of errors of measurement. That is:

$$\text{error} = (\text{true value}) - (\text{measured value}) \quad (1)$$

If error were known Equation (1) could be rearranged to

$$\text{true value} = (\text{measured value}) + (\text{error}) \quad (2)$$

and the true value could also be known. In reality the error is not known and uncertainty results. "Error" should be carefully distinguished from "uncertainty". The actual error in a measured result is a single number and, by definition is unknown and unknowable. Uncertainty, on the other hand, is a magnitude (a measure of the length of an interval) the error is unlikely to exceed.

5.3 Classification of Errors

The components of measurement error are usually placed in three classes:

- . Systematic errors
- . Random errors
- . Mistakes

WRRD STANDARD PRACTICE	Title: REQUIREMENTS FOR QUANTIFYING MEASUREMENT UNCERTAINTIES OF WRRD AND LOFT EXPERIMENTAL DATA	No.: WRRD #11
		Page 6 of 21
		Date: February 1979

These are defined below in Sections 5.3.1, 5.3.2 and 5.3.3. However, Reference 2.2, p. 4, points out that the boundaries of these classes are not perfectly clear, since obvious mistakes are often corrected by the experimenter (while small mistakes go unnoticed and therefore may be confused with systematic or random error). Similarly, an unknown systematic error often cannot be distinguished from a random error. Hence, an actual measured result usually contains unknown proportions of random and systematic errors and sometimes some small mistakes.

5.3.1 Systematic errors.

Corresponding to any real measurement system, it is possible to postulate an ideal measurement system which is frictionless, massless, has infinite resolution, zero response time, and so forth. Any tendency of actual measurements to deviate consistently from hypothetical measurements made with the postulated ideal system is a component of systematic error. Included among such systematic errors are ". . . all those errors which cannot be regarded as fortuitous, as partaking of the nature of chance. They are characteristics of the system involved in the work; they may arise from errors in theory or in standards, from imperfections in the apparatus or in the observer, from false assumptions, etc. To them, the statistical theory of errors does not apply" (Reference 2.5).

When systematic errors are constant, they are sometimes called biases or fixed errors. When systematic errors change slowly with time, they are called drifts. When systematic error results in the measurement of one variable's being dependent on another variable, (as when a pressure measurement is dependent on ambient temperature) the systematic error is called a sensitivity.

In principle, the magnitude of systematic errors can be determined by a comprehensive calibration of the measurement system. For example, drift can be determined by holding the measured variable constant and observing any change with time in the measurement. A comprehensive calibration should account for at least the following: drift, aging, wear, response time, flow regime, environmental effects, and threshold effects.

WRRD STANDARD PRACTICE	Title: REQUIREMENTS FOR QUANTIFYING MEASUREMENT UNCERTAINTIES OF WRRD AND LOFT EXPERIMENTAL DATA	No.: WRRD #11
		Page 7 of 21
		Date: February 1979

In many cases, however, comprehensive calibration (as defined above) is impossible for WRRD and LOFT instruments. In such cases the systematic error must be estimated at least in part. Consequently, systematic error is not known and contributes to uncertainty.

5.3.2 Random errors. These errors are also called experimental or accidental errors. The random error is the component of the error which causes repeated measurements of the same quantity to differ without apparent reason. The random component of e_r or can be estimated statistically if repeated measurements of the same quantity are made. In some situations it can also be estimated by more sophisticated statistical techniques such as time series analysis.

5.3.3 Mistakes. These are also called blunders or illegitimate errors or human errors or bugs. In general, mistakes are made by people, and if an experiment is repeated, the mistake may or may not be repeated. A more extended discussion of mistakes is given in Appendix B.

5.4 Uncertainty Interval

Since in practice calibrations cannot remove all systematic errors, random errors can only be estimated, and mistakes do happen, error is not known and uncertainty results. In practice, then, the measured value and an estimate of the error are combined to give an uncertainty interval which is believed to include the true value.

5.5 Confidence Level

The estimated uncertainty interval may or may not actually include the true value. The probability that it does is called the confidence level.* The confidence level is often expressed as a percentage. For example, a 95% confidence level indicates a belief that - 95 times out of 100 - the uncertainty interval will include the true value.

* A more precise statement is:
"The probability (prior to actually conducting the experiment) that the confidence interval computed, using the experimental data, will contain the true value is called the confidence level."

WRRD STANDARD PRACTICE	Title: REQUIREMENTS FOR QUANTIFYING MEASUREMENT UNCERTAINTIES OF WRRD AND LOFT EXPERIMENTAL DATA	No.: WRRD #11
		Page 8 of 21
		Date: February 1979

5.5.1 Confidence Level For Objectively Estimated Random Errors. When random error is objectively estimated statistical techniques (Reference 2.6, pp. 283-285) can be used to associate a confidence interval with a confidence level. Objective estimates of random error can be obtained by repeating experiments. Alternatively, if the measured quantity is a time-series, the random component may sometimes be estimated by time-series analysis.

5.5.2 Confidence Level for Subjectively Estimated Random Errors. If the experiment has not been repeated, or if time-series analysis is inapplicable, the random error must be subjectively estimated. When the random error has been subjectively estimated the confidence interval and the confidence level are connected using the concept of "odds". For example, a 95% confidence level is associated with 19 to 1 odds. (For details, see Section 6.6.2, Appendix A and Reference 2.2).

6.0 Practice of Data Presentation

6.1 Required Scope

The measurement uncertainty should be stated for each tabulated or plotted experimental result in reports, and for each channel in data tapes.

Uncertainties in direct measurements (such as pressures and temperatures), derived data* (such as mass flow rates and average density based on multiple beam gamma densitometers), and initial and boundary conditions should be reported.

All components of uncertainty including the effects of systematic error, random error, and mistakes (if, in the experimenter's judgment mistakes were made) must be reported for every experimental result.

6.2 Description of Measurement Device

A description of each measurement should be supplied, either as part of individual data reports or in a TREE reference document. The description should include location, details

* Referred to as "computed parameters" within LOFT.

WRRD STANDARD PRACTICE	Title: REQUIREMENTS FOR QUANTIFYING MEASUREMENT UNCERTAINTIES OF WRRD AND LOFT EXPERIMENTAL DATA	No.: WRRD #11
		Page 9 of 21
		Date: February 1979

of local geometry that may affect the measurement, and a brief description of the measurement device itself.

- 6.2.1 Zero Setting and Offsets. The description of each measurement should include a description of any procedures used to zero or offset the initial readings. In general, an initial zero setting procedure is expected to reduce the fixed ("bias") component of instrument error. Such reductions should be reflected in a reduction of uncertainty intervals, but see 3.2.1 above.
- 6.2.2 Calibration and Test Environment. Available calibration information should be reported (or referenced) for each measurement device. Where calibration conditions differ from the test environment, the calculated or estimated effects of the test environment on the validity of the calibration should be reported (or referenced).
- 6.2.3 Functional and Time Dependence of Uncertainty. Uncertainties may be functions of time, flow regime, fluid density, and rate of change of the measured variable. If the uncertainty is a function of these parameters or other aspects of the test environment, this dependence should be reported (or referenced). It may be appropriate to report different uncertainties in different parts of the measurement range. In any case, the entire measurement range experienced in a particular test should be covered when reporting uncertainties for data tapes. For reports, the entire measurement range reported in the particular document should be covered.
- 6.2.4 Responsibility to Report Factors Affecting the Accuracy of the Measurement. Other information that is pertinent to interpreting the data such as measurement ranges, amplifier saturation points, dead bands and response times should be reported. In particular, when physical constraints for measurement devices (such as mechanical stops to limit deflection) affect the results, this event should be unambiguously reported.

WRRD STANDARD PRACTICE	Title: REQUIREMENTS FOR QUANTIFYING MEASUREMENT UNCERTAINTIES OF WRRD AND LOFT EXPERIMENTAL DATA	No.: WRRD #11
		Page 10 of 21
		Date: February 1979

A comprehensive list of factors which are known to affect measurement accuracy is given in Appendix C. The effects of any such factors which affect a particular measurement should be reported (or referenced).

6.3 Description of Data Processing Procedure

Any aspects of the data acquisition and processing system which affect the interpretation of the data should be reported (or referenced). This description should include filtering, smoothing, sampling, and shifting. Also, if data compression (understood to mean the use of every n^{th} sample with $n \neq 1$) is used, its effect on the reported results should be described. Any effects of smoothing and compression on the magnitude or timing of real oscillations or rapid changes in the measured quantities should be reported.

6.4 General Requirements for Reporting Uncertainty

Most WRRD and LOFT experimental results fall into the category described by Eisenhart (Reference 2.3, p. 271) as "Neither Systematic Error nor Imprecision Negligible". ("Imprecision" is Eisenhart's term for random error.) He points out that uncertainty for this category of results cannot be specified by a single unqualified number, and strongly urges that experimental results should not be reported as $\pm b$ without precise definition of what "b" is.*

Accordingly, for WRRD and LOFT experimental data, uncertainty should be described in detail in a separate uncertainty chapter or report, and this chapter or report should be referenced to qualify the condensed indications of uncertainty given for tables, figures, and data-tape channels. Separate bounds to the systematic ("bias") and random components of uncertainty should be given in this chapter or report.

* Eisenhart points out that "b" has at least 6 different interpretations in the literature: (one-sided) limit of error, probable error, standard deviation, $2x$ (standard deviation), $3x$ (standard deviation), and average deviation. So "b" must be defined.

WRRD STANDARD PRACTICE	Title: REQUIREMENTS FOR QUANTIFYING MEASUREMENT UNCERTAINTIES OF WRRD AND LOFT EXPERIMENTAL DATA	No.: WRRD #11
		Page 11 of 21
		Date: February 1979

6.5 Consistency of Redundant Measurements

Uncertainty estimates must be consistent with multiple measurements of the same parameter. If one of a set of redundant measurements of a given parameter lies outside the estimated uncertainty interval for that parameter, it should be shown that the frequency or infrequency of occurrence of this event is consistent with the confidence level.

6.6 Form of Uncertainty Reports

6.6.1 Uncertainty Resulting from Systematic Error.

Bounds to systematic error should be stated as follows:

Where systematic error has been reliably established (i.e., by calibration traceable to a Standards Laboratory) the bounds on the systematic error should be stated in sentence form using positive wording:

" . . . the systematic errors are not in excess of . . ."

" . . . a systematic error of not more than \pm . . ."

Where systematic error is estimated (in whole or in part) from prior experience or judgement, the uncertainty statement should be qualified:

" . . . the systematic errors are (believed, estimated, considered, judged) not to be in excess of . . ."

" . . . a systematic error (believed, estimated, considered, judged) not to exceed \pm . . ."

A brief description of the method which produced the uncertainty bounds should be included or referred to. The various contributions to the uncertainty should be identified. The effects of flow regime dependence and other two-phase effects on the uncertainty analysis should be described (see Appendix C).

WRRD STANDARD PRACTICE	Title: REQUIREMENTS FOR QUANTIFYING MEASUREMENT UNCERTAINTIES OF WRRD AND LOFT EXPERIMENTAL DATA	No.: WRRD #11
		Page 12 of 21
		Date: February 1979

Where the systematic error is a combination of a number of elemental systematic errors the discussion should state explicitly the method of combination such as "the simple sum of the bounds" or "the square root of the sum of squares".

The numerical statement of the systematic uncertainty should be

$$xx + (yy \text{ percent of reading})$$

where both xx and yy are non-negative, and are expressed to not more than two significant figures. Either xx or yy may be zero.

- 6.6.2 Uncertainty Resulting from Random Error. Bounds on random error should be identified as confidence intervals at the 95% confidence level. These bounds should be reported as follows:

Where bounds on random error are based on statistical techniques, the bounds should be stated in sentence form. Paraphrasing Reference 2.3, p. 72 this sentence might be ". . . with error bounds of +3.4 meters/second derived from a computed standard deviation of 1.5 meters/second (based on 9 degrees of freedom). (The number 3.4 is equal to 2.26×1.5 , where 2.26 is the critical value of Student's t for 9 degrees of freedom at the 95% confidence level) . . ."

Where the random component of the uncertainty interval is based on the experimenter's judgment, the experimenter should choose the size and location of the uncertainty interval to correspond with his judgment that there is only one chance in twenty that the true value of the measured quantity lies outside the confidence interval. (See Appendix A). Where judgment is used, the confidence level should be stated as follows:

". . . it is estimated that the error bounds at the 95% confidence level do not exceed $\pm . . .$ "

WRRD STANDARD PRACTICE	Title: REQUIREMENTS FOR QUANTIFYING MEASUREMENT UNCERTAINTIES OF WRRD AND LOFT EXPERIMENTAL DATA	No.: WRRD #11
		Page 13 of 21
		Date: February 1979

The numerical statement of the confidence interval should be

$$zz + (w \text{ percent of reading})$$

where both zz and w are non-negative, and expressed to one or two significant figures. Either zz or w may be zero.

A brief description of the method which produced the uncertainty bounds should be included or referred to. As in 5.6.1 above, both physical and mathematical aspects of the uncertainty analysis should be described.

6.7 Reporting Uncertainty for Tabular Data

Every table of measured results must carry the uncertainty statement juxtaposed to the results. The brief uncertainty statement forms a part of the table and must refer the reader elsewhere (e.g., to a separate chapter or reference) for details of the uncertainty calculation. For example:

Temperature (K)	Length (m)	Pressure (M Pascals)
642 \pm 8*	29.3 \pm 0.1	0.134 \pm 0.002
536 \pm 8	16.7 \pm 0.1	0.270 \pm 0.003

* Total experimental uncertainty including bias and random error (95% confidence level) see (chapter or reference) for details

When the form $a + b$ is used, the footnote is essential to explain the meaning of b . Similar requirements apply to unsymmetrical intervals of the type $a \begin{matrix} + b \\ - c \end{matrix}$

WRRD STANDARD PRACTICE	Title: REQUIREMENTS FOR QUANTIFYING MEASUREMENT UNCERTAINTIES OF WRRD AND LOFT EXPERIMENTAL DATA	No.: WRRD #11
		Page 14 of 21
		Date: February 1979

Where mistakes occur*, but the data still retain utility, the mistakes should be noted. For example: "Note - Pressure measured on 0-100 M Pascal range, uncertainty band includes this effect".

6.8 Reporting Uncertainty for Graphical Data

Every figure must carry the uncertainty statement juxtaposed to the results. This may be done in any of the following ways:

- (a) by reference to a separate uncertainty analysis report or chapter
- (b) by direct graphical presentation on the figure itself
- (c) by notes appended to the figure.

6.8.1 Use of Uncertainty References on Figures. A block may appear on the figure, referencing a separate uncertainty analysis report or chapter. This method of presentation is acceptable only if the referenced report

- a - exists
- b - is available to everyone who has access to the measured results shown in the figure.

6.8.2 Direct Graphical Presentation of Uncertainty. Uncertainty information may be presented directly on the figure itself. The presentation may use such means as multiple traces, color, shading, or special symbols such as uncertainty bars. In every such case, however, the figure caption must contain the information that total uncertainty at the 95% confidence level is depicted and must contain a reference to a detailed uncertainty description. The example below illustrates the use of uncertainty bars.

Note that the uncertainty information in the caption is essential.

* See Appendix B for a discussion of mistakes and their effects on uncertainty.

520 349

WRRD
STANDARD PRACTICE

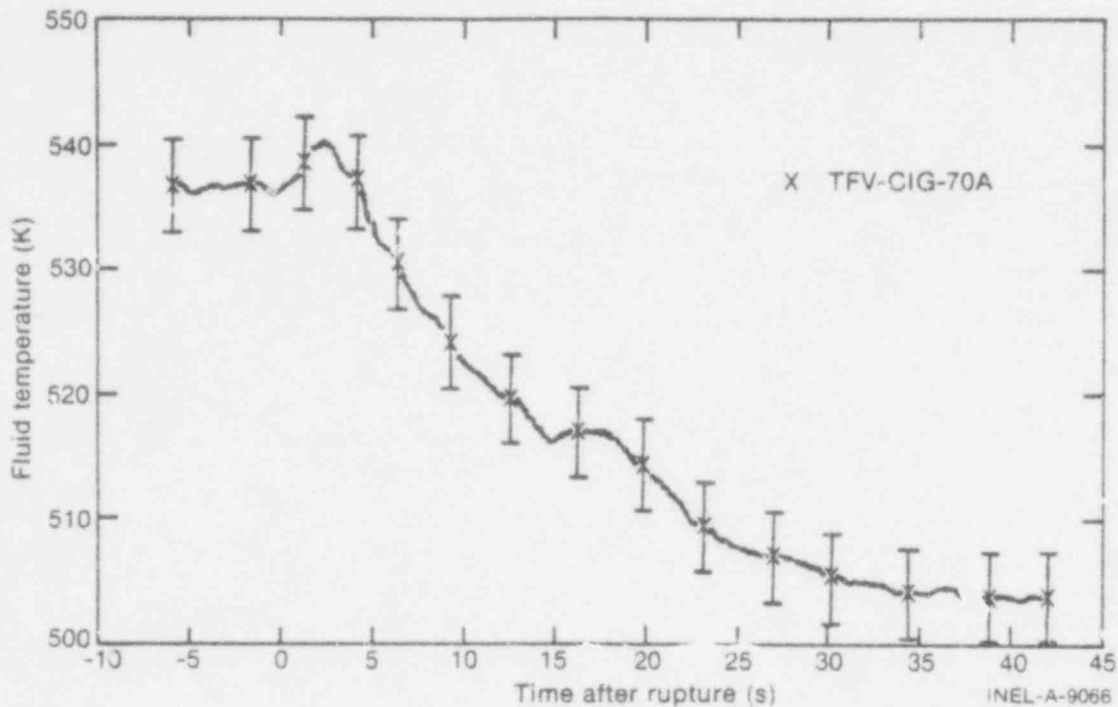
Title: REQUIREMENTS FOR QUANTIFYING
MEASUREMENT UNCERTAINTIES OF WRRD
AND LOFT EXPERIMENTAL DATA

No.: WRRD #11

Page 15 of 21

Date: February 1979

FLUID TEMPERATURE VS. TIME AFTER RUPTURE
(TOTAL UNCERTAINTY ESTIMATES AT 95% CONFIDENCE LEVEL,
See (Chapter or Reference) For Details)



The use of the above method of presentation is acceptable only if the uncertainty bars are clearly distinguishable in size from the symbols used to plot points on the figure.

The use of three traces to indicate best estimate, upper limit of confidence interval, lower limit of confidence interval, is not an acceptable method of data presentation unless the three traces are clearly distinguishable from one another and from any other traces on the figure.

WRRD STANDARD PRACTICE	Title: REQUIREMENTS FOR QUANTIFYING MEASUREMENT UNCERTAINTIES OF WRRD AND LOFT EXPERIMENTAL DATA	No.: WRRD #11
		Page 16 of 21
		Date: February 1979

- 6.8.3 Use of Uncertainty Notes on Figures. A block may be appended to the figure, giving notes on uncertainty, as in the example below. Use of such a block is preferred when the methods of 6.8.2 lead to cluttered figures. Again, the reader must be told where to find the details.

UNCERTAINTY OF TIME, TEMPERATURE

COORDINATE	UNCERTAINTY*
$t < 2.5 \text{ sec}$	$\pm 0.01 \text{ sec}, \pm 10K$
$2.5 < t < 5.0 \text{ sec}$	$\pm 0.01 \text{ sec}, \pm 15K$
$t > 5.0 \text{ sec}$	$\pm 0.01 \text{ sec}, \pm 10K$

* Total experimental uncertainty including bias and random error (95% confidence level) see (chapter or reference) for details

6.9 Requirement to Report Steady State Data

When results of transient events (such as blowdowns) are reported, the report should include figures or tables showing all measured quantities during the steady-state period immediately preceding the transient, and during any steady-state period which follows the transient. This section of the report should be sufficiently detailed to allow users of the results to compare the measured initial conditions with the specified or assumed initial conditions, and to allow data users and experimenters to compare measured quantities for consistency (in-situ comparisons).

WRRD STANDARD PRACTICE	Title: REQUIREMENTS FOR QUANTIFYING MEASUREMENT UNCERTAINTIES OF WRRD AND LFOT EXPERIMENTAL DATA	No.: WRRD #11
		Page 17 of 21
		Date:

APPENDIX A - SUBJECTIVE ESTIMATION OF CONFIDENCE INTERVALS

In applying this standard practice, experimenters will frequently need to estimate confidence intervals and to associate them with confidence levels.

Such subjective estimates are not as difficult as they might seem at first. If the reader will simply consider all the velocities, times, distances, and volumes* a driver must estimate in successfully driving a car from his home to his office, he will agree that the human brain is a first-rate estimator.

Suppose the driver's best estimate of the required time to drive to the office is 15 minutes. In his experience he seldom covers the distance in less than 10 minutes; on the other hand, he can conceive of weather or traffic delays which could stretch the trip to 25 minutes. Thus he could state his travel time as:

$$15 \begin{matrix} +10 \\ -5 \end{matrix} \text{ minutes}$$

In effect, the driver has placed an unsymmetrical 15 minute confidence interval on his best estimate of the driving time.

Associating a confidence level with subjective estimates is a little harder. As an aid in this process, the idea put forward in 1953 by Kline and McClintock (Reference 2.2) still retains validity in the 1970's:

"A useful viewpoint is that one is willing to bet with certain odds (say 19 to 1) that the error falls within the given limits" (Reference 2.6, pp 58-59).

Note that the "odds" point of view tends to place both lower and upper bounds on the size of the uncertainty interval: if the interval is too small, the experimenter may lose his bet, but if the interval is too large he may not find anyone to bet with!

In choosing the limits +10, -5, the driver should be willing and able to place a bet with 19 to 1 odds that one can make the trip in more than 10 minutes, and less than 25 minutes. He could then say, "It is estimated, at the 95% confidence level, that travel time from my home to this office is between 10 and 25 minutes."

* of gasoline

WRRD STANDARD PRACTICE	Title: REQUIREMENTS FOR QUANTIFYING MEASUREMENT UNCERTAINTIES OF WRRD AND LOFT EXPERIMENTAL DATA	No.: WRRD #11
		Page 18 of 21
		Date: February 1979

APPENDIX B - MISTAKES

Engineers have a very Victorian attitude toward mistakes. Everybody knows that mistakes occur, but almost no one will write about them in the serious literature. Like Victorian sex, engineering mistakes are dealt with euphemistically -- they are referred to as bugs, glitches, "unplanned events" -- anything but mistakes. References to mistakes tend to be quotations from off-the-record remarks at symposiums. Basing his estimate on such informal sources Hampel (Reference 2.7, p. 88) states that

"Altogether 5-10% wrong values in a data set appear to be the rule, rather than the exception"

On the other hand, Beers (Reference 2.4, pp. 5-6) states under the quaint heading "Illegitimate Errors"

"... there are three types of avoidable errors which have no place in an experiment, and the trained reader of a report is justified in assuming that these are not present.

(1) Blunders. These are errors caused by outright mistakes in reading instruments, adjusting the conditions of the experiment, or performing calculations. These may be largely eliminated by care and by repetition of the experiments and calculations.

(2) Errors of computation. The mathematical machinery selected for calculating the results of an experiment should have errors small enough to be completely negligible in comparison with the natural errors of the experiment. . . *

(3) Chaotic errors. If the effects of disturbances become unreasonably large--that is, large compared with the natural random errors--they are called chaotic errors. In such situations the experiment should be discontinued until the source of the disturbance is removed."

* The deleted sections of this quotation refer to slide-rule and logarithm table methods which are not applicable to WRRD and LOFT practices.

WRRD STANDARD PRACTICE	Title: REQUIREMENTS FOR QUANTIFYING MEASUREMENT UNCERTAINTIES OF WRRD AND LOFT EXPERIMENTAL DATA	No.: WRRD #11
		Page 19 of 21
		Date: February 1979

The opposite positions of Hampel (who expects 5-10% wrong values in a data set) and Beers (who says that the reader has a right to expect that illegitimate errors are not present in reported results) need to be reconciled for WRRD and LOFT.

Because of the cost of LOCA experiments, an experiment cannot easily be repeated just because a few measurement mistakes have occurred. Because of their visibility and importance, the results cannot be suppressed, either. Consequently, when mistakes occur, they must be reported, and their effect on the data uncertainty must be estimated. Some hypothetical examples follow:

A transducer output is expected to have a range of + 10 volts. For whatever reason, the actual output never exceeds 1.5 volts. The output is digitized by an 8 bit A to D converter. Such an occurrence should be reported in a data report by a note "effective resolution reduced from 7 bits and sign to 4 bits and sign - transducer out of range" or equivalent language.

A d.p. cell is overranged during one test and used in another test without replacement or recalibration. Suppose that experience has shown that such occurrences have resulted in biases of 0.1 psid and calibration errors of 10%. Such an occurrence should be reported in a data report by a note: "bias (95% confidence level - estimated) 0.1 psid and calibration error increased from 1% to 10% - damaged transducer" or equivalent language.

A differential pressure which should have a final value of zero is found to have a non-zero final value. The experimenter judges that there is only one chance in 20 that the measured final value is less than -0.2 psid. Such an occurrence should be reported by a note "measurement contains an unexplained negative long term offset of -0.2 psid (95% confidence level-estimated).

Experience shows that a frank, open treatment of the minority of data which contains mistakes increases the credibility of the majority of data which is free of mistakes.

WRRD STANDARD PRACTICE	Title: REQUIREMENTS FOR QUANTIFYING MEASUREMENT UNCERTAINTIES OF WRRD AND LOFT EXPERIMENTAL DATA	No.: WRRD #11
		Page 20 of 21
		Date: February 1979

APPENDIX C - A PARTIAL CATALOGUE OF EFFECTS WHICH CAN
PRODUCE SYSTEMATIC ERROR OR AMBIGUITY IN DATA INTERPRETATION

The effects listed below are drawn from the experience of a small group of data users. In the experience of this group these effects have produced systematic errors or have led to ambiguity in the interpretation of LOFT and WRRD data. The list is not complete, nor is it in order of most important or significant effects.

The list is presented to give experimental groups a starting point for identifying frequent causes of systematic error.

- (1) Instrument orientation
- (2) Instrument environment
- (3) Instrument response time
- (4) Instrument sensitivity to temperature, pressure, and radiation
- (5) Thresholds and dead bands
- (5) Drift and aging
- (7) Sensitivity to flow regime (particularly inhomogeneous two-phase flow)
- (8) Gravity effects on differential pressure (typical interpretation ambiguity: "does zero differential pressure mean that the hydrostatic head is zero?")
- (9) Thermocouple fin effects
- (10) Thermocouple wetting effects
- (11) Thermocouple thermal radiation effects
- (12) Sampling, filtering, shifting, smoothing and time-compression in data processing.
- (13) Replacement of an instrument or component without recalibration of the instrument channel

A more subtle kind of systematic error occurs when square law and logarithmic data are first filtered and then subjected to nonlinear transformations. This applies to both analog and digital filtering.

WRRD STANDARD PRACTICE	Title: REQUIREMENTS FOR QUANTIFYING MEASUREMENT UNCERTAINTIES OF WRRD AND LOFT EXPERIMENTAL DATA	No.: WRRD #11
		Page 21 of 21
		Date: February 1979

Filtering and nonlinear transformations do not commute: that is, filtering followed by a nonlinear transformation is not equivalent to a nonlinear transformation followed by filtering.

The above phenomenon may affect the following types of measurements:

- (a) Density measurements using gamma densitometers (logarithmic amplifiers)
- (b) Flow measurements using orifices and differential pressure measurements
- (c) Mass flow measurements derived by combining drag-disc measurements with gamma densitometer measurements.

520 357

ATTACHMENT II

METHODOLOGY FOR FIGURE-OF-MERIT

Decision tree analysis is a systems evaluation method in which a complex decision problem may be structured so that each important factor can be evaluated separately and then ultimately combined into a single number representing the worth of the system to the evaluators. Qualitative as well as quantitative factors can be included. In practice, using the one-factor-at-a-time procedure is often more beneficial than the actual numbers produced, because the insight developed supplants intuitive feelings. The evaluation procedure used by the task force is essentially the same as that developed at LASL by William J. Whitty, but his report has not yet been released.

In order to validly compare alternative radwaste treatment processes, each must be so specified as to meet the same processing objective. Thus, every process concept must be specified to treat the same waste material mix and to meet the same standards for the output product. This means that the evaluation of TRU waste processing concepts is site-specific. A processor or concept that works well at one laboratory may be inappropriate at another laboratory.

After the waste processing application has been specified, a set of detailed evaluation criteria is defined. These criteria represent the major areas of importance for the application as perceived by the evaluators. The major criteria may each be divided into subcriteria, and these again into sub-subcriteria. The next step is to link the capabilities of the alternative processes to the evaluation criteria. This is accomplished by devising performance measures that are called levels of performance.

As an example of these ideas, suppose that a major consideration in evaluating a waste processing facility is energy conservation. Then energy consumption would be an important criterion. The performance

measure selected could be, say numbers of kW-h per year as a quantitative measure or "high, medium, or low" as a qualitative measure. A particular process alternative under evaluation might be assigned a performance level of 10,000 kW-h or a performance level of "low".

In decision analyses with multiple evaluation criteria, the performance measures usually have dissimilar units. Consequently, the performance levels must be transformed to a common unit of measure. The ideal common unit would be one widely recognized such as monetary worth. This ideal is not possible in many situations, and the common unit is often a subjective value judgment of an evaluator as to how well a particular performance level satisfies a criterion. Naturally, the perceived value is different for different people. Thus, when more than one evaluator is involved, a consensus must be reached on each criterion.

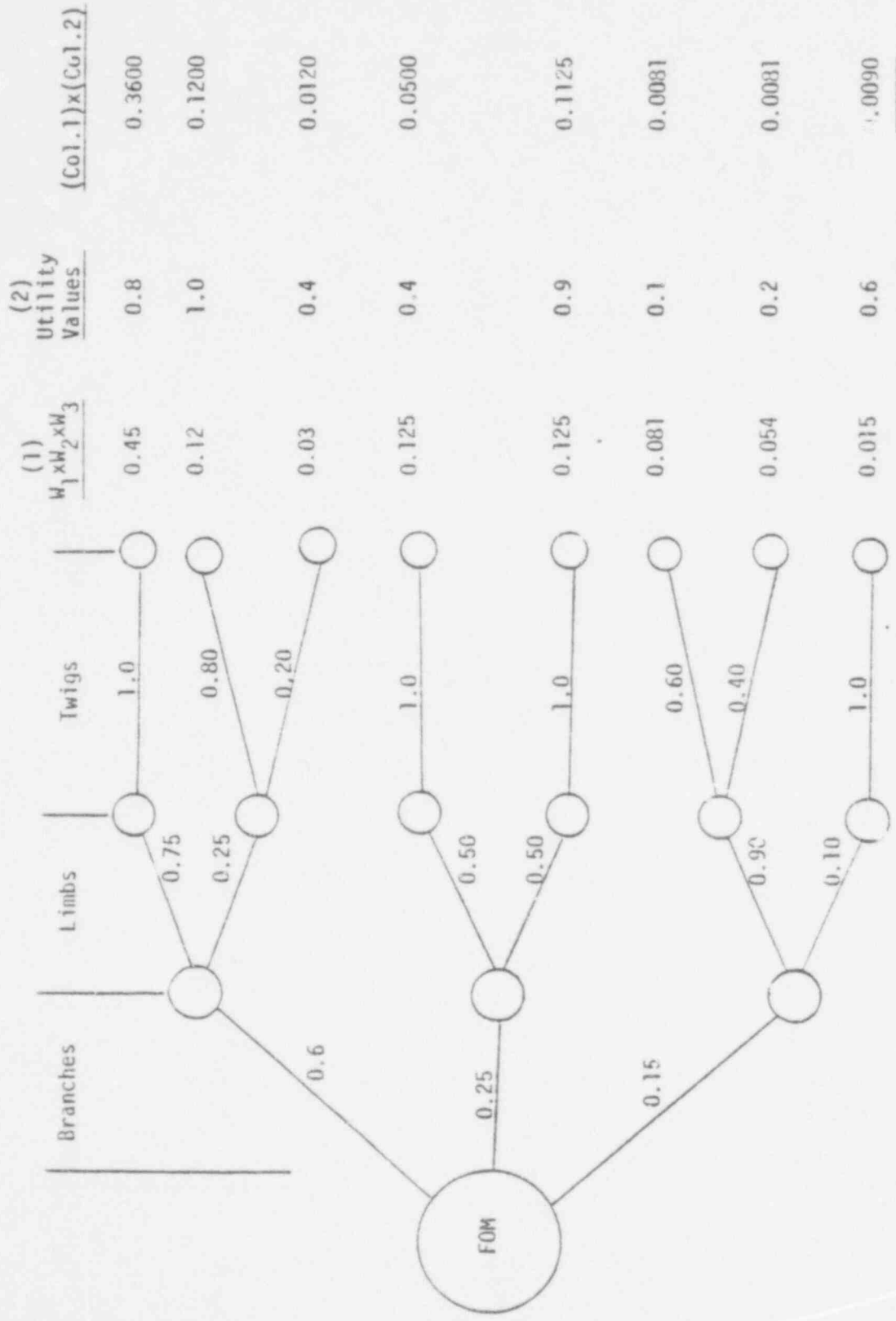
The relationship between levels of performance and a common unit of measure are called utility functions in decision problems under uncertainty. After all performance criteria are evaluated with utility function values, a single scalar value of process performance, called a figure-of-merit (FOM), is computed as a weighted combination of the utility values. The weighting factors used express the relative importance of the criteria to the success of the process alternative.

The computational procedure may be explained by means of a decision tree. A simple three-branch decision tree is shown in the figure. This example tree has three tiers of performance criteria. For purposes of illustration, these are called branches, limbs, and twigs. Each branch has a weight that reflects the importance of the branch criterion in satisfying the goal of the system. The sum of the branch weights must be unity. Likewise, the weights of the limb criteria for each branch criteria must sum to unity. The same is true for the weights of the twig criteria for each limb.

The overall weight for each twig criterion is the product of the weights of the associated branch, limb, and twig. A utility value is assigned to each twig criterion on the basis of the performance level of the particular system under evaluation. The contribution of the twig to the FOM for that system is the product of the overall weight and the utility value. The FOM for the system is the sum of the twig contributions.

In summary, the FOM computational procedure is based on partitioning an m-component problem into m one-component problems, with each being easier to solve than the original, and then to combine the m solutions. Such a computational procedure is said to be a linear additive model.

When conducting a decision analysis in the face of uncertainty, as particularly epitomized when weighting factors, performance levels, or utility functions are obtained through subjective judgments, there is the question whether differences in FOMs are significant. In such cases, it is prudent to conduct a sensitivity study of the computational procedure. One way to study sensitivity is to conduct a Monte Carlo simulation in which all weights and utility values are assumed to be statistical random variables. This is the technique used by the waste processing task force.



FOM = SUM = 0.6824

FIGURE
EXAMPLE DECISION TREE

520 361

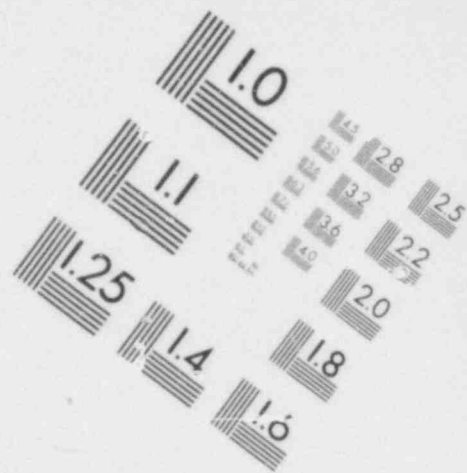
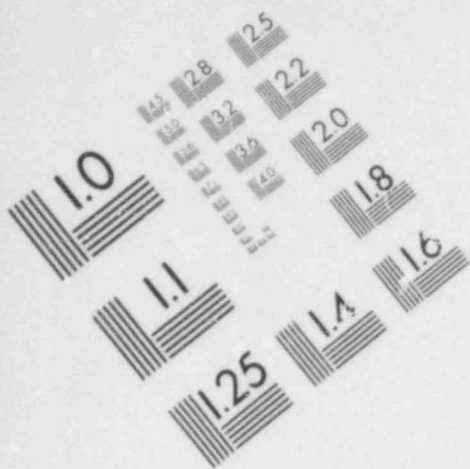
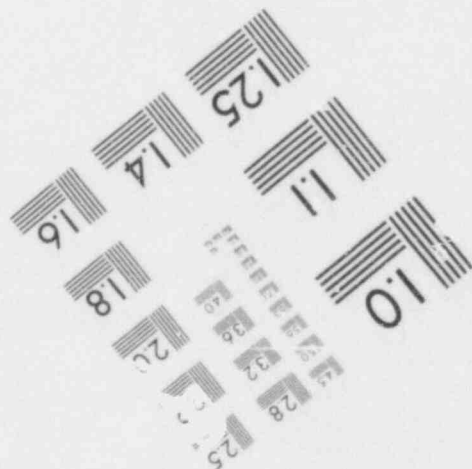
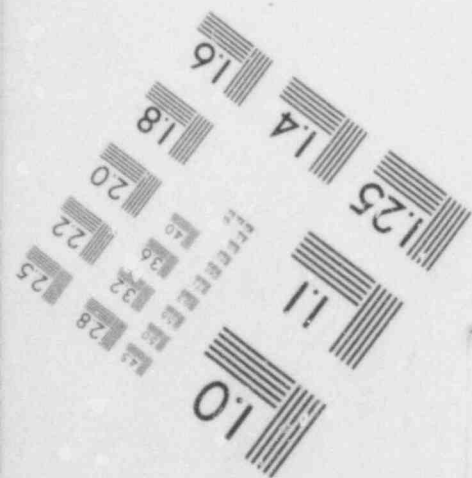
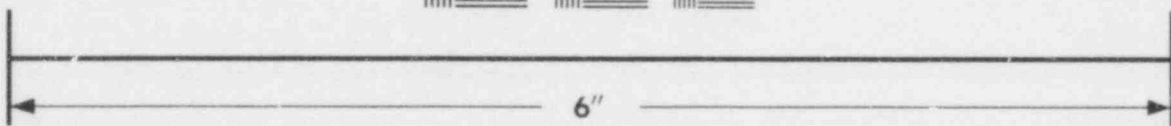


IMAGE EVALUATION
TEST TARGET (MT-3)



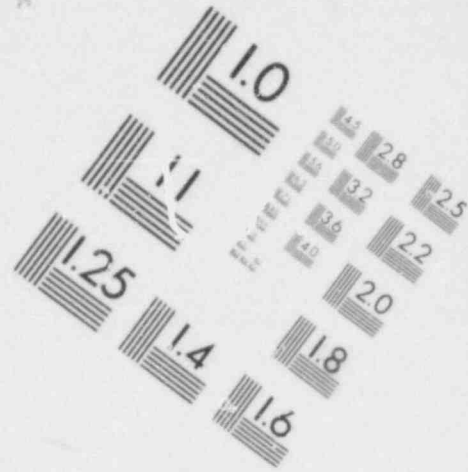
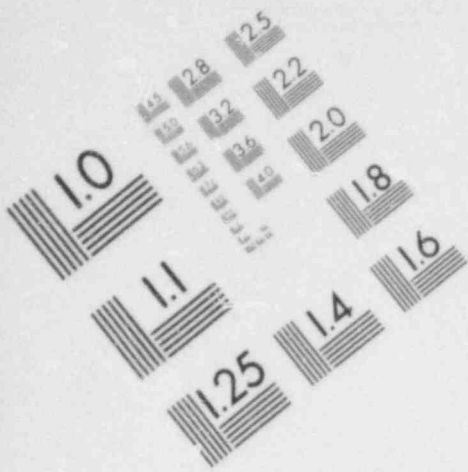
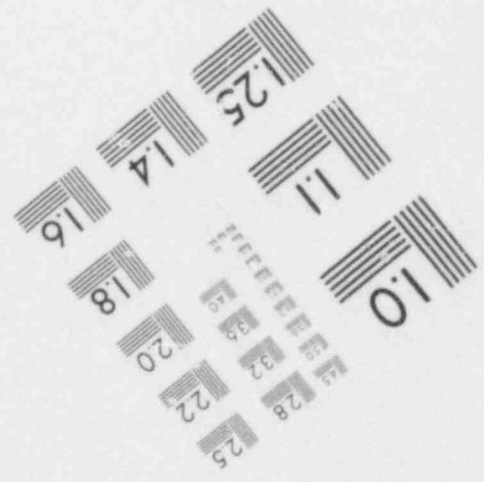
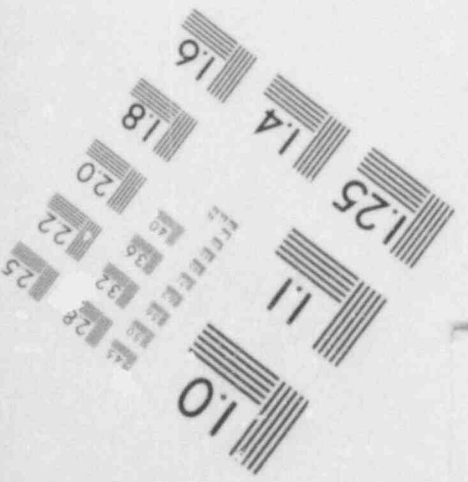
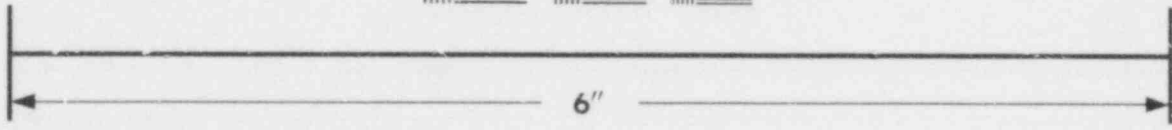


IMAGE EVALUATION
TEST TARGET (MT-3)



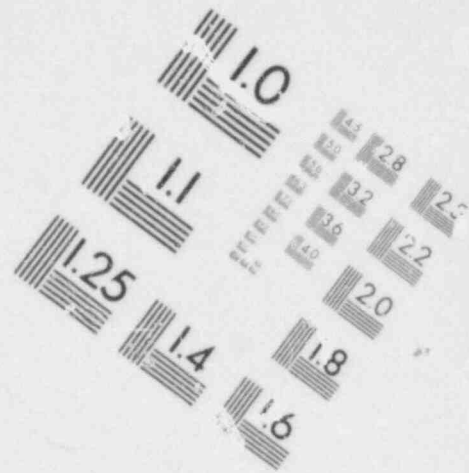
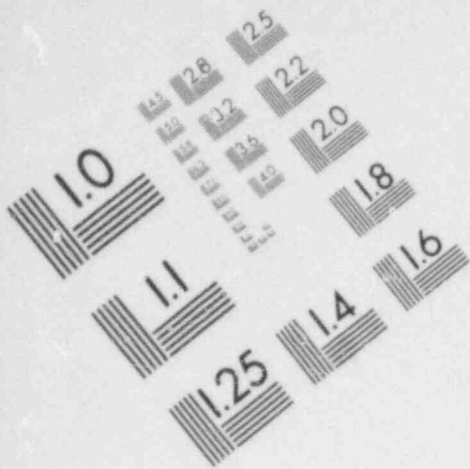
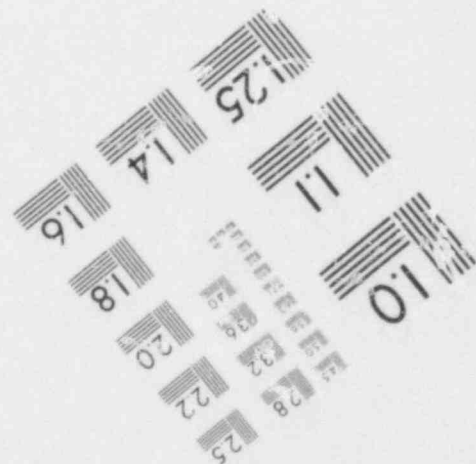
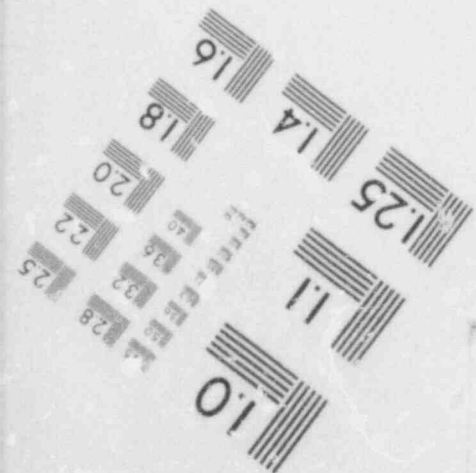
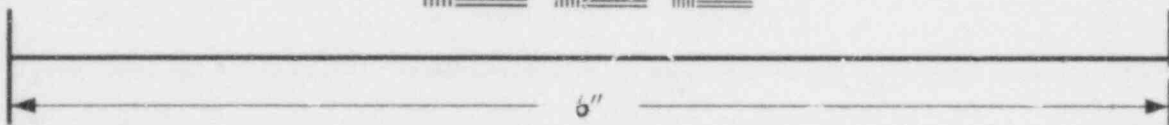


IMAGE EVALUATION
TEST TARGET (MT-3)



APPENDIX B

PERCENTILE ACCEPTANCE OF CODE

521 001

~~520 362~~

I. INTRODUCTION

As part of the development of an overall code assessment procedure, research was conducted to determine analysts' concept of "good", "bad", "acceptable", and "unacceptable". These concepts must be known if the results of code assessment are to be conveyed from one human to another and if an overall acceptance criteria is to be developed.

II. DESCRIPTION OF RESEARCH

The type of data/prediction comparisons made in code assessment are so varied and individually complex that no unique plot could be picked as representative. Therefore, a number of "thought" comparisons were assembled to test different aspects of data/prediction comparisons. Attachment B-1 shows the letter transmitted to 150 engineers at EG&G asking them to score the comparisons illustrated in Figure B-1. The letter was designed to minimize the technical detail of each plot and thus address only the person's feeling for the "goodness" of fit.

III. RESULTS

Figures B-2 through B-8 show the individual scores given for each of the thought problems. There were 118 responses to the questionnaire. Those not responding gave several reasons including 1) not qualified, 2) not enough data given (statisticians, generally, who said they could not respond unless they knew the distribution, etc.), and 3) those who said you cannot quantify this sort of thing. Those who did respond gave us data from which several conclusions can be drawn and most importantly, an indication of the variance in people's opinion on what is good and bad.

1 DISCUSSION OF SCORE PLOTS

The mean score for each problem is shown on each plot and the dividing line between pass and fail (acceptable and unacceptable) is shown at 60%. Figures B-2 and B-3 are the scores for thought problems 1 and 2. These two results show a slight tendency to reward conservatism with problem 1 obtaining a 3% better mean score than problem 2.

Figure B-4 shows an extreme range of values for a prediction that is essentially at the 4σ limits of the data. One can only surmise that the high scores were given for having the correct trend.

Figure B-5 shows the score for problem 4 where the prediction is at the limits of the data error band. The variance here is over the full range of acceptability (>60%), weighted to the low side of acceptability with less than 5% denoting failure or unacceptability of the code predictions.

Figure B-6 for the prediction which oscillates within the full width of the error bands shows a wider spread with more analysts willing to fail the code for this type of behavior. This type of behavior tended to divide the analysts into two camps (based on conversations with individuals following the evaluation); 1) those who gave the code good marks for "trying" to stay on the mean and 2) those who gave it bad marks for errors in local trends and implied numeric instabilities. This division is further amplified in Figure B-7 for the case of the prediction oscillating outside the data bands. Those who felt the code was getting the overall trend and oscillating about the mean gave it higher marks than those who put weight on local trends and implied instabilities.

Figure B-8, for the prediction oscillating within the middle of the error band, also shows the results of the two camps but was generally given high marks because of the nearness of the prediction to the mean.

2. INTERPRETATION OF RESULTS

The results of this survey were discussed with Dr. L. Mathews, an experimental psychologist at Idaho State University. Two important facts were determined. One, these results and their variability represent typical human behavior with respect to perception of worth and second, the actual scores given by the analyst do not have individual significance since they tend to be ordinal (sequential or ranking) in nature. Because of this ranking nature the best way to look at the results is from a percentile standpoint (i.e. what percentage believe the code to be better than a certain level). As the survey was taken with the information that a score less than 60% was to be considered failure, the data can be divided into pass-fail (acceptable-unacceptable) groups and percentile calculations made as a function of group mean scores.

Figure B-9 is a percentile plot of the group mean scores against the percentage of analysts finding the code results acceptable. The data show a remarkable straightline trend with the exception of one data point. This data point represents problem 7 and while the average score was high (82%), there were seven (7) analysts who felt the oscillations should fail the code and this number dropped the percentile acceptance below the trend line. One should note that in this region (95%-98%), the percentile acceptance values are very sensitive in the vertical directions.

IV. USE OF PERCENTILE ACCEPTANCE

Having Figure B-9, the numeric score produced by the quantitative assessment procedure can now be used to obtain the percentage of knowledgeable analysts who would find the results of a particular analysis acceptable. This resulting number, the percentile acceptance (PA) does not mean that a code, or the results of its application, are acceptable or nonacceptable, only that a certain percentage of knowledgeable analysts do find it acceptable. This process does two things, 1) it takes into account the variance in human perception of worth and 2) it does not force a decision on any particular analyst, (although 99 out of a 100 people may think the code is good, there is still a slot for the one person who doesn't like it).



INTEROFFICE CORRESPONDENCE

date February 2, 1979
 to Distribution
 from J. A. Dearien *JAD*
 subject QUANTITATIVE ASSESSMENT - JAD-29-79

The Code Assessment Branch at EG&G Idaho, Inc. is developing a procedure for quantifying the accuracy of computer code predictions. The procedure is based on decision theory wherein subjective or relative weightings are used to obtain quantitative results. As an expert in the field, you are being asked to supply your subjective opinion on several items which will be used in the formulation of our assessment procedure.

Attachment 1 is a representative plot of temperature vs. time from Semiscale. On this plot, you will see the mean of the data, the maximum and minimum bounds of the data (taken from the range of thermocouple readings at that level), and the code prediction. Various behavioral patterns are shown in the plot, i.e., predictions above the mean, below the mean, and outside the data bounds. It is behavior of this type we wish to quantify.

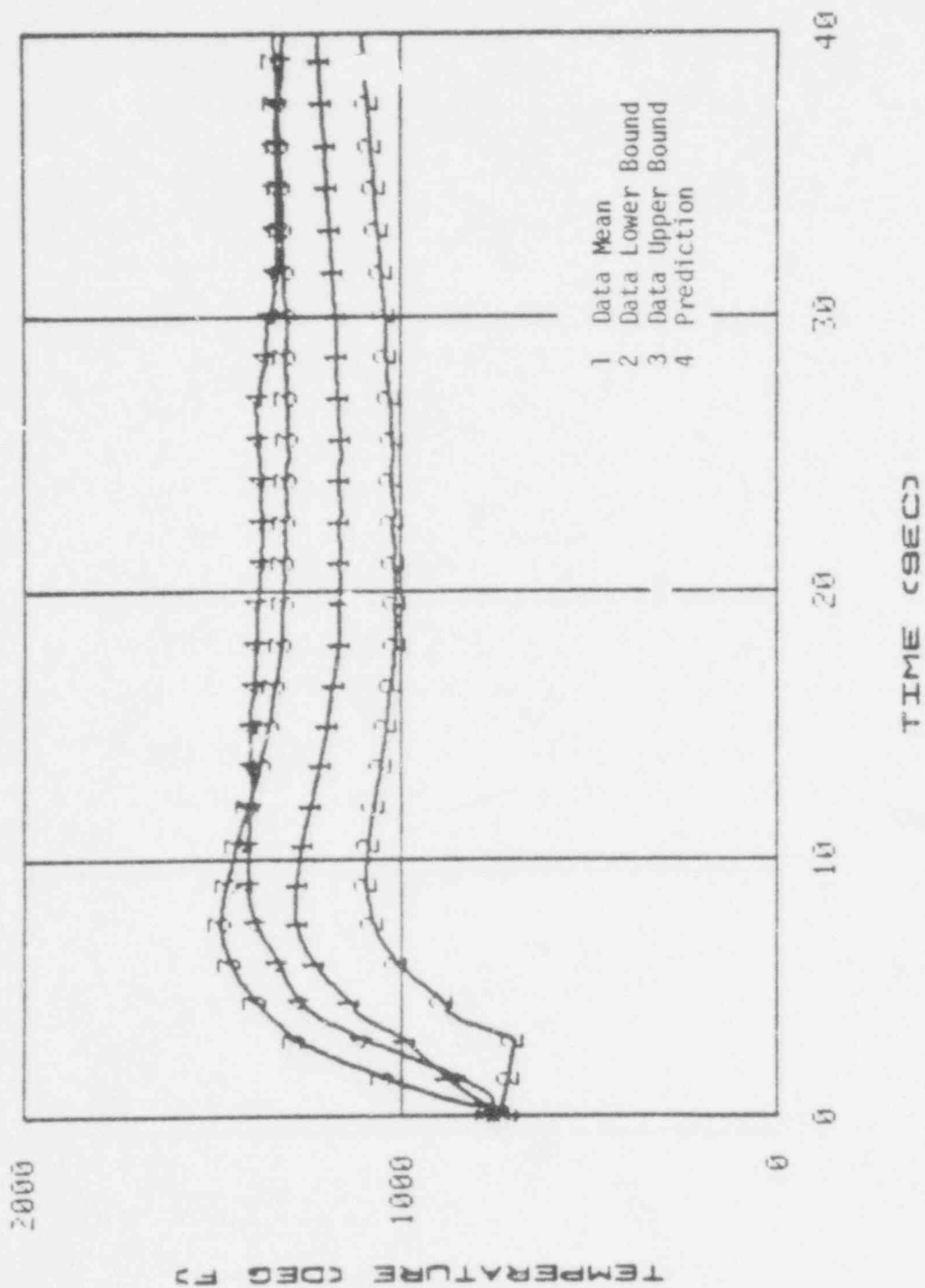
Attachment 2 is a group of idealized comparisons of code predictions and data. The data are shown as the mean of the data with error bands. The error bands represent values for which there is a 95% probability that the true value is within these bands. You are asked to score the various comparisons from 0% to 100% (100% would be the situation where the prediction overlays the mean of the data) much as you would grade a college paper, i.e., <60% = Failure, 60-70% = D, 70-80% = C, 80-90% = B, and 90-100% = A. The reason for this being that the one thing in common between all of us is our association with scholastic grades (however, please use a numeric value and not the alphabetic).

The scoring should represent your assessment of the accuracy of the code in making a best estimate prediction of the behavior. As we are attempting to develop a procedure which will have wide spread acceptance/understanding, you are requested to minimize (if not eliminate) any qualification or hedging of the scores. Comments would be appreciated.

Please return this information to P. H. Vander Hyde by February 7, 1979.

Thank you for your participation.

B-7



Score _____

(1)

Score _____

(2)

Score _____

(3)

Score _____

(4)

Score _____

(5)

Score _____

(6)

Score _____

(7)

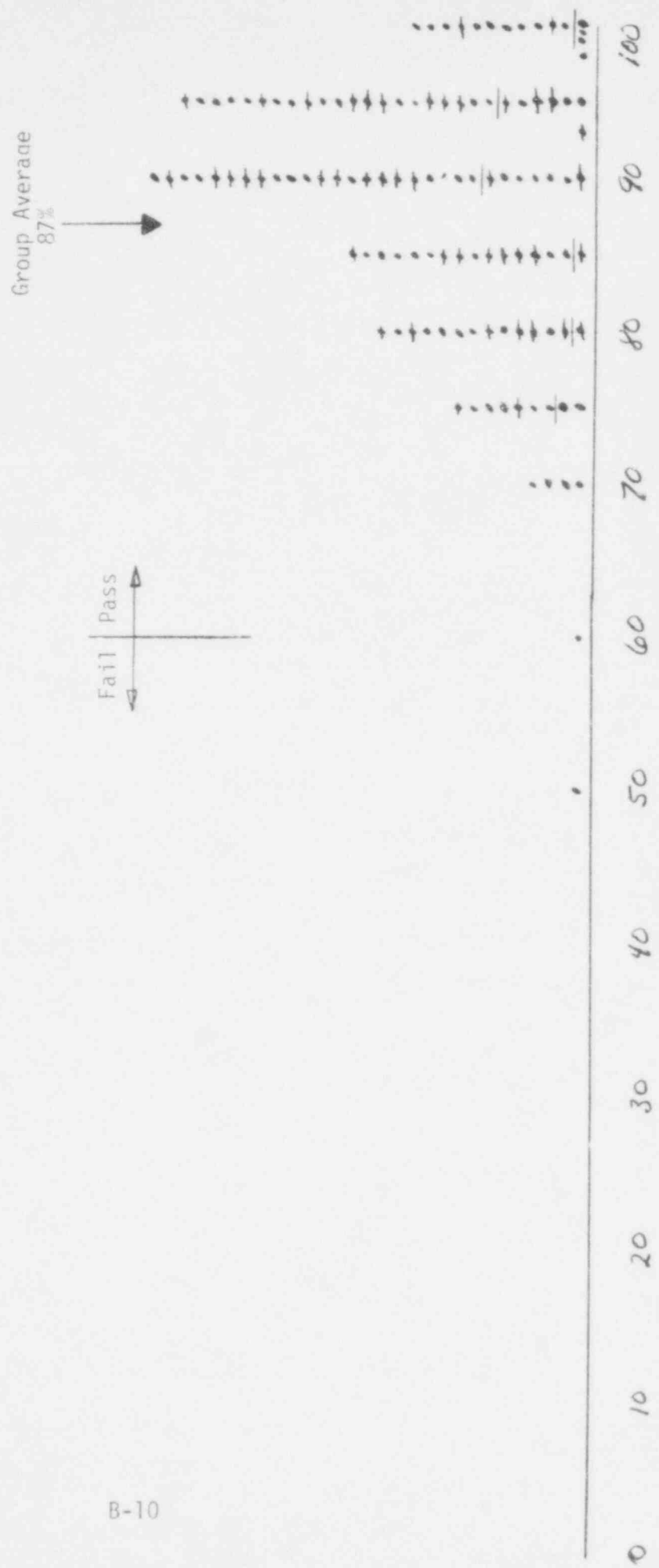
--- Data Bands
— Data Mean
--- Calculation

B-9

Name _____

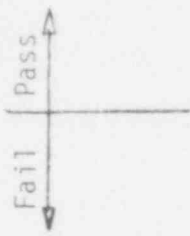
Figure B-1

521 010



Problem 1
Figure B-2

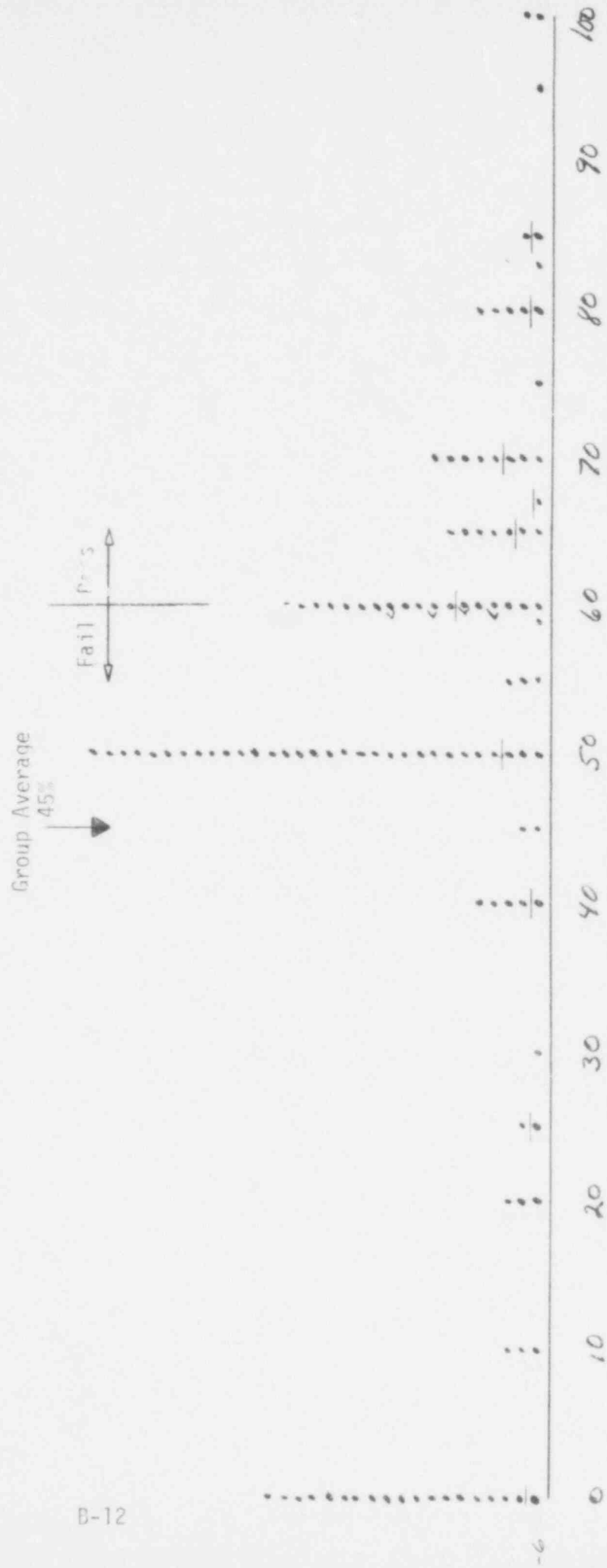
Group Average
84%



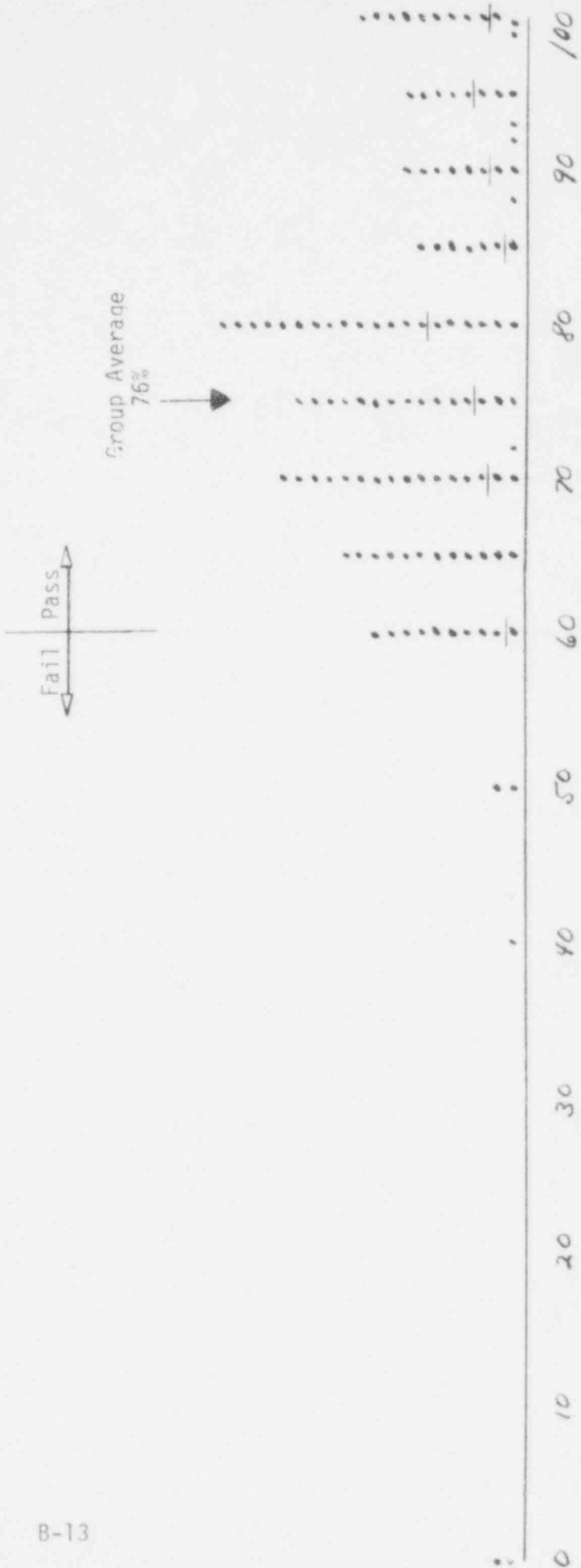
B-11



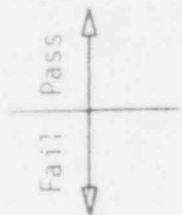
Problem 2
Figure B-3



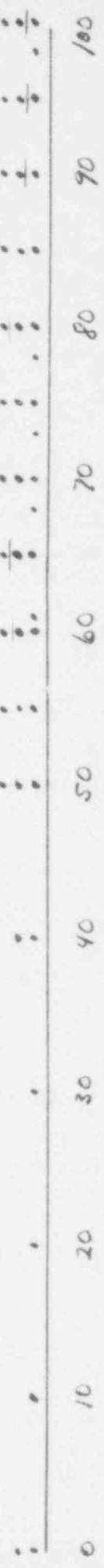
Problem 3
Figure B-4



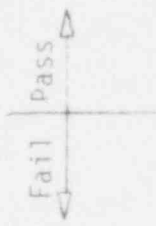
Problem 4
Figure B-5



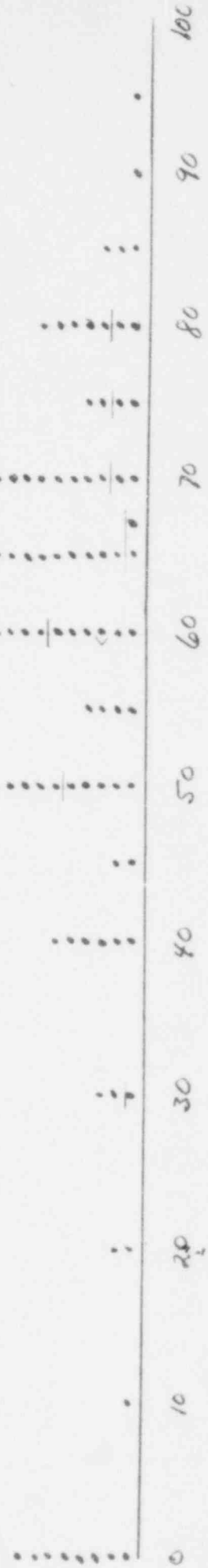
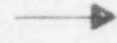
Group Average
74%



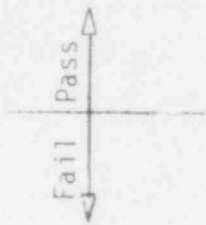
Problem 5
Figure B-6



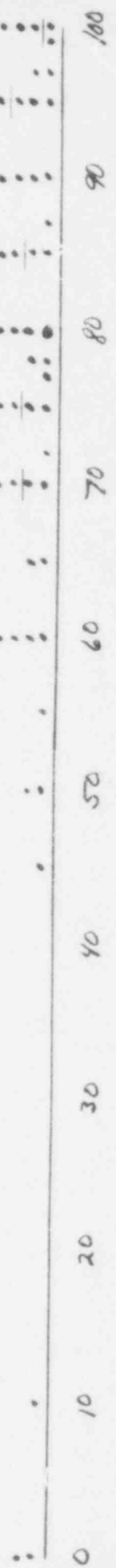
Group Average
54%



Problem 6
Figure B-7



Group Average
82%



Problem 7

Figure B-8

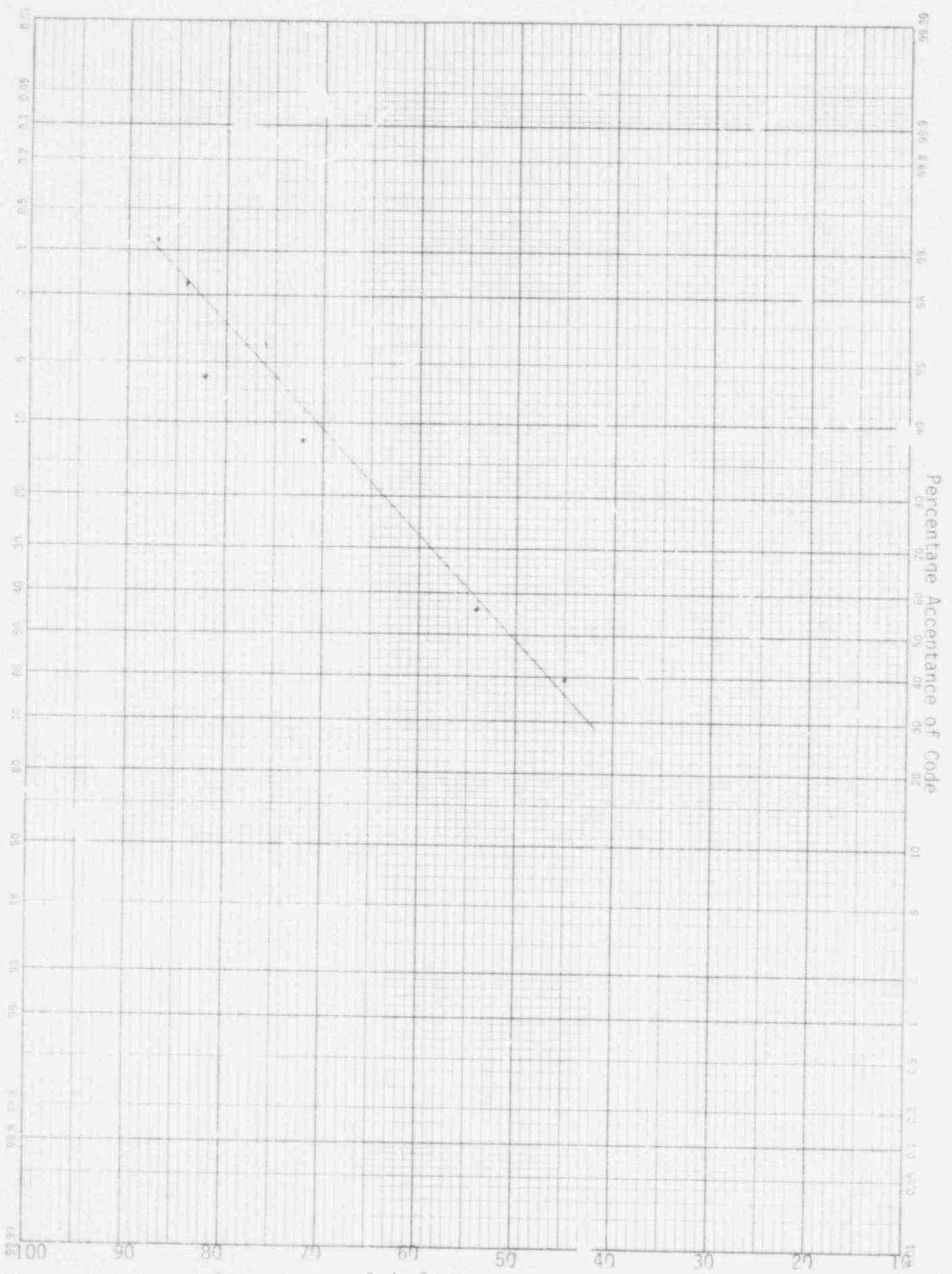


Figure B-9 Code Score
B-17

GEORGE W. KESLER CO.