APPLICATION OF DATA ANALYSIS TECHNIQUES TO NUCLEAR REACTOR SYSTEMS CODE TO ACCURACY ASSESSMENT

R.F. KUNZ, G.F. KASMALA, C.J. MURRAY, J.H. MAHAFFY
Applied Research Laboratory,
Pennsylvania State University,
University Park, Pennsylvania,
United States of America

Abstract

An automated code assessment program (ACAP) has been developed by the authors to provide quantitative comparisons between nuclear reactor systems (NRS) code results and experimental measurements. This software was developed under subcontract to the United States Nuclear Regulatory Commission for use in its NRS code consolidation efforts. In this paper, background on the topic of NRS accuracy and uncertainty assessment is provided which motivates the development of and defines basic software requirements for ACAP. A survey of data analysis techniques was performed, focusing on the applicability of methods in the construction of NRS code-data comparison measures. The results of this review process, which further defined the scope, user interface and process for using ACAP are also summarized. A description of the software package and several sample applications to NRS data sets are provided. Its functionality and ability to provide objective accuracy assessment figures are demonstrated.

1. INTRODUCTION

In recent years, the commercial nuclear reactor industry has focused significant attention on NRS code accuracy and uncertainty issues. There are several reasons for this including the inherent safety (and concomitant code reliability and licensing) concerns associated with nuclear reactors. Over a decade ago, the United States Nuclear Regulatory Commission (hereafter NRC) initiated an international effort to improve and standardize the assessment of thermal hydraulic (TH) systems codes ([1, 2], for example). Prior to that, processes for assessing the performance of TH codes were largely qualitative and subjective, and thereby difficult to use in plant safety certification. In 1984, the NRC organized the International Thermal Hydraulic Code Assessment and Applications Program (ICAP), a major goal of which was the assessment of TH codes using relevant data from a wide range of international experimental facilities. Since that time, a large amount of work has been carried out domestically and internationally in this area (see [3-9], for examples).

In 1997, the NRC contracted with the present authors to:

1) Survey available data conditioning and analysis techniques, focusing on their appropriateness in NRS code accuracy and uncertainty assessment
2) Develop software to deploy recommended techniques

This software was to be used for validation in NRC code consolidation efforts and have the potential to be expanded to play a role in determining code adequacy. The ACAP software described herein represents the outcome of this code development effort.

2. NRS CODE ACCURACY ASSESSMENT

2.1. Issues and needs

The issues associated with nuclear reactor systems (NRS) code accuracy and uncertainty assessment are numerous and complex. They include:

1) Scaling of test data.
2) Discretization, model setup and other "user issues".

3) Software reliability.

4) The move towards best estimate vs. conservative acceptability criterion in licensing decisions.

5) Key parameter selection.

6) The wide variety and complex features of the modeled transient physics and experimental data.

7) The inconsistency of measured and computed comparison quantities.

8) The inherent subjectivity of code – experimental comparisons.

9) Uncertainty in experimental measurements.

10) The large and growing available test matrix data base.

11) The lack of a generally applicable suite of code-data assessment tools.

These issues collectively motivate the need for automated code assessment in code consolidation and future development efforts. Ideally, in the future, when assessing a systems code, a single post-processor would be deployed. Based on *all* uncertainties involved, this post-processor would return, at a given confidence level, the maximum expected deviation between code results and reactor for several key parameters [3]. The methodologies embodied in this "ideal" post-processor must address each of the uncertainty components summarized above. The need for such a capability, has motivated a vast amount of research in the past decade (see [4] for a review of much of this work).

Though significant progress has been made in addressing most assessment issues, reliable and general tools to quantify NRS code accuracy are not available today. An important contribution to meeting this ideal would be a universally available assessment tool for the users of NRS codes to post-process results in a way that would return quantitative accuracy measures of code-data comparisons. Such a tool would only address some of the uncertainties in real plant analysis. However, it would be part of a process which validates a code with scaled facility data, contributing an important component to total uncertainty in full scale plant simulations.

Consistent with this view, the goal of the present work has been to initiate a software framework to automatically assess several of the NRS code uncertainty issues summarized above. In particular, the ACAP software package has been developed to objectively and quantitatively compare NRS simulations with data. This package was designed to:

- Draw upon a mathematical toolkit to compare experimental data and NRS code simulations.
- Return quantitative figures of merit associated with individual and suite comparisons.
- Accommodate the multiple data types encountered in NRS environments.
- Incorporate experimental uncertainty in the assessment.
- Reduce subjectivity of comparisons arising from the "event windowing" process.
- Accommodate inconsistencies between measured and computed independent variables (i.e. different time steps).
- Tie into data bases of NRC test data and code results.
- Provide a framework for automated, tunable weighting of component measures in the construction of overall accuracy figures of merit for a given comparison.

So the ACAP tool has been developed to address issues 6-11 summarized above. The scope of this project did *not* include the quantification of the uncertainties introduced by user training issues, discretization issues or code operational issues. Nor has the present work addressed quantification of uncertainty associated with physical models being used on a best estimate basis, nor on scaling uncertainties. However, the present investigators feel that with modest modifications the package could be applied parametrically to complement uncertainty assessment in each of these other areas.

In summary, our fundamental goal has been to develop a numerical toolkit to analyze discrete computational and experimental NR systems data, and, in particular, to use these data analysis procedures to develop code-data and code-code comparison measures. The remainder of this paper summarizes this development effort.

2.2. Categorization of NRS data

NRS data types are classified here into five categories, in order to provide a basis for assessing individual comparison methods. Specifically, scaled NR facilities are instrumented to provide a fairly wide array of key parameter and other data. These include:

I. Key parameters tables.

II. Timing of events tables.

III. Scatter plots of nominally 0-D data (e.g., see Figure 4)[1].

IV. 1-D (in space) steady state data.

V. Time record data (e.g., see Figure 5).

Each of these data types is potentially important in any particular NRS code analysis. The emphasis of the ACAP project is on the latter three. In particular, data conditioning and analysis techniques were assessed and deployed within ACAP for code-data comparisons of data Types III, IV and V. Type V data in particular provides a significant challenge for several reasons:

1) The ubiquitous appearance and relevance of these transient data in NR systems.
2) The typically long record (often $O(10^5)$ time steps) nature of these data, complicated significantly by their *non-stationarity* and *diversity in characteristic features* (long time scale damping, local quasi-periodicity, sudden changes due to active or passive phenomena, chatter (often of high amplitude), dependent variable limits (for volume fraction) between 0 and 1).
3) The fairly significant differences that often appear between computed and measured time trace data (e.g., see Figure 5).

2.3. Data analysis methods

Discrete data analysis is an important element in a wide array of technical disciplines. Techniques to analyze data samples or records lie within the scope of the three overlapping fields: probability and statistics, approximation theory, and time-series analysis. Also, the needs of several engineering and scientific communities have motivated the development of data analysis techniques, which although falling within the three general categories mentioned, are characterized by unique or extended features of relevance to the present research. In particular, methods developed in atmospheric/geologic sciences, economic forecasting, aerodynamic stability, demographics, digital signal processing, pattern (i.e., speech/optical/character) recognition and other fields have relevance to the analysis of NRS data, and could be plausibly adapted to construct systems code-data or code-code comparison measures.

A number of mathematical data analysis methods from these various fields were reviewed for their applicability in the construction of NRS code-data and code-code comparison measures. The goal of the review was to identify issues and techniques to be considered in the development of an automated simulation rating procedure. Details of that review are the subject of a forthcoming publication. These findings defined the scope, user interface and a recommended process for using ACAP. A summary of these issues and findings is provided here:

1) Most of the methods considered can be applied to provide useful quantitative measures of accuracy for at least a subset of NRS data Types III, IV and V.
2) Inappropriate use of some methods can yield incorrect results, that is, return figures of merit that are worse for more accurate simulations. This motivates:
   - Definition of a *robust* comparison measure or suite of measures as one that reliably return bet-

[1] Often these data are rendered "0-D" by collapsing data obtained at multiple space-time coordinates to a single scatter plot.

ter figures-of-merit for superior comparisons and worse figures-of-merit for inferior comparisons.

- That great care be taken in the selection of the suite of analysis tools chosen *for each particular comparison.*

3) The inherent limitations to stationary data of most available methods render straightforward application to NRS Type V data less than rigorous. Trend removal techniques can be brought to bear to preprocess the data, thereby yielding more robust comparison measures, especially when deployed in concert with time-windowing.
4) Experimental uncertainty can be effectively incorporated in code-data accuracy assessment within the framework of the "toolkit" of analysis procedures considered. Experimental uncertainty should be included with the "raw" experimental data in the code reassessment test matrix.
5) Inconsistency between the computed and measured independent variable range and basis (i.e. different time steps) motivates the incorporation of resampling and range trimming conditioners within ACAP. Such "synchronization" is required for most comparisons.
6) For Type V data, techniques that are intrinsically appropriate for *non*-stationary data analysis can be utilized in the construction of comparison measures. These include best approximation fits and, most promising in the view of the present investigators, time-frequency techniques.
7) There is a fundamental lack of rigor in applying basic statistical analysis procedures to most NR systems data. This arises due to non-stationarity of the data and the unavailability of a known distribution of error about its mean. This renders the construction of statistical *inference* measures suspect at best. Basic statistical difference and correlation measures can be deployed to construct *useful* figures of merit, but uncertainty bounds should not be inappropriately constructed.
8) The methods vary widely in range/dimensionality of their returned metrics. This complicates the definition of an overall figure-of-merit, and thereby motivated normalization and range limit scaling in constructing component figures of merit. Specifically, each individual comparison measure is redefined to range from 0 to 1.
9) As indicated in conclusion 2 above, great care must be taken in deploying comparison measures. In particular, for each experimental data set, a demonstrably robust assessment strategy must be developed. The present investigators feel that this requirement defines a *process* whereby expert assessors "calibrate" and document a suite of robust data analyses for each experimental data set in the code reassessment matrix. This assessment *configuration* will in general include preconditioning strategies, data comparison measures, figure-of-merit weighting assembly factors, and should be included with the "raw" experimental data in the reassessment matrix. Such configured assessments will then be used to define ACAP sessions in future code *re*-assessments.

In concert with the above design criteria and method assessment findings, a set of baseline techniques for code-data (or code-code) comparisons, data preconditioning, figure-of-merit-assembly and incorporation of experimental uncertainty were selected and implemented in ACAP. These are listed below. For brevity, the details of the mathematics associated with these methods are not provided in this paper - the reader is referred to forthcoming references including the formal documentation of ACAP.

3. OVERVIEW OF ACAP

3.1. Program description and mechanics

ACAP is a PC and UNIX station based application which can be run interactively on PCs running WINDOWS 95/98/NT or in batch mode on PCs as a WINDOWS console application or in batch mode on UNIX stations as a command line executable. The code will be delivered to NRC with full source code. The interactive and batch PC versions can be modified and recompiled from a WINDOWS "folder" under the Microsoft Visual C++ environment. The batch UNIX version can be modified and recompiled using any C++ compiler which conforms to the C++ draft standard (including the freely available g++/gcc compilers).

A brief summary of the operation of the ACAP is provided here. Figure 1 shows a schematic overview of the structure of the code. Experimental and computational NRS data are input through ACAP data files, which, in their simplest form, contain a table of x-y data and a few data descriptor keywords. The user specifies, either interactively or through front end script files, a suite of data conditioning and data analysis methods to be deployed in quantifying the correspondence between the measurements and the (one-or-more) simulation data sets. This suite of methods is termed the ACAP *configuration*, which can be saved in a file for later use on the current or other data sets. In interactive mode, ACAP displays the data sets and provides standard windows environment interfaces to select and adapt the mathematical methods to be deployed. The code then executes specified data conditioning processes and data comparison measures. Lastly, with user selected weighting, an overall figure-of-merit is constructed quantifying the accuracy of the individual code runs. The results of the ACAP session, including a summary of all selections made, and the component and overall figures-of-merit are output to screen and file.

ACAP is also currently being incorporated within a global auto-validation tool, being developed by Scientech Inc., under separate contract to NRC. That tool automatically runs systems code simulations for a (growing) palette of test cases and generates a prespecified series of plots which include experimental measurements and the results of the multiple simulation runs. ACAP will be invoked in batch mode from this auto-revalidation tool to provide concomitant quantification of the correspondence between each simulation run and, where available, experimental measurements.

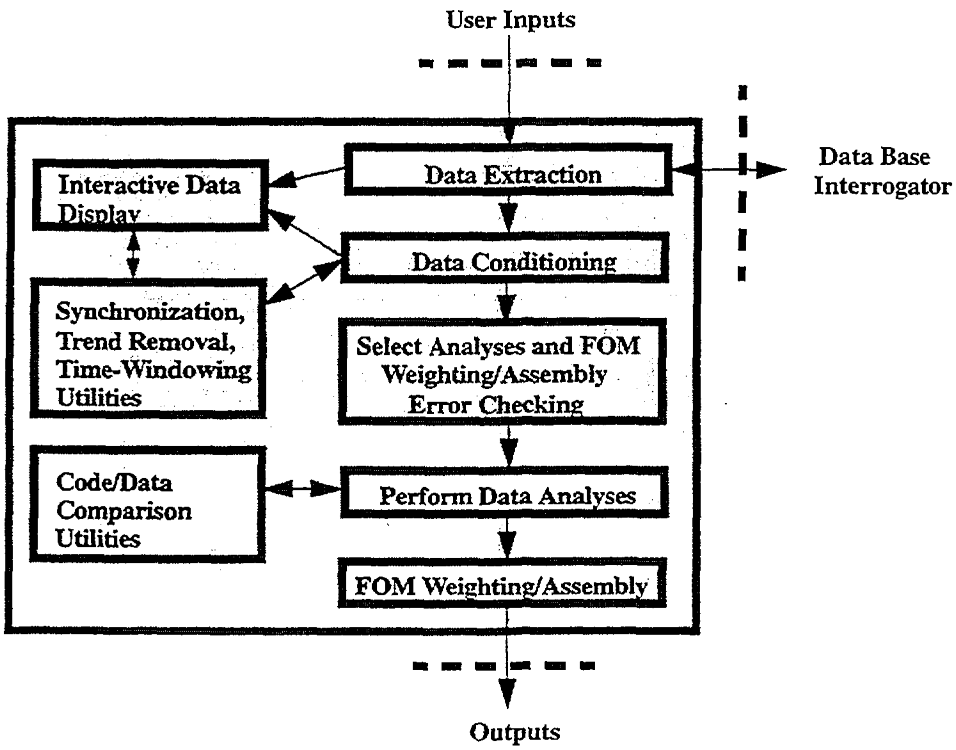

Figure 1. Schematic overview of the structure of ACAP.

3.2. ACAP Methods

Table I summarizes the methods currently installed in ACAP. There are three data conditioning utilities and sixteen data comparison utilities.

Table I. ACAP Methods

	Method	Utility Class
1	D'Auria FFT (DFFT)	Data Comparison Utility
2	Mean Error (ME)	Data Comparison Utility
3	Variance of Error (VE)	Data Comparison Utility
4	Mean Square Error (MSE)	Data Comparison Utility
5	Mean Error Magnitude (MEM)	Data Comparison Utility
6	Mean Relative Error (MRE)	Data Comparison Utility
7	Index of Agreement (IA)	Data Comparison Utility
8	Systematic Mean Square Error (SMSE)	Data Comparison Utility
9	Unsystematic Mean Square Error (UMSE)	Data Comparison Utility
10	Mean Fractional Error (MFE)	Data Comparison Utility
11	Cross-Correlation Coefficient (ρ_{xy})	Data Comparison Utility
12	Standard Linear Regression ($L_{2\text{-standard}}$)	Data Comparison Utility
13	Origin Constrained Linear Regression ($L_{2\text{-constrained}}$)	Data Comparison Utility
14	Perfect Agreement Norm ($L_{2\text{-perfect agreement}}$)	Data Comparison Utility
15	Continuous Wavelet Transform (CWT)	Data Comparison Utility
16	Percent Validated (PV)	Data Comparison Utility
A	Resampling	Data Conditioning Utility
B	Trend Removal	Data Conditioning Utility
C	Time-Windowing	Data Conditioning Utility

Among the data *comparison* utilities, the reader will likely recognize the FFT method of D'Auria [5]. Also available are a number of baseline statistical techniques (methods 2-5, 11-14), Willmott's Index of Agreement (method 7, [10]) and several adapted statistical methods utilized by the atmospheric sciences community (methods 6, 8-10, see [11] for example). Experimental uncertainty is incorporated in a fashion consistent with recent computational fluid dynamic (CFD) code validation work undertaken by Coleman and Stern [12], where a "Percent Validated" metric (method 16) is defined from the fraction of simulation data in a trace which falls within the uncertainty bands of the measurements.

In the authors' view, a particularly attractive comparison tool for NRS code accuracy assessment is the continuous wavelet transform (CWT) measure installed in ACAP (method 15), and some further discussion of this method is provided here. Wavelet transforms are time-frequency techniques which are directly applicable to non-stationary data. As such, if applied consistently, they can provide more accurate representation of local features in a time trace than global transforms (such as the FFT), especially when important features appear at widely varying time scales (as is characteristic of NRS data traces, e.g., see Figure 5). Also, a variety of CWTs are available, each targeting particular features in a signal (the *Morlet* wavelet is implemented in the first ACAP release.)

The available data *conditioning* utilities include particular choices of resampling, trend removal and time windowing methods. A number of authors have observed the usefulness of time-windowing of transient NRS data in order to isolate distinct physical processes, and thereby provide a more focused assessment of simulation strengths and weaknesses. In ACAP, the user may specify up to six time windows. For each window, a fully configured ACAP session is specified. Individual figures-of-merit are computed for each window and a global figure-of-merit is constructed based on a weighted sum of these contributions.

As mentioned above, resampling of the computed data traces is usually appropriate in order that the experiment and simulation have a consistent independent variable basis (i.e. time steps). This issue is relevant only to Types IV and V data where discretization choices and/or numerical stability issues will generally give rise to NRS predictions of dependent variables at different locations in space-time than where the data was taken. Basis consistency is required for all data comparison utilities except methods 1, 8, 15 and 16, though the DFFT and CWT methods can be more accurately deployed if samples are taken at the same time steps. Also, valid application of some trend removal processes, including running averages, require independent variable consistency. As illustrated below, ACAP provides a palette of resampling options to perform this task.

As also mentioned previously, trend removal techniques can be useful in analyzing non-stationary NRS data. For example, a computed time trace may be characterized by a relatively constant underprediction of, say, pressure and also exhibit higher frequency oscillatory differences from measurements. The authors have found that deploying a trend removal step allows these separate effects to be analyzed individually resulting in a more robust code-data comparison configuration. A running-average smoother is installed in ACAP for trend definition and the mechanics are available to separately analyze both differences in the trend itself and in the more nearly stationary "low-pass-filtered" traces. Generally, different data comparison utilities are deployed for the trend and filtered traces, consistent with their differing features and stationarity.

Another issue related to the baseline ACAP methods, mentioned above, is the widely varying range and dimensionality of the various data comparison measures. This complicates the definition of an overall figure-of-merit, and thereby motivated normalization and range limit scaling in constructing component figures of merit. Specifically, each individual comparison measure was redefined to range from 0 to 1, corresponding to worst possible and best possible agreement between a given computed trace and experiment. The process implemented to do so comprised two steps. First, all dimensional figures-of-merit are non-dimensionalized with respect to the experimental dependent variable range $|O_{max}-O_{min}|$. This "sizes" the different metrics such that $O(10^0)$ errors (i.e. order of 100 % errors) between traces will give rise to $O(10^0)$ metric values. The second step is to, where necessary, modify these "sized" metric definitions so that they independently return figures of merit between 0 and 1. Several of the comparison metrics have ranges between 0 and ∞ or $-\infty$ and ∞. For all of these except the DFFT and CWT measures, a method for achieving the desired range of [0,1] is implemented, somewhat arbitrarily, as $FOM = 1/(|\eta|+1)$, where η is the non-dimensionalized metric. The DFFT, CWT, MRE and ρ_{xy} metrics require somewhat different treatment, the form of which will be available in forthcoming references including the formal documentation of ACAP. Since chosen normalization and range limit scaling of the data comparison utilities in ACAP are somewhat arbitrary, ACAP users may wish to invoke alternate definitions or simply consider the "raw" metrics returned by the baseline methods. This latter option is available in the code, the former would require some modest C++ code modifications.

4. SAMPLE ACAP SESSIONS AND RESULTS

To date, ACAP has been deployed on a large number of test cases, as we evolve the capabilities of, and our own experience with, the tool. Several example applications are presented in this section. The purpose of these demonstration cases is to illustrate the functionality of the code and its ability to provide objective accuracy measures.

4.1. D'Auria sample experimental and calculated time traces

The functionality of ACAP is demonstrated using the D'Auria "sample" data used by the University of Pisa group in numerous recent publications ([5], for example). This data (which was digitized here from the cited reference) provides one experimental and six representative "systems code" data traces. This artificial data was originally constructed to capture several features of typical NRS transients, as well as the widely varying differences between simulation and experiment that can occur.

The data was input to ACAP and displayed graphically as reproduced in Figure 2. Four component figures-of-merit were chosen: DFFT, MSE, ρ_{xy} and CWT. These were selected and each given a

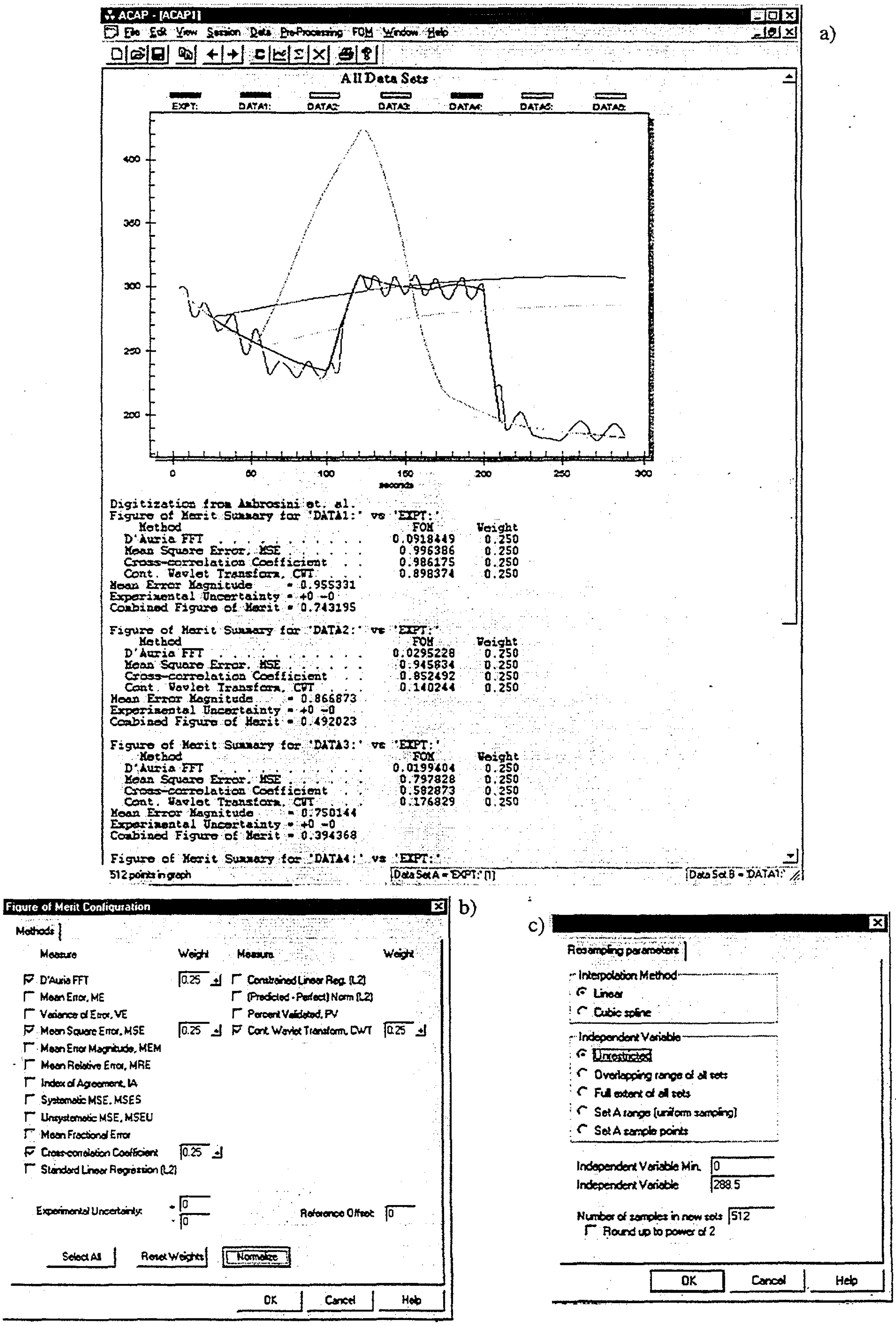

Figure 2. Elements of ACAP Interface. a) D'Auria data displayed in ACAP main window with results of comparison assessment for sample "code" results. b) Figure-of-merit configuration dialog. c) Resampling dialog.

weight of 0.25 in the *Figure of Merit Configuration* dialog box, as also shown in the figure. The assessment analysis was then run and the results displayed below the data plot. For the rather arbitrary selections made here, ACAP returns consistently superior component and overall figures-of-merit for sample trace 1. The CWT measure is illustrated in Figure 3, where the locus of points generated by the CWT for each time trace is plotted in the AA (average amplitude), 1/WF (inverse frequency) plane. The percentage of points within the illustrated acceptance boundary defines the figure-of-merit.

4.2. Type III data assessment

In order to illustrate the use of ACAP for producing figures-of-merit for Type III data, use is made of an, as yet, unpublished two-phase pressure drop analysis performed at Penn State. Several different popular empirical correlations were used to predict the two-phase pressure drop for water flowing upwards through a heated tube at 1000 psia. Comparisons were made against experimental data from [13]. Figure 4a shows a predicted vs. measured scatter plot comparison of the experimental data against Martinelli-Nelson correlation predictions. Figure 4b shows a similar comparison using results from the Freidel empirical correlation.

Visual inspection of the data illustrates that the Freidel model is clearly more accurate over the entire range of pressures analyzed. The issue here is whether this behavior can be captured quantitatively through some figure-of-merit strategy using ACAP. After importing the relevant data into the code, the ACAP session was configured to make use of the metrics that may reasonably be applied to Type III data. No data preconditioning was necessary because the data were already synchronized before being imported into the code. Table II provides a summary of the individual figures-of-merit returned by ACAP for each metric, the weighting factors used, and an overall assessment value, for each pressure drop correlation.

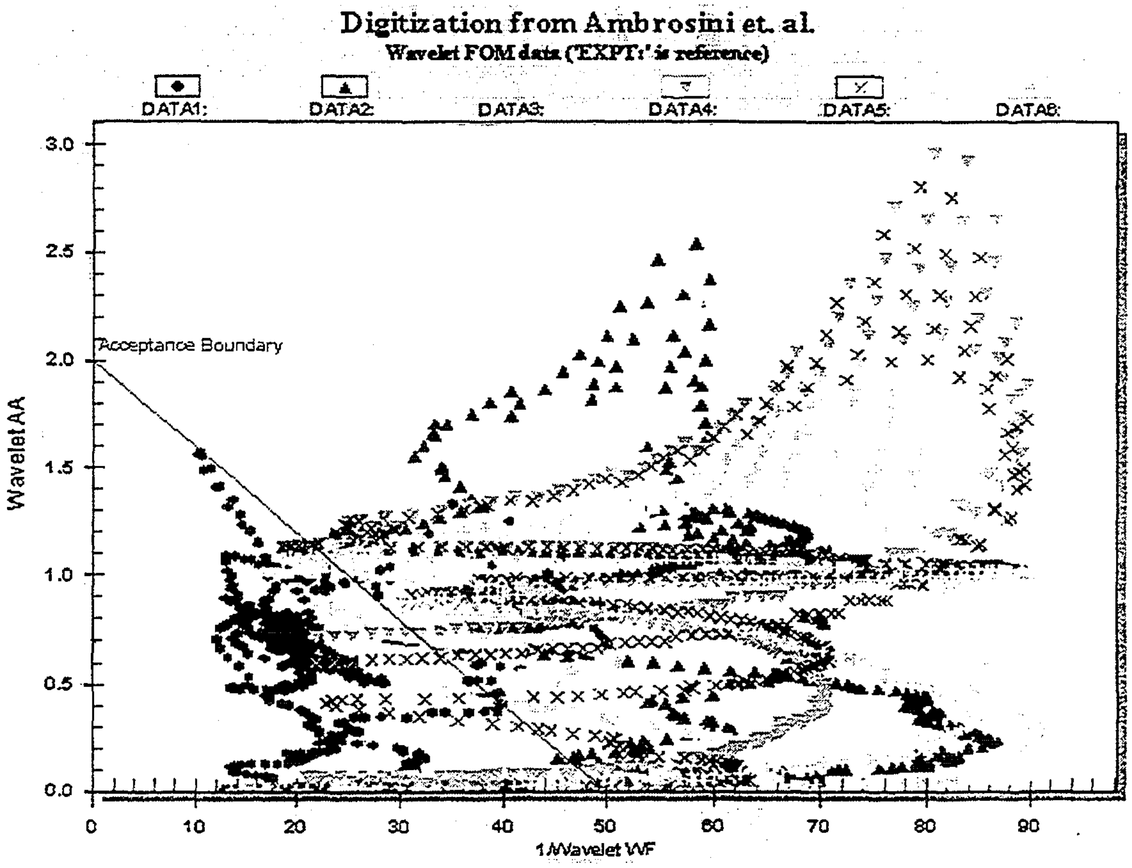

Figure 3. Display of continuous wavelet transform applied to D'Auria data, illustrating locus of points in AA-1/WF plane and acceptance boundary.

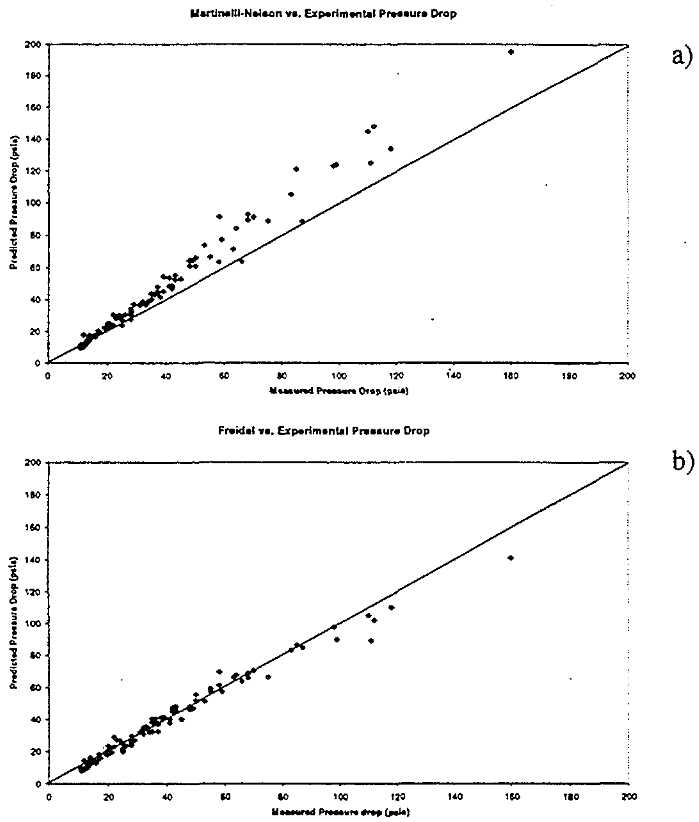

Figure 4. Sample Type III data comparisons. Predicted vs. measured scatter plot comparison of pressure drop experimental data [13] against a) Martinelli-Nelson correlation predictions and b) Freidel empirical correlation.

Table II. Comparison of ACAP Results for Presented Type III Data

Method	M-N Model	Freidel Model	Weight
Mean Error	0.948	0.995	0.077
Variance of Error	0.996	0.999	0.077
Mean Square Error	0.994	0.999	0.077
Mean Error Magnitude	0.947	0.980	0.077
Mean Relative Error	0.923	0.984	0.077
Index of Agreement	0.965	0.993	0.077
Systematic Mean Square Error	0.956	0.999	0.077
Unsystematic Mean Square Error	0.974	0.999	0.077
Mean Fractional Error	0.484	0.882	0.077
Cross-Correlation Coefficient	0.990	0.989	0.077
Standard Linear Regression	0.984	0.997	0.077
Origin Constrained Linear Regression	0.996	0.997	0.077
Perfect Agreement Norm	0.992	0.997	0.077
Combined Figure-of-Merit	**0.935**	**0.986**	

For each method, except one, the figures-of-merit are seen to be closer to unity for the Freidel case, indicating better agreement to the experimental data. The exact sensitivity of a particular metric to changes in the pressure drop correlation is seen to vary significantly. In some cases, the figures-of-merit only differ in the third decimal place while for others, the differences occur in the first or second decimal place. While it is not the purpose of this paper to present a detailed discussion of this behavior, these differences in sensitivity derive, in part, from the ability, or lack thereof, of a particular metric to capture a particular trait of the data set. For example, the closely corresponding values of the ρ_{xy} and $L_{2\text{-constrained}}$ metrics suggest that both models *correlate* quite well with data. Taken with the significant differences in MFE, which is a good measure of *bias* in the predictions, one can conclude that the shortcomings of the Martinelli-Nelson model are principally due to a consistent over-prediction of the pressure drop. These observations, and other which can be drawn from the results in Table II, illustrate the utility of selecting multiple figures-of-merit to capture different features in NRS code comparisons. Here, equal weighting was arbitrarily given to each method in constructing the overall merit value. In general, when constructing figure-of-merit configurations, the user would need to analyze the data, *identify the traits which need to be captured, and make appropriate decisions as to which metrics ought* to be used and how they should be weighted.

4.3. Type V data assessment

This next example illustrates the use of ACAP with Type V data. Figure 5 shows a comparison between the predicted and measured rod surface temperature at a particular axial level during the core heatup and reflood stages of a FLECHT SEASET vs. TRAC-B simulation [14]. Two different reflood heat transfer models within TRAC-B were deployed - the original and a newer model.

Again, after importing the relevant data into the code, methods were selected to construct a figure-of-merit. In particular, most of the metrics used in the previous example were retained and the DFFT and CWT methods were also implemented. Because the data were not synchronized before being imported, ACAP's resampling feature was used to linearly interpolate between each respective model's data points, and subsequently generate new "predicted" data points which correspond in time to those in the experimental data set. Furthermore, because the original reflood model results only ran out to about 330 seconds, the data used to construct the figures-of-merit was limited to $t < 330$ using ACAP's time-windowing feature. After these configuration steps were performed, ACAP generated the figures-of-merit for each simulation and these are summarized in Table III.

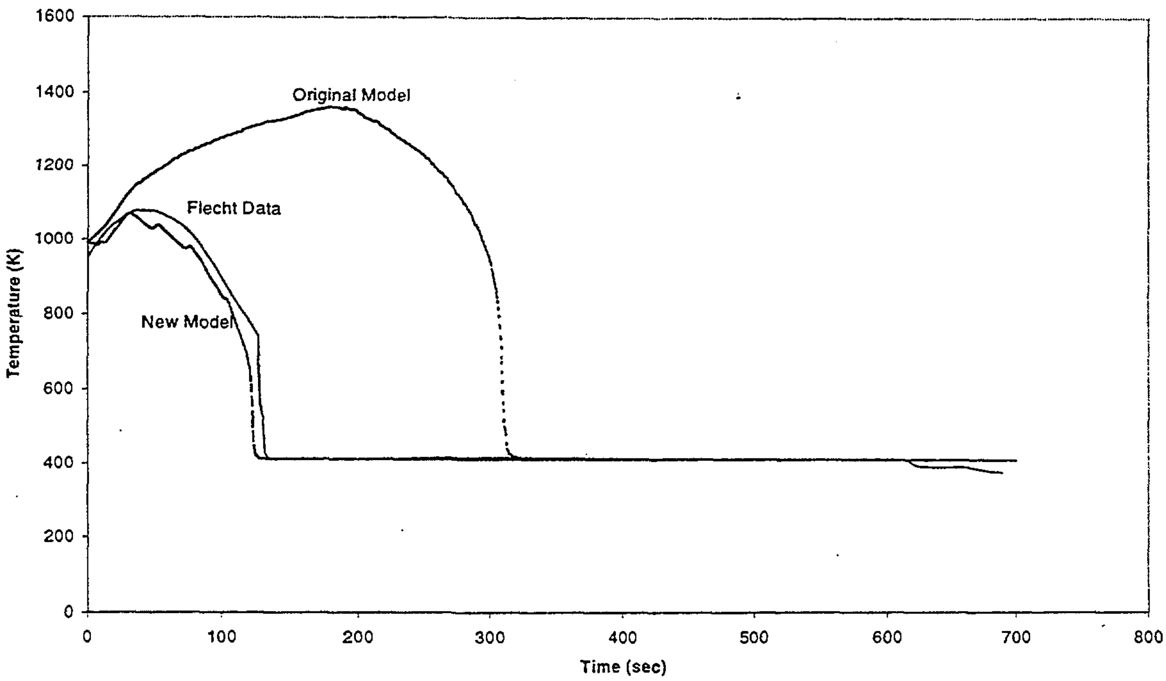

Figure 5. Sample Type V data comparisons. Predicted vs. measured rod surface temperatures during heatup and reflood of a FLECHT SEASET transient.

Table III. Comparison of ACAP Results for Presented Type V Data

Method	Original Model	New Model	Weight
D'Auria FFT	0.035	0.141	0.077
Mean Error	0.555	0.969	0.077
Variance of Error	0.779	0.996	0.077
Mean Square Error	0.519	0.995	0.077
Mean Error Magnitude	0.555	0.967	0.077
Mean Relative Error	0.689	0.985	0.077
Index of Agreement	0.421	0.992	0.077
Mean Fractional Error	0.052	0.556	0.077
Cross-Correlation Coefficient	0.037	0.988	0.077
Standard Linear Regression (L_2 Norm)	0.926	0.998	0.077
Origin Constrained Linear Regression (L_2 Norm)	0.981	0.999	0.077
Perfect Agreement Norm (L_2 Norm)	0.979	0.998	0.077
Continuous Wavelet Transform	0.665	0.864	0.077
Combined Figure-of-Merit	**0.553**	**0.880**	

For each case, the merit values do generally get better for the modified reflood model simulation, as expected. The overall figure of merit for the new reflood model is 0.880 while for the original code case, it is only 0.553.

4.4. Type V data assessment including experimental uncertainty

As a final demonstration example, a Type V data assessment is performed for a case which has an experimental uncertainty available. Integrated mass flow through an automatic depressurization system for an OSU SBLOCA case is considered (NRC12). An experimental uncertainty is known for this quantity. Figure 6a shows the ACAP interface display of the experimental data with a RELAP5 solution and an artificial systems code solution (here simply a straight line). The percent validated metric, defined above (section 3.2), is utilized to assess the relative accuracy of the two simulations. Though the artificial code solution exhibits significant differences from both RELAP5 and measured values, the RELAP5 and artificial simulations return PV values of 0.42 and 0.40 respectively. Therefore the artificial data cannot be deemed much less accurate than the RELAP5 simulation if taken in light of the experimental uncertainty. This is further illustrated in Figure 6b where the absolute error of the two simulations is plotted with the experimental uncertainty bands.

5. CONCLUSIONS

An automated code assessment program (ACAP) developed under subcontract to the United States Nuclear Regulatory Commission for use in its NRS code consolidation efforts was presented in this paper. Some background on the topic of NRS accuracy and uncertainty assessment was provided which motivated the development of and defined basic software requirements for ACAP. A description of the software package and several sample applications to NRS data sets were provided. Its functionality and ability to provide objective accuracy assessment figures were demonstrated.

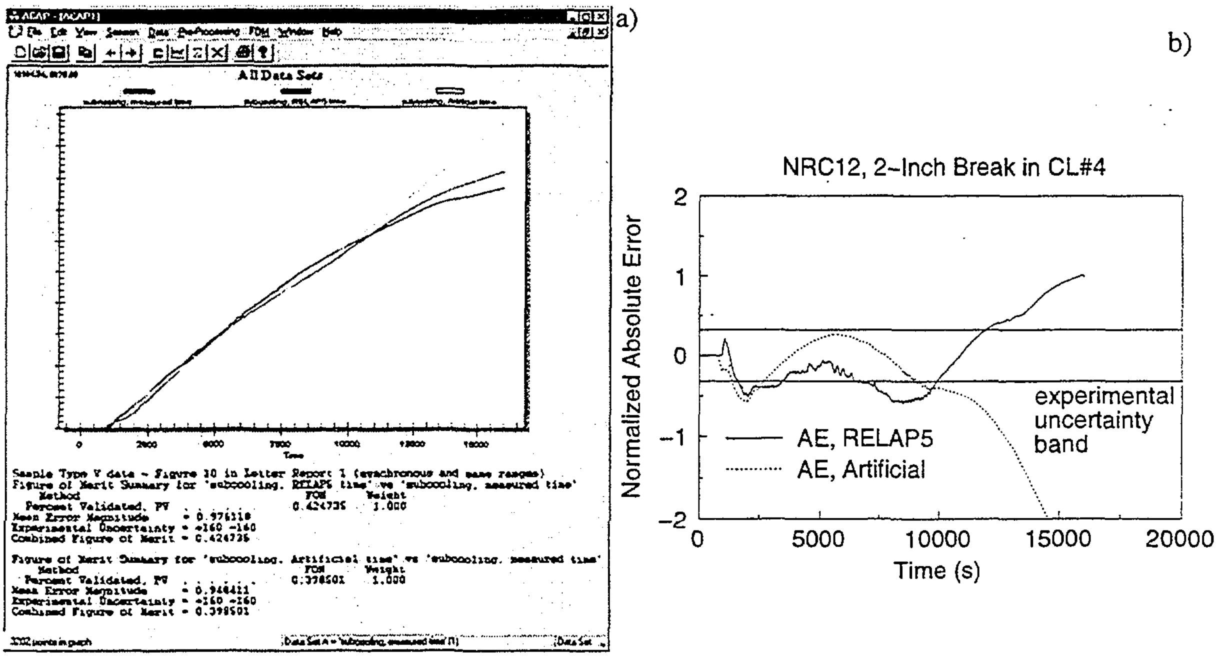

Figure 6. Sample Type V data comparisons. Predicted vs. measured integrated mass flow through an ADS vs. time for NRC12 case. a) ACAP display of data and assessment output. b) Plot of absolute error for two simulations with experimental uncertainty.

ACKNOWLEDGEMENTS

This work was performed under United States Nuclear Regulatory Commission Contract NRC-04-97-046, Task Order #3, with contract monitor Dr. Jennifer Uhle.

REFERENCES

[1] BESSETTE, D.E., ODAR, F., "The U.S. Nuclear Regulatory Commission (NRC) Program on the Development and Assessment of Thermal Hydraulic Systems Codes," NUREG/CP-0080, Vol. 1 (1986).

[2] KMETYK, L.N., BYERS, R.K., ELRICK, M.G., BUXTON, L.D., "Methodology for Code Accuracy Quantification," NUREG/CP-0072, Vol. 5 (1985).

[3] WILSON, G.E., CASE, G.S., BURTT, J.D., EINERSON, J.J., HANSON, R.G., "Development and Application of Methods to Characterize Code Uncertainty," NUREG/CP-0072, Vol. 5 (1985).

[4] D'AURIA, F., LEONARDI, M., GLAESER, H., POCHARD, R. "Current Status of Methodologies Evaluating the Uncertainty in the Prediction of Thermal-Hydraulic Phenomena in Nuclear Reactors," from Two-Phase Flow Modeling and Experimentation, 1995, ed: Celata, Shah, pp. 501-509 (1995).

[5] AMBROSINI, W., BOVALINI, R., D'AURIA, F., "Evaluation of Accuracy of Thermal-Hydraulics Code Calculation," *Energianucleare*, Vol. 7, No. 2, pp. 5-16 (1990).

[6] D'AURIA, F., GALASSI, G.M., "Code Assessment Methodology and Results," IAEA Technical Committee/Workshop on Computer Aided Analysis, Moscow (1990).

[7] D'AURIA, F., DEBRECIN, N., GALASSI, G.M., "Outline of the Uncertainty Methodology Based on Accuracy Extrapolation," *Nuclear Technology*, Vol. 109, January, pp. 21-38 (1995).

[8] D'AURIA, F., GALASSI, G.M."Code Validation and Uncertainties in System Thermalhydraulics," submitted *Journal of Progress in Nuclear Energy* (1997).

[9] SCHULTZ, R.R., "International Code Assessment and Applications Program: Summary of Code Assessment Studies Concerning RELAP5/MOD2, RELAP5/MOD3, and TRAC-B," NUREG/IA-0128 (1993).

[10] WILLMOTT, C.J., "Some Comments on the Evaluation of Model Performance," Bulletin of the American Meteorological Society, Vol. 63, No. 11, pp. 1309-1313 (1982).

[11] KU, J.Y., RAO, S.T., RAO, K.S., "Numerical Simulation of Air Pollution in Urban Areas: Model Performance," Atmospheric Environment, Vol. 21, No. 1, pp. 213-232 (1987).

[12] COLEMAN, H.W., STERN, F., "Uncertainties and CFD Code Validation," *Journal of Fluids Engineering*, Vol. 119, No. 4 (1997).

[13] MATZNER, B., J. E. CASTERLINE, E. O. MOECK, AND G. A. WIKHAMMER, "Critical Heat Flux in Long Tubes at 1000 psi With and Without Swirl Promoters," Presented at the Winter Annual Meting of the ASME, Chicago, Illinois, November 7-11 (1965).

[14] PAIGE, D.R., "Assessment of Improved Reflood Model in TRAC-BF1/MOD1", MS Thesis in Nuclear Engineering, The Pennsylvania State University, (1998).