LA-UR -78-805

CONF-780507--8

TITLE: AN APPLICATION OF STATISTICAL TECHNIQUES TO THE ANALYSIS OF REACTOR SAFETY CODES

AUTHOR(S): Michael D. McKay
John W. Bolstad
Dave E. Whiteman

MASTER

SUBMITTED TO: ANS Topical Meeting on Probabilistic Analysis of Nuclear Reactor Safety
Los Angeles, CA, May 8-11, 1978

los alamos scientific laboratory of the University of California
LOS ALAMOS, NEW MEXICO 87545

Form No. 836
St. No. 2629
1/75

UNITED STATES
ENERGY RESEARCH AND
DEVELOPMENT ADMINISTRATION
CONTRACT W-7405-ENG. 36

EMB

AN APPLICATION OF STATISTICAL TECHNIQUES TO THE ANALYSIS OF REACTOR SAFETY CODES[a]

Michael D. McKay John W. Bolstad David E. Whiteman

Los Alamos Scientific Laboratory
Los Alamos, New Mexico 87545

ABSTRACT

The objectives of this study on the application of statistical techniques to the analysis of reactor safety codes are the identification of the input variables which have a significant influence on the output variables (sensitivity analysis) and the determination of the effect of uncertainty in input values on the output variables. The Latin hypercube sampling (LHS) procedure is presented as an input value selection procedure. The partial rank correlation coefficient (PRCC) coupled with the LHS procedure is presented as a quantitative measure of sensitivity. An examination of the PRCC variability and an analysis of TRAC for a Semiscale test are presented.

I. INTRODUCTION

This paper describes the study being done at the Los Alamos Scientific Laboratory on the application of statistical techniques to the analysis of reactor safety codes. The objectives of the study are the identification of the input variables which have a significant influence on the output variables (sensitivity analysis) and the determination of the effect of uncertainty in the input values on the output variables.

Two aspects of computer code analysis are of particular interest and concern: (1) the long-running time of the codes and (2) the large number of input variables to be studied. While the first aspect limits the number of computer runs, both aspects necessitate a flexible investigation strategy in order to support a variety of analyses from the same set of computer runs (data). The investigation strategy includes both the selection of input values and the application of analysis techniques. These areas will be discussed in the following sections, together with an analysis of the TRAC[1] computer code for a Semiscale test.

---

[a]W. J. Conover, a visiting staff member from Texas Tech University, is a major contributor to the statistical work.

II. INPUT VARIABLES

Input variables are those quantities which are to be investigated. Other quantities may be required to evaluate the code, but they are not considered to be inputs in this paper.

Input variables may include:

(1) quantities related to physical characteristics, such as initial conditions;
(2) coefficients in fitted functions (engineering correlations), which are estimated from experimental data;
(3) parameters whose nominal values are derived from theory or engineering judgment;
and (4) quantities related to the numerics of the model, such as the choice of time step for solving a differential equation numerically.

Sometimes an input variable has inherent variability and can naturally be regarded as a probabilistic random variable. In other cases, when an input value has to be estimated from data, a probability distribution can describe the confidence of the analyst in his estimate. In general, it is convenient to associate with each input variable a probability distribution which determines the region of interest in the variational study.

III. SELECTION OF INPUT VARIABLE VALUES

The anticipated analyses should dictate the selection of not only the input variable values, but also the variables which are to be part of the study. The analyses could include: (1) calculation of general statistical quantities like means and standard deviations, (2) calculation of quantities needed for probability statements or uncertainty bands, (3) calculation of sensitivity functions, and (4) curve fitting of the output variable.

For a selection procedure to be effective, the generated input values should adequately span the input space. Moreover, the statistical estimators should have an inherently high degree of precision so that the number of computer runs can be kept small. The statistical sampling procedure described in the next section allows both the statistical and the sensitivity portions of code analysis to be performed on the same set of data. Also, the number of computer runs is not a function of the number of input variables as it is in one-at-a-time variation and in other systematic designs. However, particular analyses may impose lower limits on the number of computer runs.

IV. LATIN HYPERCUBE SAMPLING

The Latin Hypercube Sampling (LHS) procedure[2] for selecting values of input variables for n computer runs first divides the range of each input variable into n intervals of equal probability content. If each input variable had a uniform probability distribution, as is the case in Quota

Sampling[3], the n intervals would be of equal length. After a value is randomly sampled from each interval, the n values of each input variable are then randomly assigned to the n computer runs.

The LHS procedure possesses several desirable properties. An increase in the precision of estimators of the mean and cumulative distribution function can be expected over the corresponding estimators obtained from a completely random (ordinary Monte Carlo) sample when the output variable is a monotonic function of the input variables[4]. The reduction in variance of estimators can mean a substantial savings in terms of the required number of computer runs. In Ref. 4, the precision of an estimator of mean pressure as a function of time was examined for a random sample, a stratified sample, and a Latin hypercube samples. The standard deviations of the estimators is shown in Fig. 1. Each sampling plan contained 16 observations at each time point and was replicated 50 times.

Another desirable property of LHS is in the area of sensitivity analysis. Since the range of each input is stratified, the points in the input space remain distinct when projected into subspaces of fewer input variables. If variables are omitted in an analysis, as they might be if they were judged umimportant, the remaining variables still constitute a Latin hypercube sample.

The LHS procedure has also demonstrated a desirable property for fitting the output variable to an empirical function of the inputs, i.e., in surface fitting or response surface analysis. The surfaces fitted with the LHS procedure seem to have good predictive capabilities. The reason for this is not known at this time; however, it is believed the property arises from the way the LHS procedure spreads out the values in the input space. Two topics related to the spread of data points are currently under study, clustering and coverage. Clustering is the tendency of input values to form localized groups in the input space and subspaces of the input space. Coverage is concerned with the dispersion of the input values, and in particular, the convex hull of the points, in the input space and subspaces.

V. MODEL VALIDATION AND SENSITIVITY ANALYSIS

Model validation can be defined as the process of assessing the "truth" or correctness of a model. In actual practice, model validation means the establishment of the degree of credibility one can have in the forecasts or predictions of a model, usually for hypothesized events. One way of establishing model credibility is through the comparison of model predictions with observed data. Statistical goodness of fit tests which have a theoretical basis can be used under sets of assumptions usually containing almost-normality, almost-linearity, and almost-homogenity of variances. These assumptions tend to limit the scope of the models to being regression-like or strictly probabilistic. In the area of stochastic processes, the scope is expanded to prediction variables treated as parametric functions of time. It is not clear, however, what direction one should take in looking at goodness of fit and model validation for the more general model with multiple, time-dependent outputs which are nonlinear functions of the inputs.

Another way of establishing model credibility is the use of sensitivity analysis, by which we mean the general study of the variation of a model output as a function of variation in model inputs. Rather than being limited to local measures, such as those calculated from small, one-at-a-time perturbations about nominal values, sensitivity analysis should include global measures which characterize variability in terms of the entire ranges of values encountered in model evaluations.

Used in its general sense, sensitivity analysis has two desirable features in the area of model validation. First, by an intensive variational study, model deficiencies can be detected and subsequently corrected. Second, if a model survives a vigorous sensitivity analysis its credibility as a relevant forecasting tool is increased.

VI. THE PARTIAL RANK CORRELATION COEFFICIENT

The Partial Rank Correlation Coefficient (PRCC) is a measure of sensitivity. The PRCC is the partial correlation coefficient[5] (PCC) evaluated using rank-transformed data. The PCC measures the degree of linear association between two variables from a multivariate structure after adjusting for the linear effects of the remaining variables. Hence, the PRCC measures the degree of monotonic association in the same way that the PCC measures linear association.

The PRCC is bounded in absolute value by 1. Values near +1 indicate a strong direct (positive) association between the output and input, while values near -1 indicate a strong inverse (negative) association. Values near 0 indicate a lack of association.

The PRCC preserves the time aspect of the output variable. Hence, it can provide information about important time regimes associated with the input variables. Used in conjunction with the LHS procedure for selecting input values, the PRCC has demonstrated its value as a sensitivity measure. The PRCC is a statistic; hence, it possesses a probability distribution. At this time, the distribution of the PRCC is unknown except under the assumption of zero partial rank correlation among all inputs and the output. This assumption will not be satisfied when there are at least two important input variables in the study. Without the distribution of the PRCC, statistical inferences on the importance of the inputs cannot be made. Caution must be exercised when the PRCC is used subjectively.

To examine PRCC variability and the effect of sample size in a typical hydrodynamics problem, a study was performed on a code which models an experiment by Edwards, et. al.[6] The Edwards' experiments were blowndown studies of fluid depressurization in a straight pipe 4.1 m long with an inside diameter of 0.073 m. A glass disc at one end of the pipe was broken to initiate depressurization to atmospheric conditions. The particular experiment modeled in this study was performed at initial conditions of 7×10^6 Pa and 515 K.

The experiment was modeled with TRAC[7] using 20 computational cells as shown in Fig. 2. Eight inputs were varied and the time behaviors of calculated quantities, output variables, were recorded. The PRCCs between the inputs and each output were calculated for three samples each of size 10, 15, and 20. The PRCCs between the pressure at gauge station 1 (PGS1), and the input SLIP, a multiplier on the relative velocity between the vapor and liquid phases, are shown in Figs. 3, 4, and 5.

The three replications of sample size 20 in Fig. 3 are all comparable and indicate that SLIP (1) is unimportant for the first 0.25 seconds, (2) is important and inversely related to PGS1 for the next 0.15 seconds, and (3) lessens in importance for the last 0.10 seconds. The PRCC strongly indicates a SLIP regime change at approximately 0.25 seconds.

The three sample size 15 cases in Fig. 4 agree qualitatively with the Fig. 3 plots after 0.25s. In the first 0.25s, however, the PRCCs are somewhat unstable. Nevertheless, conclusions about SLIP in the sample size 15 cases would be the same as those from sample size 20 cases.

The sample size 10 cases in Fig. 5 clearly show the imprecision associated with the PRCC for an inadequate sample size. Since it is improbable that the quickly fluctuating importance of SLIP is real, one should assume that this sample size is too small for drawing valid inferences. The source of the oscillation is not known, but there is some indication that there is oscillation in the calculated pressure which is amplified in the PRCC. An increase in sample size seems to dampen the amplitudes observed in the smaller sample size cases.

VII. APPLICATION TO TRAC

The LHS/PRCC analysis method was applied to the TRAC code in the Analysis of Semiscale test No. 1011[8,9]. The TRAC analysis consists of the calculation of the system thermal-hydraulic response during the blowdown of the 1 1/2 loop seimscale test apparatus covering the time period 0-30s after initiation of the rupture. An analysis of this problem using best estimate inputs is presented in Ref. 1. This study extends the previous analysis to include a statistical analysis of the code.

1. Description of test apparatus.

The Semiscale test apparatus in which test 1011 was conducted had both an intact loop with active components and a blowdowm loop with simulated components. In this configuration, the operating loop represents three intact loops of a PWR and the blowndown loop represents the broken loop of a PWR.

The pressure vessel contains nine electrically heated rods 1.68 m in length. For test 1011, power to the rods was shut off prior to blowdown. The downcomer gap was .0429 m.

A simulated pump, simulated steam generator, and two rupture assemblies with blowdown nozzles comprise the blowdown loop. The simulated pump and steam generator consist of piping containing orifices to achieve the desired thermal-hydraulic resistance of these components. The full break area nozzles approximate the system volume to break area ratio for a full scale PWR. For test 1011, however, the break area was reduced to 80% of the full size break.

The operating loop, called the intact loop, has a volume approximately three times larger than the blowdown loop and represents three intact loops of a four-loop PWR. This loop has a tube-in-shell heat exchanger, centrifugal pump, and pressurizer.

2. TRAC model of experiment.

The experiment was modeled with TRAC as a set of interconnected components shown in Fig. 6. The model contains 18 components (16 components shown in Fig. 6, plus 2 break components) with a total of 122 fluid cells. Typical cell lengths used in the model are of the order of .15 - .50 m. The transient calculation is initated by instantaneously opening both breaks (adjacent to components 14 and 18). Results are calculated for the first 30s of the transient portion of the experiment.

In this study, 16 computer runs were performed in which the values of 8 input variables were selected using LHS. These input quantities, in effect, varied code models dealing with single phase pipe friction, two phase friction multipliers, orifice pressure losses, net flashing rate between liquid and vapor, slip between liquid and vapor phases, and heat transfer correlations as well as other models.

Eighteen calculated output variables were selected for analysis. These outputs included physical quantities such as volumetric and mass flow rates, pressures, differential pressures, temperatures and densities at various points in the physical system.

From the 16 computer runs, 144 (8 input variables x 18 output variables) PRCCs were calculated. The calculated PRCCs indicate the time-dependent sensitivities of the 18 output variables as a function of each of the 8 input variables. An example of the statistical analysis of the Semiscale test TRAC results is shown in Figs. 7 and 8 for the output variable lower plenum pressure (LPP). Figure 7 shows summary statistics for the lower plenum pressure (in Pascals) as a function of time. The minimum and maximum curves are the minimum and maximum of the 16 runs observed for each time point. The nominal plot, displayed for comparison, was obtained from an additional run with each input quantity set at the midpoint of its range of variation.

The PRCC analysis indicates, to some extent, which of the input variables lead to the variations in a given output variable (e.g. as shown on Fig. 7). The PRCCs between the lower plenum pressure and flashing rate (FLASH), heat transfer rate between the pipes and fluid (HTCOR), two phase friction multiplier (EXP1), and pipe roughness (RG) are shown in Fig. 8.

The PRCCs indicate that for the first seven seconds, the lower plenum pressure is strongly influenced by both the flashing rate and wall heat transfer while at later times the lower plenum pressure is influenced by the two phase friction multiplier. The small absolute value of the PRCC between the lower plenum pressure and the pipe wall roughness indicates that this input variable (RG) has little effect on the lower plenum pressure as compared to the others shown in the figure.

The positive values of the PRCCs for FLASH, HTCOR, and EXP1 indicate that an increase in the value of these inputs leads to a higher calculated lower plenum pressure. These trends are easy to understand physically since an increased heat transfer rate gives a higher fluid internal energy (and thus pressure) whereas an increase in flashing or frictional pressure drop increases the break mass flow rates which makes the pressure decay slower. Additional analyses of this semiscale test can be found in Ref. 10.

VIII. CONCLUSION

Caution should be exercised using PRCCs when the number of computer runs is not much larger than the number of input variables. In this paper, a sample size of about twice the number of inputs produced satisfactory stability in the PRCC. In the TRAC application, the PRCCs calculated using LHS agreed qualitatively with independent calculations performed by the code developers using one-at-a-time variation. However, much more information was obtained by the PRCC analysis since the time preserving aspect of the PRCCs provided information about the important time regimes associated with the input variables. Furthermore, the PRCCs aided code development by indicating which physical models contained in the code have the most influence on the code output and thus indicating a ranking for model development in various areas.

REFERENCES

1. W. H. Reed, et al , "TRAC - A New Code for LOCA Analysis," Proc. of Topical Meeting on Thermal Reactor Safety 1977 Am Nucl Soc, CONF - 770708, p. 2-475, Sun Valley, Idaho (August 1977).

2. M. D. McKay, et al., "Report on the Application of Statistical Techniques to the Analysis of Computer Codes," Los Alamos Scientific Laboratory report LA-NUREG-6526-MS (October 1976).

3. H. A. Steinberg, "Generalized Quota Sampling," Nuc. Sci. and Eng. 15, 142-145 (1963).

4. M. D. McKay, et al., "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output of a Computer Code," (accepted for publication in Technometrics).

5. R. G. D. Steel and J. H. Torrie, Principles and Procedures of Statistics (McGraw Hill Book Co., New York, 1960), Chap. 14, p. 301.

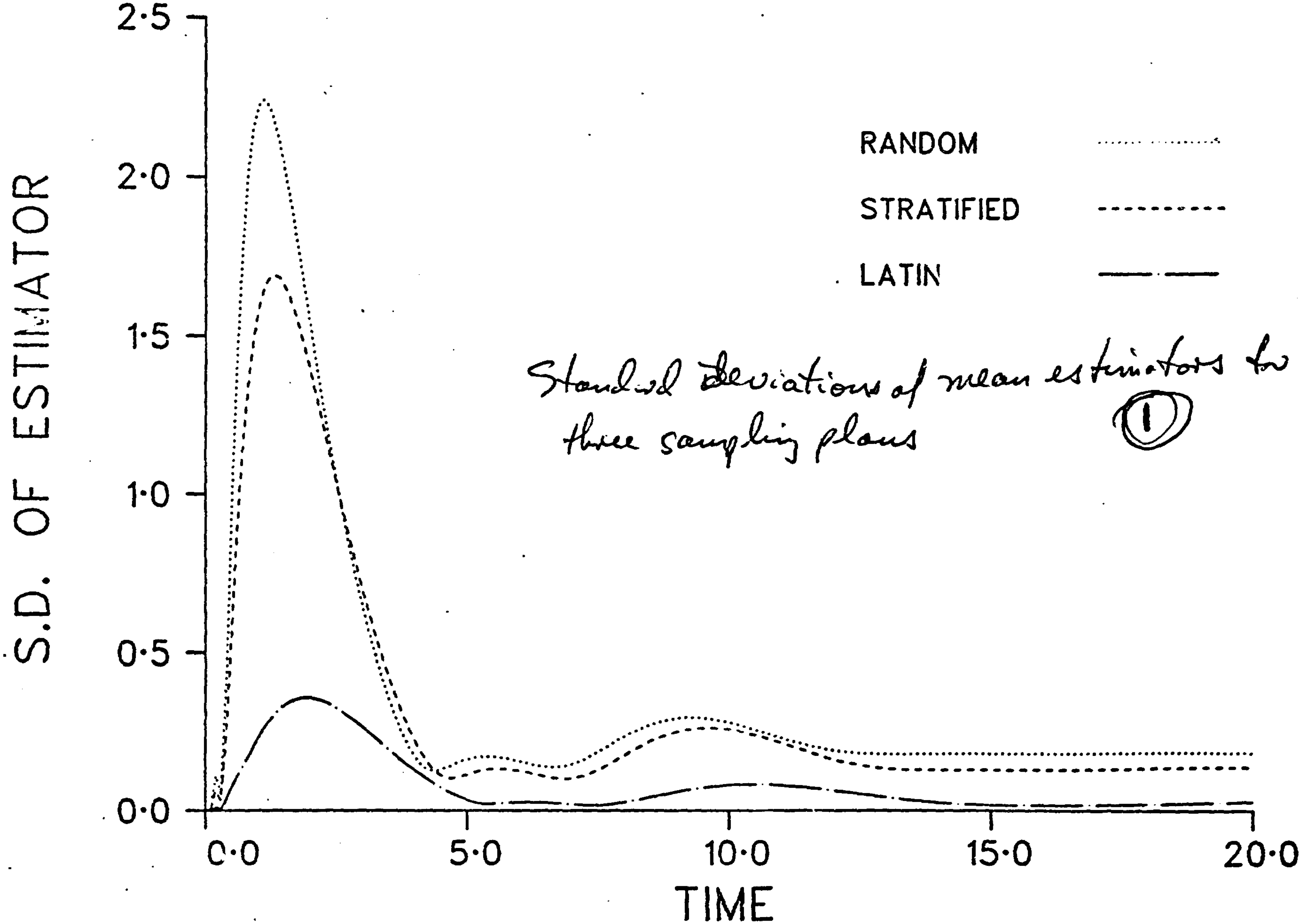
6. A. R. Edwards and T. P. O'Brien, "Studies of Phenomena Connected with the Depressurization of Water Reactors," J. British Nucl. Energy Soc. 9, 125-135 (April 1970).

7. "TRAC-2: An Advanced Best-Estimate Computer Program for PWR LOCA Analysis, Vol. I: Methods, Models, User Information, and Programming Details," Los Alamos Scientific Laboratory report (to be published).

8. R. S. Alder and P. A. Pinson, "1 1/2 Loop Semiscale System Isothermal Test Program -- Program and System Description in Support of Experimental Data Reports," Aerojet Nuclear Company report ANCR-1143 (February 1974).

9. R. S. Alder, E. M. Feldman, and P. A. Pinson, "Experiment Data Report for 1 1/2 Loop Semiscale System Isothermal Test 1011," Aerojet Nuclear Company report ANCR-1146 (March 1974).

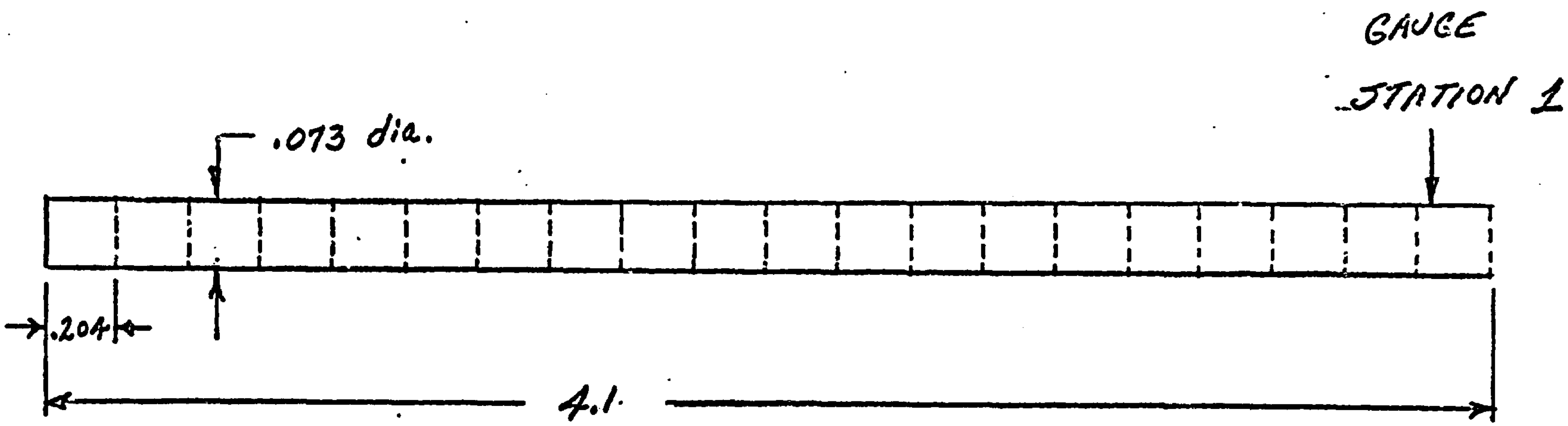
10. M. D. McKay, et al., "Application of Statistical Techniques to the Sensitivity Analysis of Standard Problem 2," Trans. Am Nucl Soc, 27, p. 606 (November 1977).



Standard deviations of mean estimators for three sampling plans

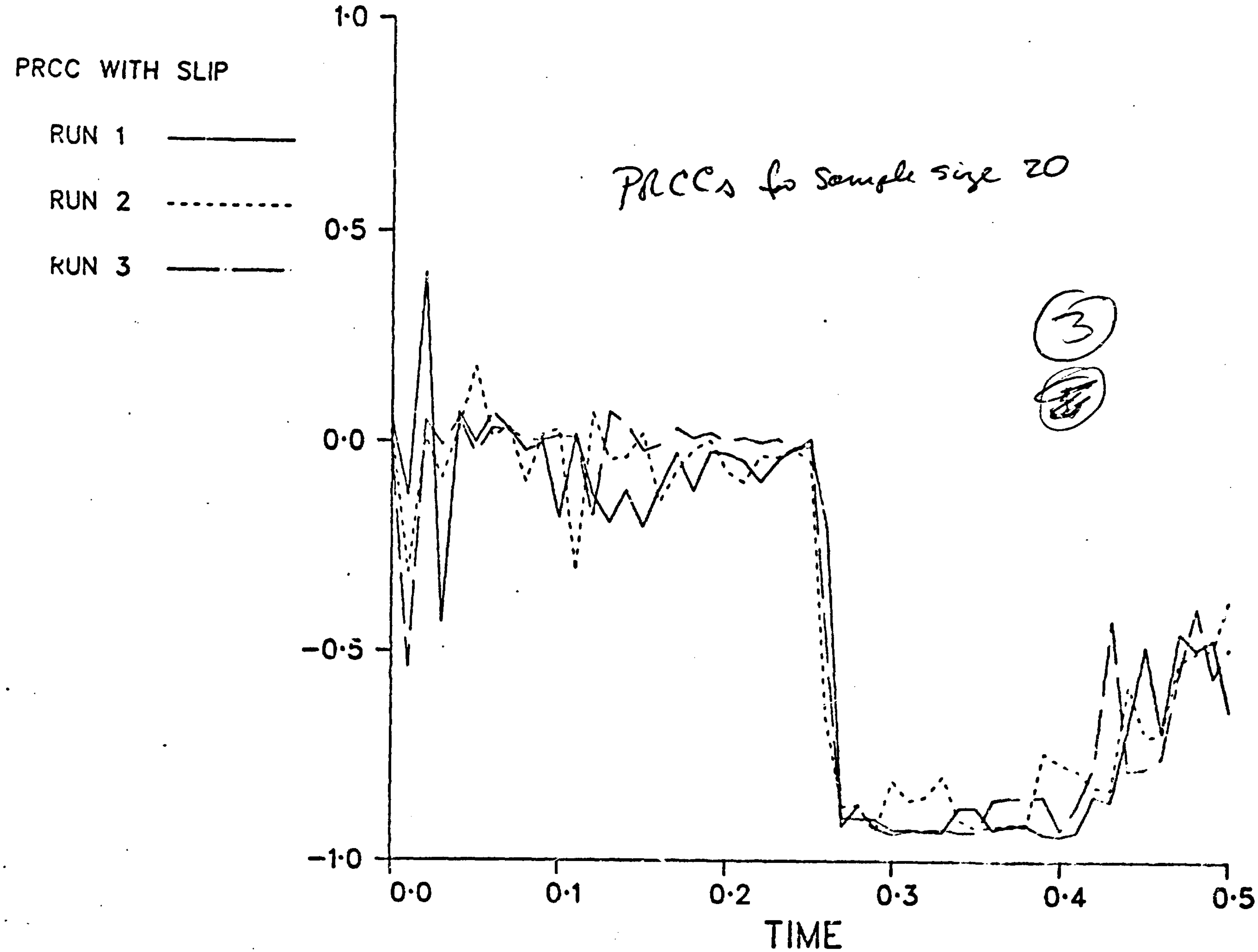
1

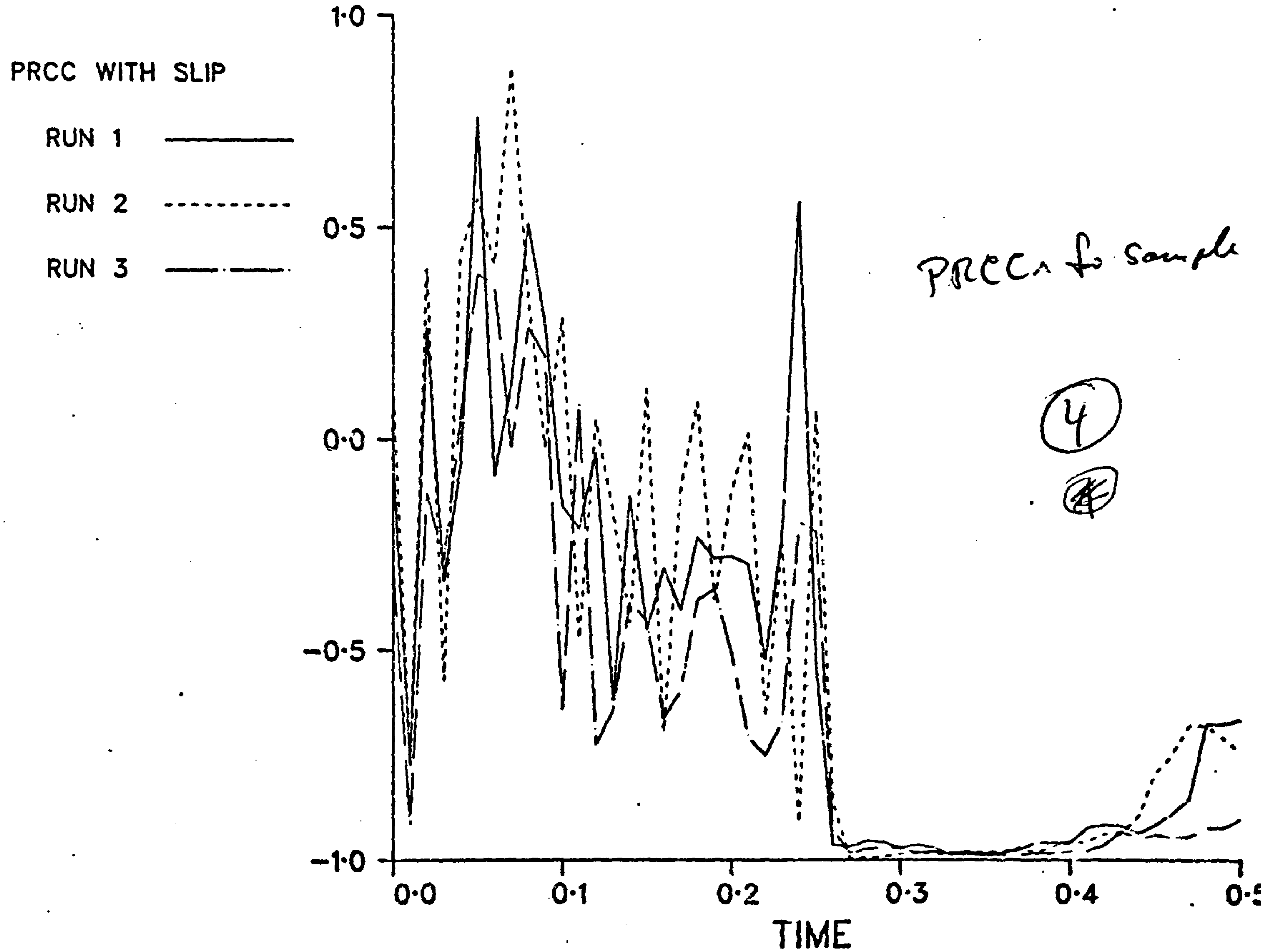
2



Edwards

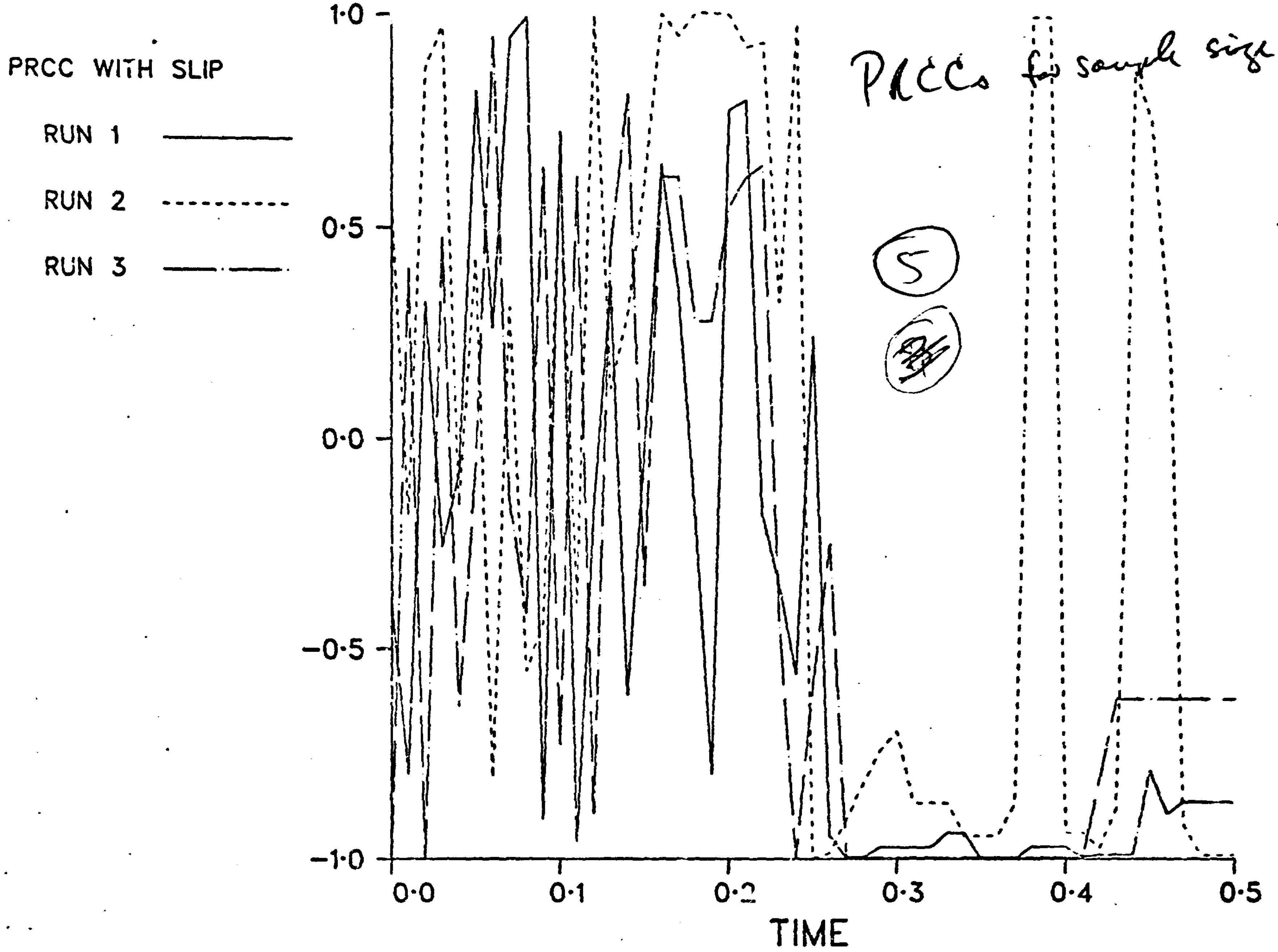
Fig. ~~1~~ 2. Model of ~~test~~ problem and physical location of experimental pressure measurement (gauge station 1). ~~All dimensions in meters.~~


PRCC WITH SLIP
RUN 1
RUN 2
RUN 3
PRCCs for sample size 20
3
1·0
0·5
0·0
−0·5
−1·0
0·0
0·1
0·2
0·3
0·4
0·5
TIME



PRCCs for sample size 15.

4



PRCCs for sample size 10

5

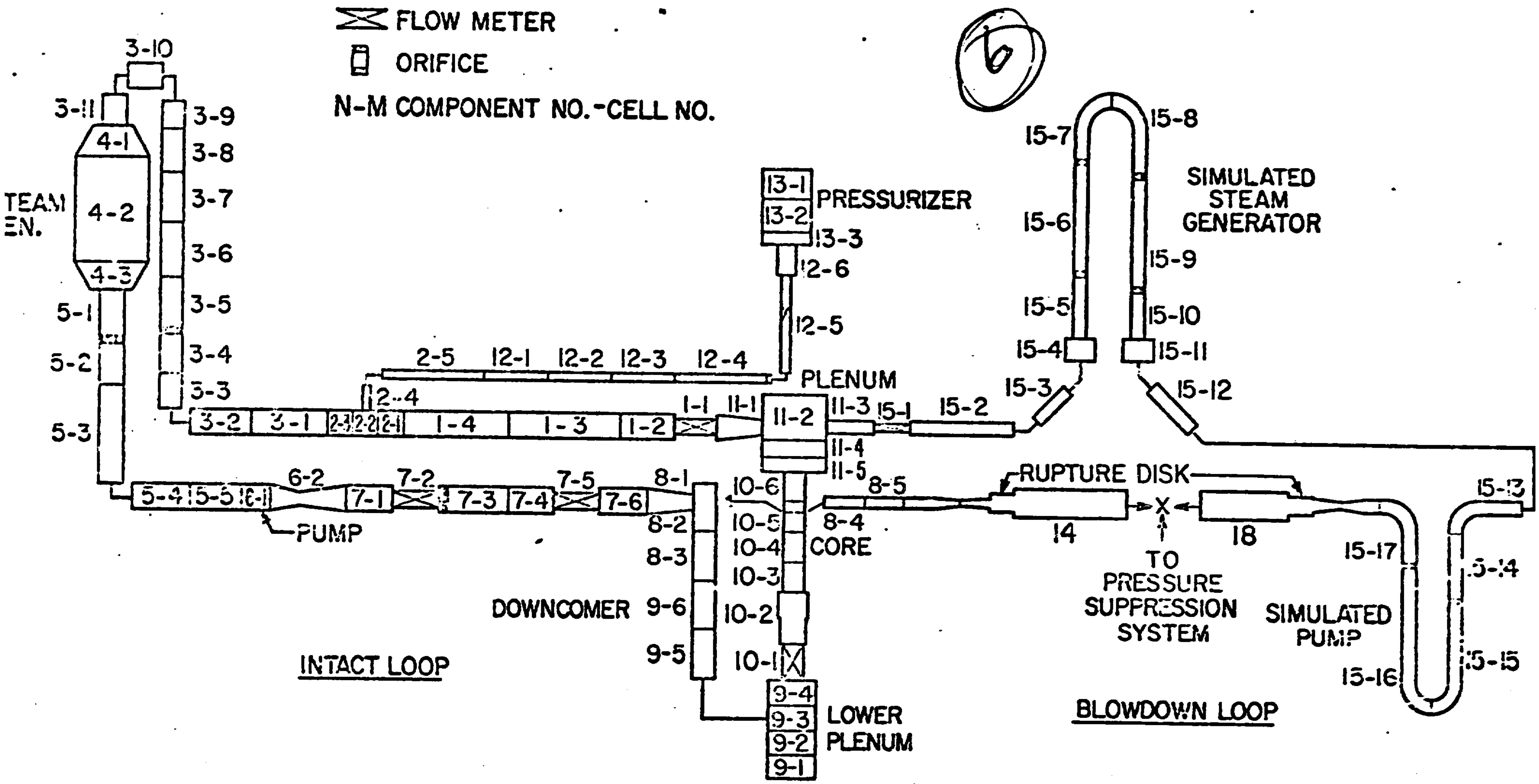
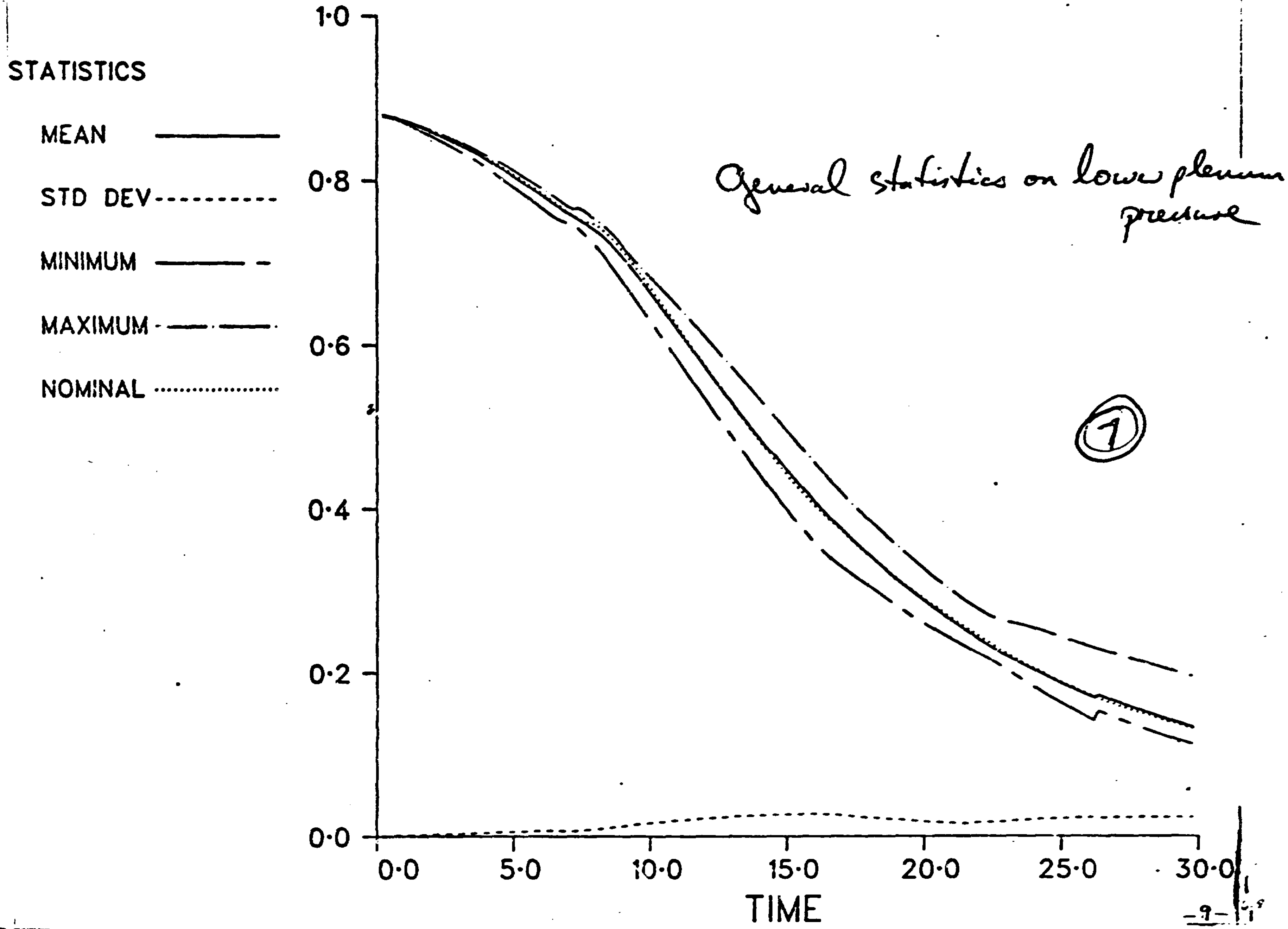
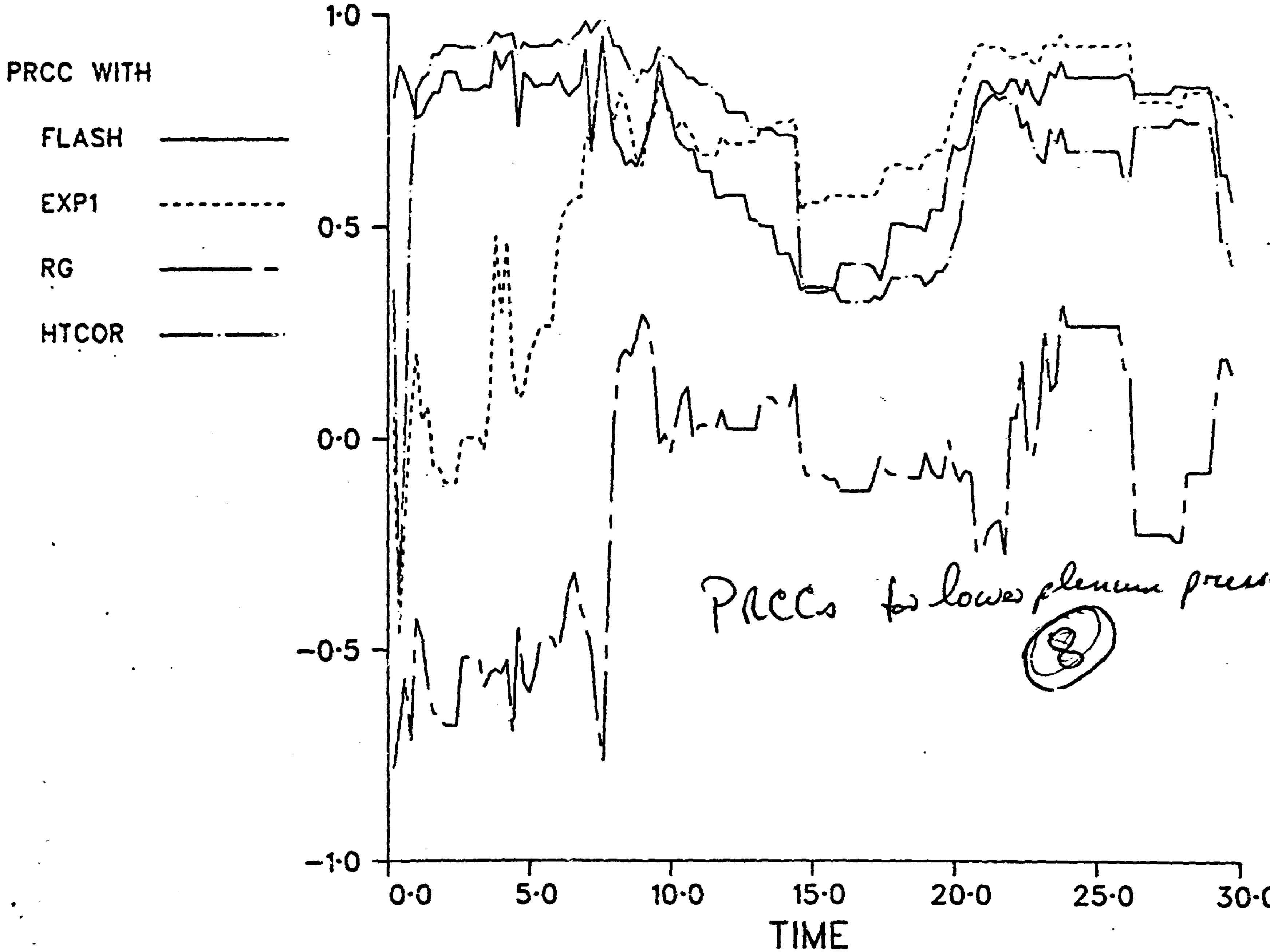


Fig. 5. (Noding diagram for TRAC Calculation of Semiscale test 1011.)


STATISTICS
MEAN
STD DEV
MINIMUM
MAXIMUM
NOMINAL
1·0
0·8
0·6
0·4
0·2
0·0
0·0
5·0
10·0
15·0
20·0
25·0
30·0
TIME
General statistics on lower plenum pressure
7



PRCCs for lower plenum pressure

8