Using Micro Saint to Predict Performance in a Nuclear Power Plant Control Room

A Test of Validity and Feasibility

Prepared by
M. T. Lawless, K. R. Laughery, Micro Analysis and Design, Inc.
J. J. Persensky, U. S. Nuclear Regulatory Commission

Micro Analysis and Design, Inc.
4900 Pearl East Circle, Suite 201E
Boulder, CO 80301

J. J. Persensky, NRC Project Manager

DISCLAIMER

Portions of this document may be illegible
in electronic image products. Images are
produced from the best available original
document.

ABSTRACT

The United States Nuclear Regulatory Commission (NRC) requires a technical basis for regulatory actions. In the area of human factors, one possible technical basis is human performance modeling technology including task network modeling. This study assessed the feasibility and validity of task network modeling to predict the performance of control room crews. Task network models were built that matched the experimental conditions of a study on computerized procedures that was conducted at North Carolina State University. The data from the "paper procedures" conditions were used to calibrate the task network models. Then, the models were manipulated to reflect expected changes when computerized procedures were used. These models' predictions were then compared to the experimental data from the "computerized conditions" of the North Carolina State University study. Analyses indicated that the models predicted some subsets of the data well, but not all. Implications for the use of task network modeling are discussed.

Contents

1 INTRODUCTION

The Nuclear Regulatory Commission (NRC) is responsible for evaluating the safety of nuclear power plant design and procedures for operating the plant during accidents. Therefore, there is a need to predict how proposed changes in procedures might impact operator performance and, ultimately, plant safety. When the need for such evaluation arises, the first approach is typically to review the literature and other available sources to determine whether there is an existing knowledge base that can be tapped, such as the knowledge of fundamental aspects of human performance or experience from other plants or similar industries. However, more often than not, the existing knowledge base is deficient with respect to it's applicability to a nuclear power plant environment. This leads to the need to study the phenomena of interest in a way that directly relates to the nuclear environment.

An obvious choice is human subjects experimentation in a realistic environment. However, experimentation with nuclear power plant operators requires extensive resources and can be difficult to conduct. Operator availability is limited and their time is in great demand. Also, time available on simulators for research purposes is limited. Furthermore, even with unlimited resources, the time required to perform experimentation in simulators may exceed the time available to make the decision. Yet, the question remains, if we can't study real operators, what are the alternatives?

One alternative is computer modeling of the human-plant system. In the past decade, a variety of tools and techniques for modeling human-based systems have emerged and have been found to be useful for studying human operator behavior in closed-loop systems. One technology that has proven useful for predicting human-system performance is *task network modeling*. In a task network model, human performance of an individual performing a function (e.g., implementing an EOP) is decomposed into a series of subfunctions which are then decomposed into tasks. This is, in human factors engineering terms, the task analysis. The sequence of tasks is defined by constructing a *task network*. This concept is illustrated in Figure 1 which presents a sample task network for dialing a telephone.

Task network modeling is an appealing approach to modeling human performance in complex systems for several reasons. First, it is a sound means for extending task analysis. Task analyses organized by task sequence are the basis for the task network model. Second, in addition to complex operator models, task network models can include sophisticated submodels of the plant hardware and software to create a closed-loop representation of the nuclear power plant control room environment. Third, task network modeling is relatively easy to use and understand. Recent advancements in task network modeling technology, including the development of the modeling system, *Micro Saint,* have made this technology more accessible to human factors
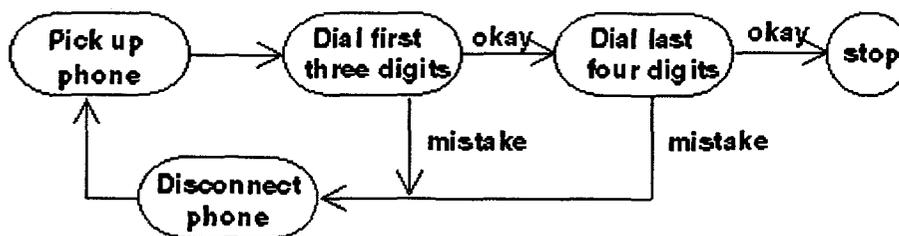

Figure 1. Example of a Task Network for Dialing a Phone

engineers. Finally, task network modeling can provide reasonable input to many types of issues. With a task network model, the human factors engineer can examine a design (e.g., control panel replacement) and address questions such as "How much longer will it take to perform this procedure?" and "Will there be an increase in the error rate?" These questions can be answered in less time and with far less effort than would be required if a prototype were developed and human subjects used.

There have been many successful applications of task network modeling as a means for predicting human performance in complex systems. At a NATO meeting comparing and contrasting human performance modeling techniques, task network modeling techniques were rated "very positive" by workshop participants in relation to other human performance modeling tools (McMillan, Beevis, Salas, Strub, Sutton, and Van Breda, 1989).

The United States Department of Defense has several hundreds task network models to study human performance in weapon systems, some of which have been subjected to validation studies. Many private organizations, including manufacturing operations (Harshell and Dahl, 1988) and the health care industry (Rayner and Laughery, 1993), are also employing task network modeling to study human-system performance. However, there has been no study showing that these models can be used to predict the human performance effects of nuclear power plant modifications. That is the focus of the study discussed herein.

2 RESEARCH OBJECTIVES

The purpose of the proposed research was to validate the use of task network modeling as a means of predicting human performance in the control room. In other words, the purpose was to prove that task network modeling can be used to answer the following question:

> "What are the expected changes in operator and plant performance based upon plant procedure and/or control room changes?"

For the purposes of the Nuclear Regulatory Commission, this is the value of this technology. To determine the utility of task network modeling in addressing this question, This research sought to prove the following three points:

1.   Valid task network models of existing nuclear power plant (NPP) control room systems can be created from a task analysis database.
2.   Once created, a task network model can be modified to reflect control room redesign or procedure changes.

3.   These modified task network models provide valid predictions of human performance times and error rates.

In sum, the goal was to evaluate how task network modeling technology can be applied to existing control room modifications.

In the next section, the details of task network modeling will be presented by showing the fundamental features and requirements for using the task network modeling system *Micro Saint*. To further illustrate the concept in the context of a nuclear power plant simulation, an example of a task network model will be presented, as well as a discussion of ways that this model might be used to address key human performance questions. The remainder of this report presents the methods, results, and discussion of the research conducted to validate the utility of task network modeling technology.

3 TASK NETWORK MODELING AND MICRO SAINT - AN OVERVIEW

Since the central technology behind this proposal is task network modeling and Micro Saint, this section provides an introduction to Micro Saint. Micro Saint is a PC-based discrete event simulation system. It will run on any DOS Windows, Macintosh, or Unix computer. Micro Saint provides all of the tools necessary to build, run, and analyze complex computer task network models.

The underlying analytic framework within Micro Saint is task network modeling. Task network modeling of human performance is a technique that has been under development since the 1970s and which has received widespread use over the past five to ten years. Essentially, the performance of an individual performing a function (e.g., operating a nuclear power plant) is decomposed into a series of subfunctions which are then decomposed into tasks. This is, in human factors terms, the task analysis. The sequence of tasks is defined by constructing a *task network*. An example to illustrate this concept is shown in earlier Figure 1.

The level of system decomposition (i.e., how finely to decompose the tasks) and the amount of the system which is simulated depends on the particular problem. For example, in a study to analyze operator workload in a helicopter crew, autonomous networks were built for the operators, the aircraft, and the threat environment. While the networks associated with the humans' tasks were far more detailed than those for the helicopter and threat environment, enough of the critical elements of the helicopter and environment were captured to permit a sound study of closed-loop human performance. In a nuclear power plant model, one might create separate models for each of the operators, one for the primary plant, and one for the secondary plant. While it would rarely be worthwhile to make the plant models as sophisticated as best-estimate thermal-hydraulic codes (e.g., RELAP),

they could be as sophisticated representations of plant behavior as the need dictates.

While the networks may be independent, performance of the tasks can be interrelated through shared variables. The relationships among different components of the system (which are represented by different segments of the network) can then communicate through changes in these shared variables. For example, when an operator manipulates a control, this may initiate an "open valve" task in a network representing the primary plant. This could ripple through to a network representing the secondary plant.

Task networks are initially defined in Micro Saint using a "point and click" approach with a window as shown in Figure 2.

Of course, the strength of task network modeling is that the dynamic aspects of task networks can be simulated on a computer. That is the purpose of the task network simulation language, *Micro Saint*. Figure 3 presents the menu within Micro Saint for defining a task. If a user provides the information required on this menu for each task in the network, he will have defined a substantial portion of a task network model.

Below are descriptions of each of these menu items.

1. **Task Number and Name** - Task identification information.

2. **Time distribution type** - Micro Saint will conduct monte carlo simulations with task performance times sampled from a distribution as defined by this option (e.g., normal, beta, exponential).

3. **Mean time** - This parameter defines average task performance time for this task.
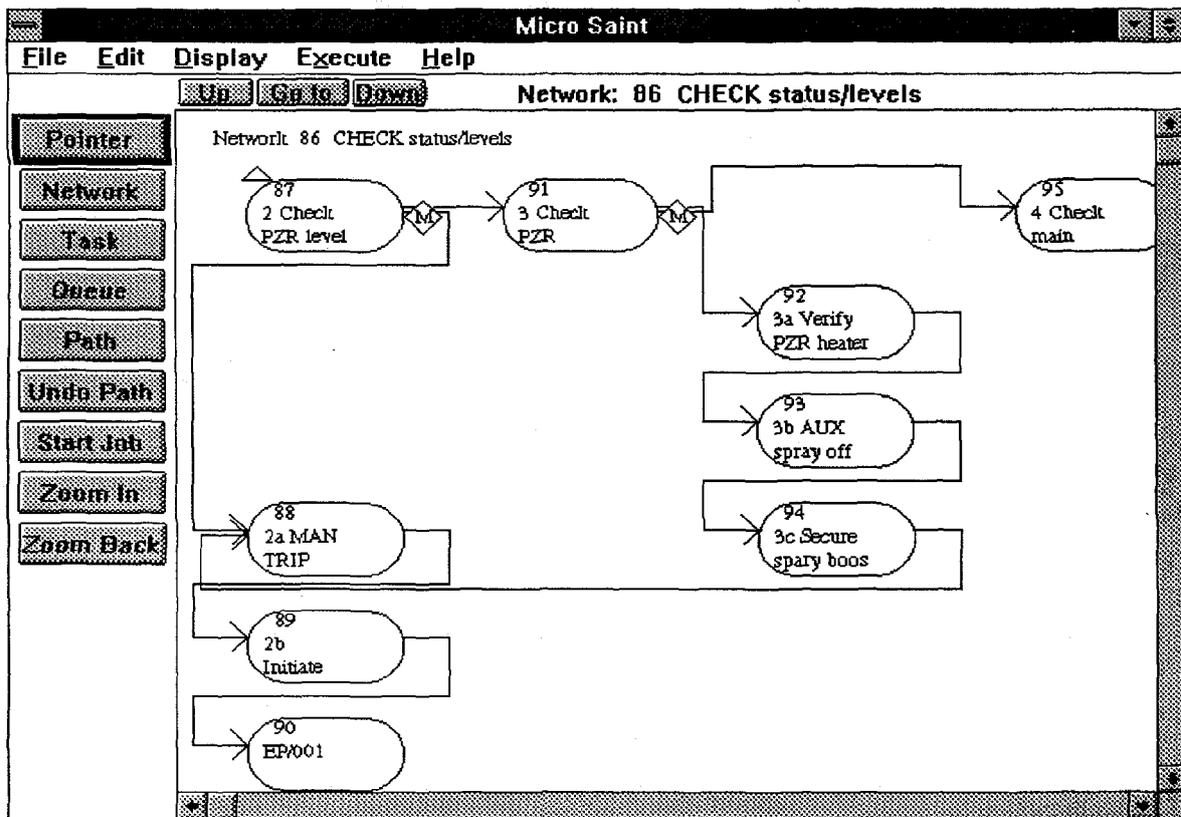
Figure 2. Micro Saint Network Development Window

4. Standard deviation - Standard deviation of task performance time.

5. Release condition - This condition (if one is defined) will hold up task execution until the condition is met. For example, a condition stating that a task won't start before another operator is available to assist may be represented by a release condition such as the following:

> operator == 1;

The value of the variable "operator" would equal zero until a task is completed in which the operator becomes available. This task would wait for that condition to be true (which would probably occur as a result of another task's completion) before beginning execution.

6. Task's beginning effect and launch effect - These fields permit the definition of how the system will change as a result of the commencement of this task. For example, if this task used an operator that other tasks might need, the following condition could be set to show that the operator was currently unavailable while he performed this task:

> operator = 0;

The difference between Beginning and Launch effects is that the Beginning effect is executed before the mean time is computed and the Launch effect is executed after the mean time is computed, a subtle but sometimes important distinction.

Beginning and launch effects are one key way in which tasks are interrelated.

7. Task's ending effect - This field permits the definition of how the system will change as a result of the completion of this task. From the previous example, when this task was complete and the operator became available, the ending effect could be set as follows:

operator = 1;

at which point the task using this as a release condition would begin.

Ending effects are another key way in which tasks can be interrelated.

8. Accept and Cancel - These buttons allow the user to close the window and either modify the model incorporating the user's changes or cancel the changes and continue.

In addition to defining tasks, it is also necessary to define decision logic when tasks may be followed by several possible tasks. To define this logic, a window as presented in Figure 4 is used.

The decision type defines what happens at the completion of this task. There are three possible decision types:

1. Probabilistic - begin one of several tasks based on a probabilistic branch

2. Tactical - begin one of several tasks based on the branch with the highest value

3. Multiple - begin several tasks at the completion of this task

Figure 3. Micro Saint Task Description Menu

Figure 4. Window for Defining Decision Logic

From these branch types, any branching logic can be developed.

The "Next Task" field defines the tasks which may follow the completion of this task . The "Routing Conditions" define the weights associated with each branch. The values can be numbers, expressions, or complicated algorithms defining the probability (for probabilistic branches) or value (for tactical branches).

A final note on this menu is that any value that appears on these screens can be not only numbers but they can also be algebraic expressions, logical expressions, or groups of algebraic and logical expressions which would form a subroutine. This is made possible in Micro Saint by the built-in parser. The function of the parser in Micro Saint is to take a string of characters defined by the user that represent a series of mathematical and/or logical

expressions and translate these expressions into executable computer code. It is the power of the parser that provides Micro Saint with the ability to address complex problems. Through the parser, a user can model many complex systems including many aspects of a nuclear power plant.

In sum, task network modeling is a logical extension to task analysis. In fact, it increases the power of task analysis in that the ability to simulate a task network with a computer permits *prediction* of human performance rather than simply the *description* of human performance that task analysis provides. What may not be apparent is the power of task network modeling as a means for modeling human performance in systems. By describing the system's activities in this step-by-step manner, complex models of the system can be developed that include a closed-loop representation of the human's interaction with the system.

There are other aspects of model development including the definition of the simulation scenario, continuous processes within the system, and queues. Further details of these feature can be obtained from the Micro Saint User's Guide.

Once a model is completed and data are collected with Micro Saint, the analyst has a number of alternatives for data analysis including exporting the data into other statistical packages, word processors, or graphics packages. Micro Saint also has built-in data analysis options to address common statistical analyses (e.g., means, histograms, line charts, bar charts).

Through the use of Micro Saint, human factors analysts have answered the following types of questions:

1. How many crewmembers are required to ensure a high probability of successful engagement in a tank (e.g., Lockett, Plocher, and Dahl, 1990)?

2. What are consequences of long working hours on a submarine crew's performance (Carson, Griffith, and Winsborough, 1994)?

3. What is the workload of an aircraft pilot during the emergencies he encounters in flight (e.g., Laughery, Drews, and Archer, 1986)?

4. What is the expected mean-time-to-repair and resulting availability of vehicles within a fleet given the mix of maintenance personnel and their skills (e.g., Laughery, Dahl, Kaplan, Archer, and Fontenelle, 1988)?

5. How long will it take an operator or crew to perform a set of activities for a new design (e.g., Chubb, Laughery, and Pritsker, 1987)?

To summarize, the goal of Micro Saint has been to develop an easy-to-use yet powerful computer package for modeling human performance in complex systems. It has been tried and tested in a number of applications.

3.1 An Example of a Task Network Model of the Nuclear Operator

Included here is a simple example of how this technology can be applied to study a nuclear environment. This example is intended to be illustrative of the basic concepts of how task network modeling can be used to study human performance in a nuclear environment. Because the models constructed for this experiment (as discussed in Section 5) are rather complex for illustrative purposes, this model has been prepared to make the concepts clearer.

This example is of an operator responding to an annunciator using a procedure requiring comparison between two meter readings. Based on these readings, the operator must either open or close a valve until the meter values are nearly the same. The operator activities for this model are represented by the task network shown in Figure 5. Also, to allow the study of the effects of different plant dynamics (e.g., control lags), a simple one node model of the line in which the valve is being opened is included in Figure 6.

The operator portion of the model will run the "monitor meters" task until the values of the variables "meter1" and "meter2" are different. The simulation could start out with these values being equal and then precipitate a change in values with a Micro Saint scenario event. This event (representing some change in the plant such as a line break or stuck valve) could be as simple as:

$$meter1 = meter1 + 2.0;$$

or as complex as an expression defining the change in the meter as a function of line break size, flow rates, etc. An issue which consistently arises in model construction is how complex the model should be. If the problem under study is purely operator performance, simple models usually suffice. However, if overall plant behavior is of interest, then the models of plant dynamics, such as meter values, are more important. Ideally, the "plant model: would include a thermal-hydraulics model with a high-fidelity
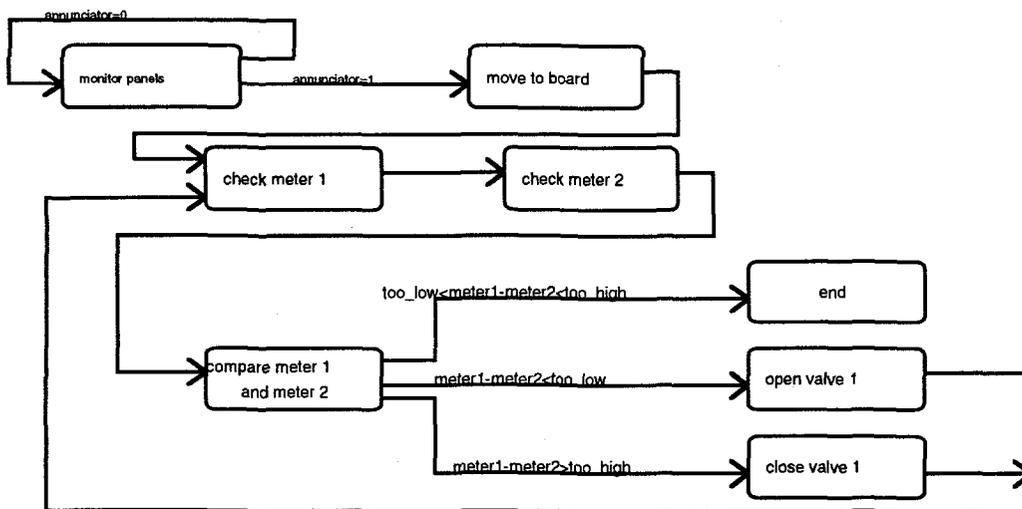
Figure 5.   Sample Task Network Model of a Nuclear Power Plant Operator
Responding to an Annunciator

representation of core neutronics, the secondary plant, as well as plant control systems. This approach is technically feasible as has been shown in other studies (Monk, 1994).
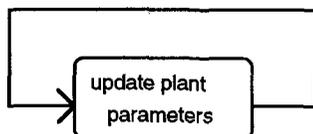

Figure 6. One Node Model of the Plant

When the transient occurs and the values of "meter1" and "meter2" start to diverge, the annunciator signal will go on. This annunciator would be triggered in the plant portion of the model by a task ending effect such as:

if meter1 <> meter2 then annunciator = 1;

In this simple model, the values of "meter1" and "meter2" are treated as being identical to the plant processes they are reflecting. However, a more sophisticated model may distinguish between the plant parameter values and those reflected on the operator displays. This would allow the study of display error and lag times.

In this example, all that is being modeled is the opening or closing of "valve1" which will, in turn, change the value of "meter1" by a commensurate amount with some type of exponential activation function. To reflect this, this simple plant model would continue to loop upon itself executing something like the following task beginning effect:

meter1 = meter1 + rate * (valve1 - meter1);

In this model, the final value for "meter1" will be "valve1" and the rate at which it approaches this value will be determined by the parameter "rate." Of course, in any model of an actual plant, equations reflecting actual plant dynamics would be included. If the plant model needed to reflect control logic, the plant model itself could be a complex task network with tasks including logic as well as equations reflecting plant dynamics.

Once the plant model sets the value of the variable "annunciator" to 1, the operator will begin his activities by moving to the appropriate board. Then, he will continue through a loop where he checks the values for "meter1" and "meter2" and either opens "valve1," closes "valve1," or makes no change. The determination of whether to make a control input is determined by the difference in values between the two meters. If the value is less than the acceptable threshold, then the operator would open the valve further. If

the value is greater than the threshold, then the operator would close the valve. This opening and closing of the valve would be represented by changes in the value of the variable "valve1" as ending effects in the tasks "open valve1" and "close valve1." In this simple model, operators do not consider rates of change in values for "meter1" and, therefore, would get into an operator induced oscillation if there was any response lag. A more sophisticated operator model could use rates of change in the value for "meter1" in deciding whether to open or close valves.

Again, this is a very simple model reflecting straightforward operator activity on one control via a review of two displays. However, it illustrates many of the basic concepts that would be used to build large models of operator teams looking at numerous controls and manipulating many displays. The central concepts of a task network and shared variable reflecting system dynamics would remain the same.

3.2 How Can Task Network Modeling Be Used to Predict Nuclear Operator Performance?

Having defined task network modeling and provided a simple example of how it can be used to model the nuclear operator in his working environment, how would such a model be used to address regulatory issues faced by the NRC? Models are developed for many reasons, one of the most important being the need to make predictions about system performance if the system were to be changed in some manner. Additionally, models can be used to study performance of a system in situations that could not be studied empirically because of safety or the cost of collecting data. Indeed, these are some of the reasons that computer codes of nuclear power plant thermal-hydraulic behavior have been developed over the years.

What questions might be addressed by task network models of the operators? Some examples are:

1. What if new controls and or displays are introduced?

2. What if the panel locations are changed?

3. What if the format and/or content of some of the key controls and displays with which the operator must interact are changed?

4. What if plant procedures are modified?

5. What if tasks are reallocated among operators or operators are unavailable as a result of other events?

6. How might operator performance be degraded because of sleep loss or work schedule?

The issues of control room redesign and procedure modification (items 1 through 3 above) as a function of new technology have been raised in several user research needs over the past decade (e.g., Coffman, Persensky, Ryan, Ramey-Smith, Goodman, Serig, and Trager, 1989). The issue of task allocation and shift staff adequacy (item 4) is currently of interest to the NRC. There have been incidences of operators falling asleep on the job as a possible consequence of work schedule that was the cause of great concern to the NRC (item 5 above). Therefore, while this list is not exhaustive of all possible ways that task network modeling could be used, it does illustrate that the questions that if could be used to answer are of regulatory interest to the NRC.

Given a task network model of a nuclear operator in a "current" control room, how might the model be used to address the above five questions? For example, the task network model could be changed by:

1. Modifying task times based on changes in the time required to access a display

2. Modifying task times and accuracy's based upon changes in the content and format of displays

3. Changing task sequence, eliminating tasks, and/or adding tasks based upon changes in plant procedures

4. Changing allocation of tasks and ensuing task sequence based upon reallocation of tasks among operators during an accident event

5. Changing task time and accuracy's based upon stressors such as sleep loss or work schedule

Again, the above list is not intended to be a definitive list of all the ways that these models may change, but it should serve to illustrate the point that network models can, in principle, be useful as a way to form technical bases for regulatory action.

Once these changes are made to the models, the models could be executed and the data collected. Measures of operator performance could be predicted including 1) response time, 2) time to complete groups of tasks (e.g., implement a procedure), 3) decisions made at key points, 4) time to put the plant into a safe state, and 5) probability of avoiding a serious event. If the models can predict the probability of avoiding a serious event, then these data can be incorporated into a probabilistic risk assessment which is an adjunct basis for some regulatory actions.

In principle, all of the above issues can be addressed with task network modeling. However, there are many factors that could limit the utility of this technology such as the availability of the required input data and the validity of the basic network approach for predicting the complex behavior that often occurs in the NPP control room. Below is a description of the study that was conducted to assess the feasibility and validity of task network modeling and Micro Saint to predict operator and plant performance, based on procedure format changes, (i.e., computer-based vs. paper-based.)

4 EXPERIMENTAL METHOD

Research Objectives Reviewed - The research was to evaluate the use of task network modeling as a means of *predicting* human performance in a nuclear power plant control room. The key question this research addressed is "Can task network modeling be used to predict changes in operator performance based upon plant procedure and/or control room changes?" To achieve this, the following three questions were addressed:

1. From the existing task analysis databases, can task network models of human performance in existing systems be created that behave in the same manner as the human?

2. Once created, can a task network model be modified to reflect control room redesign or procedure changes?

3. Do these modified task network models provide accurate predictions of human performance?

4.1 The North Carolina State University Study

To evaluate the predictive validity of Micro Saint in a nuclear operations environment, Micro Analysis and Design "shadowed" an empirical study that was investigating human performance issues in a nuclear power plant control room

environment using actual human subjects. Essentially, a model-based study of the utility of COPMA-II, a computer based procedure system, was conducted that, as identically as possible, matched an empirical study that was being performed under NRC sponsorship. In doing this, we could compare the predictions of the model-based study to the actual data obtained in the empirical study. If the answers that the model provides to the questions under study match the answers obtained by the empirical study, then it can be claimed that the model-based study would have been a good surrogate for the empirical study.

After reviewing several options of ongoing or planned studies under the control of the NRC , Micro Analysis and Design elected to shadow the study being conducted by Dr. Sharri Converse at North Carolina State University (NCSU). This study's purpose was to evaluate a computerized procedural aid called COPMA-II (Computerized Procedures Manual). The NCSU study was performed in their Scaled Pressurized Water Reactor Facility (SPWRF). The SPWRF is a scaled model of the primary and secondary loops of a nuclear power plant. What makes the SPWRF unique is that it uses freon in place of water and electrical heating in place of nuclear heating. The SPWRF performs and reacts to controls in a way that is very similar to a commercial pressurized water reactor. The procedures of the SPWRF, from "Start-Up" to "Shut Down", are very similar to procedures that Senior Reactor Operators (SRO) and Reactor Operators (RO) use every day in pressurized water reactors (see Scaled PWR Facility: Operations manual, 1993). To that extent, the SPWRF is a reasonable simulator for conducting research on human performance issues in the nuclear power plant control room.

The experimental design and data generated by the NCSU experiment was well suited as a study to shadow for several reasons. First, this research was a direct comparison between Paper procedures and COPMA procedures in the Scaled Pressurized Water Reactor Facility (SPWRF) at NCSU. This comparison in a controlled environment provided a sound source of data for modeling. Second, the value and specific impact of using the COPMA-II procedural aid were not obvious. This uncertainty of the actual effects of COPMA-II was important to assess whether modeling could evaluate an issue where the answer was not intuitively obvious. Third, the SPWRF was readily accessible to the modeling team, enabling them to make better performance estimates for the model. Fourth, the study focused on procedures that were defined, understood, and documented. Because procedures were already documented, a task analysis, which would normally be required to develop a task network model, was not required. Finally, the timing of the NCSU study was consistent with the schedule of this shadow study.

To fully understand the model-based study of COPMA-II and how the results of the model-based study were compared to the NCSU empirical study, it is necessary to understand both COPMA-II and the NCSU study. The following two sections provide background information on COPMA-II and the NCSU study.

COPMA-II. Normal operations in the SPWRF, as in nuclear power plant control rooms, require operators to interact with written procedures that detail actions and responses to the status of the plant as shown on displays within the control room. Advanced computer systems have taken paper based procedures and integrated them with the plant computers so that plant parameters are automatically transmitted to the programmed controller and the designed procedures are presented to the operator. COPMA-II is one such system.

COPMA-II presents procedures to operators using an advanced user interface that allows them to access procedures rapidly and directly. Additionally, COPMA-II can show an historic trail or memory of procedures that have been visited. COPMA-II can also translate a complex set of procedure based instructions into a linear sequence of sub tasks.

In general, COPMA-II appears to provide a set of features that may allow operators to perform control room tasks more quickly and with fewer errors. Conceptually, the added features should yield better performance. However, this is not always the case as reported by Endestad and Meyer (1993). It is this uncertainty that initially prompted the NCSU study of the COPMA-II system.

Experimental Details of the NCSU Study. This short description of the NCSU study is presented here to provide a better context for the model-based shadow study. More details of this empirical study are provided in Converse (1994).

The specific objectives of the NCSU study were 1) to determine whether the COPMA-II system enhances or degrades the speed and accuracy of reactor operators' performance under normal and/or abnormal conditions and 2) to determine whether COPMA-II provides operators with all the information they need, and whether all of that information provided is accurate. The effectiveness of COPMA-II was evaluated by the NCSU, Scaled Pressurized Water Reactor Facility (SPWRF).

The study employed eight licensed reactor operators (ROs) and eight licensed senior reactor operators (SROs) who were utility employees in normal training at the Department of Nuclear Engineering at North Carolina State University. The 16 operators were divided into eight teams of two operators, with each team containing one SRO and one RO,. Each team used both traditional hard copy procedures and the COPMA-II system to control the SPWRF, and each team operated the SPWRF under normal and accident conditions. Performance under normal conditions was evaluated by having the ROs perform a routine change of power maneuver. The two accident scenarios presented to operators were a small break loss of cooling accident (LOCA) and a steam generator tube rupture (SGTR).

Data on error rates, performance times, and subjective estimates of workload were collected for each procedure type (Paper vs. COPMA-II) and trial type (Normal vs. LOCA vs. SGTR). The data that were most relevant to the comparisons with the task network models were the performance time data. The data were divided into two parts based on the hypothesized effect of paper versus COPMA-II effects - the *preliminary procedures* and the *final procedures*. The preliminary procedures set starts with the initialization of the procedure (i.e. start load maneuvers, initiate the LOCA, initiate SGTR) and ends at a point where procedures change abruptly. The final procedures set starts with the new change in procedural instructions and ends when status is brought back to normal operating conditions. (See appendix A to review the procedures and to identify the specific point at which data were collected).

Based on an initial evaluation of the crews, the eight teams of operators were equally qualified at the beginning of the experimental trials.

The results of this study showed no significant differences between COPMA-II and paper procedures for the change of power data. For the accident scenario, time to initiate a response was significantly slower for COPMA-II than for paper procedures. While response initiation time was slowest for COPMA-II trials, the error rate for COPMA-II trials was approximately half the error rate for performance with paper procedures. The interaction between procedure type and team number was significant for accident scenario accuracy. Three teams committed significantly fewer errors with the COPMA-II procedures, while the use of COPMA-II did not significantly affect accuracy for the remaining teams. In Section 5 of this report, we will present more details on the performance time data as it compares to the task network model predictions.

4.2 The Model-Based Shadow Study

As NCSU was designing experiments and collecting data for the above empirical study,

Micro Analysis and Design was conducting a shadow study using task network modeling as the means for estimating the effects of COPMA-II on performance. The goal of this shadow study was also to predict the effects on performance of the use of COPMA-II. However, rather than assessing these effects through experimentation with human subjects, these effects would be predicted by building models of the same procedures and scenarios and then conducting experiments with these models.

This section describes the procedures used in this model-based shadow study. While the goal of this study was to evaluate the effectiveness of COPMA-II, the measure of success was *not* the predicted performance of the crew using COPMA-II but, rather, *the degree of similarity between the results of the model-based study and the results of the NCSU study*. The more similar the model predictions and the experimental data, the better the argument that task network modeling is a reasonable alternative to human subjects experimentation for determining the effects of human factors issues on crew performance.

Building the models. The first step in the modeling study was to collect the data necessary to build the task network models. As stated previously, much of the procedure documentation that was available for the SPWRF could be used for designing the model and setting model parameter values. These procedures show the flow of tasks for every procedure in the SPWRF in sufficient detail to permit the construction of the task networks.

As stated previously, the three NCSU experimental conditions included all procedures involved in 1) Load Maneuvers, 2) Small Break Loss of Coolant Accident (LOCA) and, 3) Steam Generator Tube Rupture (SGTR). The procedures involved in Load Maneuvers are normal operating conditions. LOCA and SGTR procedures are the accident scenarios. These procedures require different interactions by the SPWRF operators and therefore are hypothesized

to show performance differences between paper and COPMA procedures. The task networks for the models for these three scenarios were developed using the available documented procedures. No difficulties were encountered in either obtaining the necessary data nor in translating these data into the networks of crew activity.

Task performance parameters, such as task time and error rates, are necessary elements of a task network model. They are also generally less available than the data required to develop just the network portion of the model. Typically, these data are collected for a baseline model through observation of the system under study. For this study it was determined that the Paper procedures data from the NCSU experiment could be used to set these parameters in the task network models that were built for the Paper procedures. However, Paper procedures performance data from NCSU were not detailed at the task level, only at the level of "preliminary procedures" and "final procedures," each of which involved numerous tasks. Therefore, to estimate the individual task performance parameters, estimates were made of the relative amounts of time that a task would take as a portion of the overall procedure. Then, the NCSU data regarding the time for the overall procedure was used as the basis for determining specific values for the individual task parameters. To illustrate the concept the following simple example is presented. A procedure involves three tasks that are performed in sequence. It is estimated that task 1 takes 10% of the time, task 2 50%, and task 3 40% of the time for the total procedure. If the empirical data indicated that the total procedure time was 100 seconds, then the time parameters would be set to 10 seconds, 50 seconds, and 40 seconds for tasks 1, 2, and 3, respectively. In these models, it was rarely that simple since procedures were not linear (i.e., they involved branching at some points) and task variability also had to be estimated. However, this basic approach was used to build individual task parameter estimates from the NCSU performance time data which was collected at a

grosser level. This is not an unusual practice in model development and is commonly referred to as fine-tuning a model.

Model performance parameters for the COPMA-II conditions were estimated at the individual task level using expert judgment. This was done by starting with the individual task times that were generated during the development of the Paper procedures model as discussed in the paragraph above. These task times and variances were then modified based on the estimated impact of COPMA-II on each individual task. In many tasks it was estimated that COPMA-II would have no effect on performance. In complex procedural changes, COPMA-II was estimated to yield shorter performance times. Alternatively,

conditions in this manner. Since the goal of this study was to examine the validity of modeling *as it might be used in an actual study in the future*, the approach used to generate COPMA-II time estimates was realistic and appropriate.

To summarize, 1) the source of data for all of the task networks were the documented procedures, 2) the source of data for task performance parameters for all models of the scenarios using paper procedures was the empirical data from the NCSU study with expert judgment used to adjust individual task parameters so that the empirical results were matched, and 3) the source of data for task performance parameters for all models of the scenarios using COPMA-II was expert judgment.

Table 1. SPWRF Procedures and Associated Conditions

SPWRF Procedures	Condition
Load Maneuvers	Paper Procedures
	COPMA-II Procedures
Small Break Loss of	Paper Procedures
Coolant Accident(LOCA)	COPMA-II Procedures
Steam Generator Tube	Paper Procedures
Rupture(SGTR)	COPMA-II Procedures

based on the Endestad and Meyer (1993) study, tasks that were very simple and straight forward would not benefit from the added functionality of COPMA-II. It was also expected that paper procedures would be better in simple procedural changes such as those where an operator can see the next procedural activity on paper by looking ahead and/or marking the next procedural page and flipping back and forth.

In essence, the procedure for making time estimates for tasks in the COPMA-II condition was expert judgment supported by any available data - the same procedure followed in many model-based studies. In using modeling and simulation, the analyst must often generate task time and accuracy estimates for experimental

Table 1 summarizes the SPWRF procedures for which empirical data were collected at NCSU and for which models were developed with Micro Saint.

4.3 Model Execution

After all models were developed and parameterized for each of the six conditions presented in Table 1, each model was run 5000 times. Data were collected on performance time and variance for the preliminary and final sets of tasks - exactly as they were in the NCSU study.

5 RESULTS

The analyses examined how closely the data generated by the models would have provided the same "answers" as the empirical study about the effectiveness of COPMA-II. Three different measures of the similarity between the models and the empirical study were analyzed:

1. The ability of the model to predict statistically significant effects that match the experimental findings.

2. The ability of the model to predict effects in the same direction as the experimental data.

3. In absolute terms, the comparison between model predicted values and actual experimental values.

To generate and interpret these measures, three different analyses were performed. The first analysis compared the models' predictions of COPMA-II performance against the actual empirical data for the paper based study. In fact, this is the type of model-based analysis that one might conduct for the implementation of a computerized procedures system. Actual human performance data on crew performance for Paper procedures would be available or easily collected, but the model would be used to generate estimates of the crew's performance with a new computerized procedures system.

The second set of analyses compared the statistical conclusions (e.g., which conditions differed significantly between paper procedures and COPMA-II) of the model study to the statistical conclusions from the empirical study. If the statistical conclusions of the empirical NCSU study closely match the statistical conclusions of the model-based study, then one could argue that lower cost model-based research would lead to the same conclusions and, therefore, an acceptable replacement. In these analyses for

each scenario, the models' predictions for either a statistically significant improvement, no effect, or a significant decrement in performance are compared to the statistical conclusions of the empirical research.

The third set of analyses are comparisons of the model predictions vs. the empirical data for the Paper and COPMA-II conditions. These analyses present the most direct answer to the question "Are the model predictions significantly different from the empirical data?"

In each of the analyses below, there are six sets of data that can be used to compare the differences between performance with Paper procedures and performance with COPMA-II. These conditions represent the three scenarios (load maneuvers, LOCA, and SGTR) each of which has a preliminary and final set of data.

5.1 NCSU Paper Procedures Data vs. Model COPMA-II Predictions

These analyses compare the models' predictions of COPMA-II to the actual empirical data for the paper based procedures. T-tests were conducted comparing the NCSU empirical data for the Paper procedures to the model predictions for COPMA-II. This comparison yields the prediction of COPMA-II effectiveness that would have been obtained if this had been an actual model-based study. In this type of study, we would have empirical baseline data from observing actual operators in the control room, but we would need the model to generate the data for the experimental condition - the use of COPMA-II in this case. Therefore, these results reflect the predictions we would have made about the effectiveness of COPMA-II if *only* model derived data were available about COPMA-II.

These means and standard deviations for each of the six data sets are presented in Table 2. Below are descriptions of these comparisons for each of the three scenarios.

results do not support this hypothesis. Comparisons between 5000 runs of the SGTR COPMA-II models and the NCSU data showed no statistically significant differences in performance

Table 2. Paper versus COPMA-II: Comparisons of Time in Seconds to Perform the NCSU Paper Procedures Data versus Model Predictions of Time in Seconds to Perform the Procedures Using COPMA-II

Scenario	Preliminary Procedures		Final Procedures	
	Paper- NCSU Study Data	COPMA-II- Model Predictions	Paper- NCSU Study Data	COPMA-II- Model Predictions
	Mean/SD	Mean/SD	Mean/SD	Mean/SD
Load	33/12.95	44.52/17.35	314.25/195.9	320.2/207.8
LOCA	377.75/55.2	445.4/79.7	431.25/242.3	490.4/264.1
SGTR	390.25/175.5	427/246.8	1050.8/656.6	1159.5/759

Load Maneuvers: Paper versus COPMA-II. The control condition of Load Maneuvers was hypothesized in the NCSU study to show no differences between paper and COPMA-II procedures. The model results support this hypothesis. Results of 5000 runs of the Load Maneuvers COPMA-II models showed that the model predicted no statistically significant differences from the NCSU Paper procedures data in either the preliminary set of procedures (t=1.5050, p=.1545) or the final set of procedures (t=-.0589, p=.9538).

LOCA: Paper versus COPMA-II. This accident scenario was hypothesized to show superior performance with COPMA-II versus Paper procedures. The model predictions do not support this hypothesis. Comparisons between 5000 runs of the LOCA COPMA-II models and the NCSU data showed no statistically significant differences in performance from the NCSU Paper procedures in either the preliminary set of procedures (t=1.3956, p=.2123) or the final set of procedures (t=.3295, p=.7530).

Steam Generator Tube Rupture: Paper versus COPMA-II. This accident scenario was also hypothesized to show superior performance with COPMA-II versus Paper procedures. The model

from the NCSU Paper procedures in either the preliminary set of procedures (t=.2451, p=.8145) or the final set of procedures (t=.3095, p=.7674).

5.2 NCSU Predictions

This set of analyses compares the conclusions of the model study to the conclusions from the empirical study with respect to the conditions in which COPMA-II was found to result in statistically significant effects. Table 3 shows the NCSU study data and those differences that were found to be statistically significant. Below, we compare the conclusions drawn from the NCSU data to the conclusions drawn from the model-based data.

Load Maneuvers: Paper versus COPMA-II in the NCSU Study. It was hypothesized that there would be no differences between Paper and COPMA-II procedures. An analysis of the NCSU empirical data on the Load Maneuvers condition showed that Paper procedures performance was significantly faster than performance with COPMA-II procedures in the preliminary set of procedures (t=2.2308, p=.0426). The final set of empirical data on the paper procedures showed no differences between

Table 3. Time in Seconds to Perform the Procedure Using Paper versus COPMA-II:
NCSU Study Results
* indicates significant differences in NCSU data between Paper procedures and COPMA-II

Scenario	Preliminary Procedures		Final Procedures	
	Paper- NCSU Study Data	COPMA-II - NCSU Study Data	Paper- NCSU Study Data	COPMA-II - NCSU Study Data
	Mean/SD	Mean/SD	Mean/SD	Mean/SD
Load	* 33/12.95	* 58.62/29.79	314.25/195.9	361.25/174.3
LOCA	* 377.75/55.2	* 575.5/80.6	431.25/242.3	1463/893
SGTR	390.2/175.5	476.7/88.62	1050.8/656.6	618.5/286.7

COPMA-II and Paper procedures performance (t=.5070, p=.6201).

These results are different from the model predictions presented in Section 5.1 where no significant differences were predicted in either the preliminary or final set.

LOCA: Paper versus COPMA-II in the NCSU Study. It was hypothesized that this accident scenario would show superior performance with COPMA-II versus Paper procedures. An analysis of the NCSU empirical data for the LOCA condition showed that performance with Paper procedures was significantly faster than performance with COPMA-II procedures in the preliminary set of procedures (t=4.0536, p=.0067). The final set of procedures showed no statistically significant differences between COPMA-II and Paper procedures performance (t=2.2301, p=.0673), although the actual difference was substantial.

These results are different from the model predictions presented in Section 5.1 where no significant differences were predicted in either the preliminary or final set.

Steam Generator Tube Rupture: Paper versus COPMA-II in the NCSU Study. It was hypothesized that this accident scenario would show superior performance with COPMA-II versus Paper procedures. An analysis of the NCSU empirical data for the SGTR condition showed that using COPMA-II performance did

not significantly differ from performance with Paper procedures in either the preliminary set of procedures (t=.7884, p=.4605) or the final set of procedures (t=-1.2068, p=.2792).

These results are consistent with the model predictions presented in Section 5.1 where no significant differences were predicted in either the preliminary or final set.

Discussion. As discussed in Section 5.1, the comparisons between COPMA-II *model runs* and NCSU Paper procedures data showed no statistically significant performance differences whatsoever, whereas the results from the NCSU empirical study, as presented in Table 3 and described above, found superior performance with Paper procedures in two conditions: Load Maneuvers-Preliminary set and, LOCA-Preliminary set. However, in these two comparisons where the model did not show statistically significant differences that the data did, the model did predict a performance change in the same direction.

5.3 Direct Comparisons of NCSU Data vs. Model Data

The final analyses compare NCSU data directly with the model predictions. These comparisons are a direct test of the ability of the models to predict data that would be obtained in human subjects experiments.

Table 4. Time in Seconds to Perform the Procedures Using Paper versus COPMA-II: Results of Model Runs and NCSU Study

Scenario-Data Source	Preliminary Procedures		Final Procedures	
	COPMA-II	Paper	COPMA-II	Paper
	Mean/SD	Mean/SD	Mean/SD	Mean/SD
Load-NCSU Data	58.62/29.79	33/12.95	361.25/174.3	314.25/195.9
Load-Model Predictions	44.52/17.35	36.61/13.91	320.2/207.8	317.29/204.8
LOCA-NCSU Data	575.5/80.6	377.75/55.2	1463/893	431.25/242.3
LOCA-Model Predictions	445.4/79.7	385/61.1	490.4/264.1	457.4/261.6
SGTR-NCSU Data	476.75/88.62	390.25/175.5	618.5/286.7	1050.8/656.6
SGTR-Model Predictions	427/246.8	401.6/249.8	1159.5/759	1091.1/754.6

Table 4 reformats the data from Tables 2 and 3 to show all NCSU data and all model predictions. Each condition (Load Maneuvers, LOCA, SGTR) is represented for COPMA-II and Paper procedures for both preliminary and final sets of procedures. Figures 7 and 8 present these data graphically.

Table 5 highlights the significant results from the t-tests comparing the COPMA-II data from the NCSU study to the COPMA-II predictions made by the model. Figures 9 and 10 present these data graphically.

Table 5. COPMA-II from the NCSU Study versus COPMA-II from Model Runs
* indicates significant differences between NCSU and model data
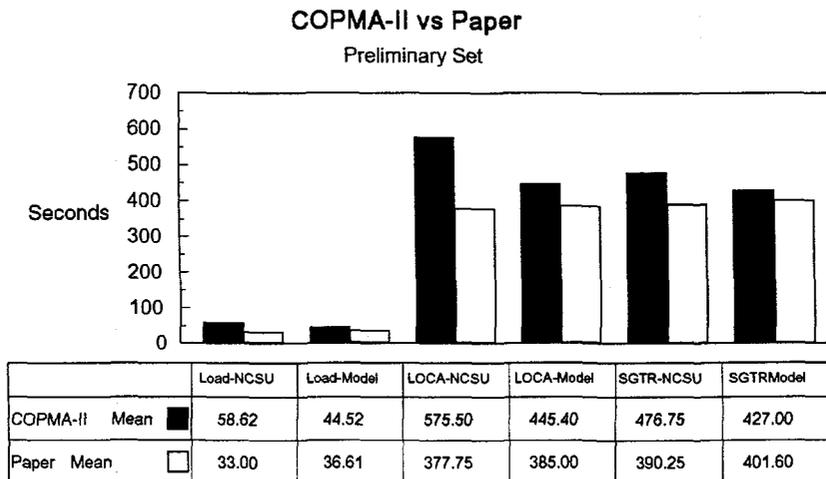
Scenario	Preliminary Procedures	Final Procedures
Load		
LOCA	*	*
SGTR		

COPMA-II vs Paper
Preliminary Set

		Load-NCSU	Load-Model	LOCA-NCSU	LOCA-Model	SGTR-NCSU	SGTRModel
COPMA-II Mean ■		58.62	44.52	575.50	445.40	476.75	427.00
Paper Mean □		33.00	36.61	377.75	385.00	390.25	401.60

Figure 7. Paper versus COPMA-II: *(Mean Times)* Preliminary Set

COPMA-II vs Paper

Final Set

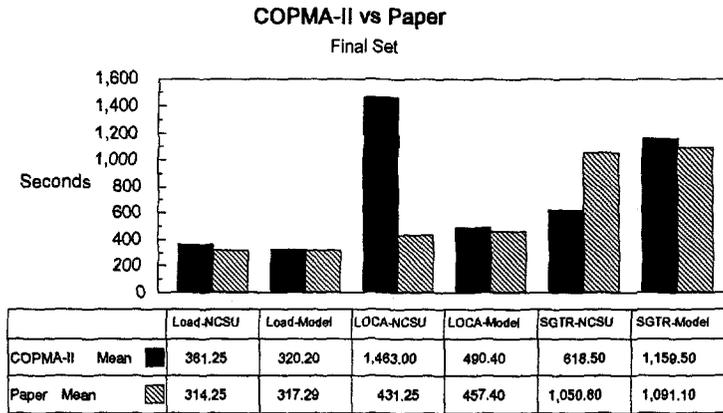|		Load-NCSU	Load-Model	LOCA-NCSU	LOCA-Model	SGTR-NCSU	SGTR-Model
COPMA-II	Mean ■	361.25	320.20	1,463.00	490.40	618.50	1,159.50
Paper	Mean ▧	314.25	317.29	431.25	457.40	1,050.60	1,091.10

Figure 8. Paper versus COPMA-II (*Mean Times*) Final Set

NCSU Data vs Model Prediction

COPMA-II Preliminary Set

Figure 9. NCSU Data versus Model Prediction - COPMA-II, Preliminary Set
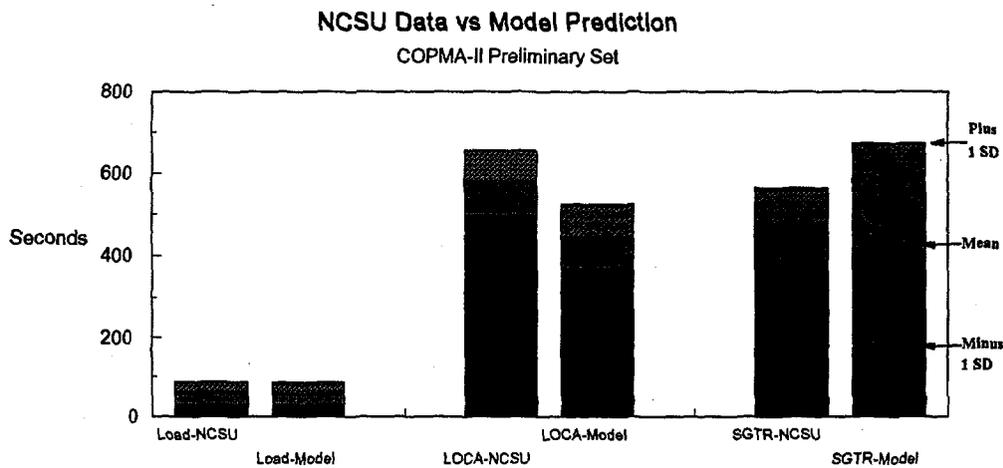
NCSU Data vs Model Prediction

COPMA-II Final Set

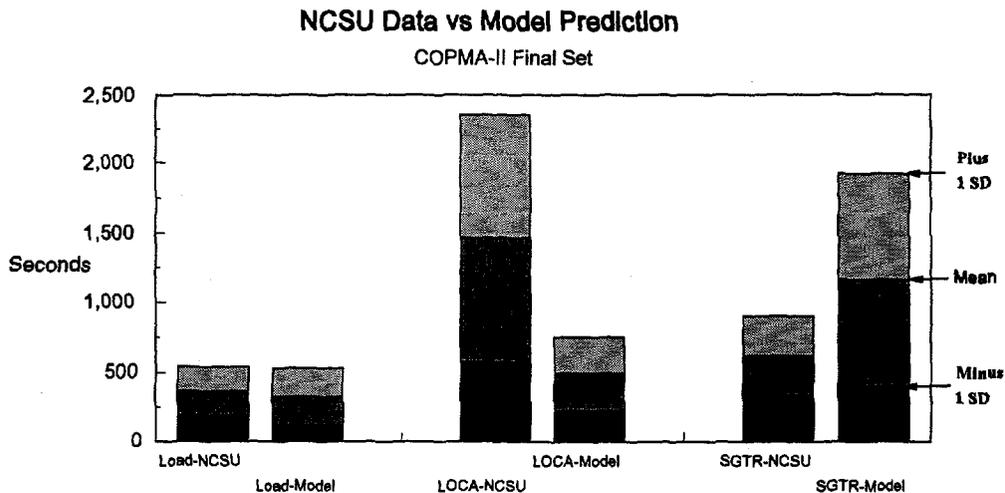Figure 10. NCSU Data versus Model Prediction (*Standard Deviations*) COPMA-II - Final Set

19   NUREG/CR-6159

Load Maneuvers showed no significant differences between observed COPMA-II performance in the NCSU study and model prediction of COPMA-II performance for either the preliminary set of data (t=1.1568, p=.2667) or for the final set of data (t=0.4288, p=.6746).

COPMA-II performance in the NCSU study was significantly different from the COPMA-II model predictions with the LOCA procedures. This is true for both the preliminary set of data (t=3.2464, p=.0059) and for the final set of data (t=2.9554, p=.0104). For the SGTR, there were no significant differences between observed COPMA-II performance and the model

predictions of performance for either the preliminary set of data (t=0.5366, p=.600) or the final set of data (t=-1.8866, p=.0801).

Also, it is worthwhile to compare the Paper performance data from the NCSU study to the Paper performance generated by the models. These data should not differ since the Paper results from the NCSU study were used to calibrate the models. As expected, there are no statistically significant differences. These data are presented graphically in Figures 11 and 12.
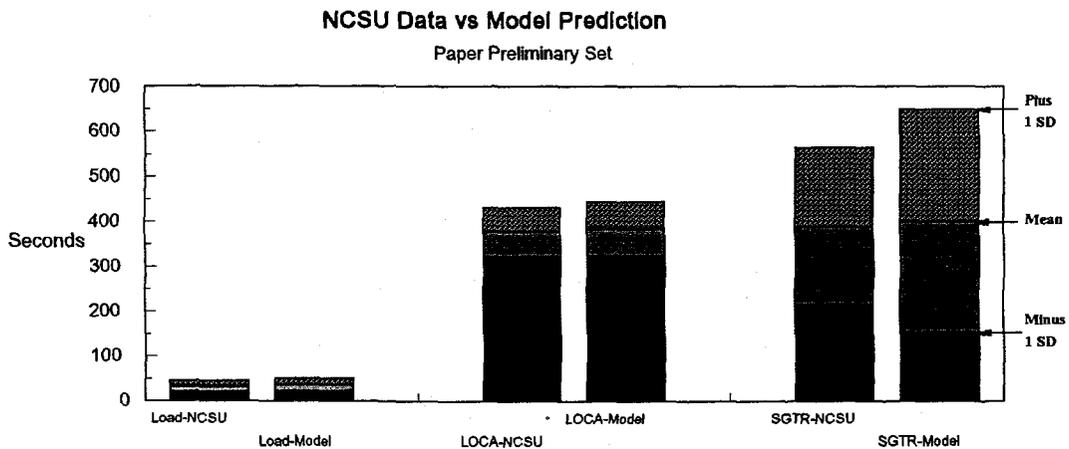

Figure 11. NCSU Data versus Model Prediction (*Standard Deviations*) - Paper Preliminary Set
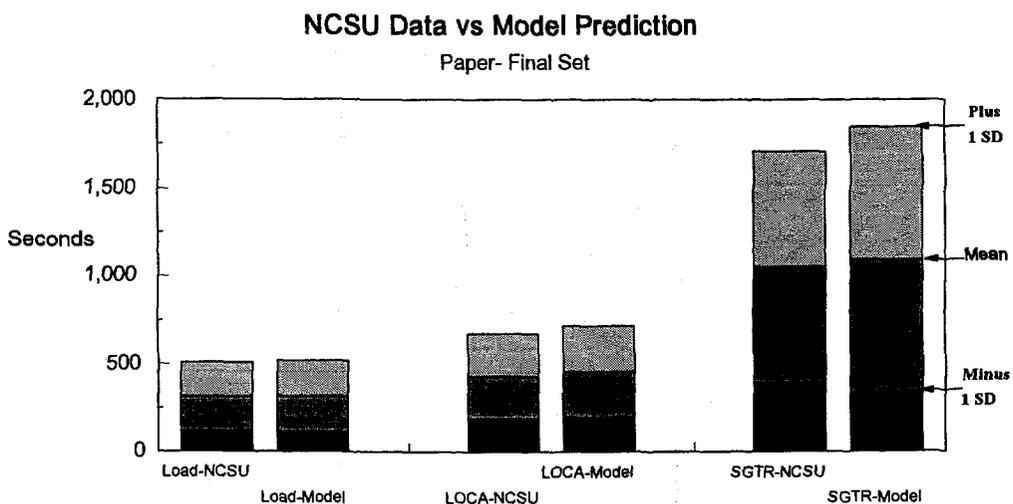

Figure 12. NCSU data versus Model Predictions (standard deviations),
Paper - Final Set

6 DISCUSSION AND CONCLUSIONS

This section is separated into three sections relating to each of the three questions of feasibility, modifiability, and validity. Within each section the strengths of the task network modeling approach, the weaknesses, and the issues which are still outstanding will be presented.

Question 1. *Feasibility* **- Can valid task network models of existing systems be created from the existing task analysis data bases?**

Strengths of Task Network Modeling Approach

Models can easily be built for procedures from existing data - The process of defining the task networks for the procedures studied was very straightforward. In essence, the researchers were able to use the procedure documentation to define the basic tasks and task networks for individual procedures and then use the events in the simulation to trigger the commencement of these procedures. Collection of time data for operator activities for the baseline procedure was also straightforward. In sum, the process for constructing network models of operator procedures was clear and the models themselves were intuitively understandable.

Time to build and modify the models for procedures is short because of the detailed documentation of procedures - Modeling and simulation is viable only to the extent that the models can be developed in a reasonable period of time. For the nuclear power plant procedures used in this study, the time to gather the basic data for model construction was short thanks to the well-documented nature of the procedures themselves. Regardless of whether the procedures are in an easy-to-use form, they are still a complete documentation of the steps that the operator must follow, including contingencies, in the execution of plant procedures. As such, they are an ideal and nearly complete data source for developing operator task networks.

The task time data took longer to obtain. There was no obvious data source from the simulated nuclear facility studied nor is there a known obvious data source at most actual nuclear power plants. What was required for setting the baseline model parameters was the use of the empirical data collected in this study on crew performance on the baseline conditions (i.e., using paper procedures). This data collection was not difficult nor would it be for other procedures or other plants. The technique used in this study, videotaping, represents a low-cost approach to collecting performance time data that could be used elsewhere.

Task error data was not collected nor incorporated into the models, although it could be. Obtaining estimates of low probability errors would demand either 1) extensive data collection to obtain a stable estimate of error or 2) the use of other sources of human error data to obtain error estimates on similar tasks.

To summarize, the building of the baseline model involved a combination of the use of preexisting data sources, empirical data collection, and expert judgment. Based on experience in other domains in which the researchers have used modeling to address design evaluation issues, the balance between the use of hard data and judgment is quite reasonable, if not relatively better, than usual. The time required to collect all of these data and build it into a model was roughly one man-month.

Developing the model forced a level of analysis and rigor that experimentation did not require, but that was worthwhile - Whereas high level data could be collected and analyzed in the experiment at an abstract and high-level, model development required a more detailed analysis of what was occurring during the procedures and, therefore, what might be affected by COPMA-II. As the researchers were building the baseline model, a number of issues related to the possible utility of COPMA-II arose that were driven by the level of

detail of the analysis such as tasks that might be adversely affected and communication problems that might arise or be alleviated by COPMA-II. If this study were an evaluation of COPMA-II, these issues would have been explored either through the model or other data sources. The point is that by having to study the system in detail to construct the model, these issues emerged naturally. However, setting up the NCSU experiment did *not* force this detailed analysis of the COPMA-II system and the data analysis of the NCSU study looked at performance at a very macro-level, therefore missing many of the more micro issues that were identified during model development. Therefore, while the NCSU study did provide some insights as to the value of COPMA-II overall, it did not provide some of the possible insights that *the process of building the model itself* might provide.

It was not possible to decompose the NCSU data to the level that would have allowed a review of what was causing the observed effects of COPMA-II. However, with the model as a basis for analysis, this would have been a more straightforward process.

This is a common anecdote for modeling and simulation - that the process of building the model often provides the analyst with the most useful insights about how the system operates and, therefore, how it will be affected by the experimental factors. Models also provide a view of the system that allows the analyst to understand how the factors are affecting the system. In other words, by examining experimental data in light of a model, one can gain a better understanding of the underlying factors causing differences instead of simply a "yes, it made a difference" or "no, it did not." The value of simulation in this context should not be overlooked.

Weaknesses of the Task Network Modeling Approach

With respect to being able to develop a model of nuclear power plant operating crew performance

in executing procedure, there were no significant weaknesses identified. In effect, the task network modeling approach embodied in Micro Saint was designed for simulating human crews performing these types of tasks. It is therefore no surprise that models of nuclear crews were straightforward to develop.

Outstanding Issues

What about non-procedural tasks such as diagnosis, strategy development, and problem solving? The entire study focused on the use of modeling to study the crews as they executed procedures. However, even with symptom-based procedures, some higher order cognitive behavior is central to the success of the crew in the control room. A complete model of nuclear crew performance would need to be able to simulate these higher-order aspects of human performance.

While Micro Saint and task network modeling has the capacity for embedding models of higher level cognitive behavior into crew performance models, they are not inherently part of the technology. There are ongoing projects for embedding cognitive models into task network models, but the technology is by no means mature. Issues that should be considered are 1) how important are these higher level aspects of cognitive behavior to crew performance in different operating tasks and 2) depending upon the answer to the previous question, what are the boundaries where a network modeling approach would cease to provide useful answers to human factors questions in the nuclear power plant environment?

Question 2. *Modifiability* **- Once created, can a task network model be modified to reflect control room redesign or procedure changes?**

Strengths of Task Network Modeling Approach

Identification of changes required of the model were straightforward - Given a baseline model, it was easy to identify where and how the model needed to be changed to reflect the differences between paper procedures and COPMA-II.

Again, this reflects the wealth of information that was available on the COPMA -II system including access to the engineering experts who were linking it to the SPWRF.

There is every reason to believe that the information available in this study was typical of the type of information that should be available for any equipment or procedure upgrade for an actual plant. The information submitted to the NRC prior to plant or procedure changes would normally be sufficient to assess where and how operator task sequence and demands would change.

Reasonable estimates of time differences for affected tasks were possible without empirical data - The approach used to generate new time estimates for tasks affected by COPMA-II was expert judgment. No empirical data were employed. As the Results section indicates, these estimates at the individual task level resulted in predictions of overall scenario effects that were accurate in predicting the direction of the effect (improvement or decrement) five out of six times and the same statistical conclusions four out of six times.

Weaknesses of the Task Network Modeling Approach

Making time estimates for the new models lacked the rigor and structure that would provide a more defensible model for decision making - While the use of expert judgment provided sound, although not perfect, predictions, they would be difficult to defend as part of any regulatory action than if a more scientific or technical basis were used for estimation.

The reality is that expert judgments are often the basis for estimating model parameters in the real modeling world. There are structured and more rigorous methods for making these estimates, such as the Delphi technique. However, in the end, studies using models to estimate the effects of a system change often rely on the estimates of experts. To ensure that these subjective estimates

do not result in inaccurate conclusions, what is often done is a sensitivity testing of the model predictions to the accuracy of the subjective estimates. If the results are sensitive to a subjective estimate, then more attention can be given to validating those estimates. If the model results are not sensitive to the range of possible values for subjectively estimated parameters, then no further validation is required. This approach is common for many applications of modeling to real-world problems.

It is worth noting that, in the world of human performance modeling, there are advances being made in the development of "first principle" models for estimating human performance parameters such as time and error rates. Additionally, the availability of improved data bases on human performance is increasing. Method time measurement (MTM) data is now available for many types of tasks. Data bases specific to the nuclear domain have been intermittently under development over the past ten years. As these sources of time and error parameter estimates improve, the need to rely on subjective estimates will decrease.

Difficult to predict effects on performance may not be picked up with network modeling - The reductionist approach represented by task network modeling does not always lend itself to an analysis of the types of human performance variability that are critical to success in the real-world. In this study, the operator behaviors were fairly structured and, therefore, the reductionist task network modeling approach was sufficient. However, many tasks may be subject to significant individual/team differences with respect to how the task is performed. In fact, based upon the data and anecdotal reviews of the teams, there were substantial team differences even in the well-structured tasks studied here. As the crew tasks become more emergent (i.e., crew strategies change in unpredictable ways as the event emerges), it will be more difficult to capture these emergent strategies in a task network model. Theoretically, a network model can be constructed with appropriate branching and task

logic to reflect the emergent task environment. Practically, this can become quite difficult.

Outstanding Issues

Again, what about non-procedural tasks? - The same issues that apply to the development of models for higher-order cognitive behavior apply to how would the models be modified to predict the system design or other changes to be studied with the model.

How can the process of revising performance parameter estimates be improved? - As discussed above, acceptance of the model predictions as a basis for regulatory action will hinge on the analysts' ability to justify their estimates. The more theoretically or empirically sound these estimates, the better the results and the more likely the results will be accepted. While new technologies are emerging and being incorporated into task network modeling tools, the question that still must be addressed is "What is a sufficient basis for estimation of task parameter changes?" This question would be difficult to answer empirically as there are far too many variables that would relate to model validity. Rather, it is largely a policy question. For example, would the use of expert judgments be a sufficient basis for estimating parameter changes in the absence of better data sources or would more scientifically or empirically determined estimates be needed? Perhaps the best point of departure for the analysis would be to examine the basis for estimation for two modeling approaches that are already in wide use at the NRC - probabilistic risk assessment and the nuclear reactor safety codes. In both of these types of models, estimates are made regarding parameters of the model to reflect the plant issues under study. What is an acceptable basis for these estimates? Then, how can these current practices be used to define requirements for estimating parameter changes in task network models?

What will be the acceptance by NRC regulators and industry? - While the technical issues raised above are important, an issue that must be confronted directly is how is this technology perceived by the regulators and the industry? Practically, to use task network models to develop and evaluate technical bases for regulatory action, the approach must have sufficient face validity to be accepted by these communities. What degree of validity must be demonstrated for this technology to be accepted?

Question 3. *Predictive Validity* **- Do these modified task network models provide valid predictions of human performance times and error rates?**

Strengths of Task Network Modeling Approach

The task network models predicted the same results as were observed in the human performance data in most, but not all cases - The following points summarize the data with respect to predictive validity:

1. Some evidence supporting the ability of the model to predict statistically significant effects was provided by the observation that the model matched the experimental findings in four of the six conditions.

2. In five of the six conditions, the model predicted effects in the same direction as the experimental data.

3. In absolute terms, the model values matched the actual experimental values in four of the six conditions.

These results are encouraging, but not conclusive evidence of the strength of task network modeling in predicting nuclear operator performance in all situations.

An interesting analysis would be to review the model and experimental data at a more detailed level to assess the source of the differences. The NCSU data divided procedures that were taking on average approximately fifteen minutes into two groupings. Therefore, the model comparisons had to be at this level as well. The models had the

most difficulty predicting situations where the NCSU data had large variances. In some cases, the NSCU data variance equaled more than half of the mean. The models did not predict these large variances. This implies that there are some activities being performed by the crews in the experiment that are either not carefully controlled or are highly variable. Only a more careful analysis of the data would clarify which situation led to model/data differences.

This study highlights the potential synergy between experimentation and model-based applied research - Another common use of simulation is to take the results of a small experiment and use a model to extrapolate the data to other tasks. For example, the data from this relatively small experiment could be used to estimate the effects of COPMA-II on individual types of tasks. To further estimate the utility of COPMA-II on other tasks, rather than incurring the high costs of human subjects experimentation, these estimates could be use as input to much larger modeling studies that look at the effects of COPMA-II across many different procedures and operating scenarios. Then, as high risk areas are identified through the models, human subjects experimentation could be used for final validation if deemed necessary. In other words, the experimentation helps to build the models and estimate how the parameters are affected by the factor being studied (e.g.,

COPMA-II) and then the models are used to extend these results. This flow of information between models and experiments could be as shown in Figure 13.

To use this approach, a much finer level of data analysis for the human subjects experimentation would need to be employed.

Weaknesses of the Task Network Modeling Approach

The predictive validity was too low to prove the predictive validity of task network modeling - As stated above, the results were good, but not good enough to declare a clear success of the modeling approach.

Outstanding Issues

A more detailed analysis of the individual time data would be illuminating with respect to how our individual task estimates compared to actual data - As discussed above, future validation studies should include data analysis at a higher resolution than permitted in this study. Then, the real strengths and weaknesses of modeling vs. experimentation can be properly assessed.

7 SUMMARY

In summary, the following could be said about this study of task network modeling:

1. Models were straightforward to develop for nuclear power plant procedures.

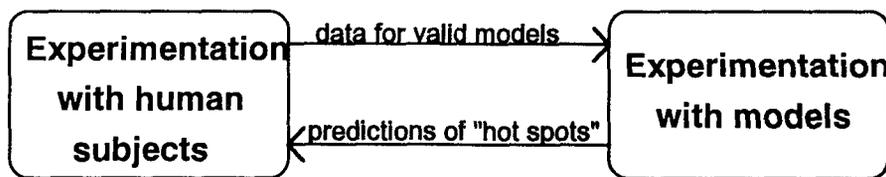2. Defining how the models had to be changed to reflect computerized procedures was also straightforward.

Figure 13. Possible synergy between human-subjects and model-based experimentation

3. The predictive validity of task network modeling as shown in this study is encouraging, but not sufficient for establishing it as a standard for forming technical bases for regulatory action.

All in all, the above results are encouraging. Given that this was intended to be an exploratory study to evaluate feasibility with an eye towards evaluating validity, the results indicate that further research on this technology is warranted.

8 REFERENCES

Carson, S.G., Griffith, W.E., and Winsborough, M., "Command and Control Team Performance Model: Trafalgar Class Submarine." Interim Technical Report, Defense Research Agency - Alverstoke, United Kingdom prepared by Micro Analysis and Design, Inc., Boulder, Colorado. April 1994.

Chubb, G.P., Laughery, K.R., and Pritsker, A.B., "Simulating Manned Systems." Chapter in the *Handbook of Human Factors*, Edited By Gavriel Salvendy, Wiley-Interscience Publication, New York, New York. 1987.

Coffman, F., Persensky, J., Ryan, T., Ramey-Smith, A., Goodman, C., Serig, D., and Trager, E., Human Factors Regulatory Research Program Plan. NUREG-1384 prepared by the Nuclear Regulatory Commission, October 1989.

Converse, S.A., Perez, P., Myer, S., and Crabtree, W., "Evaluation of the Computerized Procedures Manual." Paper presented at the Enlarged Halden Program Group Meeting, Storefjell, Norway, March 1993.

Harshell, J. and Dahl, S., "Simulation Model Developed to Convert Production to Cellular Manufacturing Layout." *Industrial Engineering*, December 1988.

Laughery, K.R., Dahl, S, Kaplan, J., Archer, R., and Fontenelle, G., "A Manpower Determination Aid Base upon System Performance Requirements." *Proceedings of the 32nd Annual*

Human Factors Society Meeting, Anaheim, CA. October 1988.

Laughery, K.R., Drews, C., and Archer, R., (1986). A Micro Saint Simulation Analyzing Operator Workload in a Future Attack Helicopter. In *Proceedings of NAECON Meeting, Dayton*, Ohio, May 1986.

Lockett, J.F., Plocher, T., and Dahl, S, "Crew Reduction in Armored Vehicles Ergonomic Study." In *Contemporary Ergonomics 1990*, Proceedings of the Ergonomics Society Meeting, Loughborough, Leicestershire, United Kingdom. April 1990.

McMillan, G.R., Beeves, D., Salas, B.E., Strub, M.H., Sutton, R., and Van Breda, L., "*Applications of Human Performance Models to System Design*." New York, New York, 1989.

Monk, T., "Programmer's Guide to the DEPTH Micro Models." Software documentation prepared by Micro Analysis and Design, Inc. for the Air Force Armstrong Aeromedical Research Laboratories, Dayton, Ohio. April 1994.

Rayner, A. and Laughery, K.R., "Simulation in the Health Care Industry." Presented at the Simulation in Health Care Meeting in London, England, April 1993.
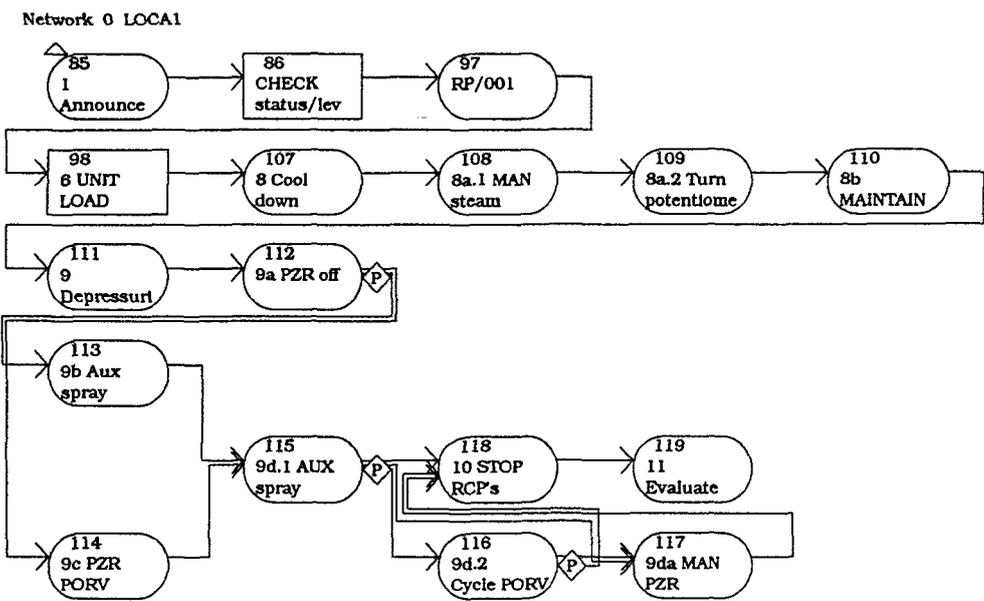
Implementation." Paper presented at the Enlarged Halden Program Group Meeting, Storefjell, Norway, March 1993.

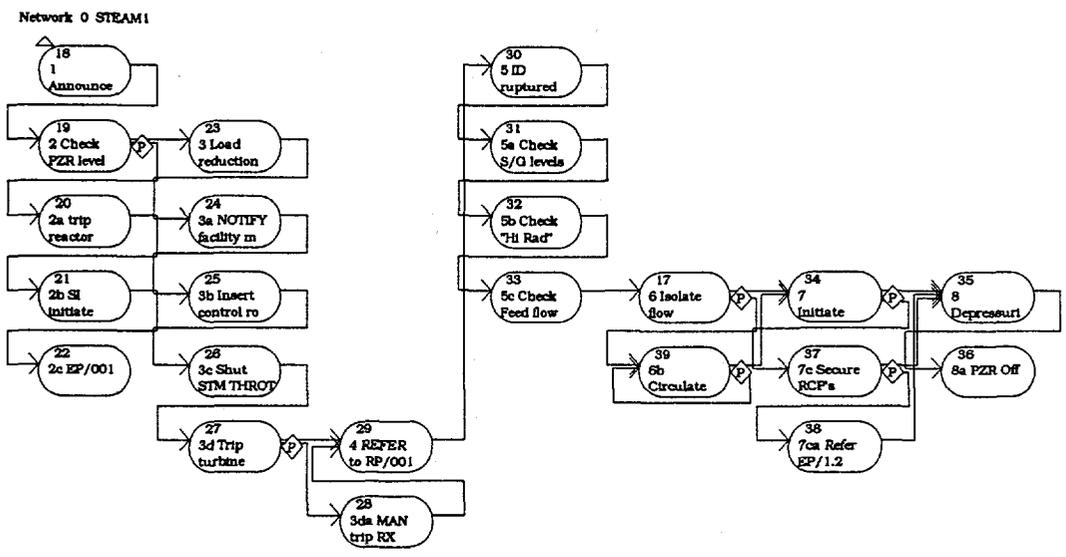# APPENDIX A - TASK NETWORK DIAGRAMS OF THE MODELS USED IN THE STUDY
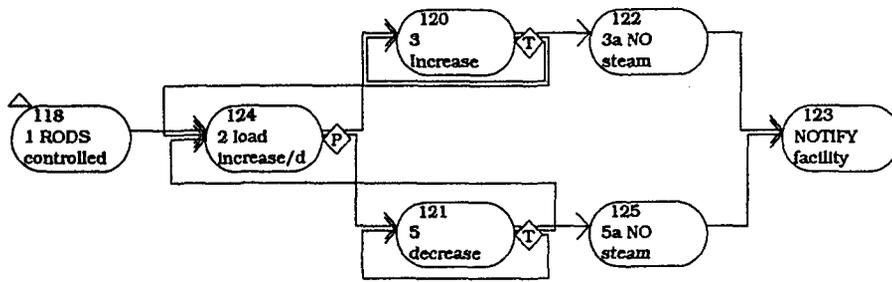
Task Network Diagram For The Load Maneuvers Models

Task Network Diagram For The Loss of Coolant Accident (LOCA) Models

Network 0 STEAM1

Task Network Diagram For The Steam Generator Tube Rupture Models