Proceedings of the 1996 Winter Simulation Conference
ed. J. M. Charnes, D. J. Morrice, D. T. Brunner, and J. J. Swain

DISCRETE EVENT SIMULATION AS A TOOL TO DETERMINE NECESSARY NUCLEAR POWER PLANT OPERATING CREW SIZE

Ron Laughery
Beth M. Plott
Thomas H. Engh
Shelly Scott-Nash

Micro Analysis and Design, Inc.
4900 Pearl East Circle, Suite 201E
Boulder, Colorado 80301, U.S.A.

ABSTRACT

There are not always sufficient resources or time available to identify human factors issues early enough for development of detailed technical bases using empirical experimentation with human subjects. Consequently, analytical approaches are needed to augment the experimental approach for human factors regulatory decision making at the US Nuclear Regulatory Commission. One analytical approach, computer modeling of human performance, is being investigated by the NRC Office of Nuclear Regulatory Research. As an example of the types of supporting research required, we discuss two specific studies pertaining to the use of Micro Saint, a discrete event simulation package, as a means of evaluating the effects of crew size on safety in a nuclear power plant setting. Both studies provided data that permit an evaluation of the practicality and validity of using models built in Micro Saint for the specific purpose of studying staffing issues, as well as the value of modeling of human performance in general.

1 INTRODUCTION

Current shift staff requirements in nuclear power plants in the United States have been based primarily upon experience. While computer models have been used to predict thermal hydraulic and hardware performance of the plant, no such tools were available for considering the human element. There are many reasons to consider human performance when making predictions of any system's performance. For the NRC, the need to make sound licensing decisions is paramount and a key determinant of plant safety is human performance. The number of operators in a control room, the tasks those operators perform, and how well operators perform tasks are all significant factors in plant safety. Experience, expert judgment, experiments, and modeling are all tools used to make predictions of human performance in a variety of non-nuclear environments. When these tools are used together, the resulting "consensus" prediction carries more weight than a prediction derived from only one of these tools. To strengthen both the analytical power and believability of human factors analysis, it is our position that if models of human performance can be shown to be valid predictors , these tools should be one of the methods used for predicting human performance in nuclear power plants.

A specific area of NRC human factors analysis that is ripe for the use of human performance modeling is the area of determining shift staffing requirements. In 1993, the Nuclear Regulatory Commission (NRC) began several research projects seeking to establish a technical basis for criteria for minimum shift staffing levels for licensed and non-licensed operators at nuclear power plants. Specifically, the NRC sought to either confirm the adequacy of requirements of 10 CFR 50.54 (m) or establish a basis for modification. These requirements deal with the minimum number of licensed operators required to operate a nuclear power plant. The required number of licensed operators varies based upon the plant design. As we will discuss, there are computer modeling tools available that permit the analysis of staff requirements as a function of task and equipment design. Because this was an area of specific interest to the NRC, laboratory research was planned to address the issue empirically. This combination of 1) a need for answers on the staffing issue, 2) the availability of models that were directly relevant, and 3) the potential for empirical research that could be used to validate the models provided an ideal opportunity to seriously explore the use of human/system modeling at the NRC.

The research program that the NRC and Micro Analysis and Design have been pursuing involves leveraging the shift staffing empirical research with the goals of 1) the development of models designed to predict human/system performance as a function of staff makeup and then 2) comparing the predictions from these models to experimental data. The comparison of model predictions to experimental data is the essence of any model validation exercise. If the models predict reasonably accurately, then the models can be used to significantly augment empirical research and, when empirical research is impossible, to serve as a means for performing sensitivity studies.

The remainder of this paper provides an overview of this model validation research program. First, we will discuss the human/system modeling technology that is being used and evaluated - Micro Saint. Second, we will provide overviews of two of the research projects that are underway to evaluate the validity of models built in Micro Saint to predict human/system performance in the nuclear power plant environment.

The goal of this paper is to describe a framework for the investigation into the use of modeling as a tool to support regulatory decision making and to discuss some experiences we have had in attempting to fill out this framework.

2 A BRIEF DESCRIPTION OF THE APPROACH TO HUMAN/SYSTEM PERFORMANCE MODELING - TASK NETWORK MODELING AND MICRO SAINT

One technology that has evolved over the past 20 years for predicting human/system performance is *task network modeling*. In a task network model, human performance of an individual performing a function (e.g., performing a procedure) 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 simple task network for dialing a telephone.

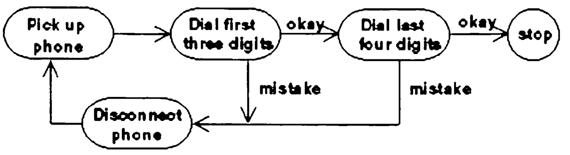

Figure 1: Example of a Task Network for Dialing a Phone

Task network modeling is an approach to modeling human performance in complex systems that has evolved for several reasons. First, it is a reasonable means for extending the human factors staple - task analysis. Task analyses organized by task sequence are the basis for the task network model. Second, as we will show in the example below, task network models can include sophisticated submodels of the plant hardware and software to create a closed-loop representation of the human/machine system. Third, task network modeling is relatively easy to use and understand. With Micro Saint, a human factors analyst can build computer models of task networks that can involve many operators interacting with a complex system in complex ways. For more details of task network models and how they perform, please see McMillan et al. (1989) and Micro Analysis and Design (1993).

2.1 An Example of a Task Network Model of a Process Control Operator

This simple hypothetical example illustrates how many of the basic concepts of task network modeling can be applied to studying human performance in a process control (e.g., nuclear power plant operation) environment. The simple human task for our model is of an operator responding to an annunciator. The procedure requires that the operator compare two meter readings. Based on the relative values of these readings, the operator must either open or close a valve until the two meter values are nearly the same. The operator activities for this model are represented by the task network in Figure 2. 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 3.

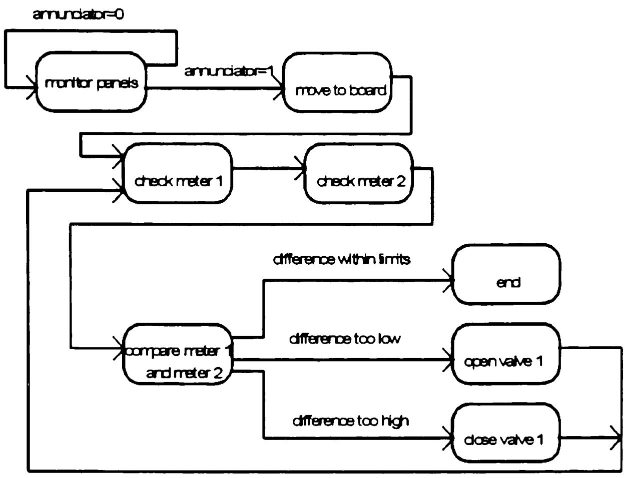

Figure 2: Sample Task Network Model of a Process Control Operator Responding to an Annunciator

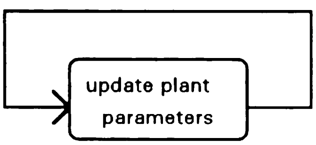

Figure 3: One Node Model of the Plant

The operator portion of the model will run the "monitor panels" task until the values of the variables "meter 1" and "meter 2" are different. The simulation could begin with these values being equal and then precipitate a change in values based on what is referred to as a scenario event (e.g., an event representing the effects of a line-break on plant state). This event could be as simple as:

meter 1 = meter 1 + 2.0;

or as complex as an expression defining the change in the meter as a function of line break size, flow rates, etc.

When the transient occurs and the values of "meter 1" and "meter 2" start to diverge, the annunciator flag will be set in the model to "on." Then, the operator will move to the appropriate board. The operator model will continue through a loop where the operator checks the values for "meter 1" and "meter 2" and either opens "valve 1," closes "valve 1," 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 a model parameter. In this simple model, operators do not consider rates of change in values for the meter they are viewing 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 meter values when deciding whether to open or close valves.

Again, this is a very small model reflecting simple operator activity on one control via a review of two displays. However, it illustrates how large models of operator teams looking at numerous controls and manipulating many displays could be built via the same building blocks used in this model. The central concepts of a task network and shared variables reflecting human/system dynamics remains the same.

2.2 Addressing Staffing Issues with Task Network Modeling and Micro Saint

The technology described above is the basis for evaluating many aspects of human/system design. The specific nuclear power plant issue in which we were interested was the evaluation of staffing requirements both within and outside of the nuclear power plant control room.

The idea of using Micro Saint to address staffing and manpower issues in not new. Drews et al. (1985) reported the use of Micro Saint to evaluate the number of crew members required in a helicopter being designed by the Army. The Army had hoped that the helicopter could be safely and effectively operated with fewer crewmembers with extensive use of automation in the cockpit. However, the Micro Saint analysis of crew workload helped to ascertain that this staffing reduction would not be possible. The methods that were used during this research were later embedded into several Micro Saint-based tools for evaluating staffing requirements for Army systems and missions (e.g., Fontenelle and Laughery, 1988). These tools are now being used routinely to evaluate staffing requirements in the Army (e.g., Plott, 1995) and, in fact, have successfully undergone the Army's rigorous process for model verification, validation, and accreditation (Allender et al., 1995).

The details of the specific techniques for studying staffing issues with Micro Saint are too detailed to delve into in this paper, but are well documented in the references above. Briefly, we have developed a method for estimating individual and crew workload by building workload estimation metrics into a Micro Saint model. The underlying theory of workload that we employ is the Wickens' Multiple Resource Theory (Wickens, Sandry, and Vidulich, 1983). By embedding a representation of this theory within the task network, we can determine points at which each human's workload is unreasonably high and, therefore, when task performance can be expected to change and potentially suffer (e.g., task delays, increased error rates). Through the model, we can study the possibility of reallocating tasks to other available, qualified operators when workload appears to be too high in an attempt to "level out" workload across the crew. However, if this leveling out still cannot solve problems of excessive workload demands or undesired task delays because of high workload, then the model tells us that we are designing a system that will probably not perform as the designers intended because they are expecting more of the human operators than the operators have the capacity to provide.

This approach, which is identical to that developed and accredited by the US Army, is the approach that we were testing in the series of studies described briefly below. Because these studies are still ongoing, we can not state definitive success. However, the trends that we are seeing in the data are consistent with what we have found in the other applications discussed above.

3 VALIDATION STUDY 1 - THE USE AND EVALUATION OF MICRO SAINT MODELS IN ASSESSING OUTSIDE THE CONTROL ROOM STAFFING REQUIREMENTS

Background - In August 1993, the Nuclear Regulatory Commission (NRC) contracted with Brookhaven National Laboratories (BNL) to establish a technical basis for criteria for minimum shift staffing levels of licensed and non-licensed operators at nuclear power plants. As part of this effort, a detailed, on-site data collection effort was performed. The goal of the data collection was to assess the adequacy of current nuclear power plant (NPP) staffing practices for performing the activities necessary for responding to and mitigating emergency events. A total of seven NPPs were visited and information was collected regarding current staffing practices for both licensed and non-licensed shift personnel. The plants visited represented different plant types, vendor types, NRC regions, number of units, and plant ages. The two person research teams sent to each site consisted of both an operational and a human factors specialist. Four methods were used to collect data: plant documentation review, table-top analyses, walk-throughs of specific outside of the control room tasks, and interviews with individuals from different groups and organizational levels at the site. At the conclusion of the data collection exercise, the data was synthesized into a timeline or GANTT chart for each plant and scenario. Two different scenario events were examined at each of the seven NPPs.

The next phase of the effort was to develop Micro Saint models of these events. The models were built and modified to reflect system, personnel, timing, and resource changes. This effort was divided into two phases. For the first phase, five base models covering different scenario events were developed from the task analysis and timing data collected during the site visits. Often, this involved a more detailed analysis of the process followed by the operators. In the second phase, the baseline models were modified to reflect different plant parameters and staffing configurations. Only subject matter expertise and plant specific procedures were used to modify the models -- the data collected on-site was not used in the model modification process. The modified models were then executed and compared to the collected data to conduct model validation. Figure 4 illustrates the process used. This evaluation provided an excellent test of the models' validity for evaluating staffing.

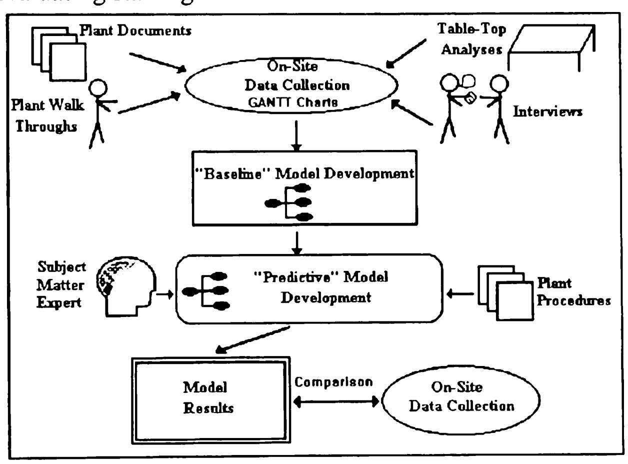

Figure 4: Model Development Process

Results -

<u>Baseline Model Development</u>

As an indication of possible staffing shortages, the models were constructed to predict and record when tasks were delayed in starting due to operator unavailability. Since the collected data was used as input to construct the baseline models it was expected that the model output would be closely aligned with the on-site data. This was largely true. When the model predicted task delays due to operator unavailability for more tasks than the data collection was able to capture, we conducted an item-by-item analysis to determine why. We found that the variability between the data collected and the model output was due to:

- the GANTT chart analysis represented a synthesis of table top data and walk-through data from two separate observers who sometimes disagreed
- the model development task analysis of the scenarios was conducted at a more detailed level than was used for the GANTT chart analysis
- when building the model, the source data from the data collection sometimes had to be redefined into a task framework that more clearly defined operator tasks, task sequences, and task interdependencies
- task data that were not obtained in the data collection were estimated by a subject matter expert (SME)

In sum, the construction of the baseline models required a more detailed analysis of the process than the table-top analyses and walk-throughs. Therefore, the opinion of the SME was that, in all likelihood, the model predictions were more realistic than the "data."

Predictive Model Development

After the baseline models were built, we altered them to reflect different plant parameters and staffing conditions using only subject-matter expertise and plant specific procedures for the other plants and scenarios. The actual data collected on-site was not used to construct the models. The new predictive models were then executed and the results compared to the data collected at the site visits.

The feasibility of using the modeling approach was evaluated by comparing the simulated response of crews at different plants with the data collected at those plants. Plant-specific task network models were developed for the new plants by modifying the existing baseline models for a similar type of plant (PWR, BWR) and scenario. The task networks were modified on the basis of expert judgment to account only for differences in plant-specific equipment or procedural requirements between the baseline plant and the new plant. The completion times for the baseline plant were used for corresponding tasks on the modeled plant. Completion times were only modified when there were plant-specific differences between the baseline plant and the modeled plant that would clearly impact the performance of the task. In all cases, the task networks were modified without referring to the on-site data collected from the new plant for operator tasks. After the task network was constructed, the on-shift personnel resources were selected for the modeled plant.

After the task network was constructed and the shift staffing level was established, the Micro Saint model was run for 100 iterations using normalized statistics for task completion times. The results were tabulated and then compared to the on-site data that had been collected by the Brookhaven team at the modeled plant. This analysis compared the observed start times of the various tasks from the on-site data collection effort to the predicted mean start times and standard deviations from the Micro Saint model.

Again, we found that virtually all observed differences were attributable to the higher resolution of the model's analysis than the analysis that was performed at the site. Both phases of the model development process provided insight into the staffing practices and procedures used at the plants that were not readily apparent from the collected data. In general, three types of insights were gained that were not apparent from either the on-site table-top analysis or the high level GANTT chart analysis. These areas are:

- *Deviation from the procedures* - The extended "predictive" models were built assuming that the crews would follow the operating procedures of the plant. Instances were identified where operators deviated from station procedures when the table-top and walk-through analyses were conducted.
- *Delay of tasks due to lack of staff availability* - During the data collection, the Shift Supervisor (SS) occasionally lost track of the personnel resources available to respond to the event. In some cases, the SS thought that there were available Auxiliary Operators (AOs) and directed them to perform tasks when all AOs were engaged in completing previously assigned tasks. In other instances, the SS didn't realize that an operator was available. Some of these instances were caused by the SS under-estimating the time that it would actually take to complete a task in the plant (as compared to the plant walk-through data). In addition, the models simulated several administrative tasks that were not included in the table-top simulation and on-site data collection. Although these tasks have been observed to occur during actual events they were beyond the capability of the table-top methodology, but easily addressed by the model.

- *Close consideration of the link between inside and outside control room activities* - The activities in a nuclear power plant during an event are closely tied to the status of the plant. The nature of the table-top procedure does not allow for the close link between plant status, control room activities, and outside control room activities to be considered in detail whereas the model did.

Summary - This study of the validity of task network modeling to study staffing issues demonstrated how models can be used in conjunction with empirical data to improve the basis for decision making. The models were able to use the empirical data as a starting point and then *immediately improve the value and accuracy of the empirical data.* Additionally, when subjected to the predictive validation test with new data, the model predictions were, in some ways, better representations of expected human/system performance than the empirical data.

4 STUDY 2 - THE USE AND EVALUATION OF MICRO SAINT MODELS IN ASSESSING INSIDE THE CONTROL ROOM STAFFING REQUIREMENTS

The work reported below is ongoing as of this writing, so the discussion below will serve as more of a status report than a final description of the results. For more information on the final results of this study, please contact any of the authors.

Background - In 1994, the NRC entered into a bilateral research agreement with Halden Reactor Project in Halden, Norway. The purpose of this agreement was to conduct research into the staffing requirements of advanced control rooms. In 1995 and continuing into 1996, Halden staff are conducting empirical studies with operating crews at the Lovissa Nuclear Power Plant in Finland. The first phase of these studies, which were completed in 1995, involved the comparison of 3 vs. 4 person operating crews in a conventional control room under five different emergency scenarios. Measures of human/system performance for these crews were collected and are being analyzed.

Method - As with the other study, we are using one set of the data as a baseline set for building and calibrating the models. This set of data is the 4-person crew configuration for two of the four scenarios. The scenarios that were calibrated involved a steam generator tube rupture (SGTR) and a loss of off-site power (LOOP). We have completed this phase. Then, using analyst and subject-matter expert input, we will change these models to reflect what we *expect* to occur when the staff is reduced to a 3-person crew configuration. The subject matter expert inputs to the model will reflect expected changes in task time and allocation to operators. The model will then be used to predict overall human/system performance as a function of these expected changes when reducing the crew from four to three people. Finally, we will compare our model predictions to the actual data that was obtained in the 3-person phase of the study. This second portion of the phase is expected to be completed in mid 1996.

The measures of crew performance that we are predicting with the models and that were collected in the Halden study include 1) the time required to perform critical groups of tasks in the scenario and 2) subjective workload. These measures were selected for modeling due to their relevance in assessment of the changes in crew size. The time a critical task is delayed due to a lack of resources can impact the probability of a serious plant problem occurring. The subjective workload is a relevant measure because of links shown between workload and the error rate of humans. By modeling the control room operators' workload, it is hoped to determine portions of the scenario where the operator is likely to make errors.

Results - Since only the first phase of this study - model development and calibration have been completed, we report only those results here.

The initial model was developed with purely SME input. In other words, we performed a table top analysis using Lovissa operators and developed the task networks, task times and error rates, and the task interdependencies. Then, we used the four person crew empirical data from Lovissa (i.e., the baseline data set) to calibrate the model. Therefore, one measure of the model's validity was the amount of calibration that was required to make the model match the baseline data. In this study less than 5% of the tasks in the models required any modification.

With respect to getting the model time lines to match the data time lines, very little calibration had to be done. Calibration of the models primarily consisted of synchronizing plant events from the experimental data, from the Lovissa project, to events that occurred in the models. The task times predicted by the SMEs proved to be highly accurate. The majority of calibration involved removing tasks that were in the model as a result of the task analysis and were not actually performed in the experiment. Adjustments were also made with regard to the number of times repeating tasks were performed so the number of repetitions more closely reflect the data obtained experimentally.

The portions of the model that predicted crew performance, measured by task completion time, were statistically the same as the data we had obtained from the SMEs when building the models. After calibration, the correlation coefficient for the SGTR scenario times was 0.97 in comparing the model times to the times collected for the 4 person crew. For the LOOP scenario, a correlation coefficient of 0.99 between the model and experimental data was obtained.

With respect to the workload predictions, the model workload scales were different from the data collected in the study. The workload values in the model were predicted by SMEs using McCracken and Aldrich's (1984) visual, auditory, cognitive and psychomotor (VACP) scales. The experimental workload values were obtained by administration of the NASA Task Load Index (TLX) scale. The TLX asks subjects to rate their workload based upon mental demand, physical demand, temporal demand, performance, effort and frustration. Therefore, the calibration of the model has involved using regression analysis in an attempt to build a map defining the relationships between model workload predictions (VACP) and the scales used by Halden (TLX). Preliminary results show that the regression models provide a statistically significant estimate of the individual TLX scales. The VACP scores only predict a percentage of the subjective TLX workload scales. The multiple correlation coefficients are on the order of 0.4 to 0.6 for the individual TLX scales.

The next step in this study is to make predictions about crew performance when the number of operators in the control room is reduced from four to three. Modifications in the models related to which operators perform specific tasks, and problem detection probabilities will be made in conjunction with the reduction in the number of operators.

Summary - Preliminary results of this study are quite encouraging with respect to being able to predict task timelines without any empirical data as input. It appears that in this case SME input alone was a valid basis for model design and parameter estimation. The validity of the models to predict workload is yet to be determined. Early results indicate that the VACP and TLX scales are related, but not highly.

5 SUMMARY

In all, we find these results quite encouraging with respect to the potential of modeling human/system behavior as a means of developing technical bases for regulatory decisions. We have shown how we can improve and supplement the empirical data collection process and provide sound results. However, significantly more research and evaluation will be required to assess whether this modeling technology is valid for assessing all staffing issues, much less the many other human factors regulatory issues that are faced by the NRC. The answer, undoubtedly will be "for some things it works, and, for others, probably not."

REFERENCES

Allender, L., T. Kelley, L. Salvi, D. B. Headley, D. Promisel, D. Mitchell, C. Richer, and T. Feng. 1995. Verification, Validation, and Accreditation of a Soldier-System Modeling Tool. In the *Proceedings of the 39th Human Factors and Ergonomics Society Meeting, October 9-13, San Diego, CA* 1219-1223. Human Factors and Ergonomics Society, Santa Monica, CA.

Drews, C., K. R. Laughery, K. Kramme, and R. Archer. 1985. LHX Cockpits: Micro SAINT Simulation Study and Results. Report prepared for Texas Instruments - Equipment Group, Dallas, Texas.

Fontenelle, G.A. and K.R. Laughery. 1988. A Workload Assessment Aid for Human Engineering Design." In *Proceedings of the 32nd Annual Human Factors Society Meeting, Anaheim, CA.* 1122-1125. Human Factors and Ergonomics Society, Santa Monica, CA.

McCracken, J. H. and T. B. Aldrich. 1984. Analysis of selected LHX mission functions: Implications for operator workload and system automation goals. Technical Note ASI479-024-84. Anacapa Sciences, Inc., Fort Rucker, AL.

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

Micro Analysis and Design. 1993. *Micro Saint for Windows User's Guide*. Boulder, Colorado: Micro Analysis and Design.

Plott, B. 1995. Software User's Manual for WinCrew, the Windows-Based Workload and Task Analysis Tool. Aberdeen Proving Ground, MD: U.S. Army Research Laboratory,

Wickens, C. D., D. L. Sandry, and M. Vidulich. 1983. Compatibility and Resource Competition Between Modalities of Input, Central Processing, and Output. *Human Factors* 25:227-248.

AUTHOR BIOGRAPHIES

Ron Laughery is the President of Micro Analysis and Design in Boulder, Colorado. He received B.S., MS., and Ph.D. degrees from the State University of New York at Buffalo. For over twelve years, he has led the development of the discrete-event modeling product Micro Saint. He has also applied Micro Saint to modeling human performance, health care systems, manufacturing systems, and process reenigneering.

Beth M. Plott is a Principal Engineer at Micro Analysis and Design. Ms. Plott holds a Bachelor's degree in Chemical Engineering from the University of Colorado and is currently pursuing a Master's degree in Environmental Engineering from Colorado State University. She has eight years experience in software development and computer modeling. Ms. Plott has developed several simulation models in diverse areas such as manufacturing, health care, retail systems, and nuclear power plants.

Thomas H. Engh is a Human Factors Engineer with Micro Analysis and Design in Boulder, Colorado. He received B.S. and M.S. degrees in industrial engineering from Purdue University in 1991 and 1993 respectively. His interests include the modeling of human performance and user interface design.

Shelly Scott-Nash is a software analyst with Micro Analysis and Design in Boulder, Colorado. She received a B.A. in Math and Computer Science from the University of Colorado in 1987. Her interests include simulation software design and application.