

March 25, 1998

FOR: The Commissioners

FROM: L. Joseph Callan /s/  
Executive Director for Operations

SUBJECT: RESPONSE TO STAFF REQUIREMENTS MEMORANDUM OF OCTOBER 24, 1997, REGARDING IMPROVEMENTS IN SENIOR MANAGEMENT ASSESSMENT PROCESS FOR OPERATING REACTORS (M970919C)

## PURPOSE:

This Commission paper responds to the staff requirements memorandum (SRM) of October 24, 1997, in which the Commission requested that the staff provide the following:

1. A series of graphs providing both false alarm and detection probabilities as a function of the observation period and various decision criteria. The performance model/algorithm should be peer-reviewed by a broad audience of qualified individuals or organizations, including NRC staff, and not limited to just ACRS (9600087).
2. Information regarding benchmarking trend and regression models against something other than decisions emanating out of past SMM processes, an evaluation of the use of proprietary indicators in the staff's assessment process, and a discussion of World Association of Nuclear Operators' (WANO's) performance assessment process as compared to NRC's performance indicators and assessment processes (9600087).
3. Staff's plans regarding the request from the Institute of Nuclear Power Operations (INPO) that the NRC review the INPO/WANO indicators and current arrangements between INPO and NRC regarding the sharing of, or access to, proprietary performance data and indicators (9600087).
4. Information regarding the resource commitments expended to date on the improvement of the current SMM process, an estimate of the anticipated resources needed to complete this effort, and an estimate of the resources needed to maintain and refine the SMM process once it is in place in its final form (9700268).

In addition to the above items, the Commission asked the staff to provide an appropriate feedback mechanism to enable the staff to assess the reliability and validity of the improved SMM process. In accordance with [SECY 98-045](#), the staff proposes to supersede the SMM process with a new integrated performance assessment process. The staff will assess its plans to address the reliability and validity of the new assessment process in the Commission paper forwarding the final proposed integrated assessment process on October 30, 1998.

The Commission also asked the staff to provide plans for additional stakeholder interaction on assessing management effectiveness and utilizing economic indicators. A Commission paper delineating the policy issues associated with assessing licensee management and organizational (M&O) effectiveness and proposing options for dealing with such issues will be completed by March 27, 1998. After obtaining Commission feedback on the policy issues, the staff will determine the need for and timing of stakeholder interaction. The staff intends to release the economic indicators for public comment along with the new integrated assessment process discussed above. The new integrated assessment process and M&O assessment options will be the subject of a Commission meeting on April 2, 1998.

## BACKGROUND:

In an SRM dated June 28, 1996, the Commission directed the staff to assess the SMM process and evaluate the development of indicators that can serve as a basis for judging whether a plant should be placed on or deleted from the "watch list." In response, the consulting firm, Arthur Andersen, performed a study for the staff on the effectiveness of the SMM process and issued a report, "Recommendations to Improve the Senior Management Meeting Process," on December 30, 1996. On September 19, 1997, the staff briefed the Commission on (1) the status of improvements to the SMM process, and (2) the staff's plans to perform an integrated review of the NRC assessment process for operating plants. Subsequently, the Commission issued SRM M970919C on October 24, 1997, asking the staff to provide information on several issues, including those noted above.

The Office of Nuclear Reactor Regulation (NRR) issued SECY-97-294 on December 19, 1997, responding to the following issues in SRM M970919C:

- How the staff intends to use such information as performance trends and economic plots to reach decisions at the SMM, and the staff's plans for responding to an ACRS letter dated September 10, 1997, regarding the SMM Template and the SMM process in general.
- Benefits that the SMM process and watch list provide in terms of public health and safety, and whether the SMM is the dominant driver of the NRC's inspection, assessment, and regulatory processes.

The Office for Analysis and Evaluation of Operational Data (AEOD) was assigned the responsibility for responding to the remaining issues. The AEOD responses to these issues are given below.

## DISCUSSION:

ITEM 1: A SERIES OF GRAPHS TO BE PROVIDED BY THE STAFF FOR FALSE ALARM AND DETECTION PROBABILITIES AS A FUNCTION OF THE

## OBSERVATION PERIOD AND VARIOUS DECISION CRITERIA, AND BROAD PEER REVIEW OF THE PERFORMANCE MODEL / ALGORITHM.

The Commission asked the staff to provide a series of graphs that provide both false alarm and detection probabilities as a function of the observation period and various decision criteria. The Commission suggested that graphs could be used to provide a basis for selecting appropriate decision criteria and duration period. The performance model/algorithm should be peer-reviewed by a broad audience of qualified individuals or organizations, including staff, and not limited to just ACRS.

### Response:

#### a. A Series of Graphs that Provide Both False Alarm and Detection Probabilities as a Function of the Observation Period and Various Decision Criteria

The development and use of the performance trend model is only one part of a larger process which considers both quantitative and qualitative information using the Plant Performance Template to improve the identification of plants for discussion at the SMM. The trend model is used to identify candidate plants for consideration in the overall SMM process. It is not intended to be the precise definitive identifying element. Therefore, the nature of the trend plots will result in the occurrence of false positives (where the model indicates a candidate plant for discussion but the historical record indicates it was not discussed during the SMM) and false negatives (where the model does not indicate a candidate plant for discussion while the historical record indicates it was discussed). While the model parameters were chosen to minimize the number of false positives and false negatives, it is the use of the trend model in conjunction with the Plant Performance Template in the integrated process that is relied upon to correct these mis-identifications.

Attached are the graphs of the type requested by the Commission ([Attachment 1](#)). These graphs are based on a trend model which monitors eight variables. These graphs provide both detection and false alarm probabilities as a function of various observation periods and decision criteria. The titles of the graphs identify the observation period and decision criteria. Each graph contains three curves: (1) Identified for Discussion; (2) False Positive; and (3) False Negative.

The "Identified for Discussion" curve shows the probability that a plant will be identified as a candidate for discussion versus the number of variables (hits) exceeding the industry value for the noted observation period and decision criteria. The percentage of the plants identified by the model as candidates for discussion at each hit level which were not actually discussed are indicated by the "False Positive" curve. Additionally, the percentages of the total number of plants not identified by the model but actually discussed are indicated by the "False Negative" curve. For example, in Figure 1 (6 Quarter Average - 1 std dev), the percentage of the total number of plants which this model would identify as candidates for discussion if two of the variables exceeded a standard deviation from the industry (i.e., a hit level of 2) would be 20 percent. Of these, approximately 15 percent would not be discussed at the SMM (false positives). Additionally, approximately 5 percent of the total population would not be identified by the model for discussion, but would be discussed (false negatives). For a population of 100 plants, this means that 20 plants would be identified as candidates for discussion, of which 3 are false positives, and 5 additional plants (false negatives) would be discussed but would not have been identified by the model.

Based in part on the model performance demonstrated by these curves, the staff chose the model represented in Figure 1 for trial use during the January 1998 SMM cycle. As shown in Figures 1-3, there was similar performance for the four, six, and eight quarter average plots. The six quarter period used in this model was chosen because it nominally represents a refueling cycle. We are currently evaluating the use of a four quarter unweighted average for future work. Based on the staff's analysis of the variability of the data, the single standard deviation was selected. At standard deviations of two or greater, fewer plants were identified for discussion and the false negative rate was somewhat higher.

As shown in Figure 1, the false positive and false negative curves are closest at a hit threshold of three. However, a threshold value of two hits was selected due to the advantage of reducing the false negative indications without significantly increasing the number of false positives. The impact of giving greater weighting to the most recent performance was also examined and found not to have sufficient differences in performance to justify its use over the equal weighting analysis.

The candidate plants identified for the January 1998 SMM cycle using this model agreed well with the actual discussion list. While there were a few false positives and false negatives in the candidate list, all of the plants placed on the watch list at the January 1998 SMM were identified by the trend model using these screening criteria.

#### b. Peer Review of the Performance Model/Algorithm

As part of the overall SMM process improvement effort, the performance model/ algorithm will receive a broad review, including review by industry and the public. A draft report summarizing the model development effort was issued in November 1997 for internal NRC comment, including the ACRS. Model development continues, incorporating the comments received along with comments generated during the recently completed SMM cycle. A revised report will be included as part of the package to be issued for public comment. Additionally, the report will be specifically provided to INPO and NEI for peer review and comment. These two organizations have been, or are developing, similar plant performance monitoring indices. The statistical processes employed in the trending methodology are simple and well known, and will receive a broad review along with the overall SMM improvement process.

## ITEM 2: INFORMATION REGARDING BENCHMARKING TREND AND REGRESSION MODELS AGAINST SOMETHING OTHER THAN DECISIONS EMANATING OUT OF PAST SMM PROCESSES, AND EVALUATION OF THE USE OF PROPRIETARY INDICATORS IN THE STAFF'S ASSESSMENT PROCESS, AND A COMPARISON OF WANO'S PERFORMANCE ASSESSMENT PROCESS WITH THE NRC'S PROCESSES.

The Commission asked whether the trend and regression models can be benchmarked against something other than decisions emanating out of the past

SMM processes. The staff should consider and evaluate the use of proprietary indicators in its assessment strategy, and provide a discussion of WANO's performance assessment process, including the significance of their "figure of merit," and the means by which it is derived and the performance indicators that are utilized, as compared to NRC's performance indicators and assessment processes.

**ITEM 3: STAFF'S PLANS REGARDING INPO'S REQUEST THAT THE STAFF REVIEW INPO/WANO INDICATORS. ARRANGEMENTS BETWEEN INPO AND NRC TO SHARE PROPRIETARY PERFORMANCE DATA AND INDICATORS.**

The Commission stated that the staff should inform the Commission of its plans regarding INPO's request that the NRC take a hard look at the INPO/WANO indicators. In its response, the staff should also discuss its past activities, experiences, and conclusions regarding its attempts to obtain, evaluate, and correlate proprietary performance data and indicators, such as those maintained by INPO, to the results of NRC's assessments, and of any current arrangements between INPO and NRC regarding the sharing of, or access to, performance data.

**Response (Items 2 and 3):**

The Arthur Andersen study indicated that the plants discussed at the SMM constituted an appropriate grouping of poorly performing plants. Therefore, the staff has used the plants discussed at the SMM, rather than those subsequently placed on the watch list, as the benchmark in developing the performance trending methodology.

The staff understands that the industry and INPO believe that the plant-specific INPO Performance Indicator Index, which is derived from the 10 WANO Performance Indicators, is the best measure of overall plant performance. The staff reserves judgment on the ability of the INPO index to measure plant performance until it has had the opportunity to review the INPO index in detail. The staff believes that the trending methodology could be compared to this index as one means of validating the models.

The NRC has been receiving data from INPO for five of the WANO indicators. As described during a Commission briefing by AEOD staff on August 22, 1995, one of the WANO indicators, Collective Radiation Exposure, is used directly in the NRC Performance Indicator (PI) program. Three other WANO indicators have similarities with two of the NRC PIs. In addition, as part of the industry voluntary alternative to the proposed reliability and availability data rule, the staff is receiving the supporting data for the WANO Safety System Performance Indicator (SSPI).

The staff has requested from INPO the remaining WANO indicators along with the INPO Performance Index so that comparison and benchmarking could be done. Recently, the staff met with INPO staff and received a briefing on the WANO Performance Indicators and the INPO Performance Indicator Index. Subsequent to this meeting, the staff again requested access to this information. INPO responded in a February 2, 1998, letter indicating that INPO would agree under certain conditions to grant NRC access to all of the U.S. plant-specific WANO performance indicator data, but that they would not provide access to the INPO Performance Indicator Index.

Since access to the INPO Performance Indicator Index is not available, the staff could formulate a method for combining the WANO indicators into an overall plant value once all of the WANO Performance Indicator data are obtained. However, to determine the INPO Performance Indicator Index requires access to the weighting and grading factors used by INPO to convert the individual WANO indicators into the Performance Indicator Index. Thus, the staff's effort would not replicate the INPO Performance Indicator Index. Alternatively, the staff could compare the WANO indicators individually to the SMM discussion plants as was done for the individual indicators used in the staff's performance trending methodology models. Absent the required index information, putting the WANO indicators through a similar process of evaluation as was done for the current trending methodology would take an estimated calendar year to complete, involving approximately 1.0 FTE of staff effort and \$300K in contractor support after initial receipt of all of the data for the 10 WANO indicators. However, neither of these options would be sufficient to benchmark the performance trending methodology to a widely accepted definition of poor performance.

If the staff's continuing efforts to obtain the INPO Performance Indicator Index are successful, it will determine if there exists a correlation with SMM discussion plants to see if there were any significant differences between the two. If there were, the staff would benchmark the trending methodology variables to the INPO Performance Indicator Index to understand the differences. It is estimated that this effort, based on the staff's limited knowledge of the INPO Performance Indicator Index, would require approximately 0.3 FTE of direct staff effort, \$100K in contractor support, and 6 months to complete.

Besides the INPO Performance Indicator Index, the staff has been unable to identify any other broadly agreed upon definition of poor plant performance that has a quantitative basis for benchmarking the trending methodology. Therefore, the staff presently has only the SMM discussion plants as an overall plant performance benchmark to test the validity of the trending methodology.

**ITEM 4: INFORMATION REGARDING RESOURCE COMMITMENTS RELATED TO SMM PROCESS IMPROVEMENTS**

The Commission asked the staff to provide information regarding the resource commitments expended to date on the improvement of the current SMM process, an estimate of the anticipated resources needed to complete this effort, and an estimate of the resources needed to maintain and refine the SMM process once it becomes final and is in-place. The resource commitment information should include data from both Headquarters and the Regional Offices. The staff should continue to ensure intra- and inter-office involvement in the development of the SMM process. The degree of involvement by the Regional Offices should be provided to the Commission.

**Response:**

The resource commitments associated with the improvements to the SMM process are included in Table 1 ([Attachment 2](#)) and summarized below.



AEOD	<b>CONTRACTS</b>						
	Original study by Arthur Andersen (AA)	348K	---	---	---	---	---
	Statistical analysis by INEEL to support original AA study	52K	---	---	---	---	---
	Support by AA to the development of Plant Performance Template	130K	---	250K	---	---	---
	Support by AA to the development of trending methodology	266K	---	---	---	---	---
	Statistical analysis by INEEL to support developing trending methodology	86K	---	---	---	---	---
	Complete trending methodology, revise to reflect comments, and comparisons with WANO performance indicators	---	---	30K	---	---	---
	Contractor support to maintain and refine trending methodology	---	---	---	---	30K	---
	INEEL support to develop financial database and financial indicators	160K	---	50K	---	---	---
	Contractor support to maintain and refine financial database	---	---	---	---	30K	---
<b>AEOD STAFF SUPPORT</b>							
Technical support for developing main categories and subcategories associated with the Plant Performance Template and contract oversight	---	2.0	---	1.0	---	---	
• Management oversight	---	0.5	---	0.3	---	---	
Technical support for trending methodology development and contract oversight	---	1.2	---	0.8	---	---	
Technical support to provide trending information; refine the methodology; and contract oversight	---	---	---	---	---	0.8	
• Management oversight	---	0.2	---	0.2	---	0.2	
Technical support for development of financial database, financial indicators, and contract oversight	---	1.2	---	0.8	---	---	
Technical support to provide financial database; refine financial indicators; and contract oversight	---	---	---	---	---	0.8	
• Management oversight	---	0.2	---	0.2	---	0.2	
NOTE: The estimates provided above reflect efforts to provide information to the SMM process above the baseline efforts.							
NRR	Input to Plant Performance Template, attending workshops, developing guidance for regional offices	---	0.2	---	0.2	---	0.2
	• Management oversight	---	0.1	---	0.1	---	0.1
RES	<b>CONTRACTS</b>						
	INEEL work on management and organizational effectiveness	---	---	250K	---	---	---
	<b>RES STAFF SUPPORT</b>						
	Input to Plant Performance Template (Human Performance and Management Effectiveness categories), participation in workshops, developing guidance for evaluating risk significance of PIM entries, and contract oversight	---	0.3	---	0.3	---	---
• Management oversight	---	0.1	---	0.1	---	---	
Regional Offices	Input to the development of Plant Performance Template, participation in workshops, and review and comment on Template categories	---	0.2	---	0.2	---	---
	• Management oversight	---	0.1	---	0.1	---	--