27
5-31-81
240715
MASTER

ORNL-5744

OAK
RIDGE
NATIONAL
LABORATORY

UNION
CARBIDE

Contribution to Screening
Methodology

F. G. Perey

OPERATED BY
UNION CARBIDE CORPORATION
FOR THE UNITED STATES
DEPARTMENT OF ENERGY

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency Thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

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

Printed in the United States of America. Available from
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Road, Springfield, Virginia 22161
NTIS price codes—Printed Copy: A04; Microfiche A01

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

ORNL-5744
Special

Contract No. W-7405-eng-26

Engineering Physics Division

CONTRIBUTION TO SCREENING METHODOLOGY

F. G. Perey

Date Published - March 1981

DISCLAIMER

This book was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

This Work Sponsored by
Department of Energy
Energy Information Administration

OAK RIDGE NATIONAL LABORATORY
Oak Ridge, Tennessee 37830
operated by
UNION CARBIDE CORPORATION
for the
DEPARTMENT OF ENERGY

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

THIS PAGE
WAS INTENTIONALLY
LEFT BLANK

TABLE OF CONTENTS

	<u>Page</u>
Abstract	v
A. Introduction	1
B. Application of Durston's Estimator to LEAP Model 82 . . .	3
C. Analysis and Reformulation of the Problem	9
D. Solving the Reformulated Screening Problem	20
E. Summary and Conclusions	27
F. Acknowledgements	29
References	30
Appendix	31
Figures	37

THIS PAGE
WAS INTENTIONALLY
LEFT BLANK

ABSTRACT

The screening method for determining objective approximate values for the sensitivity coefficients of the output of a computer code to the input data is analyzed. An "estimator" proposed to solve such problems is used to investigate one of the outputs of a simplified model of the Long-Term Energy Analysis Program (LEAP) of the Energy Information Administration (EIA). The results are presented and compared to the true values of the sensitivity coefficients. The estimator produces results which appear more arbitrary than approximate. The screening problem is analyzed from the point of view of a theory of logical inference, and it is argued that no objective, approximate and useful solution exists for a code treated as a "black box" (that is to say that no assumptions, prior or otherwise, are made about the code, and we have only the ability to run the code) since no degree of confidence can be placed on the solution. Requirements on additional information needed to produce a solution which could qualify as objective, approximate and useful are investigated. It is shown that such information can in principle be obtained, and the problem then admits a solution. The emphasis in this report is on the conceptual aspects rather than on the mathematical details of the problem and its solution.

A. Introduction

This final report discusses the use of "screening methodology" as a tool for computer code evaluation. Determining the uncertainties in the outputs of a code and the origin of those uncertainties is one aspect of code evaluation. As indicated in the interim report [1], sensitivity analysis provides one of the elements useful in determining the uncertainties in the outputs of the code. Sensitivity analysis, among other things, determines the derivatives of the outputs of the code with respect to the input parameters: the sensitivity coefficients. With the sensitivity coefficients one can propagate to the outputs of the code the uncertainties in the input parameters. Linear sensitivity analysis is concerned only with the linear dependence of the outputs on the input parameters. This report discusses only linear sensitivity analysis.

If no convergence problems exist in determining the numerical values of the output of a code, given the input parameter values, the first-order derivatives can be obtained by finite difference methods. However, finite difference methods are impractical for codes that have long running times and many input parameters. Using this brute force method may be advantageous for codes that have many outputs of interest since many derivatives are obtained at each rerun of the code. However, the "adjoint method" can be used very effectively in situations where sensitivity coefficients are needed for only a few of the outputs. If this method is used, a second code must be written to solve the "adjoint equations" for the output of interest, and the sensitivity coefficients for every input parameter can be obtained running the "forward" and "adjoint" codes only once each. This method is being pursued for LEAP [1].

The two complementary approaches described above are aimed at determining the exact values of the sensitivity coefficients. A third approach, the screening method, has been advocated in the literature [2] (and references quoted therein). This method is claimed to determine, at much lower costs, only the approximate values, or approximate relative values, or to only approximately identify the large sensitivity coefficients. The screening method makes use of a "base calculation" with the input parameters set at their nominal value. More calculations are then made

by perturbing, in a known fashion, some or all of the input parameters from their value in the base calculation. If the input perturbations are small or if the outputs are linearly dependent on the input parameters, the observed differences in the outputs of the code from the base calculation outputs are a linear function of the known perturbations in the input values and the unknown sensitivity coefficients. If the number of recalculations, N , made is less than the number of input parameters, M , one cannot uniquely solve the set of equations obtained for the sensitivity coefficients because there are only N linear equations in M unknowns, and $N < M$. The screening problem addresses the possibility of whether a perturbation scheme can be devised for the M input parameters so that the results observed in the N reruns can be used to obtain approximate values for all sensitivity coefficients, or approximate relative values, or to only approximately identify the large sensitivity coefficients when $N < M$ and nothing else is known about the code. The reason for restricting the information used in the screening problem to the data obtained by running the code $N + 1$ times is to ensure obtaining an objective result; that is to say not use any code specific human judgement which may be fallible. When $N = 0$, we have no data that are relevant to the magnitude of the sensitivity coefficients; when $N = M$, and we choose our perturbations adequately, we have all the data needed to calculate all the sensitivity coefficients exactly. When $0 < N < M$, we are in an "intermediate situation"; we have some data relevant to the magnitude of the sensitivity coefficients, but not enough to calculate any of the coefficients exactly, unless we choose a perturbation scheme in which the same N parameters are perturbed in each run. The screening problem is based on the assumption that in such an intermediate situation we may conceivably have enough data to obtain approximately the values of the sensitivity coefficients or their relative values or to identify the input data most likely to have large sensitivity coefficients. This idea is not recent and is related to the field of experimental design [2]. The various proposed schemes for solving the screening problem, which have been applied to complex computer codes, are based on statistical estimation procedures in which each rerun is considered a random sample from some distribution. From the attention given to this problem in the

literature many people obviously believe that the screening problem, as defined above, is meaningful and their only question is: What is the correct estimator which corresponds to a particular sampling distribution?

When we began our work for LEAP [1], the above arguments appeared plausible, and if they were true, the "screening methodology" appeared to be potentially useful (at least until adjoint methods were developed to get the exact sensitivity coefficients) and to require only the knowledge of how to run the code. The first phase of the work at ORNL for the LEAP evaluation project used an "estimator" proposed by Durston *et al.* [2] (referred to hereafter as Durston for ease of reference), which was based on experimental design methods and which they had applied to a large nuclear safety code. Although in the case of Durston this estimator appears to have been partially successful, this estimator was largely a failure when applied to a simplified model of LEAP, Model 82 [1]. The next section describes our application of this estimator to LEAP Model 82 and shows its failure. We next analyze, from a semantic point of view, the screening problem as defined in this section and argue that it is an "ill-posed" mathematical problem. We extend deductive logic in the manner proposed by Jaynes [3] and, even with the notion of probability being a measure of the logical degree of truth conditional upon the given data, the problem as posed does not admit a useful solution. Our analysis suggests that the screening problem could be reformulated so that it admits an objective, useful solution. In the last section we discuss the conceptual problems associated with solving the reformulated screening problem.

B. Application of Durston's Estimator to LEAP Model 82

Durston's report [2] should be consulted for full details of the different rationales that led the authors to propose their estimator. We will present, in some detail, only one of their arguments, but first we introduce the notation used in this report.

For simplicity of notation, we consider the code under study to produce a single numerical result denoted by r . Let the M input data be denoted x_j , where $j = 1, 2, \dots, M$; we shall refer to these data as parameters. For the base calculation, the values of the parameters are

x_j^0 , and the output is r^0 . We call x_j^i the value of the x_j parameter at the i^{th} recalculation that produces the result r^i . The term a_{ij} denotes the perturbation made at the i^{th} recalculation in the x_j parameter value:

$$a_{ij} \equiv x_j^i - x_j^0 . \quad (1)$$

If we make N recalculations, perturbing the M parameters at each rerun, the quantities a_{ij} can be considered the elements of an $N \times M$ matrix A . The term $r_i^!$ denotes the differences observed at the i^{th} recalculation from the base calculation result r^0 :

$$r_i^! = r^i - r^0 . \quad (2)$$

The N quantities $r_i^!$ are considered the elements of a vector $R^!$. Given the parameters x_j^i , we assume the results $r_i^!$ to be sufficiently precise and accurate. If we denote by s_j the sensitivity coefficients of the result r with respect to the parameters x_j and view them as components of the vector S , the base case calculation and the N recalculations yield a system of N linear equations in the M unknown constants s_j , which we write:

$$R^! = AS . \quad (3)$$

Sometimes it is more convenient to work with the relative sensitivity coefficients defined as the logarithmic partial derivatives. Equation (3) still holds in this case, but the $r_i^!$'s are then fractional changes in the output, and the a_{ij} 's are the fractional changes in the parameters x_j at the i^{th} recalculation. In the rest of this section, we use relative sensitivity coefficients because Durston did so. We recall that all of the perturbations are small and, therefore, deal strictly with the linear behavior of r about r^0 .

The screening problem, which can easily be formulated in terms of Eq. (3), presumes that if we select an $N \times M$ matrix A in some appropriate fashion and observe the N values that make up the vector $R^!$, we can obtain an approximate ranking of the M sensitivity coefficients s_j that are unknown, $N < M$, and no information other than Eq. (3) is used.

Durston notes that we can formally write the general solution of Eq. (3) as

$$S = A^{\dagger}(AA^{\dagger})^{-1} R' + V, \quad (4)$$

where $A^{\dagger}$ denotes the transpose of A . The first term of Eq. (4) is a particular solution, S_p ,

$$S_p = A^{\dagger}(AA^{\dagger})^{-1} R', \quad (5)$$

which can in principle always be calculated if A and R' are given. The vector V is only known to satisfy the relation,

$$AV = 0, \quad (6)$$

as can be verified by substituting Eq. (4) into Eq. (3).

There are many particular solutions to Eq. (3), but if we view Eq. (6) as defining an M - N manifold in the M space of the s_j 's, assuming the rank of A to be N , then S_p is the only particular solution that has no components in the M - N manifold (i.e., S_p is orthogonal to V). In other words, Eq. (3) defines a subspace in which the end point of the vector S lies.

Essentially, Durston argues that, if A is selected at random (i.e., in a way unrelated to the code), the vector V does not depend on the particular code being studied. However, because R' is a function of the code and V is independent of the code, the only information obtainable about the sensitivity coefficients is contained in S_p ; therefore, Durston states that using S_p as an estimate of S is reasonable. Durston now states that the estimate of S (S_p) will be unbiased if the elements of A are chosen randomly and independently from a distribution having zero mean and unit variance. He suggests that, using the concept of a "two-level design" from the theory of experiment design, one can select with equal probability the a_{ij} 's from a distribution consisting of values ± 1 . Other distributions could be used without affecting their derivation.

Durston notes that S_p satisfies the necessary requirement that it produces the correct result, if A is of rank N when $N = M$ because the matrix A is square. It also gives a reasonable result if R' is zero (i.e., all the components r'_i are zero); that is, $S_p = 0$. Durston also argues that, if the elements a_{ij} are taken randomly and independently from a distribution having zero mean and unit variance, then the r'_i 's, by virtue of the central limit theorem, will be distributed approximately

as a Gaussian distribution if M is large. Consequently, because one can consider the N recalculations as a sample of r' of size N , an unbiased estimate of the sensitivity coefficient is:

$$\hat{S} = \frac{M}{N} S_p \quad (7)$$

We used Durston's estimator $\hat{S}$ to perform screening for the LEAP Model 82. We considered only 80 parameters in the input as variables. We wanted to obtain the sensitivity coefficients of these variables for one output of the code. The elements of the matrix A were chosen according to Durston's suggestion (i.e., taken randomly and independently ± 1). We verified that each row of the $N \times M$ matrix A was orthogonal to the other thereby ensuring that the rank of the A matrix was N . Therefore, no rerun results could be obtained as a linear combination of other reruns, and we obtained the maximum amount of information from N reruns.

In the Appendix, we provide the A matrix, where $N = 20$, the values of r'_i given by the code, and the values of $\hat{S}_i$ obtained using Durston's estimator, Eq. (7). To verify the usefulness of the method, we also calculated, by finite differences, all 80 sensitivity coefficients; these are also given in the Appendix.

Figure 1 plots the true ranking of the sensitivity coefficients s_j against the approximate rankings based on the $\hat{S}_j$'s. The ranking was based on the absolute value of the sensitivity coefficients in both cases, and we show, by an open circle, when the estimated sensitivity coefficient has the same sign as the true value and, by a triangle, when it has an opposite sign.

The symbols on Fig. 1 appear to be scattered over the whole area. This is surprising because the true value of the sensitivity coefficients varies by about three orders of magnitude, and 25% of all the runs needed to calculate the sensitivity coefficients exactly were made in the screening. The scatter of points on Fig. 1 indicates that, in the case of LEAP Model 82, the estimator, Eq. (7), is not useful. We would have considered the results shown in Fig. 1 acceptable if all the sensitivity coefficients had been of approximately the same magnitude since then the notion of approximate ranking may not be too meaningful.

We believe that for an estimator to be useful in a code that has large differences in the magnitude of the sensitivity coefficients, as was the case here, the results of screening, when plotted on a figure such as Fig. 1, should give points in the upper left corner with few if any points in the lower left and upper right corners of the figure. However, this interpretation may be naive and needs to be investigated more closely.

We found our results surprising at first because Durston's report led us to expect a greater degree of success from this method. However, closer study of Durston's results for the code DEMOD4 showed that they do not necessarily contradict ours. In Durston's study, the code had 661 input parameters, and the screening included 63 reruns. Table I of Durston's report compares the 21 highest-ranked sensitivities from their estimator with their true values. Only 8 of the 21 highest-ranked parameters were actually found to have large sensitivity coefficients; all the others had negligible sensitivity coefficients. Except for an essentially random sample of 15 sensitivity coefficients from the 661, there is no indication that the 8 large sensitivity coefficients are larger than some other coefficients that were ranked lower by the estimator. In the text, the authors appear impressed by the "efficiency" of their estimator in detecting large sensitivity coefficients; with slightly less than 10% of the total number of reruns required to calculate all the sensitivity coefficients, five of the six highest-ranked coefficients are found to be "large" after exact calculation. This was also found to be true for 7 of the top 13.

Our results appear far less impressive. We do not think it is possible to reach firm conclusions only on the basis of our results. Figure 2 is a plot similar to Fig. 1, except that it was obtained using only the first 10 reruns used for Fig. 1 (i.e., 12.5% of all the reruns needed to calculate all the sensitivity coefficients). On the basis of the percentage of runs made, the results shown on Fig. 2 are more comparable to Durston's results. Figures 1 and 2 are not independent results because all the data used to generate Fig. 2 were also used to generate Fig. 1.

In an actual case, Figs. 1 and 2 cannot be drawn because we do not know the true value of all the sensitivity coefficients; therefore, we must emphasize that Figs. 1 and 2, although quite instructive, are somewhat misleading about the usefulness or lack of usefulness of the method.

In our opinion, the greatest deficiency of Durston's estimator is not that it does not "perform better" than indicated by our results, but that there is no way of knowing how well it has performed in a specific case. No claim was made that the estimator would produce an exact ranking, only that it would produce an approximate one. Until we define exactly what is meant by the term "approximate," we must claim that the estimator performed as claimed because it produced a ranking that we could not expect to be exact. In this sense, a random ranking selected from any distribution could also be called an approximate ranking; we would perceive it to be unbiased, but totally useless. The point at issue is not that we should compare the results obtained with the estimator, based on N reruns of the code, with a purely random ranking, where no runs have been made. If we had performed N reruns by perturbing only one parameter at a time, we would know exactly N sensitivity coefficients, but would know nothing about the other $M-N$ sensitivity coefficients. For our labor, we would have obtained the value of some sensitivity coefficients. However, if we perform the N reruns in the manner required by the estimator and apply it, we do have an approximate result for all the M sensitivity coefficients, but what can we do with it? Can we act rationally on the basis of the approximate ranking? Do we have any means of estimating the risk involved for making a decision based on the results? We use "risk" here in the sense of decision theory. We claim that the risk cannot be estimated because we have no idea of the degree of confidence we should place on the results obtained, no matter how we choose to define degree of confidence.

If screening is viewed as a method for obtaining guidance in making an exact calculation of sensitivity coefficients that may be large, in our opinion, Durston's work should not be judged in the way the authors concluded. They performed 84 reruns of the code (63 for the screening and 21 to calculate the 21 top-ranked coefficients exactly) to identify 8 "large" sensitivity coefficients. If they had exactly calculated 84

sensitivity coefficients selected at random to obtain unbiased results, would they have performed any better? This question cannot be answered with any stated degree of confidence unless we know the distribution of the sensitivity coefficient values.

C. Analysis and Reformulation of the Problem

In this section we analyze the logical structure of the screening problem. We argue that proposed solutions to the problem fail to meet the objective for a fundamental reason. We then adopt an extension of deductive logic and prove that the problem as posed does not admit a useful solution. We then reformulate the problem which will yield an objective and useful solution under some circumstances.

Since given a set of N linear equations in M unknown sensitivity coefficients cannot yield the exact values of the sensitivity coefficients, the screening problem was formulated to demand approximate values of the sensitivity coefficients or their approximate ranking. We claim that obtaining these approximate results is not useful if we have no idea of how approximate they are. The question raised is a semantic one. What is meant by an approximate value? The American Heritage Dictionary of the English Language gives as a definition of "approximate": "1. Almost exact, correct or perfect. 2. Very similar; closely resembling. 3. Close together; near." Therefore saying that we have an approximate value means that this approximate value bears some correspondence to the true value and that it is not arbitrary in the sense of "1. Determined by whim, or caprice. 2. Based on or subject to individual judgement or discession." The discussion in the previous section resulted because of our comparison of the proposed solution with the truth. On a fully subjective basis, we perceived that the adjective approximate did not properly describe the achieved results despite the rationale used to obtain them. Some coefficients had been correctly ranked higher than others, but the opposite was also true. Although we expect this to occur in an approximate result, in view of the fact that there were differences of several orders of magnitude between the coefficients, the adjective arbitrary could also have been used to describe the results. Of course this is a matter of degree because very approximate is, to us, almost

arbitrary. As we pointed out, in practice we cannot calculate all the exact sensitivity coefficients for a specific code. If we invest time and effort to perform screening on a given code, we expect some benefit from it to qualify the exercise as useful. The benefit may not be large and may not be estimable before the screening, but for the results to be useful, the benefit must be estimable after the screening and be finite. In short, the adjective approximate has no operational meaning until some measure is provided of how approximate the results are. This measure of how approximate the results are must also have some operational meaning with respect to the true values.

Unless experience has shown that a particular algorithm provides for a class of codes results which are "close" to the truth, and we can prove that the specific code we are using belongs to that class, we must provide an objective measure of how approximate our results are when we propose an approximate solution. We are not aware that with respect to screening there are claims that we have reached this level of "empirical knowledge." We must conclude that currently proposed solutions to this problem do not yet meet the goal they were supposed to meet. This goal was to provide an objective and useful sensitivity vector for the code under investigation that would not necessarily be the true sensitivity vector, only on the basis of N reruns of the code, with M input parameters and $N < M$. The failure of these algorithms to meet the goal lies in the fact that they do not provide a measure of how approximate the approximate results are.

It is convenient to separate the two different problems that we must solve in order to obtain the final result. We must first choose an $N \times M$ matrix, A , because it is not given to us. Having chosen a specific A matrix, and having observed the fact that it produces R' when applied to the code, we must then choose a specific vector S_c among all the vectors S which satisfy the equation $R' = AS$. The current proposed solutions to the problem call S_c the approximate solution. Since we do not want to waste any running time, and the code has a definite sensitivity vector, this places the requirement on the matrix A to be of rank N . The A matrix may have to fulfill other requirements, but eventually a specific A matrix is chosen and only one. Assuming that we have chosen such a

matrix A to obtain the final result, we will have to choose a vector S_c given the equation $R' = AS$. If we select a particular S_c , it can only be called approximate and useful if we do the selection according to a logic which indicates why we choose this vector over all other possible vectors and provides a measure of how approximate this S_c is. This measure must be defined, a rule provided for assigning it, and we must prove that, according to the definition of measure we chose and its associated rule, the value of S_c is preferable over any other vector S which satisfies the observed data $R' = AS$.

We conjecture that the logical weakness in Durston's argument, which we have presented in some detail in the previous section, is the following one. However the matrix A is chosen, on the basis of only the data $R' = AS$, we can only conclude that the vector S for the code under study must satisfy the relation

$$S^T S \geq R'^T (A A^T)^{-1} R' \quad (8)$$

This deduction would be useful if we were interested in setting a lower limit to the norm of the true sensitivity vector. But it is not useful if we are interested in the absolute or relative magnitudes of the individual components of S . Therefore, when we select among all possible vectors S the only one which fulfills the equality sign in relation (8), we should not be surprised if we find little or no correspondence between its components and the true values. These components may or may not be close to the true values for the code under study. Since the matrix A is a random matrix, we will in general obtain a random pattern when we compare the result of their choice of S_c to the true value, as we observed in the previous section for LEAP Model 82. There is one exception to this rule, as we will show later.

Durston provides another derivation for his estimator, which we will not describe in detail, based on a least-squares procedure. We think that the weakness of this derivation is the assertion that prior to acquiring the data $R' = AS$ it is rational to assume all s_i 's to be zero since we presume them to be positive or negative, with equal plausibility. We do not understand the logical basis for saying that when we know nothing about a number which could be positive or negative the value zero is less arbitrary than any other value.

Some solutions to the screening problem are based upon regression methods (see, for instance, Vaurio [4]) and appear to provide confidence limits for the approximate solutions they propose. This is misleading, however, because the confidence limits provided are valid only on the assumption that the results of the code are dominated by a small group of parameters, less than N in number. They provide no basis for testing this assumption. In any event, their solution requires additional information about the code, the validity of their assumption, which goes beyond the observed data $R' = AS$.

The results presented so far, and our discussion of them, lead us to infer that the various points of view used in obtaining proposed approximate solutions to the screening problem suffer at least from ambiguity when applied to this problem. As we have argued, the difficulty is largely a semantic problem and is related to the meaning attached to key words used in defining the problem, specifically the adjectives "objective" and "approximate." Our criticism is valid only to those who agree with our definition of these words. In the case of "objective" we mean that an objective result is one based only on demonstrable or observable facts and is, therefore, uninfluenced by emotion, surmise, or personal prejudice. We define a demonstrable fact as one derivable from observable facts using only the laws of deductive logic. The solution to the screening problem cannot be derived, from the demonstrable or observable facts, using only the laws of deductive logic, since when $N < M$ the problem is said to be an underdetermined, or ill-posed, problem of deductive logic.

The use of the words "approximate" or "likely" to describe the solution expected to the problem implies the use of an extension of deductive logic to arrive at it. There is no general agreement, today, on how the laws of deductive logic should be extended to perform inductive reasoning and "solve" ill-posed problems of deductive logic, such as the one we are considering. Proposed solutions to the screening problem that we have discussed so far have been based on "statistical inference," as an extension of deductive logic, and appear, in this case, to lead to ambiguities, as we have stated earlier. We do not express any opinions as to the validity of statistical inference to treat problems of this

type, nor if the laws of statistical inference have been properly used by those who propose solutions to this problem using it.

We will now adopt a different method of extending the laws of deductive logic to perform inductive reasoning. We will essentially adopt the point of view of Jaynes [3] where probability, taken to be a scalar, is a measure of logical degree of truth conditional upon the observable and demonstrable facts given in the statement of the problem. This measure of logical degree of truth is assigned on the basis of a rule, called by Jaynes a desideratum of consistency, which we will phrase later. Jaynes' theory is a Bayesian theory of probability and we assume that the reader is familiar with the concept of Bayesian probabilities. There are many Bayesian theories of probabilities; they differ both in how probability is defined and in how it is assigned. We find Jaynes' theory objective in the sense that probabilities are based only upon demonstrable and observable facts, not opinions, and the desideratum of consistency requires for its application only the use of deductive logic and appears to be a direct consequence of the definition of probability. Of course, in Jaynes' theory probabilities are not directly observable, since they are degrees of logical relationship between assertions, and in the sense that they do not have physical existence they are subjective. However, to have some usefulness they must lead to observable quantities. We do not assume the reader is familiar with Jaynes' theory.

We shall call observable and demonstrable facts "testable facts," following Jaynes. Assuming that the concept of probability as a measure of logical degree of truth conditional upon testable facts is understood, we phrase the desideratum of consistency: "In two different problems, given the same testable facts, we should assign the same probability." This rule appears to follow directly from the definition of probability used since it is a measure of what the testable facts tell us logically about what we are interested in, which we know must be either true or false. We must emphasize that this theory of logical inference, because of its definition of probability and the rule used for assigning it, is not a theory where the conclusion is based on some assumptions regarding testable facts which are not available to us. It is rather a theory

where the conclusion is based only upon the testable facts given in the statement of the problem.

We shall now prove, using Jaynes' theory of logical inference, that the screening problem as posed, where the only testable fact is the observable fact that $R' = AS$, regardless of how A has been selected, has no useful solution in the sense that there exists no rational choice of a specific value we have called S_c .

As was implicit in our discussions, we are considering an abstract M -dimensional space where the coordinates of a point represent the components of a sensitivity vector. In this space, when we refer to a point, we are referring to a specific sensitivity vector. On the basis of deductive logic alone, given $R' = AS$, we know that the sensitivity vector of the code, and there is only one, is a point in the M - N dimensional subspace of the M dimensional space. When we consider a point of that subspace, we therefore know that it is either the sensitivity vector of the code or it is not the sensitivity vector of the code. This is as far as deductive logic allows us to go. Introducing the notion of probability we assign a number to represent the degree of truth to the statement: "The sensitivity vector is in a volume element dS about a point S ," where S is a point in the M - N dimensional subspace. When S is outside the M - N dimensional subspace, we know with certainty that the above assertion is false. Presumably the degree of truth of the assertion, when S is a point inside the M - N dimensional subspace, is a function of where the point is in that subspace, and we write

$$f[S|R'(A)] dS, \quad (9)$$

to represent how this probability, a number, varies as a function of S . Our notation is such that we indicate by $|R'(A)$ upon what testable facts this probability is conditional. Our first task is to find what this function $f[S|R'(A)]$ is. There is one and only one rule in the theory for assigning the numerical value of (9); it is the desideratum of consistency. Its use requires another and different problem where we would have the same testable facts, and we must assign the same probability in the two problems. Our testable fact is that $R' = AS$ and our problem is to find what is the sensitivity vector S . To clearly indicate

that we are dealing with a different problem, let us use the symbol S' to indicate the sensitivity vector of another and different code. The only two requirements we place on that second code are 1) that it also has M input parameters and at least one output quantity, for which S' represents the sensitivity of this output with respect to its M input parameters, and 2) that when we use the same matrix A it produces the same vector R' . We have for these two different codes the same testable facts: $R' = AS$ in the original code and $R' = AS'$ in the second code. Using deductive logic only we conclude that S and S' must be related as follows:

$$S' = S + \Lambda \quad (10)$$

where Λ is an M dimensional vector which satisfies the relation $A\Lambda = 0$. We actually do not need to physically have a code which satisfies the above requirements; the conclusion we have reached is based upon deductive logic and on the existence of our original code, an observable fact, and the other observable fact $R' = AS$. This second different code whose existence we have postulated above we have proven could exist. Our second problem then is what is the sensitivity vector S' , given that, when we apply the same matrix A to this different code, we observe the same vector R' .

The desideratum of consistency states that given the same testable fact in this second problem we must assign the same probability in this second problem as we did in the first one. We therefore write

$$f[S'|R'(A)] dS' \quad (11)$$

as the probability for the truth of the assertion "the sensitivity vector is in a volume element dS' about the point S' " where S' is a point in the M - N dimensional subspace, of an abstract M dimensional space representing the vector S' , defined by $R' = AS'$. It should be emphasized that we have now two different M dimensional spaces, one we can call the S space and the other the S' space. The desideratum of consistency merely states that given the same observable fact $R'(A)$ in the two different spaces we should assign in the S' space the same probability as we do in the S space. This is so because probability is defined as a measure of what $R'(A)$ tells us. The relations (9) and (11) represent numbers and what

the desideratum of consistency says is that these numbers in the two different problems are the same and we write:

$$f[S|R'(A)] dS = f[S'|R'(A)] dS' \quad (12)$$

We have emphasized the fact that (9) and (11) are the same statements made in two different M dimensional spaces. We know from deductive logic that Eq. (10) defines a transformation from the S-space to the S'-space. What we precisely mean by this is the following. We can visualize both the S-space and the S'-space as being embedded in yet a third M-dimensional space which we will call the T-space. What Eq. (10) tells us is that we can map the S-space into the S'-space by means of a translation in the T-space by the vector Λ . In the T-space the M-N dimensional subspace defined by $R' = AS$ in the S-space is the same M-N dimensional subspace as the M-N dimensional subspace defined by $R' = AS'$ in the S'-space. This is so because the translation vector Λ has the property that $\Lambda\Lambda = 0$, as we have deduced before. This means that Λ is, in the T-space, any vector which lies in the M-N dimensional subspace defined by $R' = AS$ and/or $R' = AS'$ since we have only one such M-N dimensional subspace.

What the equality (12) says is that we assign to every element of volume dT , in the M-N subspace, a number which is a measure of degree of truth of the statement that the sensitivity vector is in that volume element. That number can be expressed by the same functional form in the S-space and the S'-space. Since we know the transformation law between the S-space and the S'-space, Eq. (10), we can take this number expressed in the S'-space, expression (11), and formally transform it into the S-space representation. Since the Jacobian of the transformation Eq. (10) is unity, we have:

$$f[S'|R'(A)] dS' = f[S + \Lambda|R'(A)] dS \quad (13)$$

Using Eq. (12), which we had obtained from the desideratum of consistency, we get:

$$f[S + \Lambda|R'(A)] = f[S|R'(A)] \quad (14)$$

The equality (14) must be true for any Λ which lies in the M-N dimensional subspace defined by $R'(A)$, and we have therefore proven, if we accept the definition of probability given above and its associated

desideratum of consistency, that $f[S|R'(A)]$ is a constant over the whole subspace defined by $R' = AS$.

We must emphasize what it is that we have just completed proving. On the basis of deductive logic alone we encoded the runs of the code which were performed into a mathematical expression $R' = AS$, which must be satisfied by the sensitivity vector of the code. This mathematical expression is said to define an M - N dimensional subspace because $R' = AS$ for every point of it, and we can only conclude that any point in this subspace could be the true sensitivity vector. The only thing that our extension of deductive logic has bought us is the ability to say that every point in that subspace is equally likely to be the true sensitivity vector for the code. This conclusion is hardly surprising to us if we only focus on the mathematical expression and look at it as an abstraction and the method of proof a type of well known "invariance argument." We should note that, although our theory is a Bayesian theory, we have in this proof completely bypassed the use of prior probabilities and Bayes theorem. We claim the result to be as objective as the fact upon which it rests, $R' = AS$, and therefore should be noncontroversial to anyone who accepts the notion of probability as a conditional measure of logical degree of truth.

There is but one difficulty with the above result and it deals with its interpretation, or decoding of the mathematical result, in terms of the "physical problem" we originally started with. We had assumed that we dealt with a code which has a physical existence and quite explicitly stated that it was well behaved in the sense that we had no convergence problem and the quantities r_j were sufficiently accurate. This implies that none of the sensitivity coefficients s_j are infinitely large, where we use the word "infinite" in the sense of a mathematical abstraction. Assuming that we deal with a well behaved computer code, the true sensitivity vector can only be finite. This is a piece of information that we have neglected until now. It should be clear, however, that the invariance argument still holds if we bound our M dimensional S -space, and S' -space, by any "surface." Our density function $f(S|R'(A))$ will be constant within this "surface" inside the M - N dimensional subspace defined by $R' = AS$. If we do have a bounded M - N dimensional subspace, then we

can make a choice of a particular value of S inside as soon as we specify the "loss function." We assume the reader familiar with decision theory. If we take the conventional quadratic loss function, then the value S_c that we should choose as the true value for the sensitivity vector of the code will be the first moment (center of gravity) of the volume and the expected loss will be given by the second moment (moment of inertia) of the volume since the density function is uniform inside the volume.

It is now clear that the conventional definition of the "screening problem" leaves out a critical element of this problem. There is some information, the fact that the true sensitivity vector is finite, that we have failed to specify in any quantitative manner. This failure prevents us from determining not only a rational choice for the sensitivity vector knowing that it satisfies $R' = AS$, but also of determining the level of confidence in that choice.

Whatever point of view one chooses to try to solve the "screening problem" as it has been defined, it does not admit any useful solution, even in our extension of deductive logic, until we specify additional information. We personally prefer to view the problem as an ill-posed one of inductive logic because it suggests what kind of information is required to transform it into a well-posed problem of logical inference where, given a specific loss function, we can make an objective choice for the sensitivity vector and indicate a specific degree of confidence in our choice. Of course this solution can only be called objective if this additional information required to solve the new problem can be called objective. Our major objection to the usual approach of treating the screening problem as a statistical one is that all proposed solutions following this approach are based upon some specific assumptions. In the case of the Durston *et al.* estimator, as we have shown in the interim report [1], it is equivalent to assuming that the sensitivity coefficients have a normal distribution with zero mean. In the case of Vaurio [4], his estimator is based on the assumption that there are fewer than N "large" sensitivity coefficients. The price which must be paid when we make an assumption which is specific to the code under study is that one cannot obtain any degree of confidence upon the solution since there is no way to specify a degree of confidence upon an assumption. The results obtained

based upon an assumption are only valid if the assumption made actually holds for the code under study.

Our analysis of the screening problem in this section is consistent with the stated objective we find in the literature. The aim is to assign an objective value to the magnitude of the sensitivity coefficients of a code given that we can only rerun the code N times when there are M input parameters and $N < M$. We are supposed to treat the code as a "black box" in the sense of using only information obtained by rerunning it N times. Our main argument has been that our conclusion can only be objective if regardless of the rationale used for selecting the input matrix A , having chosen one such matrix A and used it to perturb the input of two different codes, if we obtain the same vector R' , we should come to the same conclusion regarding the magnitude of the sensitivity coefficients of the two codes. To us this seems the only logical thing to do — the only consistent thing to do. The fact that we could just as well have selected another matrix A is totally immaterial to the conclusion based upon a single A matrix. We are not interested at all in the distributions of vectors R'_i which might or might not be obtained from different matrices A_i selected according to some common technique, since this is not what we are asked. The specific vector R' that we have to base our conclusion upon is only a function of the specific matrix A chosen and not a function of the rationale which made us select this matrix A . We have shown that if we adopt such a point of view we cannot favor any particular point of the subspace defined by the only data available, $R' = AS$. In our terminology every point of the subspace is equally probable to be the true value. If we ignore the fact that no sensitivity coefficient can be infinite, a contradiction from the fact that we assume the code to be well behaved, then there is no way to choose any point of the subspace as the true value, whatever loss function we choose. The expected loss will be infinite because our subspace is infinite. In our opinion this is not a legitimate attitude to take if we are told that the code is well behaved. We could take the point of view that the subspace defined by $R' = AS$ gives us the likelihood function for the true vector S , and the difficulty is that it is uniform over an infinite subspace. It is then clear to us what the difficulty is with the conventional statement of the screening problem.

We are accepting the fact that the code is well behaved but cannot come to a rational conclusion until we state explicitly what this means. To say that the code is well behaved implies that the subspace defined by $R' = AS$ is not infinite. We will not be able to choose a vector S in the subspace until we provide some definite boundaries for it and specify a loss function as well.

In the next section we therefore remove from the statement of the conventional screening problem the fact that the solution shall only be based upon the data $R' = AS$ and allow us to have access to information which will enable us to objectively determine some boundaries in the space S where the true S sensitivity vector for the specific code under study lies.

D. Solving the Reformulated Screening Problem

In this section we analyze some aspects of solving what we will call the reformulated screening problem which includes not only consideration of "observable" facts but also what we will call "demonstrable" facts which could be obtained for the specific code under study. We reserve the label observable facts for those which can be obtained directly from rerunning the code and are generated only from the ability to run the code treated as a "black box." Our previous definition of the screening problem required the use of only observable facts. We have used the notation $R'(A)$ for those observable facts to indicate that R' is observed when we use the matrix A to perturb the input parameters. The observable facts $R'(A)$ are related to the unknown sensitivity vector of the code by the relation $R' = AS$. From our analysis of the previous section we conclude that if the rank of A is N and less than M , the number of input parameters, no objective useful approximate solution can be found for a code treated as a black box. Only the exact sensitivity vector can be found if we raise the rank of the matrix A to be equal to M . We now assume that for a particular code the cost of raising the matrix A to equal M is considered for some reason prohibitive but we can afford to run the code with a matrix A of rank $N < M$. We have indicated in the previous section that if we could objectively bound the $M-N$ dimensional subspace defined by $R' = AS$, if we in addition specified a "loss function,"

we could generate a solution which we could call useful approximate and objective. This information which would allow us to bound the M-N dimensional subspace cannot be based upon observations of the output of the code but would be just as objective as observations if it was obtained from deductions based upon the known equations programmed in the code or from the considerations used to derive the equations programmed in the code. We shall call such types of facts that could be used and may be available "demonstrable" facts. What we shall do in this section is concentrate on some of the conceptual aspects, as opposed to the mathematical details, involved in finding the solution to this reformulated screening problem. Two major topics are discussed at some length: the notion of demonstrable facts and the choice of observable facts. We indicate how the demonstrable facts and the goal we seek, the "loss function," determine which observable facts we should obtain. That is to say, the demonstrable facts and the loss function chosen determine uniquely which A matrix should be chosen to obtain the observable facts: $R' = AS$. What we are in effect doing in this section is looking at the problem from the opposite direction we looked at it in the previous section. There we assumed that we were given a set of observable facts, $R'(A)$, and found that no useful solution could be found unless we had what we now call some specific demonstrable facts. What we now do is consider that, because they cannot be obtained in as mechanical a fashion as the observable facts, we should first look at the demonstrable facts and then possibly determine if the observable facts should not be a function of whatever demonstrable facts we have.

The concept of demonstrable facts, which is a central element of our reformulated screening problem, needs to be clearly understood. Given a particular code under study, a demonstrable fact is a mathematical relation, an equality or inequality, involving one or more sensitivity coefficients, which we can derive from the known equations programmed in the code or more likely from the considerations upon which these equations were obtained. This relation is evaluated at the x_j^0 's and therefore consists of numbers and the unknown values of the sensitivity coefficients. The only criteria which must be met by this relation is that it can be certified to be true, or hold, for the true value of the sensitivity

coefficients. This relation need not be linear in the sensitivity coefficients; it is only "simpler" in that case. It cannot be a probability statement since such statements cannot be called true in the deductive logic sense. For this reason also an approximate value of a sensitivity coefficient is not a demonstrable fact, nor is an opinion a demonstrable fact. The "ultimate" demonstrable fact is a proof that a sensitivity coefficient is precisely equal to a number. If we can generate enough such ultimate facts, our screening problem disappears. (It is because of our limited mathematical skills that we have a screening problem.) We presume that it may always be easier to generate inequalities involving the sensitivity coefficients than equalities, given the equations programmed in the code. The generation of demonstrable facts is obviously a matter of time, mathematical skills, inventiveness and also possibly a matter of experience. We do not believe, since we currently lack experience at performing this task, that we can safely emit an opinion, at this time, concerning the magnitude of the undertaking in general nor for a given code such as LEAP. We grant that often it will not be trivial. However, in view of the alternative of having to calculate exactly all of the sensitivity coefficients, by difference methods or writing a full adjoint code, we believe that the potential usefulness of solving the reformulated screening problem should not be ignored in some instances.

We perceive that the major obstacle to solving the reformulated screening problem is the generation of demonstrable facts. As we have already pointed out, they cannot be obtained in a mechanical fashion as are the observable facts. In many instances we think that an upper and/or lower bound on some sensitivity coefficients or expressions involving them could be established by means of approximations. We must emphasize that when we approximate the equations we must still be able to prove that the true values of the sensitivity coefficients satisfy our inequalities. Until experience is gained with solving the reformulated screening problem, we do not think that too much emphasis should be placed upon achieving tight limits on sensitivity coefficients, assuming we are seeking these. Although our final answers must be a function of some of these limits, we should not lose sight of the fact that the final

results will also depend upon the observable facts, and these may in some instances eliminate the dependence of the results upon some of the demonstrable facts, making these redundant facts.

Of course for a given code the demonstrable facts are not unique any more than were the observable facts in the original problem since the A matrix was not given to us. It is largely because the demonstrable facts are not unique that the problem of finding them for a given code may be easier than getting the true value of the sensitivity coefficients, the "ultimate" demonstrable facts which are unique. Since there are no unique sets of demonstrable facts for a given code, it does not make any sense to say that the reformulated screening problem admits a unique solution.

We have not placed any specific requirements on the demonstrable facts. It is clear from the previous section what role they play when combined with the observable facts in determining the solution to the problem: they serve to bound the $M-N$ dimensional subspace defined by the observable facts, $R' = AS$. For a solution to exist, the demonstrable facts do not need to define by themselves a bounded region in the M dimensional manifold in which the true value of S must lie. The invariance argument which we have used in the previous section, to prove that conditional only upon the observable facts every point in the subspace they define has equal probability for being the true value of S , can be carried out in an identical fashion to arrive at the same conclusion using only the demonstrable facts. So long as all the facts we have, whether demonstrable and/or observable, merely constrain the true value of S to be in a region of the M dimensional space, and these are the only facts we have, then conditional upon them we must assign equal probability to every point in the region for being the true value. With our definition of probability and the desideratum of consistency, this conclusion follows necessarily from the fact that the M components of S , the s_i 's, are defined as the solution of a linear system of M equations in M unknowns. This, of course, does not mean that the individual components of S have a uniform marginal probability for being the true value between some limits. These marginal probabilities are almost never uniform and depend upon how the M dimensional region defined by the facts projects along the particular

axes. When we say that every point in a region of the M dimensional space has equal probability for being the true value of S , we mean that conditional upon our facts the joint probability density function for the components of S is a constant. If the region of the M dimensional manifold consistent with the demonstrable facts, which we shall call the D -domain, is unbounded, then this places a constraint on the observable facts for which a useful solution can exist. A useful solution can only exist if the true value of S is known to lie inside a bounded region of the M dimensional space. Let us call the $M-N$ dimensional subspace defined by the observable facts, $R' = AS$, the O -subspace of the M dimensional manifold. All that is required for a useful solution to exist is that the intersection of the D -domain and the O -subspace must be bounded. If the D -domain is bounded, then a useful solution exists without any observable facts. However, the size of the domain where the true value of S is known to lie will always be reduced by the addition of observable facts and therefore the usefulness of the solution will always be increased by the observable facts.

We now turn our attention to the question: given a specific set of demonstrable facts for the code of interest, and consequently the D -domain corresponding to them, what specific set of observable facts should we generate? That is to say: how do we choose an A matrix to generate them? The answer to this question is not only a function of the specific D -domain but also a function of what we will have to do after we obtain the observable facts. We will have to make yet another decision which is the choice of specific vector S_c to use as the true value inside a domain which will be determined by the observables we do not know yet. However, because at the time we face the choice of an A matrix we have a knowledge of the D -domain, we are not ignorant of what the observable facts might be for any given A matrix. We therefore face a typical parameter optimization problem: having to decide upon some parameters, the elements of the matrix A , based upon minimizing some function of these parameters and the observable facts they may generate.

When, in the last section, we reached the point of deciding what value to choose, when we had a uniform probability density function inside a closed domain, we assumed the reader familiar with decision

theory. We then stated that if we take a quadratic "loss function" this will lead us to select the center of gravity of the bounded domain and the "expected loss" will be given by the moment of inertia of the domain. The situation is in fact a little more complicated not only because we have to distinguish between statistical decision theory and Bayesian decision theory, but because we are not choosing a scalar but an M dimensional vector. Since our theory is a Bayesian one, we must use a Bayesian decision theory. Since there are many Bayesian theories, we cannot assume that the reader is necessarily familiar with how the concepts of "Bayesian decision theory" are applied in what we have called Jaynes' theory of logical inference which we presume the reader is not familiar with. We shall therefore very briefly give a few words of explanation.

Given that we are concerned about the true value of a quantity (α), we first obtain the probability density function for the true value of this quantity conditional upon all the relevant testable facts (TF). This so-called posterior conditional pdf we write as $(\alpha|TF)$. We then determine, based upon the consequences to us, in the specific application, for choosing a value (α_c) and the true value happens to be different, a function of the difference between these values. This is the "loss function" which we write: $L(\alpha - \alpha_c)$. The only rational decision for α_c that we can make is the one which minimizes the losses whatever the true value of α happens to be. Mathematically we choose α_c so as to minimize the "expected loss":

$$\int L(\alpha - \alpha_c) (\alpha|TF) d\alpha \quad . \quad (15)$$

As is well known, if we take the loss function to be proportional to $(\alpha - \alpha_c)^2$, a quadratic loss function, then we must choose for α_c the first moment of $(\alpha|TF)$, its center of gravity, and the expected loss becomes the second moment of $(\alpha|TF)$, its moment of inertia about the center of gravity. We apologize to the readers for which the above is well known and are elementary considerations.

There is a straightforward generalization for a quadratic loss function when we are dealing with a vector, S , instead of a scalar α . It is to take the loss function proportional to $(S - S_c)^{\dagger} (S - S_c)$, the square of the norm of the difference between the value we choose, S_c , and the

true values, S . In the interim report [1] we point out that if S is a sensitivity vector for a code there are some reasons for not choosing such a loss function. But for simplicity in this discussion we shall take as a loss function the quadratic form $(S-S_c)^{\dagger} (S-S_c)$. As can easily be verified from the generalization of the expression (15) to several dimensions, the choice of S_c which we will then make is the expectation value of the joint density function, or the center of gravity of the domain since our density function is uniform inside it. The expected loss will be given by the trace of the covariance matrix of the joint density function or the trace of the moments of inertia tensor of the domain about its center of gravity. We are therefore interested in using a matrix A which will minimize the trace of the moments of inertia tensor. Clearly this trace is a function of the shape and dimension of the final domain and these are determined by both the elements of the matrix A which we are free to choose and the observable R' which we do not know yet. However, we are not totally ignorant of what a specific matrix A will produce as the observable R' since we are knowledgeable about the D -domain where the true vector S must lie, and we know at this stage that inside it we have a uniform probability density function for the true value of S . We can view each row of the A matrix, each rerun of the code, as determining a "surface" (a $M-1$ dimensional subspace in the M dimensional manifold) where we know the true value of the sensitivity coefficient must lie and on this "surface" the probability density function for the true value of S is uniform. What the elements of the row of the A matrix do, and we control their value, is determine uniquely the direction cosines of the M dimensional vector which is orthogonal to this "surface." What the component of the vector R' , which corresponds to this row of the A matrix, does is determine the "position" of this "surface" in the M dimensional space. Before we execute the rerun corresponding to this row of the A matrix, we do not know exactly what the "position" of the "surface" to be generated by it will be, but we can calculate, using the D -domain, a probability density function for its true value. We can again, using a quadratic loss function for it, choose a value for this component of R' . In the above fashion we can express the trace of the moments of inertia tensor of the final domain as a

function of the elements of the matrix A. We must then choose consistent with all of the facts available and our goal values for the elements of the A matrix which minimize the trace of the moments of inertia tensor.

We therefore think that the conceptual aspects of finding the solution to the reformulated screening problem, to include both the demonstrable facts and some observable facts, are well in hand. We do grant, however, that for any specific code the mathematical problems of finding demonstrable facts and the A matrix needed to generate the observable facts may well be nontrivial. The purpose of this section was mostly to argue that in principle the reformulated problem not only admits a useful and objective solution, but we have also a well laid out procedure for determining it.

Although in this report we have avoided presenting mathematical details which would not be essential to the main arguments, some of these have been investigated and are presented in the interim report [1]. We were able to show that the estimator of Durston *et al.* is appropriate if the sensitivity coefficients have values which are normally distributed about zero. However, in order to obtain the degree of confidence in the screening results, we need to know the variance of this distribution.

We conclude by some remarks concerning the prospects of solving the reformulated screening problem in practice. We think that in general it may be difficult to obtain the needed demonstrable facts directly from a knowledge of the equations programmed in the code. However, many codes are developed starting from simple models to which successive refinements are added. In these cases it may be a lot easier to generate the observable facts. We conjecture that this feature is the basis of many so-called engineering and economic judgements which are often used to decide for a given output of the code what are the important parameters and often are the basis of practical sensitivity studies.

E. Summary and Conclusions

In this report we have analyzed "screening methodology" as a tool for determining approximate values for sensitivity coefficients of outputs of complex computer codes to the input data. In view of the apparent failure of a proposed algorithm to solve this problem in the case of LEAP

Model 82, we have investigated the nature of solutions claimed to be based only upon results obtained by rerunning the code a number of times less than the number of input parameters. We have found these solutions are based upon explicit or implicit assumptions regarding the code. Since the validity of these assumptions is usually not known for the specific code under study, no degree of confidence can be obtained in these results. Therefore, we think that these solutions to the problem are not as objective or useful as we might believe.

We have analyzed the screening problem from the point of view of a theory of logical inference due to Jaynes rather than the usual one of the theory of statistical inference. We find that from this point of view the problem does not admit a useful solution if we do not specify additional information about the code. This means that no objective and useful solution exists to the screening problem where the code is treated as a black box. If we can derive a sufficient number of facts, called demonstrable facts, from this additional information about the code, these can be combined with results obtained by running the code, called observable facts, to yield a solution which qualifies as being objective, approximate and useful. Given these demonstrable facts, we have shown conceptually how to generate the set of observable facts which produces the solution consistent with our stated objective.

We believe that the screening problem reformulated to include demonstrable facts, which therefore no longer treat the code as a black box, and our proposed solution to it meet all the criteria for being called an objective and useful solution to the problem of obtaining approximate values for the sensitivity coefficients of a given computer code. This is so by virtue of the facts that: a) only facts meeting the criteria for being objective are used, b) the notion of probability used to introduce a measure which provides meaning to the word "approximate" is objective, and c) the solution is useful in the sense that only decisions consistent with our avowed purpose for solving the problem are made, and the size of the possible consequences to us for deciding to use the solution given as the true value is provided. We made no claim that in any given problem anyone may be satisfied with the size of the possible consequences to him for using the solution as the true value. This is a

value judgement, and when these risks are considered too large they can always be decreased by seeking more constraining demonstrable and/or observable facts.

There are obviously other ways of reformulating and solving the screening problem in such a way that the solution can be called objective, approximate and useful. At this time we think that some of the criteria for calling the solution objective are well understood: the facts upon which the solution rests must be demonstrable and/or observable. In deference to the usual formulation of the screening problem, we have limited the class of demonstrable and observable facts to those pertaining only to the specific computer code being investigated. Regarding the meaning of approximate and useful the situation is far less clear at the moment, it seems to us. It seems conceivable that other concepts could be introduced than the ones we have used and yet meet what we perceive to be the same degree of objectivity. The question is a rather fundamental one and epistemological. The virtue of the concept of probability we have used is that it has been extensively analyzed over the last three centuries and subjected to extensive scrutiny. The desideratum of consistency we have used and the manner in which we have applied it are based upon very specific suggestions of Jaynes and appear to us, as demonstrated in this report, to yield an unambiguous solution to some problems of logical inference. This theory avoids some previously perceived difficulties in solving problems of logical inference that were thought to be associated with the concept of probability itself, but we think at the moment are only connected to the way it was previously used.

F. Acknowledgements

I wish to thank Collin Durston and Ted Krieger of Brookhaven National Laboratory for interesting discussions concerning their studies and the permission to quote their results and report. I thank also J. E. Horwedel for the running of the code LEAP Model 82, providing the data presented in this report. I acknowledge many discussions with my colleagues at ORNL on the theory of logical inference, particularly G. de Saussure, Jim Marable, R. W. Peelle and Ed Oblow, who initiated these screening studies for LEAP. I also thank Paul Werbos and Doug Hale of the EIA for

constructive comments on an early draft of this report. This work would not have been possible without the continued support and encouragement of F. C. Maienschein.

References

1. R. G. Alsmiller *et al.*, "Interim Report on Model Evaluation Methodology and the Evaluation of LEAP," draft, April 1980, ORNL/TM-7245.
2. C. Durston, T. Krieger, and D. C. Albright, BNL-NUREG 23430, Brookhaven National Laboratory.
3. E. T. Jaynes, IEEE Trans. on Systems Science and Cybernetics 4, 227 (1968).
4. J. K. Vaurio, Trans. Am. Nucl. Soc. 32, 296 (1979).

APPENDIX

We provide here the numerical data upon which Figs. 1 and 2 are based. The parameters of Model 82 of LEAP for which the sensitivity coefficients for the quantity of domestic oil produced in the year 2020 were obtained are given with an identifying numeral in Table I. In Table II the A matrix used in the screening is tabulated. In Table III are presented the differences observed in the code output when using the matrix A of Table II. In Table IV are given the results of the exact calculation of the sensitivity coefficients ranked according to their absolute value and the corresponding values obtained by screening and their ranking based on the screening.

TABLE I
LEAP MODEL 82 PARAMETER IDENTIFICATION

<u>Parameter #</u>	<u>Name</u>	
<u>Demand Process</u>		
1	Behavioral lag	
2	Price sensitivity	
<u>Conversion Process for Oil Heat</u>		
3	Specific capital cost	
4	Variable operating cost	
5	Thermal efficiency	
<u>Resource Process for Domestic Oil</u>		
6	Equity finance fraction	
7	Initial resource cost	
8	Base price	
9	Undiscovered recoverable resource at base price	
10	New resources at twice base price	
11	Capital cost fraction	
12	Deposit decline rate	
<u>Allocation Process</u>		
13	Base price sensitivity	
14	Behavioral lag	
<u>Time-Dependent Parameters that Occur in both the Conversion Process for Oil Heat and the Resource Process for Domestic Oil</u>		
15-24	Price of labor	Values in all 10 model years
25-34	Price of material	Values in all 10 model years
35-44	Return on equity	Values in all 10 model years
45-54	Return on debt	Values in all 10 model years
<u>Residential Demand</u>		
55-63	Personal income series	Values in all model years other than 1975
64-70	Design sensitivity oil heat	Values in years 1975, 1980, 1985, 1990, 1995, 2000 and 2020 (Note: The code does not use the input values for years 2005, 2010 and 2015)
<u>Oil Prices from Imports</u>		
71-80	Oil price	Values in all 10 model years

TABLE II
PERTURBATION MATRIX A USED IN SCREENING

```

1000010010101111110
11000110111001110101
0010110110010000000
00110100001000011011
01010011001111100110
01101011001001110110
10101011011101111001
00010000000111101100
10001100110111010110
10111001001000101011
01000100001110100010
00101000101000000011
01111000000101101101
01000111001011111111
01101011011001101000
01111110000100001010
11010100101011011011
11111010100011101101
00001011001111100101
01011100011011101010
10100011100111011001
00001000010100110001
01001111010111110001
0011110100000011111
00101010110111010011
01011111011110100011
00011000010110011010
01010010110101100000
01000011111010000111
11001000110100101000
10101101011010010110
11100100001111111111
10101011000100100000
10101110011010101001
10011010011010011110
11001001110000101010
11011010011011011001
11001010111000000000
1010101001111110110
1001101010000101010

```

(continued in next column)

```

11000010011010110110
0101111011001001011
10101100110000000100
00101001101001100011
10100011000001011110
10110001000100101101
01011101011111111100
00001000010000011100
11001010000001010001
01000101010001100100
10011110000001110011
0000101101110100100
11001011110011000110
10100001110011100110
00000011000100111110
00011111000000101110
10110101000010110111
00110000111110000111
01001100111101000000
11000101110010001001
11001101000000010101
00110101010011101000
11111001100100110111
11000010101010011111
10001011011110000010
0111010000000011010
00100110111100100001
10101100110010101100
11100101111010010011
11111011100000110010
11101101010011111110
10100111111111011010
00101001110011100000
11011101111010011010
00100010001000011100
01111010111100110111
10101101111010101100
10100000011100100001
0100011111011100101
00001010010100010010

```

NOTE: For convenience of presentation, we give the matrix A^+ .
A 1 corresponds to a 1% increase in the value of the
input parameter and a 0 corresponds to a 1% decrease in
the value of the parameter.

TABLE III
 CHANGES IN PERCENT IN THE QUANTITIES OF DOMESTIC OIL
 YEAR 2020 OBSERVED WITH THE MATRIX A OF TABLE II

Rerun #i	R_i^1
1	-1.7180
2	0.0139
3	0.3096
4	-1.4588
5	-0.9960
6	3.4650
7	-0.9415
8	-2.0191
9	3.1976
10	1.3867
11	-4.0345
12	-1.2207
13	1.8464
14	-1.9887
15	-3.4480
16	1.3515
17	-2.1282
18	1.2777
19	0.6234
20	-0.7358

TABLE IV
RELATIVE SENSITIVITY COEFFICIENTS FOR DOMESTIC OIL QUANTITIES
YEAR 2020

True Ranking	Parameter #	True Value	Approximate Value	Approximate Ranking
1	10	-0.9502	-1.468	1
2	7	-0.9081	-1.400	2
3	5	-0.8026	-0.644	14
4	69	0.3882	0.548	19
5	25	0.3731	1.044	3
6	9	0.3515	0.732	11
7	66	0.2950	0.836	8
8	56	0.2949	-0.084	71
9	30	-0.2310	-0.376	32
10	70	0.2123	0.152	59
11	58	0.1889	0.376	33
12	68	0.1889	0.868	7
13	12	-0.1642	0.152	60
14	65	0.1598	-0.680	12
15	55	0.1581	-0.140	64
16	13	0.1573	-0.396	30
17	59	0.1529	-0.100	70
18	57	0.1430	0.552	18
19	67	0.1430	0.316	37
20	29	0.1395	0.400	29
21	31	-0.1362	0.324	36
22	60	0.1348	0.432	27
23	34	0.1256	-0.028	76
24	61	0.1196	0.272	44
25	32	-0.1027	-0.220	50
26	62	0.0994	0.192	54
27	44	0.0953	-0.656	15
28	63	0.0830	-0.152	58
29	2	-0.0782	-0.012	79
30	1	-0.0720	-0.036	75
31	27	-0.0703	0.332	35
32	28	-0.0679	0.024	77
33	26	-0.0648	-0.280	40
34	39	0.0513	-0.224	48
35	8	0.0502	-0.172	56
36	36	-0.0472	-0.524	21
37	40	-0.0426	0.024	78
38	37	-0.0362	-0.588	17
39	45	0.0360	-0.200	52
40	14	0.0357	-0.048	74

TABLE IV, cont.

True Ranking	Parameter #	True Value	Approx. Value	Approx. Ranking
41	14	0.0351	0.472	23
42	20	-0.0288	-0.344	34
43	50	-0.0264	0.144	62
44	33	-0.0252	-0.412	28
45	75	0.0240	0.276	43
46	19	0.0208	-0.676	13
47	41	-0.0206	0.192	55
48	43	0.0149	0.876	6
49	49	0.0133	0.012	80
50	51	-0.0119	-0.148	61
51	21	-0.0119	0.112	69
52	54	0.0117	0.220	49
53	22	-0.0117	0.132	65
54	38	-0.0096	-0.196	53
55	52	-0.0093	0.120	67
56	47	-0.0082	-0.532	20
57	46	-0.0075	-0.876	5
58	17	-0.0072	-0.128	66
59	24	0.0066	0.384	31
60	16	-0.0058	0.448	25
61	42	-0.0058	0.304	38
62	48	-0.0055	0.440	26
63	35	-0.0050	-0.232	46
64	23	-0.0042	0.468	24
65	73	0.0034	0.492	22
66	18	-0.0031	0.204	51
67	3	0.0029	0.616	16
68	76	-0.0029	0.140	63
69	77	-0.0027	-0.068	73
70	53	-0.0024	0.288	39
71	72	0.0023	-0.120	68
72	79	0.0017	0.956	4
73	78	-0.0015	-0.792	10
74	71	-0.0014	0.280	41
75	4	0.0013	0.076	72
76	11	0.0008	0.232	47
77	74	-0.0005	-0.156	57
78	15	0.0003	-0.832	9
79	6	0.0000	-0.280	42
80	64	0.0000	0.264	45

Fig. 1. Comparison of true and approximate rankings of the absolute values of the sensitivity coefficients. Number of screening runs $N = 20$. Open circles indicate that true and approximate sensitivity coefficients have the same sign; open triangle, opposite sign.

ORNL-DWG 80-16419

Fig. 2. Comparison of true and approximate rankings of the absolute values of the sensitivity coefficients. Number of screening runs $N = 10$. Open circles indicate that true and approximate sensitivity coefficients have the same sign; open triangle, opposite sign.

INTERNAL DISTRIBUTION

1-2.	L. S. Abbott	76.	J. V. Pace
3.	R. G. Alsmiller, Jr.	77.	S. C. Parikh
4.	F. S. Alsmiller	78.	C. V. Parks
5.	S. J. Auerbach	79-88.	R. W. Peelle
6.	J. Barhen	89.	F. G. Perey
7.	J. Barish	90.	R. B. Perez
8.	D. E. Bartine	91.	R. Pevey
9.	M. A. Bjerke	92.	Y. Ronen
10.	D. Bjornstad	93.	M. W. Rosenthal
11.	D. G. Cacuci	94.	R. W. Roussin
12.	S. Cantor	95.	R. T. Santoro
13.	G. de Saussure	96.	D. L. Selby
14.	D. J. Downing	97.	R. Sharma
15-39.	EPIC	98.	L. Stewart
40.	G. F. Flanagan	99.	C. Thomas
41.	T. B. Fowler	100.	B. C. Toney
42.	W. E. Ford, III	101.	J. Trimble
43.	W. Fulkerson	102.	D. K. Trubey
44.	D. A. Gardiner	103.	D. R. Vondy
45.	D. Gilai	104.	E. Wacholder
46.	H. Goldstein (Consultant)	105.	J. J. Wagschal
47.	M. Goldstein	106.	C. F. Weber
48.	D. Greene	107-116.	C. R. Weisbin
49.	P. M. Haas	117.	L. W. Weston
50.	E. Halbert	118.	J. R. White
51.	M. C. G. Hall	119.	G. E. Whitesides
52.	L. Hill	120.	T. J. Wilbanks
53.	E. Hirst	121.	M. L. Williams
54.	J. Horwedel	122.	R. Q. Wright
55.	D. T. Ingersoll	123.	A. Zucker
56.	J. M. Kallfelz	124.	P. Greebler (Consultant)
57-66.	A. S. Loeb	125.	H. J. C. Kouts (Consultant)
67.	J. L. Lucius	126.	W. B. Loewenstein (Consultant)
68.	B. J. Jacobs	127.	R. Wilson (Consultant)
69.	F. C. Maienschein	128-129.	Central Research Library
70.	J. J. Manning	130.	ORNL Y-12 Technical Library -
Document Reference Section
71.	J. H. Marable	131.	Laboratory Records
72.	B. F. Maskewitz	132.	ORNL Patent Office
73.	B. D. Murphy	133.	Laboratory Records - RC
74.	D. Nguyen		
75.	E. M. Oblow		

EXTERNAL DISTRIBUTION

134. Office of Assistant Manager for Energy Research and Development,
DOE-ORO, Oak Ridge, TN 37830
135. R. L. Basmann, Department of Economics, Texas A&M University, College
Station, Texas 77843
136. M. Becker, Dept. of Nuclear Engineering, Rensselaer Polytechnic
Institute, Tibbets Avenue, Troy, NY 12181
137. D. J. Beecy, Acting Director, Division of Policy & Strategic Planning,
Office of Fossil Energy, MS C-125, U. S. Department of Energy,
Washington, DC 20545
138. W. G. Bentley, Florida Power and Light Company, P.O. Box 529100,
Miami, FL 33152
139. T. E. Browne, Energy Supply Studies Program, Electric Power Research
Institute, P.O. Box 10412, Palo Alto, CA 94303
140. D. Cato, Office of Energy Information, U. S. Dept. of Energy, Federal
Building, 12th and Pennsylvania Avenue, NW, Washington, DC 20461
141. E. Cazalet, Decision Focus, Incorporated, 5 Palo Alto Square, Suite
410, Palo Alto, CA 94303
142. R. Gene Clark, Office of Energy Information, U. S. Department of
Energy, Federal Building, 12th and Pennsylvania Avenue, NW
Washington, DC 20461
143. S. Cohen, Energy Information Administration, U. S. Dept. of Energy,
12th and Pennsylvania Ave. NW Washington, DC 20461
144. Charles Cook, Office of Energy Information, U. S. Department of
Energy, Federal Building, 12th and Pennsylvania Avenue, NW, Washington,
DC 20461
145. C. Cornett, Energy Information Administration, U. S. Department of
Energy, 12th and Pennsylvania Ave., NW, Washington, DC 20461
146. C. W. Craven, Science Applications, Inc., 800 Oak Ridge Turnpike,
Building C, Suite 100, Oak Ridge, TN 37830
147. Robert Crockett, Energy Information Administration, U. S. Department
of Energy, 12th and Pennsylvania Ave., NW, Washington, DC 20461
148. Ned Dearborn, Energy Information Administration, U. S. Department of
Energy, 12th and Pennsylvania Ave., NW, Washington, DC 20461

149. L. E. DeSimone, San Diego Gas & Electric Company, P.O. Box 1831, San Diego, CA 92112
150. R. C. Diehl, Aerojet Energy Conversion Co., P.O. Box 13222, Sacramento, CA, 95813
151. Mitchell Durst, Office of Energy Information, U. S. Department of Energy, Federal Building, 12th and Pennsylvania Avenue, NW, Washington, DC 20461
152. Energy Systems Library, Brookhaven National Laboratory, Building 475, Upton, Long Island, NY 11973
153. R. T. Eynon, Office of Energy Information, U.S. Department of Energy, Federal Building, 12th and Pennsylvania Avenue, NW, Washington, DC 20461
154. J. Eyster, Office of Energy Information Validation, U.S. Department of Energy, Federal Building, 12th and Pennsylvania Avenue, NW, Washington, DC 20461
155. R. C. Fair, Department of Economics, Yale University, 206 Elm Street, New Haven, Connecticut 06520
156. J. Finucane, Office of Energy Information Validation, U.S. Department of Energy, 12th and Pennsylvania Avenue, NW, Washington, DC 20461
157. A. Ford, Los Alamos Scientific Laboratory, P.O. Box 1663, Los Alamos, NM, 87545
158. D. Freedman, Professor of Statistics, University of California, Berkeley, CA 94708
159. S. I. Gass, College of Business and Management, University of Maryland, College Park, Maryland 20742
160. Donald Gantzer, U.S. Department of Energy, Federal Building, Room 4447, 12th and Pennsylvania Avenue, Washington, DC 20461
161. O. Gildersleeve, Electric Power Research Institute, 3412 Hillview Ave., Palo Alto, CA 94303
162. C. R. Glassey, University of California Berkeley, 1163 Keeler Avenue, Berkeley, CA 94708
163. Neil Goldman, Energy Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
164. L. Gordon, Office of Energy Information Validation, Federal Building, 12th and Pennsylvania Avenue, NW, 20461

165. H. J. Greenberg, U. S. Department of Energy, 12th and Pennsylvania Avenue, NW, Washington, DC 20461
166. M. Greenberger, Stanford University Graduate School of Business, Room 383, Stanford, CA 94305
167. E. Greenspan, University of Illinois, Nuclear Engineering Program, 214 Nuclear Engineering Laboratory, Urbana, IL 61801
168. James Gruhl, Energy Model Analysis Program, Model Validation Group, MIT Energy Laboratory, Cambridge, Massachusetts 02138
169. N. Gruhl, 142 Plain Road, Wayland, Massachusetts 01778
170. D. E. Hale, Office of Energy Information Validation, U.S. Department of Energy, 12th and Pennsylvania, NW, Washington, DC 20461
171. John Hansen, Charles River Associates, 200 Clarendon Street, Boston, MASS 02116
172. C. Harris, Center for Management & Policy Research, 1625 "I" Street, NW Washington, DC 20006
173. J. Hausman, Economics Department, E-52, Massachusetts Institute of Technology, Cambridge, MASS 02139
174. W. W. Havens, Columbia University, 520 West 120th Street, New York, NY 10027
175. W. W. Hogan, Harvard University, 79 Boylston Street, Cambridge, MASS 02138
176. Mary Hutzler, Energy Information Administration, 12th and Pennsylvania Avenue, NW, Washington, DC 20461
177. R. H. F. Jackson, Center for Applied Mathematics, U.S. Department of Commerce, National Bureau of Standards, Washington, DC 20234
178. C. W. Kilgore, Energy Information Administration, U.S. Department of Energy, 12th and Pennsylvania Avenue, NW Washington, DC 20461
179. D. Kresge, 53 Church Street, Cambridge, MASS 02138
180. G. M. Lady, Energy Information Administration, U.S. Department of Energy, 12th and Pennsylvania Avenue, NW, Washington, DC 20461
181. Joel Lambert, National Bureau of Standards, Building 101, Washington, DC 20234
182. W. Langley, FJSRL, U.S. Air Force Academy, Colorado Springs, CO 80840

183. C. J. Lee, Department of Economics, Knox College, Galesburg, IL 61401
184. A. Linden, Jr., Energy Information Administration, U.S. Department of Energy, 2000 M Street, Washington, DC 20545
185. T. V. Long, II, University of Chicago, The Committee on Public Policy Studies, 1050 E. 59th Street, Chicago, IL 60637
186. H. McDaniel, Office of Energy Information Validation, U.S. Department of Energy, 12th and Pennsylvania Avenue, NW, Washington, DC 20461
187. M. D. McKay, Los Alamos Scientific Laboratory, P.O. Box 1663, Los Alamos, NM 87545
188. D. L. McNichol, Energy Information Administration, U.S. Department of Energy, 12th and Pennsylvania Avenue, NW, Washington, DC 20461
189. R. H. Males, Director, Energy Systems, Environment and Conservation Division, Electric Power Research Institute, 3412 Hillview Avenue, P.O. Box 10412, Palo Alto, CA 94303
190. R. A. Marshalla, Decision Focus, Inc., 5 Palo Alto Square, Suite 410, Palo Alto, CA 94303
191. J. S. Maybee, University of Colorado, Mathematics Department, Boulder, CO 90309
192. John Mitrisin, Energy Information Administration, U.S. Department of Energy, 12th and Pennsylvania Avenue, NW, Washington, DC 20461
193. Thomas Mooney, Energy Information Administration, U.S. Department of Energy, 12th and Pennsylvania Avenue, NW, Washington, DC 20461
194. G. H. Moore, Center for Energy Studies, University of Texas, Austin, TX 78712
195. T. H. Morlan, Energy Information Administration, U.S. Department of Energy, 12th and Pennsylvania Avenue, NW, Washington, DC 20461
196. F. H. Murphy, Energy Information Administration, U.S. Department of Energy, 12th and Pennsylvania Avenue, NW, Washington, DC 20461
197. E. L. Nelson, Energy Information Administration, U.S. Department of Energy, 12th and Pennsylvania Avenue, NW, Washington, DC 20461
198. D. M. Nesbitt, Decision Focus, Inc., 5 Palo Alto Square, Suite 410, Palo Alto, CA 94304

199. David Nissen, Chase Manhattan Bank, 1 Chase Manhattan Plaza, New York, NY 10015
200. E. Oatman, Electric Power Research Institute, 3412 Hillview Avenue, P.O. Box 10412, Palo Alto, CA 94303
201. Richard P. O'Neill, Energy Information Administration, U.S. Department of Energy, 12th and Pennsylvania Avenue, NW, Washington, DC 20461
202. Jerry Peabody, Energy Information Administration, U.S. Department of Energy, 12th and Pennsylvania Avenues, NW, Washington, DC 20461
203. John Pearson, Energy Information Administration, U.S. Department of Energy, 12th and Pennsylvania Avenue, NW, Washington, DC 20461
204. R. Phillips, Decision Focus, Inc., 5 Palo Alto Square, Suite 410, Palo Alto, CA 94303
205. B. Poole, Energy Information Administration, U.S. Department of Energy, 12th and Pennsylvania Avenue, NW, Washington, DC 20461
206. J. Rambo, Lawrence Livermore Laboratory, University of California, Livermore, CA 94550
207. Lillian Regelson, Energy Information Administration, U.S. Department of Energy, 12th and Pennsylvania Avenue, NW, Washington, DC 20461
208. Richard Richels, Electric Power Research Institute, 3412 Hillview Avenue, P.O. Box 10412, Palo Alto, CA 94303
209. Fred S. Roberts, Rutgers University, Department of Mathematics, New Brunswick, NJ 08903
210. M. E. Rodekahr, Energy Information Administration, U.S. Department of Energy, 12th and Pennsylvania Avenue, NW, Washington, DC 20461
211. D. J. Rose, Dept. of Nuclear Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139
212. L. Rubin, Electric Power Research Institute, 3412 Hillview Avenue, P.O. Box 10412, Palo Alto, CA 94034
213. Roy Shanker, Resource Planning Associates, 1901 "L" Street, NW, Washington, DC 20036
214. John B. Shewmaker, Energy Information Administration, U.S. Department of Energy, 12th and Pennsylvania Avenue, NW, Washington, DC 20461

- 215. Scott Sitzer, Energy Information Administration, U.S. Department of Energy, 12th and Pennsylvania Avenue, NW, Washington, DC 20461
- 216. C. Smith, Energy Information Validation, U.S. Department of Energy, 12th and Pennsylvania Avenues, NW, Washington, DC 20461
- 217. James Sweeney, Energy Modeling Forum, Stanford University, Stanford, CA 94305
- 218. Takashi Takayama, Energy Information Administration, U.S. Department of Energy, 12th and Pennsylvania Avenue, NW, Washington, DC 20461
- 219. J. E. Todd, Energy Information Administration, U.S. Department of Energy, 12th and Pennsylvania Avenue, NW, Washington, DC 20461
- 220. Noel Uri, Energy Information Administration, U.S. Department of Energy, 12th and Pennsylvania Avenue, NW, Washington, DC 20461
- 221. J. Weatherly, Institute of Gas Technology, 3424 S. State Street, Chicago, IL 60601
- 222. A. M. Weinberg, Institute for Energy Analysis, Oak Ridge Associated Universities, Oak Ridge, TN 37830
- 223. P. Werbos, Office of Energy Information Validation, U.S. Department of Energy, 12th and Pennsylvania Avenue, NW, Washington, DC 20461
- 224. D. O. Wood, Energy Laboratory #-38, 418, Massachusetts Institute of Technology, Cambridge, MA 02139
- 225-228. L. Zutovsky, Energy Information Administration, U.S. Department of Energy, 12th and Pennsylvania Avenues, NW, Washington, DC 20461
- 229-255. Technical Information Center (TIC)