CONF-860990-8

ADGEN: A SYSTEM FOR AUTOMATED SENSITIVITY ANALYSIS OF PREDICTIVE MODELS*

F. G. Pin, J. E. Horwedel, E. M. Oblow and J. L. Lucius

Oak Ridge National Laboratory
P.O. Box X
Oak Ridge, TN 37831

CONF-860990--8

DE87 002186

Invited paper - Eighth Annual DOE Low-Level Waste Management Forum, Denver, Colorado (September 22-26, 1986).

*Research sponsored by the Office of Defense Waste and Transportation Management of the U.S. Department of Energy, under contract No. DE-AC05-84OR21400 with Martin Marietta Energy Systems, Inc.

ADGEN: A SYSTEM FOR AUTOMATED SENSITIVITY ANALYSIS OF PREDICTIVE MODELS

F. G. Pin, J. E. Horwedel, E. M. Oblow and J. L. Lucius

ABSTRACT

A system that can automatically enhance computer codes with a sensitivity calculation capability is presented. With this new system, named ADGEN, rapid and cost-effective calculation of sensitivities can be performed in any FORTRAN code for all input data or parameters. The resulting sensitivities can be used in performance assessment studies related to licensing or interactions with the public to systematically and quantitatively prove the relative importance of each of the system parameters in calculating the final performance results. A general procedure calling for the systematic use of sensitivities in assessment studies is presented. The procedure can be used in modelling and model validation studies to avoid "over modelling", in site characterization planning to avoid "over collection of data", and in performance assessments to determine the uncertainties on the final calculated results. The added capability to formally perform the inverse problem, i.e., to determine the input data or parameters on which to focus additional research or analysis effort in order to improve the uncertainty of the final results, is also discussed.

INTRODUCTION

This session of the forum is aimed at addressing several questions of major importance concerning the assessment of site and system performance for Low Level Waste (LLW) disposal facilities. In this paper, we will restrict our discussions to four of these questions which are particularly pertinent to the analytical part of performance assessments:

- How good is the answer?
- How can the answer be improved?
- What should be changed?
- How do all the pieces fit?

These questions, which could be asked generically in any performance assessment study, hint at some very specific analysis problems in the particular context of LLW management. In this paper, we will briefly recall the problems from which these questions emerge and propose solutions under the form of new procedures and state-of-the-art analysis methodologies for use in performance assessment of LLW disposal facilities.

GENERAL STATEMENT OF THE PROBLEM

The two general objectives of a performance assessment for a waste isolation system are to evaluate the level of confidence which can be placed in a facility (i.e., estimate the risks involved) in meeting prescribed performance measures, and to pass this level of confidence along to the public and/or the regulatory or licensing authorities. Very special attention is required in reaching these objectives as three aspects of radioactive waste management have become increasingly critical with increased public awareness and involvement. These aspects requiring special attention can be stated as follows: (1) uncertainties in performance assessments equate to risks to public health, (2) remedial actions should be possible to plan and implement (including in the analytical phase) if assessment results are unsatisfactory, and (3) qualitative results are insufficient in performance assessments of radioactive waste isolation systems. With the increasing scale of our performance assessment studies for waste isolation facilities, these concerns have become crucial as no methods suitable for large-scale systems are available to fully answer them.

The data characterizing geologic media and waste behavior are generally associated with large uncertainties. These uncertainties result from the extreme heterogeneity of the subsurface materials, our lack of understanding of all the complex geochemical and geohydraulic phenomena, and, still today, the inaccuracy of even the most refined state-of-the-art measuring techniques. When these data are used in design studies or performance assessments of large-scale systems, their large uncertainties can propagate through the calculational schemes to lead to very large uncertainties on the performance measures or results of interest. As we mentioned earlier, large uncertainties in low level radioactive waste disposal assessments generally equate to added risk to the environment or public health. Thus, to know how good the answers of our design studies or risk assessment analyses are, we should evaluate the uncertainties on the final results or performance measures concurrently with the calculated results. Moreover, this evaluation should provide for the assurance, preferably quantitative, that all potential sources of uncertainty in a given problem have been systematically identified and that their effects (i.e., sensitivity) on the final results have been taken into account in the analysis. Finally, to allow us to improve the answers if needed, the method for evaluating the sensitivities and uncertainties should provide for the capability to formally perform the inverse problem, that is, given an unacceptable (i.e., too large) uncertainty on a final performance measure, identify and prioritize all data or input parameters that contribute to that uncertainty, pointing out those needing added attention or research effort in the site characterization or data generating processes, in order to decrease the uncertainty on the calculated values.

Statistical methods and perturbation analysis have traditionally been used to perform sensitivity and uncertainty analyses. When applied to large-scale studies such as those related to radioactive waste isolation, these conventional methods have serious drawbacks: not only is their cost prohibitive, but they generally cannot handle all parameters involved in a study. They also rely on qualitative expert opinion for selection of "key" parameters, thus preventing formal solution of the inverse problem. Deterministic methods, on the other hand, such as the adjoint formulation, allow systematic and quantitative screening of the parameter space, thus, providing all the information necessary to perform a complete sensitivity analysis including the inverse problem. These methods, however, have not enjoyed the wide-spread use they deserve, mainly because they require a large initial analytical investment and because their implementation, which is specific to a given model or code, has been considered too costly to practically permit more than first order results to be obtained.

A system which makes use of the strengths of the adjoint method and of computer calculus, has been designed to automate the costly and time-consuming processes of deterministic sensitivity analysis. This system, named ADGEN, is presented in the next section. It uses a computer pre-compiler named EXAP (EXtended Arithmetic Processor), and will allow cost-efficient, quantitative and systematic sensitivity analyses to be performed which otherwise might not have been undertaken. Once a complete sensitivity analysis is achieved using ADGEN, it can be coupled with conventional statistical methods within an overall sensitivity and uncertainty analysis procedure which should allow one to answer the concerns underlying the questions addressed in this session.

THE ADGEN SYSTEM

The ADGEN system has been designed to automatically generate adjoint solutions of computer codes. The overall automated process can be simply described using the following example. Let

$$\overline{y} = \overline{F}(\overline{y},\overline{c}) \tag{1}$$

represent, in vector form, the set of equations and storage operations programmed in a FORTRAN code. The components of the vector $\overline{y}$ on the left-hand side of the equation are the stored value of the variables being solved for, $\overline{c}$ represents the user-specified model data or parameter set, and $\overline{F}$ is a functional that defines the model equations. Typically, several results which are some function of the solution to Eq. (1) are of particular interest to the model user. Let

$$R = h(\overline{y}) \tag{2}$$

define a typical result, R, where R is a single number and h represents the known functional dependence of R on $\overline{y}$. For notational ease, let α_i denote a generic parameter which can be a component of the vector $\overline{c}$ or any other variable used in the code.

The basic problem in any sensitivity study is to find the rate of change in the result R arising from changes in any of the model parameters. For the generic parameter α_i, then, the quantity of interest is the numerical value of $dR/d\alpha_i$ given analytically by

$$\frac{dR}{d\alpha_i} = \frac{\partial h}{\partial \overline{y}} \frac{d\overline{y}}{d\alpha_i} \quad . \tag{3}$$

Since the functional dependence of R on $\overline{y}$ through $h(\overline{y})$ is defined analytically by the model user, only $d\overline{y}/d\alpha_i$ needs to be generated in order to evaluate Eq. (3). The procedure needed to get $d\overline{y}/d\alpha_i$ is to differentiate Eq. (1) and rearrange it to yield the following set of coupled equations to solve for $d\overline{y}/d\alpha_i$,

$$\left(I - \frac{\partial \overline{F}}{\partial y}\right) \frac{d\overline{y}}{d\alpha_i} = \frac{\partial \overline{F}}{\partial \overline{c}} \frac{d\overline{c}}{d\alpha_i} \tag{4}$$

where I is the identity matrix.

If Eq. (4) were solved directly for $d\overline{y}/d\alpha_i$, the result could be used in Eq. (3) to evaluate $dR/d\alpha_i$. This method of sensitivity analysis is called the "direct" approach and is a classical methodology which has received a great deal of attention in the literature.[1,2] Its main drawback arises in large-scale applications where the size of the vector $\overline{c}$ (and therefore the number of α_i's whose sensitivities need to be evaluated) becomes prohibitively large. Since Eq. (4) must be solved each time a new α_i is defined, the computational expense puts this method out of reach as a practical tool for large-scale sensitivity studies. Its practical value is therefore restricted to smaller-scale analytical problems or other cases where $\left(I - \frac{\partial \overline{F}}{\partial y}\right)$ can easily be inverted.

Since the ultimate objective of a large study, however, is still the evaluation of $dR/d\alpha_i$, the intermediary step of solving for $d\overline{y}/d\alpha_i$ and its inherent computational inefficiency can be avoided. For such problems the "adjoint" approach is far more applicable. In this methodology, use is made of the fact that Eq. (4) is linear in $d\overline{y}/d\alpha_i$, and an appropriate adjoint equation can therefore be developed specifically to evaluate Eq. (3) as

$$\frac{dR}{d\alpha_i} = y^{*tr} \frac{\partial \overline{F}}{\partial \overline{c}} \frac{d\overline{c}}{d\alpha_i} \tag{5}$$

$$\left(I - \frac{\partial \overline{F}}{\partial y}\right)^{tr} \overline{y}^{*} = \left(\frac{dh}{dy}\right)^{tr} \quad , \tag{6}$$

and the superscript "tr" represents the transpose of the vector or matrix.

The simplicity of the adjoint approach lies in the fact that Eq. (6) needs to be solved only once to get any and all sensitivities in the problem. This is a result of Eq. (6) being independent of the choice of α_i. The particular choice of α_i is only reflected in the evaluation of Eq. (5) which involves only simple vector products. In essence, the adjoint approach reduces the computational effort needed to evaluate $dR/d\alpha_i$ from solving many coupled linear equations to the evaluation of several vector products. For large scale systems with many hundreds or even thousands of parameters, this represents orders of magnitude in computational efficiency.

The major drawback of the adjoint approach has traditionally been the large analytical investment necessary to derive the many derivatives represented by $\partial\overline{F}/\partial\overline{y}$ and $\partial\overline{F}/\partial\overline{c}$. This drawback is now eliminated with the automatic differentiation capability of GRESS.[3,4,5,6] Thus, ADGEN uses a GRESS-like precompiler, named EXAP, to calculate all required derivatives in Eqs. (5) and (6), and automatically sets up Eq. (6) as a large matrix equation. The solution $\overline{y}^{*}$ is then calculated using standard back-substitution techniques and is used in Eq. (5) to obtain all sensitivities of interest through straightforward vector calculations. The efficiency of the overall system increases with the number of required sensitivities, i.e., the number of parameters, and problems that were practically unapproachable with the "direct" approach can now be done in routine fashion.

The flow chart of the ADGEN system is shown in Fig. 1. In the first step of the procedure, the FORTRAN code of interest is submitted to the EXAP precompiler. EXAP reads the FORTRAN code text, searches for arithmetic statements, enhances these for gradient calculation, and then generates a new FORTRAN source program which can produce the required derivatives for adjoint generation. The code enhancement is performed using automated computer calculus and consist of an automatic addition of lines of coding. These added lines of coding are calls to EXAP library routines which, for each storage operation, calculate the analytical derivatives of the stored quantity with respect to any variable used in its computation. The modified code is then compiled into a pseudo-machine code (P-code) for use during program execution. The two output files of this step are, therefore, the enhanced FORTRAN code and the binary P-code file.

In the second stage of the ADGEN procedure, the enhanced FORTRAN code is combined with the EXAP interpretive library and the P-code file to form the complete executable program. The library contains the P-code interpreter, a series of derivative calculation routines, and a set of utility subroutines which are used to automatically output the derivatives in a large matrix format. The resulting enhanced program will execute identically to the original FORTRAN code with the option of also calculating derivatives and setting up the adjoint equations in a large matrix form.

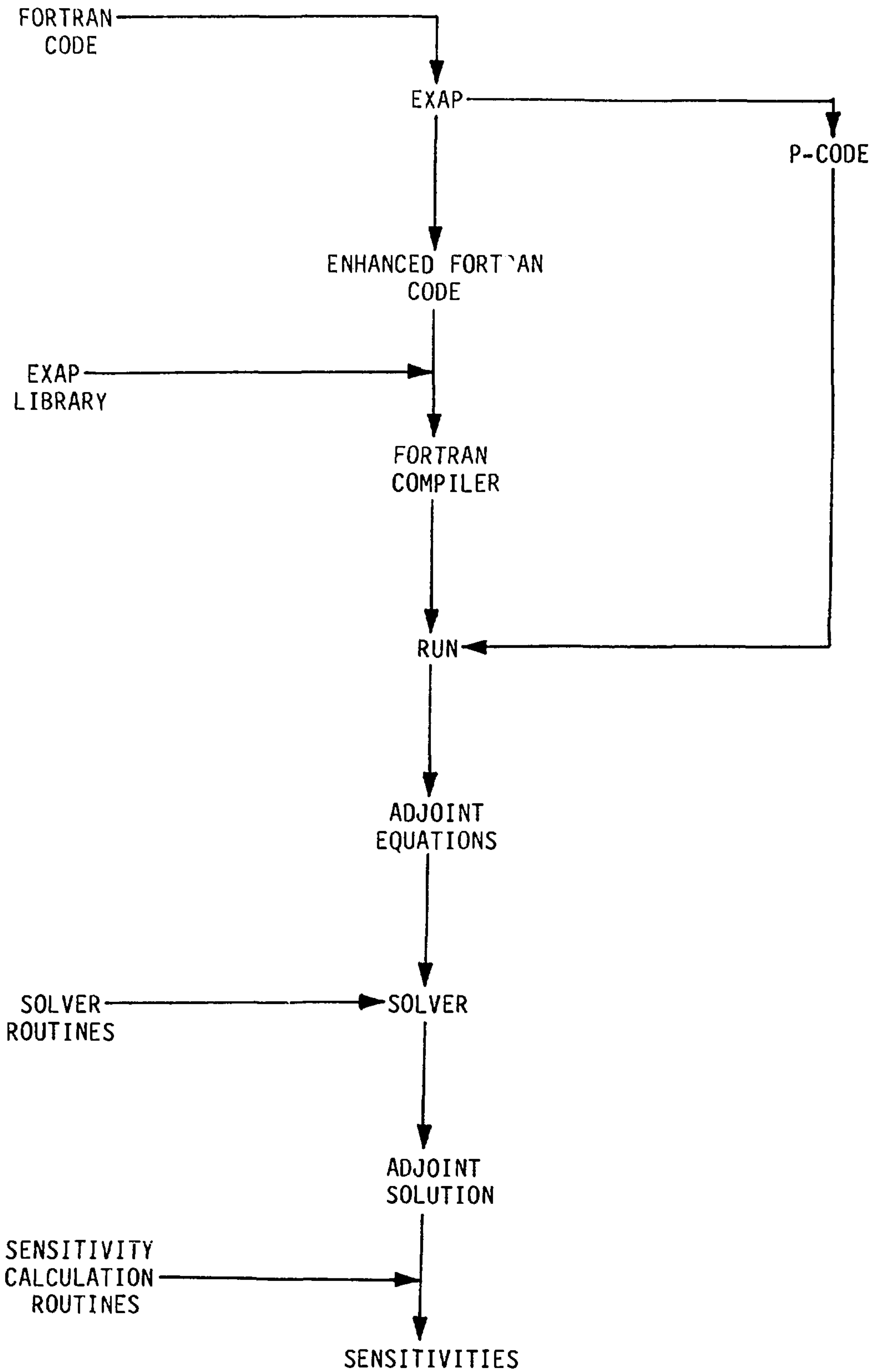

Fig. 1. Flow Chart of the ADGEN System

The next step of the procedure calls upon the routines of an efficient back-substitution solver to calculate the solution of the matrix equation [see Eq. (6)] and upon sensitivity calculation routines to implement the calculations of Eq. (5) for any parameter or real variable in the original code.

A preliminary version of the overall system (Version A) has been completed with the intent to check the feasibility and practicality of the ADGEN procedure. Version A includes a preliminary precompiler limited to the most common FORTRAN syntax, a simulated library, and simplified adjoint matrix generator, adjoint solver and sensitivity calculation modules. Version A has been successfully tested on sample problems and has provided the design requirements for the development of a full-scale system suitable for large scale applications.

COMPLETE SENSITIVITY AND UNCERTAINTY ANALYSIS PROCEDURE

The following approach which makes use of the new ADGEN sensitivity analysis tool and of conventional statistical methods is proposed to answer, for the analysis part of performance assessments, the concerns underlying the questions addressed in this session. The procedure embodying this approach is shown schematically on the diagram of Fig. 2. Following a preliminary data acquisition phase, conceptual models of the site, disposal units, and source terms are developed. These conceptual models generally are refined during the site characterization phase and the data acquisition process should proceed until all data necessary to support the numerical models of the overall system are developed. At this stage of the analysis, all data acquired, be they qualitative or quantitative, are defined as unrefined and are associated with uncertainties which should both be evaluated and become part of the data base for the site. It should be realized that this data uncertainty evaluation task is by no means trivial but is the necessary evil for improving our knowledge of the accuracy of our performance assessment results.

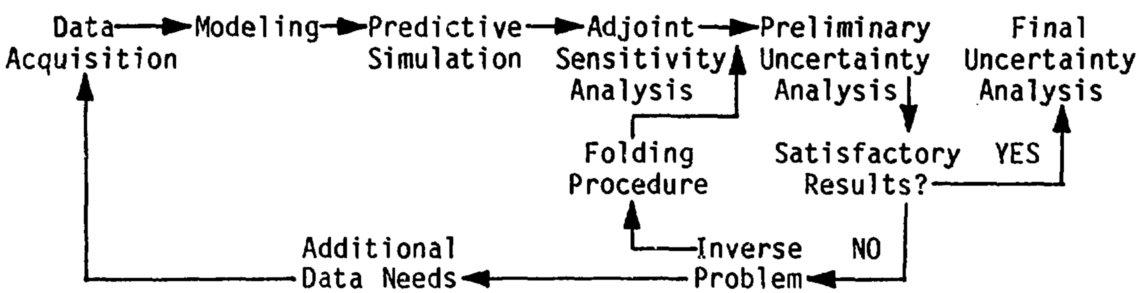

Fig. 2. Flowchart of a complete sensitivity and uncertainty analysis procedure for waste isolation problems.

Once the numerical models necessary to simulate all site specific phenomena of interest have been developed, appropriate verified computer codes are selected (alternatively developed and verified if necessary) to perform the predictive simulations. To this point, little innovation has been proposed. The next step in the procedure outlined in Fig. 2., however, is the one which has traditionally not been performed in large-scale studies because of cost and time constraints. The complete adjoint sensititivity analysis, now feasible using the ADGEN system, serves several purposes. First, it allows one to thoroughly check the numerical codes by quantitatively verifying that the influence of each parameter on the results of the analysis corresponds to what was intended in the conceptual models, thus adding to the verification, validation and reliability of the codes. Second, the complete sensitivity analysis allows a formal (i.e., based on systematic and quantitative results) reduction in scope of the uncertainty problem by separating the parameters that have a proven influence on the results from those that have none. Last and most importantly, it allows one, through a ranking of the products of the sensitivities and the data uncertainties, to identify all leading sources of uncertainty for a given result, thus further reducing the uncertainty problem to a reduced set of "key" parameters to be further considered in the detailed treatment of the problem uncertainties. Note that this reduced set of parameters is an exhaustive list of the parameters which have been formally shown to have an influence on the results of interest and that, as opposed to the qualitative expert opinion used for this task in the conventional approaches, the automated sensitivity analysis procedure has provided a systematic and quantitative methodology for the successive reductions in scope of the uncertainty problem.

At this point in the study, preliminary or partial uncertainty calculations generally make it clear if acceptable uncertainties on the final results are obtainable given the current set of data uncertainties. If satisfactory results can be obtained, then the final uncertainty analyses can be completed at minimum cost using the reduced set of parameters and conventional statistical techniques since these methods, whose costs are directly proportional to the number of parameters considered, are practically feasible only for reduced sets of parameters. If, on the other hand, the results are unsatisfactory either because of their actual value or their associated uncertainties, then the results of the complete sensitivity analysis procedure allow the inverse problem to be formally performed since sensitivities with respect to <u>all</u> parameters have been calculated and <u>all</u> leading sources of uncertainty have been identified. The results of the automated procedure, therefore, clearly point out which parameters are the most critical in interpreting the results of a specific performance assessment in the context of regulatory standards or licensing requirements, and where additional research or data acquisition would be most beneficial for the purpose of improving the assessment results or reducing the overall uncertainty in the calculated results.

Since it is not always possible to improve the quality of the data specifically measured for input to a performance assessment, an alternative method, referred to as a "folding procedure" on Fig. 2, can be used to reduce the overall uncertainty in the calculated results. The methodology[7] allows one to take into account in the system uncertainty analysis any additional knowledge or measured data which may not be explicitly used in the calculational scheme but are related to some physical properties of the system. This folding procedure, described in detail in Maerker et al. (1985), combines the added knowledge of the data with that used in the calculations by means of a generalized least-square adjustment procedure, resulting in the reduction of the uncertainties in the responses as well as in some of the most important parameters.

In summary, the procedure shown in Fig. 2. allows evaluation of all parameters and identification of all major sources of uncertainty in a given problem to support a formal reduction in scope of the uncertainty problem. In this respect, the procedure allows one to avoid "over collection of data" and "over modeling" by focussing research and analysis efforts where needed most. The procedure also allows formal performance of the inverse problem to improve, if necessary, the assessment results and reduce their associated uncertainties. To that affect, use can also be made of an original folding methodology which takes into account in the system uncertainty analysis any additional knowledge of physical properties of the system which have not been explicitly used in the calculational scheme. The new deterministic methodology based on the ADGEN system is totally compatible with, and fully complementary to the conventional statistical methodologies. The conjugated use of these methodologies is recommended in any complete performance assessment study of large-scale systems such as waste isolation facilities.

CONCLUSIONS

A new computer system, named ADGEN, which can generate adjoint solutions of FORTRAN computer codes, has been designed. The system makes use of the EXAP precompiler to automatically calculate partial derivatives in computer codes. The system uses these derivatives to set up the adjoint equations in large matrix form and solves these equations using powerful back-substitution techniques. From the adjoint solution, sensitivities of all results with respect to all data and parameters are efficiently calculated. A simplified version (Version A) of the ADGEN system had been produced and tested on sample problems. Using this version, the feasibility of the overall system for automated generation of adjoint solution of FORTRAN computer codes has been successfully demonstrated. A new deterministic procedure for improving large-scale performance assessment of waste isolation systems has also been presented. The procedure makes use of the ADGEN system to provide, in an automated, systematic and quantitative manner, all the information necessary to support a complete sensitivity and uncertainty analysis. This approach should allow one to better respond to the essential questions of how good and reliable performance assessment results are and, if necessary, how these results can be improved.

REFERENCES

1. Frank, P. M., Introduction to Sensitivity Theory, Academic Press, NY (1978).

2. Tumovic, R. and M. Vukobratovic, General Sensitivity Theory, American Elsevier, NY (1972).

3. Oblow, E. M., "GRESS, GRadient Enhanced Software System, Version D User's Guide," ORNL/TM-9658 (1985). Available from NTIS, 5285 Port Royal Road, Springfield, VA 22161.

4. Oblow, E. M., F. G. Pin, and R. Q. Wright, Nucl. Sci. Eng. **94**, 46-65 (1986).

5. Pin, F. G., B. A. Worley, E. M. Oblow, R. Q. Wright and W. V. Harper, Nucl. Chem. Waste Manag. **6** (1986) (in press).

6. Worley, B. A., R. Q. Wright and F. G. Pin, Nucl. Sci. Eng. **94**, 180-191 (1986).

7. Maerker, R. E., B. L. Broadhead and J. J. Wagschal, Nucl. Sci. Eng. **91**, 369-392 (1985).