An Overview of the IAEA Safety Series on Procedures for Evaluating the Reliability of Predictions made by Environmental Transfer Models*

F. Owen Hoffman

CONF-8710169--3

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

DE88 007174

and

Eduard Hofer

Gesellschaft für Reaktorsicherheit, mbH
D-8046 Garching
Federal Republic of Germany

ABSTRACT

The International Atomic Energy Agency is preparing a Safety Series publication on practical approaches for evaluating the reliability of the predictions made by environmental radiological assessment models. This publication identifies factors that affect the reliability of these predictions and discusses methods for quantifying uncertainty. Emphasis is placed on understanding the quantity of interest specified by the assessment question and distinguishing between: (1) stochastic variability and (2) lack of knowledge about either the true value or the true distribution of values for the quantity of interest. Among the many approaches discussed, model testing using independent data sets (model validation) is considered the best method for evaluating the accuracy in model predictions. Analytical and numerical methods for propagating the uncertainties in model parameters are presented and the strengths and weaknesses of model intercomparison exercises are also discussed. It is recognized that subjective judgment is employed throughout the entire modelling process, and quantitative reliability statements must be subjectively obtained when models are applied to different situations from those under which they have been tested.

* Research sponsored by the International Atomic Energy Agency and the Office of Health and Environmental Research, U. S. Department of Energy under contract DE-AC05-84OR21400 with Martin Marietta Energy Systems, Inc. Publication No. ____, Environmental Sciences Division, ORNL.

MASTER

INTRODUCTION

The International Atomic Energy Agency is preparing a new Safety Series on practical approaches for evaluating the reliability of predictions made by environmental radiological assessment models (IAEA, in preparation). The need for such a document is obvious. Mathematical models are used extensively to evaluate the acceptability of planned or accidental releases of radioactivity. Such models typically produce only a single prediction of concentration, dose or risk to an individual or population group, and until recently, statements about the reliability of these predictions have been merely qualitative.

The new Safety Series publication introduces methods for obtaining quantitative reliability statements. Quantitative statements reflect the uncertainty in model predictions and indicate the extent to which model predictions are expected to either under- or overpredict. When compared to limits of concern (i.e., dose limits or derived concentration limits) quantitative reliability statements provide a measure of confidence with which model predictions can guide decisions and indicate the situations in which this confidence must be improved.

Numerous individuals have contributed to the IAEA document. Special recognition is given to: H. G. Paretzke and R. H. Gardner, who along with the authors served the IAEA as special consultants, and the IAEA Scientific Secretaries, G. S. Linsley and I. Savolainen.

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.

GENERAL CONCEPTS

Factors Affecting Reliability

Five groups of factors are recognized that affect the reliability of model predictions (Fig. 1). Among these, the most important is the formulation of the assessment question (specification of the problem). The assessment question affects the scale of time and space as well as the processes and mechanisms that should be taken into account by the model. Changes in the assessment question may enhance the significance of previously assumed minor processes. Thus, errors and uncertainties associated with previously unimportant processes may suddenly dominate results when the model is applied to a new question. Even slight modifications to an assessment question may cause a model with an outstanding reliability record to grossly mispredict.

The discussion about the issues of quality assurance for calculating and documenting results are discussed only briefly in the IAEA document. Nevertheless, it is recognized that large errors can be introduced by simple typographical and programming mistakes. Unfortunately, quality assurance procedures are frequently compromised in the face of pressing deadlines and emergencies.

Important Distinctions

The evaluation of reliability in model predictions requires a thorough understanding of the assessment question. For each question there is a quantity of interest (e.g., concentration, dose rate, committed dose, committed collective dose, etc.) and an associated reference unit (kg, m³, year, individual, population, etc.). Distinctions are made in the IAEA document about whether the quantity of interest is a stochastic variable or whether it is invariant with respect to the reference unit. This distinction determines whether the assessment question has a probabilistic or a deterministic answer. The

document therefore discriminates between two fundamentally different types of uncertainty: Type A and Type B.¹

Type A uncertainty is due to stochastic variability with respect to the reference unit of the question asked of the model.

Type B uncertainty is due to a lack of knowledge about items that are invariant with respect to the reference unit of the question asked of the model.

The presence of Type A uncertainty in the quantity of interest requires a probabilistic answer to the question. A probabilistic answer represents stochastic variability in the quantity of interest in the form of a distribution.

In practice, deterministic as well as probabilistic answers can be determined only imprecisely because of Type B uncertainties. Type B uncertainties suggest a range, not of variability but of alternative, possibly true deterministic or probabilistic answers to the assessment question. Four possible prediction formats are discussed in the IAEA document depending on the extent to which Type A and Type B uncertainties are present (Fig. 2).

• APPROACHES FOR EVALUATING RELIABILITY

Model Validation

In the IAEA document, the process of testing model predictions against independent data sets is referred to as "model validation." Model validation is considered the best method for evaluating the accuracy in model predictions.

For assessment questions in which the quantity of interest is invariant with respect to its reference unit, a single predicted value is compared with a sin-

¹ In previous drafts of the IAEA document and in papers summarizing its contents (Hofer and Hoffman 1987), the terms Type 1 and Type 2 uncertainties have been used. These terms have been changed to Type A and B to avoid confusion with Type I and Type II errors in statistical hypotheses testing.

gle value which is often an estimate obtained from a number of measurements. Examples are average concentrations or time-integrated concentrations for a specific time period and location. The presence of Type B uncertainty requires that subjective confidence bounds be placed on the model prediction. Statistical techniques are then employed to the confidence that the true value is encompassed by uncertainty placed about the deterministic model prediction.

For assessment questions in which the quantity of interest is a stochastic variable, comparisons are made of the empirical distribution of observed values with the predicted distribution (assuming the use of a probabilistic model to address these questions). The presence of Type B uncertainty requires placement of subjective confidence bounds on the predicted distribution. Statistical techniques are then employed to evaluate the confidence that certain characteristic values (i.e., mean, median, mode, 95th percentile, etc.) of the true distribution are encompassed by the uncertainty placed on those of the predicted distribution.

For both types of questions, the validation process is repeated as many times as necessary to encompass the range of conditions over which the model may be applied. Analyses are made of the potential for the model to over- or underpredict the true values and the extent to which these results confirm quantitative standards of performance that have been assigned to the model (e.g., "the model should not underpredict by more than a factor of three nor overpredict by more than one order of magnitude"). For quantities of interest that vary as a function of time and space, the validation process analyzes the temporal and spatial trends in the over- or underpredictions. These trends indicate the capacity of the model to simulate the dynamics and spatial resolution of a system as well as to produce accurate results for a specific time and location.

Parameter Uncertainty Analysis

Validation data are not likely to ever be available for the full range of conditions in which models may be applied. In the absence of results from validation tests, an analysis of the reliability of model predictions must rely on methods that translate uncertainty in individual model parameters into an

estimate of uncertainty in the model prediction. The IAEA document discusses several methods, among which are variance propagation equations, moment matching equations, and numerical methods involving the use of a computer. For the types of models used in environmental radiological assessment, numerical methods are the most robust. The numerical methods given the most detailed discussion are Simple Random Sampling and Latin Hypercube Sampling. Both methods are "Monte Carlo" procedures .

In the IAEA document, emphasis is given to the treatment of Type B uncertainties as Type A uncertainties are assumed to be properly addressed using a probabilistic model. Among Type B uncertainties, most attention in the past has been given to uncertainty in the estimation of parameter values. As a consequence, present methods are particularly suited for the analysis of the effect of parameter uncertainties on the model prediction.

The steps recommended for conducting parameter uncertainty analyses are as follows:

- 1) List all potentially important uncertain parameters.
- 2) Within the context of the assessment question, specify the maximum conceivable range for each parameter.
- 3) Over this range, specify a subjective probability distribution.

(This step usually requires a considerable amount of judgment, given that data are seldom collected from experiments explicitly designed to address specific assessment questions. The probability distribution is to represent the degrees of belief for the appropriate parameter value to lie within any specific part of the the range specified in step 2).

- 4) Account for correlations between parameters by introducing suitable restrictions, quoting conditional degrees of belief, or by estimating correlation coefficients, respectively.
- 5) Use analytical or numerical procedures to generate a "subjective probability distribution" of predicted values from the joint "subjective probabil-

ity density function" over the combined range of parameter values (Fig. 3).

- 6) Derive quantitative statements ("subjective confidence intervals") representing the effect of parameter uncertainty on the model prediction.
- 7) Rank the parameters with respect to their contribution to the uncertainty in the model prediction. (This step provides incentive for additional research and gives priority to those parameters that have the greatest influence on the uncertainty in the model prediction).
- 8) Present and interpret the results of the analysis.

The results of a parameter uncertainty analysis determine a subjective level of confidence (degree of belief) that the value to be predicted is within a specified range or, that it is in compliance with specified limiting values.

Once a subjective probability distribution of predicted values has been established, the result should lead to one of three basic conclusions:

- 1) At a high subjective level of confidence, the value to be predicted is in compliance with the limiting value (i.e., dose limit, source upper bound, derived concentration limit, etc.);
- 2) At a high subjective level of confidence, the value to be predicted is not in compliance with the limiting value, or
- 3) The subjective levels of confidence for violation of and for compliance with the limiting value are of the same order of magnitude.

Additional studies are necessary to improve the knowledge base for the most important parameters in the model prior to making decisions about compliance with the limiting value.

Analytical and numerical methods for ranking important parameters are discussed in detail in the IAEA document. Among the measures used for ranking are: ratios of variances, simple correlation coefficients, partial correlation coefficients, standardized partial regression coefficients, rank correlation coefficients, partial rank correlation coefficients, and

standardized partial rank regression coefficients (see Hoffman and Gardner 1983, Draper and Smith 1981 and Freund and Minton 1979 for a description of these methods). For complex models, it is recommended that several of these methods be used as no one method provides the best results for all model types and applications.

It is important to note that the results of a parameter uncertainty analysis account for only those uncertainties that have been quantified in the analysis. Therefore, additional terms must be added to the model to represent uncertainty beyond those associated with estimation of parameter values (i.e., suspected structural deficiencies in the conceptual and computational model or bias in the quoted parameter distributions). Different quantitative uncertainty statements and parameter rankings are expected for each question asked of the model. Therefore, different quantitative uncertainty statements and parameter rankings should be associated with the assessment of dose to members of critical populations subgroups than for exposures involving large populations. Differences are also expected for uncertainty statements associated with generic assessments than for assessments that are site-specific. For this reason, the uncritical acceptance of published values of distributions for model parameters should be avoided. Caution must be exercised to ensure that the estimates of parameter uncertainty conform to the conditions defined by the assessment question.

Because subjective judgment must be used to specify ranges and probability distributions for uncertain parameters, the results of a parameter uncertainty analysis will change with the state of knowledge of the experts involved. For this reason, the IAEA document recommends that individuals be consulted who have intimate familiarity with the model, the data available for the uncertain parameters, and the type of question being asked of the model. In the absence of such expertise, the individuals performing the analysis should specify distributions of sufficient width to reflect their lack of knowledge about these parameters and to ensure that their subjective confidence intervals will encompass the true value. The analysis can then be refined through acquisition of additional information when these subjective confidence intervals approach a level of concern (i.e., dose or concentration limit).

Model Intercomparison

Another procedure for evaluating the reliability of model predictions is the intercomparison of different models. This method is often criticized because it offers no measure of accuracy without independent test data. When discrepancies exist between model predictions, the question arises as to which set of results is closest to being correct. This question is extremely difficult to answer without independent test data.

Intercomparison exercises, however, are very useful for identifying obvious errors and weaknesses and for sharing information. They also provide a forum for learning about alternative modelling approaches on how different model structures, parameterizations and numerical characteristics influence predictions. They are especially useful when each of the models has been subjected to a parameter uncertainty analysis and when at least one of the models has been tested against relevant data. Care, however, must be taken to ensure that similarity in the predictions of different models is not simply due to similarity in the data bases for estimating parameter values. It is not at all uncommon to find a variety of different models using similar references for their parameters; thus, despite differences in their structure, their results may be nearly identical (Hoffman et al. 1984).

*

CONCLUSIONS

The IAEA document recognizes that models, at the very best, are only approximations of real systems; therefore, truly "valid" models do not exist. Models can only be considered "valid" when sufficient testing has been performed to ensure an "acceptable" level of accuracy. The level of accuracy considered "acceptable" is determined subjectively and will vary from case to case. Reliability is defined as a measure of confidence in model predictions. This measure of confidence can be expressed quantitatively or qualitatively. Although many of the procedures recommended in the IAEA document are

still in a state of development and refinement, they should facilitate the transition from a qualitative to a quantitative assessment of reliability.

In the absence of relevant test data, or data obtained from experiments designed to answer a specific assessment question, the uncertainty associated with model predictions must be determined. Well established procedures from probability calculus and statistics can be used to propagate subjective estimates of parameter uncertainty into quantitative uncertainty statements about the model prediction. However, these uncertainty statements must still be recognized as being subjective. Quantitative estimates of uncertainty will change depending on the assessment question and the "state of knowledge" of the experts involved.

Once quantitative uncertainty statements are given, the following questions arise:

- 1) Is this uncertainty acceptable?
- 2) What are the most important sources that contribute to this uncertainty?
- 3) Can this uncertainty be effectively reduced?

Quantitative uncertainty statements can be used to defend decisions about the acceptability of potential and actual releases of radioactivity and to indicate priorities for future research. Through such research, the predictive accuracy of environmental transfer models will increase as well as our overall understanding of natural systems.

REFERENCES

- Draper, N. R. and H. Smith. 1981. *Applied Regression Analysis*. John Wiley and Sons, Inc., N. Y.
- Freund, R. J. and P. D. Minton. 1979. *Regression Methods - A Tool for Data Analysis*. Marsel Ekker, Inc., New York, and Basel.
- Hofer, E. and F. O. Hoffman. 1987. Selected examples of practical approaches for the assessment of model reliability - Parameter uncertainty analysis.

sis. In: Proceedings of a Workshop on Uncertainty Analysis for Systems Performance Assessments, OECD/NEA, Seattle, U.S.A., 24th-26th February 1987.

Hoffman, F. O. and R. H. Gardner. 1983. Evaluation of uncertainties in environmental radiological assessment models. In: *Radiological Assessment: A Textbook on Environmental Dose Assessment*. J.E. Till and H. R. Meyer, eds. U. S. Nuclear Regulatory Commission, Washington, D. C. NUREG/CR-3332, ORNL-5968, Ch. 11, pp. 11-1 to 11-55 (1983).

Hoffman, F. O., U. Bergström, C. Gyllander and A.-B. Wilkens. 1984. Comparison of predictions from internationally recognized assessment models for the transfer of selected radionuclides through terrestrial food chains. *Nuclear Safety* 25: 533-546.

International Atomic Energy Agency (draft report 1987-09-29). Procedures for evaluating the reliability of predictions made by environmental transfer models. IAEA, Division of Nuclear Fuel Cycle. (to be published as an IAEA Safety Series).

Figure 1. The five classes affecting the reliability of model predictions

1. Both Type A and Type B uncertainties are negligible

2. Type A is negligible but Type B is not

3. Type B is negligible but Type A is not

4. Neither Type A nor Type B are negligible

Figure 2. Prediction formats accounting for the presence of Type A and Type B uncertainties.

THE STATE OF KNOWLEDGE ABOUT UNCERTAIN
PARAMETERS OF A MODEL IS DESCRIBED BY
SUBJECTIVE PROBABILITY DISTRIBUTIONS F_p OR
THEIR DENSITY FUNCTIONS $f_p = dF_p/dp$

Fig. 3 An illustration of a subjective probability distribution of predicted values that is produced from subjective probability distributions specified for each uncertain model parameter.