ORNL-5507

A Model Evaluation Methodology Applicable to Environmental Assessment Models

D Lynn Shaeffer

ORNL-5507
Dist. Categories UC-11,
UC-41, UC-79p

Contract W-7405-eng-26

HEALTH AND SAFETY RESEARCH DIVISION

A MODEL EVALUATION METHODOLOGY APPLICABLE TO ENVIRONMENTAL ASSESSMENT MODELS

D. Lynn Shaeffer

Date Published: August 1979

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

CONTENTS

ACKNOWLEDGMENTS

The author would like to thank C. T. Garten, Jr., F. O. Hoffman, C. A. Little, C. W. Miller, R. V. O'Neill, and F. S. Sanders for their helpful discussions.

A MODEL EVALUATION METHODOLOGY APPLICABLE TO ENVIRONMENTAL ASSESSMENT MODELS

D. Lynn Shaeffer

ABSTRACT

A model evaluation methodology is presented to provide a systematic framework within which the adequacy of environmental assessment models might be examined. The necessity for such a tool is motivated by the widespread use of models for predicting the environmental consequences of various human activities and by the reliance on these model predictions for deciding whether a particular activity requires the deployment of costly control measures. Consequently, the uncertainty associated with prediction must be established for the use of such models. The methodology presented here consists of six major tasks: model examination, algorithm examination, data evaluation, sensitivity analyses, validation studies, and code comparison. This methodology is presented in the form of a flowchart to show the logical interrelatedness of the various tasks. Emphasis has been placed on identifying those parameters which are most important in determining the predictive outputs of a model. Importance has been attached to the process of collecting quality data. A method has been developed for analyzing multiplicative chain models when the input parameters are statistically independent and lognormally distributed. Latin hypercube sampling has been offered as a promising candidate for doing sensitivity analyses. Several different ways of viewing the validity of a model have been presented. Criteria are presented for selecting models for environmental assessment purposes.

INTRODUCTION

A model evaluation methodology has been developed to provide a systematic framework within which the adequacy of environmental assessment models might be examined. This task is motivated by the widespread use of mathematical models developed as tools for predicting whether various human activities (such as the deployment of nuclear other industrial facilities) might result in violation of policies set forth by the National Environmental Policy Act (NEPA) of 1969 and the various requirements of the Environmental Protection Agency (EPA), the Nuclear Regulatory Commission (NRC), and other Federal and state agencies.

BLANK PAGE

Often these models are utilized without any foreknowledge of the confidence which can be placed in their predictive capabilities. Such a course of action may result in potentially serious consequences. For example, environmental assessments which are overly conservative may result in the needless prohibition of industrial expansion due to the prediction of deleterious effects on humans or the environment. Conversely, irreversible environmental degradation and/or serious threat to human health may result from the decision based on incorrect model predictions to permit certain industrial expansion. A knowledge of the uncertainty in the model predictions may have the effect of reversing either of these decisions. Therefore, it is of vital concern that the uncertainties associated with the predictions of environmental assessment models be identified and also quantified if possible.

A systematic approach is needed to effectively accomplish this task of model evaluation because of the complexity of the various subsidiary issues which must be addressed in answering the question of model adequacy. In addition, the diversity of types of existing models, along with their associated data, are not all amenable to one universal method of analysis. Consequently, a variety of approaches is required for scrutinizing the different kinds of models. Although many of the examples provided in the following discussion are radiological assessment models, this methodology is suitable for the evaluation of all assessment models.

The model evaluation methodology depicted in Fig. 1 consists of six major tasks: (1) model examination, (2) algorithm examination, (3) data evaluation, (4) sensitivity analyses, (5) validation studies, and (6) code comparison studies. *Model examination* involves questioning whether anything fundamental was omitted in the initial conceptualization of the model. *Algorithm examination* attempts to determine whether appropriate numerical schemes have been adopted to represent the model in the form of a computer code. *Data evaluation* is intended to ascertain the quality and the quantity of the data available for use with the model. *Sensitivity analyses* are studies directed toward identifying those parameters which are most influential in determining model predictions. The identification of these key parameters enables priorities to be established for performing experiments. *Validation* refers to the attempt to determine

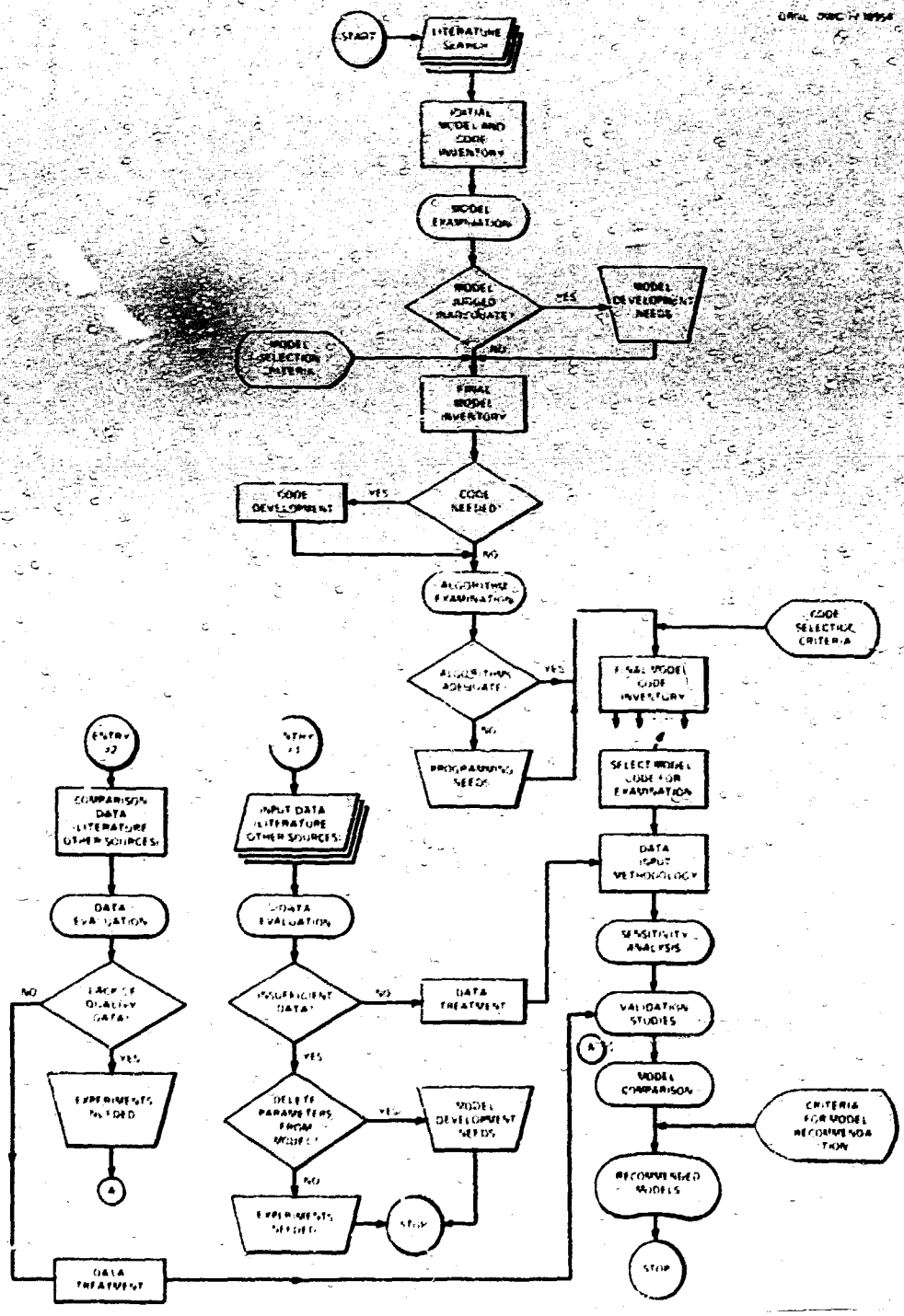

Fig. 1. Flowchart of a methodology for evaluating environmental assessment models.

the degree to which model predictions agree with field or laboratory measurements of those same quantities. *Code comparison* is the process of comparing the strengths and weaknesses of various codes designed to perform similar tasks.

Often there are more than one model or computer code available for application to a particular kind of assessment. In this case the assessor may desire to know the advantages and disadvantages of using one model as opposed to another. As a result, the capabilities of the available models and codes must be compared and evaluated. This analysis requires the existence of a model and code inventory. Hoffman et al. (1977a, 1977b) have performed a survey for models used in environmental radiological assessments. Espey et al. (1978) have tabularized models of potential use in hydrologic assessments. Once an initial inventory of models exists the six major tasks of the model evaluation methodology may be applied in an attempt to provide information which will allow a decision maker to select those models best suited for the application being considered.

MODEL EXAMINATION

Model examination is concerned with determining whether a model adequately represents the phenomena, processes, and actions that are of interest to model users. This determination involves answering such questions as: Are all the relevant physical, chemical, and biological phenomena properly included in the equations? Are the equations representing the various parameters valid within the intended range of applicability of the model? Are the equations mathematically well-behaved within the intended range of model applicability? Are the boundary conditions and initial conditions appropriate? Is the approach to the solution the most appropriate one for the intended model appplication?

Note that the model examination process is aimed at determining the adequacy of the model only over the intended range of model applicability. This range of applicability is defined by the documentation of the model and its associated computer code. This approach differs from that

adopted by those modelers who construct models for the purpose of understanding the fundamental behavior of a system being modeled and who often desire to know the behavior of the model when subjected to abnormal input conditions. From the point of view of the methodology presented here, "abnormal conditions" will mean any condition outside the realm of applicability of the model and will therefore not be of interest.

The model examination procedure may uncover deficiencies in the model. If these deficiences are deemed serious, a recommendation might be made to submit the model to major conceptual modifications. However, often a model may appear to be seriously deficient but nevertheless is not adversely affected by such deficiencies when applied to specific situations. Some models, such as linear systems models and other lumped parameter models, often work quite well for no apparent reason (O'Neill, 1971). If the model being examined is judged to adequately represent the phenomena of interest, the model is selected for inclusion within the final model inventory. Any model included in this final inventory which is not represented by a computer code, but for which one is needed, must then enter the code development stage. On the other hand, if the model is already represented by a code, then the model is subjected to the algorithm examination stage.

The manner in which the input data are used in the sensitivity and validation studies will depend upon the kind of model being evaluated. Three different kinds of environmental assessment models have been identified. These are the multiplicative chain models (MCMs), linear systems models, and complex models. A multiplicative chain model may be defined as one describable by a mathematical relationship of the form

$$z = A \frac{x_1 x_2 \cdot \cdot \cdot x_m}{y_1 y_2 \cdot \cdot \cdot y_n}, \qquad (1)$$

where z might be a pollutant concentration or a measure of impact on health (such as dose), A is a constant, and the x_i's and y_i's represent measurable environmental quantities such as concentration. Also, the ratio x_i/y_i might be a bioaccumulation factor or a concentration

ratio* (Vanderploeg et al., 1975). The MCMs are frequently used to estimate the transport of pollutants through terrestrial and aquatic environments. In the radiological field, the computer code FOOD (Baker et al., 1976; Baker, 1977) is an example of the former and the computer code ARRRG (Soldat et al., 1974) is an example of the latter.

The linear systems models are characterized by a system of first-order, linear, coupled differential equations which describe transfers with time of mass or energy from one environmental compartment to another. The coefficients in the differential equations represent transfer rates which are generally assumed to obey first-order kinetics. The solution of a linear systems model describes the content (or concentration) of a given compartment as a function of time. Sometimes this kind of model may be solved analytically by resorting to the techniques of ordinary differential equations or Laplace transforms. More frequently, however, the model must be solved numerically on a computer. The existence of convenient software packages, such as CSMP (IBM, 1975), i.e., Continuous System Modeling Program, eliminates the need for developing detailed numerical schemes of solution and therefore makes the programming task easier. An example of a linear systems model which describes the transport of radionuclides through a terrestrial environment is the computer code TERMOD (Booth et al., 1971; Booth and Kaye, 1971).

Finally, a complex model is defined as one which does not fall into either of the other two categories. These may be, but not necessarily, characterized as nonlinear models, distributed system parameter models (Polis and Goodson, 1976), or models for which there exists no tractable analytic solution relating the inputs to the outputs. Examples in the radiological field are the computer codes ADPIC (Lange, 1973), AIRDOS (Moore, 1977) and SERATRA (Onishi 1977a, 1977b).

---

*The bioaccumulation factor is defined as the ratio of the concentration per unit weight of a substance contained in one entity (e.g., fish) to the concentration per unit weight or volume in a different entity (e.g., water).

ALGORITHM EXAMINATION

Algorithm examination involves questioning whether the numerical technique selected for the computer code is the most appropriate one for the model application and whether there are any inherent numerical problems. Care must be exercised to ensure that the numerical solution obtained is a unique one. There are many codes which are capable of giving almost any answer depending on how various parameters governing the execution of the routine are adjusted. Special problems occur with techniques designed for solving differential equations. For example, some methods are known to produce numerical dispersion. Conservation of mass and numerical stability are problems which frequently require meticulous attention and often cannot be satisfied simultaneously. Special methods are generally required for solving differential equations which are "stiff," that is, where the eigenvalues range over many orders of magnitude because of numerical instabilities which are often encountered in some of the simpler methods such as the Runge-Kutta algorithm.

Selection of an algorithm may involve trade-offs. Is an increase in computational time worth the gain in accuracy? Some numerical techniques commonly used in the solution of partial differential equations (PDE) have been extensively studied, and their associated pitfalls have been documented (Long and Pepper, 1976). The finite element method (Mitchell, 1977; Pinder and Gray, 1977) for solving PDEs has come into use over the last few years and appears to be superior to the finite difference method in many applications (Lee et al., 1976; Vinokur, 1976; Long and Pepper, 1976).

The final code inventory consists of those codes which have been developed in the code development stage and those codes which were judged adequate in the algorithm examination stage.

DATA EVALUATION

Data Quality

The meaningfulness of the sensitivity analyses and validation studies will be determined by the quality of the available data. Two

kinds of available data are needed. One kind is the data needed for input to the computer codes. The other kind is the data (i.e., comparison data) to which the predictions of the models must be compared. The aim of the data evaluation process is to determine the quality of both kinds of data.

Data quality refers to how well the data represent the processes being measured. The data will be effected by the presence of measurement errors. These errors are of two types, systematic and random. Systematic errors are those associated with the particular instruments or techniques being used. For example, an instrument which is improperly calibrated or zeroed will produce a systematic error. Thus, in a measurement sample, systematic errors will usually enter into each measurement with the same algebraic sign. The elimination of systematic errors from an experiment will depend on the experience and the care of the experimenter.

Random errors occur as a result of unknown variations in the experimental conditions. These errors may result from small errors in the judgment of the observer or from inherent random fluctuation in the processes being measured. This randomness can often be expressed in terms of a probability law, for example, one governed by normal or lognormal statistics. Thus statistical methods can often be relied upon to deal with random errors.

Data quality is expressed in terms of accuracy, precision, and completeness. Measurements which have small systematic errors are said to have high accuracy. Statistical measures of central tendency (e.g., mean, median, and mode) relate to the accuracy of the data. If the random errors are small, then the data have high precision. Statistical measures of dispersion (e.g., variance, standard deviation, coefficient of variation) relate to the precision of the data. Completeness of the data is important for determining reasonable estimates of the measures of central tendency and dispersion. The larger the number of measurements in a sample, the better will be the estimates of these quantities.

More often than not, the model developer and model user do not have direct access to the data. The data may have been massaged and passed through several hands before reaching the user. Frequently the data are

be available only in the literature. Information is usually lost and the data may have suffered degradation in this data handling and transmittal process. The model user should not use data blindly but should attempt to ascertain all that is known about the quality of the data.

Data handling may fall into four different categories, data collection (i.e., the experimental stage), data reduction (mathematical manipulation), data presentation (journal articles, reports, etc.), and data usage. In the context of these four categories, the following considerations should be given to evaluating data.

With regard to data collection the following questions should be asked. Are the experimental methods acceptable? Have data been measured for the range of conditions for which the model was designed? Is there any indication that systematic errors are present or that blunders have been committed? Is there a significant bias in the data resulting from an experimental detection limit? Are field conditions realistically replicated in laboratory experiments? For example, there exists evidence in the literature showing that vegetation-to-soil concentration ratios for plutonium and americium measured in the laboratory may differ by as much as four orders of magnitude from actual field measurements (Romney et al., 1976; Schulz et al., 1976). Hydrologic dispersion effects measured in the laboratory tend to be significantly different from field measurements (Schwartz, 1977; Staley et al., 1977). Atmospheric deposition velocity measurements under laboratory conditions have not successfully duplicated field situations because of the difficulty of simulating meteorological and biological phenomena (Heinemann et al., 1976).

Concerning data reduction, the following questions are important. What was the quality of the data used to determine an appropriate form for the probability density function? What was the sample size? Have proper tests of significance been applied? Was a frequency distribution derived from just a few data points? What data were thrown away? What are the confidence intervals? How were data averaged? There appears to be confusion in the minds of many researchers concerning the difference between the arithmetic mean and geometric mean (median) when dealing with lognormal statistics (Aitchison and Brown, 1969). The former is

always larger than the latter. The difference in the two may be large depending on the variance of the data. The statistics of data are sometimes abused when statistical quantities are combined to form algebraic relationships. Atchley et al. (1976) have revealed some problems involved with forming the ratio of statistical quantities. In addition, a significant bias may occur when forming the ratio of two variables having censored probability density functions resulting from experimental detection limits (Shaeffer and Little, 1978). This bias may lead to the calculation of an average value for the ratio which is significantly different from the true average of the ratio.

Data presentations in journal articles, reports, etc., are sometimes very ambiguous or misrepresentative of other data. The author may fail to indicate whether a datum is the result of a single measurement or the average of a number of measurements. Sometimes a mathematical representation of a frequency distribution is given without showing the data. In other cases data are presented without any statement about error bars and the uncertainty of the measurements. Additionally, a researcher may justify the use of a certain value for a particular parameter with the citation of a reference in which either there is no such parameter value or there is a lack of evidence that the cited value is applicable to the researcher's purpose. Such inaccuracies need to be clarified before a judgment can be made about the quality of the data.

Data usage is often inconsistent. Information contained in a data base for a computer code should be cross-checked with other data bases and with data in the literature. Not only can copying mistakes be made but sometimes data are used in ways for which they were not intended. An attempt should also be made to discern whether a model developer may have a bias toward particular data. Sometimes a model is developed at a particular installation in such a way as to take advantage of data which are collected at that installation. Other institutions may have different kinds of data available and therefore adopt a different modeling approach. This kind of situation may lead to the development of conflicting philosophical approaches to modeling. In addition, researchers sometimes become emotionally attached to their own models, a condition which obscures objectivity and hampers the attempts of others to properly evaluate such models.

Finally, if data required by a particular model are unavailable, then a recommendation that experiments be performed to gather the appropriate data may be required. Without such data the model may have to be rejected in favor of a less desirable model, but one for which all the necessary input data exist. On the other hand, if a less desirable model does not exist and if these experiments prove to be prohibitively expensive, then the model may have to be redesigned to eliminate those parameters for which data do not exist. If the elimination of all these parameters is not possible without destroying the model itself, then perhaps the model should be preserved by restructuring it to require the minimum number of parameters for which data are nonexistent. Hopefully, the redesigning of the model will not render it inapplicable to those situations for which it was originally developed.

Statistical Treatment and Description of the Data

The amount of data available will often determine how the data are described and subsequently used to perform sensitivity and validation studies (Fig. 2). The availability of statistical information sometimes permits the development of a probability density function (p.d.f.). If a range of values is available for the data but not enough data points exist to define a p.d.f., it may be possible to assume a p.d.f. and calculate a mean and variance from the given data. Most environmental data may be described by either a lognormal, Weibull, or normal distribution function (Pinder and Smith, 1975; Speer and Waite, 1975; Toy and Lindeken, 1975). Pinder and Smith (1975) investigated the frequency distribution of ^{137}Cs concentrations in soils, plants, and animals for thirty-three different cases and found that the lognormal was the best fit of the three p.d.f.'s mentioned above in about 55% of the cases. The Weibull distribution function, which occurred in 36% of the cases, was the next most prevalent fit. Apt (1976) fit the Weibull distribution function to the temporal and spatial distributions of atmospheric radioactivity data. Shaeffer and Hoffman (1979) have shown that short-term averages of deposition velocities for ^{131}I appear to be lognormally distributed. Ellett and Brownell (1964) conclude that the gamma

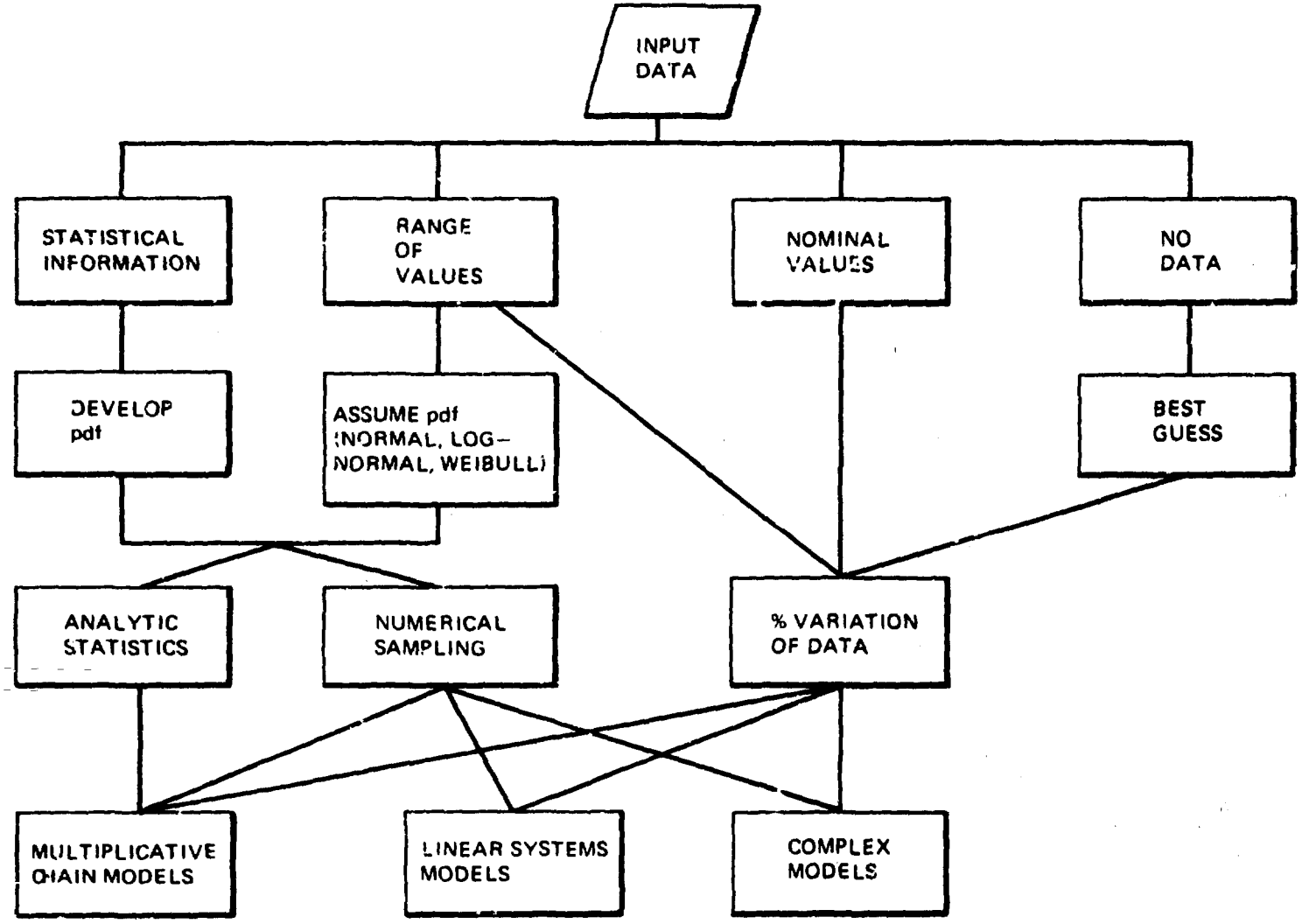

Fig. 2. Schematic representation of a methodology for inputting parameter values, supported by varying quantities of data, into three basic types of models.

distribution is better than the lognormal as a model for the frequency distribution of concentrations of ^{137}C in muscle samples. However, Eberhardt and Gilbert (1973) show that in food chain kinetics there is no practical difference between the lognormal and gamma distributions.

Occasions may arise in which there are insufficient data to develop a p.d.f. and insufficient information about the characteristics of an environmental parameter to permit the adoption of any particular p.d.f. Tiwari and Hobbie (1976b) used a method based on the principle of maximum entropy to develop p.d.f.'s when only a limited amount of information is available. This procedure leads to a p.d.f. which agrees with the available data but which is maximally noncommital to missing information.

SENSITIVITY ANALYSIS

The purpose of a sensitivity analysis is to determine the amount of variation in one or more model output quantities caused by a variation of a model input parameter. This kind of study makes possible the identification of those key parameters which are most important in determining model predictions. These key parameters are the parameters for which the greatest accuracy and precision are needed to reduce the uncertainty in model predictions. Thus experiments might be recommended because of a lack of data for a key parameter and/or because the data manifest large imprecision. Also, a model might be simplified mathematically if a sensitivity study reveals that model predictions are unaffected by a particular input parameter.

Two general approaches, analytic and numerical, may be taken in a sensitivity analysis. The analytic approach is possible only if the exact mathematical relationship between the inputs and the outputs of a model are known. However, even if this functional relationship is known, the analytic approach may not be a tractable one. In the analytic approach, the functional dependence of the output quantities on the input variables is examined to ascertain the strength of this dependence. Perturbation theory (Tomovic, 1963) is the usual analytic methodology adopted. Aoyama et al. (1977) have applied perturbation theory to a sensitivity analysis of a model for predicting the transport to the

ocean surface of radioactive waste disposed deep in the ocean. An analytic statistical treatment is possible for certain simple models (Bevington, 1969; O'Neill, 1971). In the next section a method is developed for investigating error propagation through models involving a multiplication of several factors which are statistically independent and lognormally distributed.

If an analytic approach is not possible, then a numerical approach must be adopted. The numerical procedure may be either statistical or nonstatistical. The most common method used is the nonstatistical procedure of varying the value of an input parameter by a certain percentage on either side of a nominal value while holding all other parameters constant (Miller et al., 1976; Vithayathil and Hirsch, 1975). Values of a sensitivity coefficient are then calculated to express the sensitivity of an output quantity to each input quantity.

In order to calculate a sensitivity coefficient, a fractional deviation D is first determined,

$$D_{ij} = \sum_{k} \frac{y_i(x_j, t_k) - y_i(x_j^o, t_k)}{y_i(x_j^o, t_k)} , \qquad (2)$$

where

D_{ij} = sum of the fractional deviations over the discrete times t_k of the output variable y_i due to a perturbation in the input variable x_j. (Time independent outputs will not require the summation over time);

$y_i(x_j, t_k)$ = value of the ith output variable at time t_k due to the perturbed input variable x_j; and

$y_i(x_j^o, t_k)$ = value of the ith output variable at time t_k due to the unperturbed input variable x_j^o.

The relative sensitivity coefficient is then defined as

$$S_{ij} = \frac{\partial D_{ij}}{\partial u_j} , \qquad (3)$$

where u_j is related to the perturbation in the nominal value of the j^{th} input by

$$x_j = x_j^o(1 + u_j) \quad . \tag{4}$$

Equation (3) can be rewritten as

$$S_{ij} = \frac{\partial D_{ij}}{\partial x_j} \frac{\partial x_j}{\partial u_j} \tag{5}$$

$$= x_j^o \frac{\partial D_{ij}}{\partial x_j} \quad . \tag{6}$$

Substitution of Eq. (2) into Eq. (6) gives

$$S_{ij} = \sum_k \frac{x_j^o}{v_i^o} \frac{\partial y_i}{\partial x_j} \quad , \tag{7}$$

where

$$y_i = y_i(x_j, t_k) \quad , \tag{8}$$

and

$$y_i^o = y_i(x_j^o, t_k) \quad . \tag{9}$$

The disadvantages of using this method are essentially fourfold. First, if the range of values for an input parameter x_i is unknown then the percentage variation to be chosen for the study is completely arbitrary. However, a complete understanding of the nature of the model response requires a knowledge of that response at all perturbation

levels over which the parameters are likely to range. Second, a code having a large number of input parameters will require a large number of computer runs when each input variable is individually perturbed while holding all other input parameters constant. Third, misleading results may be obtained when the output is a nonlinear function of the input. Fourth, correlation between the various input parameters is not taken into account.

An attempt has been made by McKay et al. (1976) and McKay and Bruckner (1976) to overcome the second and third difficulties. Their research has involved the use of Latin hypercube sampling (LHS) to study the statistical output of a computer code as a result of statistical inputs. The LHS procedure involves dividing each input variable into equal probability intervals. For each run, one interval from each input variable is sampled according to the p.d.f. of that variable. Thus all input variables are changed simultaneously and each interval of an input is sampled only once.

Computer studies show (McKay et al., 1976, Fig. 2.6) that LHS offers approximately four times the precision afforded by random sampling. In other words the same precision can be obtained by LHS with one-fourth the number of runs required by random sampling. Consequently, the utilization of LHS may result in considerable savings of total computer run time required for sensitivity studies of computer codes which have a large number of inputs and outputs.

A nonparametric partial rank correlation coefficient (PRCC), such as Spearman's (Conover, 1971; Siegel, 1956), may be used as a sensitivity coefficient. This type of coefficient has the advantage of removing the effects of all but one input parameter on a given output variable even though all input parameters are varied simultaneously from one run to another.

Although the correlation coefficient is a concept derived from linear regression analysis, the use of the PRCC appears to work quite well in cases for which the output is a nonlinear function of the input. Difficulties may be encountered when the output has a singularity or when the output is not a monotonic function of the input. Under these

circumstances one may not be able to claim that LHS is superior to random sampling (M. D. McKay, personal communication, 1977).

Neither the nonstatistical sensitivity coefficient nor the PRCC gives an absolute measure of key parameters. Rather, the coefficients relating the various inputs to a particular output must be evaluated according to their relative magnitude. In other words, the relative importance of the input quantities may be determined by ranking the values of the sensitivity coefficients. Sometimes, the use of intuition, occasionally referred to as face validity (Hermann, 1967), is helpful in evaluating whether a coefficient is as large or as small as it should be. Often a particular input is known to be important. A value of the sensitivity coefficient which indicates the contrary may have resulted from a programming error. Thus intuition, or face validity, may be helpful in identifying programming errors through the use of sensitivity coefficients.

Weber (1976), Bohac* et al. (1974) and Miller et al. (1978) have performed sensitivity analyses of the Gaussian plume atmospheric model. Onishi (1977a, 1977b) performed a sensitivity analysis of a stream-flow and an estuary model. Garten (1978) presents results of a sensitivity analysis of the terrestrial model found in Nuclear Regulatory Guide 1.109.

Propagation of Statistics through Multiplicative Chain Models

The fact that the values for many of the input parameters of a model are statistical in nature means that model outputs are frequently statistical quantities. Therefore, model predictions do have uncertainties associated with them. A realistic assessment of the effects of a postulated set of actions such as a release from a nuclear facility on the environment or on humans requires a knowledge of the uncertainties of the predictions of the models designed to simulate these actions. In fact, it would be nice if these uncertainties could be obtained

*For a response to the work of Bohac et al. (1974) the reader is referred to Gifford (1974).

analytically rather than through the use of a computer. Unfortunately, linear systems models and complex models are not amenable to analytically propagating the statistics through the equations. O'Neill (1971) used Monte Carlo computer techniques to study the uncertainty in model predictions of several kinds of models as a result of systematic bias due to model complexity and to measurement errors associated with model parameters. Tiwari and Hobbie (1976a) performed a Monte Carlo simulation of a simple aquatic ecosystem. Gardner et al. (1976) develope a Monte Carlo simulation technique to analyze the statistics of the outputs of linear systems models, given that the model [illegible] describable by a multivariate normal distribution.

However, it is possible to analytically propagate [illegible] statistics through the multiplicative chain models. As discussed above, many environmental parameters do appear to obey lognormal statistics. The statistical concepts developed below have been applied to determining the uncertainties in the model predictions of the concentration of ^{131}I in cow's milk due to transport via the pasture-cow-milk pathway for short-term (hours or days) and long-term (weeks or months) releases from nuclear facilities (Shaeffer and Hoffman, 1979).

The confidence that can be placed in a given parameter, either input or output, is determined from the concept of confidence interval, which is simply a probability statement,

$$P(z_l < z < z_u) = 1 - \alpha , \quad 0 < \alpha < 1 , \tag{10}$$

where z_l and z_u are determined by the confidence level α and all the statistical quantities involved in the determination of z. Equation (10) states that the probability is $1 - \alpha$ that the interval from z_l to z_u includes the value z. [The confidence intervals for the input data and the comparison data are determined in the data treatment stage (Fig. 1); confidence intervals for the model outputs may be determined in either the sensitivity or validation studies.] Equation (10) can also be written as

$$1 - \alpha = \int_{z_l}^{z_u} f(z)\, dz \,, \tag{11}$$

where $f(z)$ is the p.d.f. for z.

For the multiplicative chain model, Eq. (1), where all the input parameters (the x_i's and y_i's) are assumed to be statistically independent and lognormally distributed, z may also be shown to be lognormal (Aitchison and Brown, 1969). Then, by definition, $\ln z$ is normally distributed with mean μ and variance σ^2. The p.d.f., $f(z)$, is defined in terms of μ and σ^2 by

$$f(z) = \frac{1}{\sigma\sqrt{2\pi}z} \exp\left[-\frac{(\ln z - \mu)^2}{2\sigma^2}\right] . \tag{12}$$

The parameters μ and σ are determined by the corresponding parameters μ_{x_i}, σ_{x_i}, and μ_{y_i}, σ_{y_i}, defining the lognormal distributions of the x_i's and y_i's, respectively, in Eq. (1). The relationship between all these quantities is given (Aitchison and Brown, 1969) by

$$\mu = \ln A + \sum_{i=1}^{m} \mu_{x_i} - \sum_{i=1}^{n} \mu_{y_i} \,, \tag{13}$$

$$\sigma^2 = \sum_{i=1}^{m} \sigma_{x_i}^2 + \sum_{i=1}^{n} \sigma_{y_i}^2 \,. \tag{14}$$

Substitution of (12) into (11) gives

$$1 - \alpha = 1/2\ (\mathrm{erf}\ t_u - \mathrm{erf}\ t_l) \,, \tag{15}$$

where

$$t_u = \frac{1}{\sqrt{2}\sigma} (\ln z_u - \mu) , \tag{16}$$

$$t_l = \frac{1}{\sqrt{2}\sigma} (\ln z_l - \mu) , \tag{17}$$

and the notation erf t refers to the error function (Abramowitz and Stegun, 1972), the values of which are tabulated. Equations (15) through (17) permit the calculation of the confidence level associated with a given interval of z.

Frequently there is a need to know the probability P ($z \leq z_u$) that a contamination level will not exceed a certain limit z_u. This limit may be associated with a regulatory standard or with a decision as to what level of ecological or health effect is socially tolerable. Mathematically this probability may be expressed as

$$P(z \leq z_u) = 1/2 \ (1 + \text{erf } t_u) , \tag{18}$$

which is obtained from Eqs. (15) through (17) by setting z_l to zero. The limit z_u is the contamination limit which must not be exceeded. Alternatively, P ($z \leq z_u$) may be considered as the probability of not over-predicting a particular contamination level (e.g., radiological dose).

In assessment work the average or most probable value of each of the parameters involved in a calculation is often chosen. The average, median, and most probable values are all equal for a variable which is normally distributed. However, all three values are different for a variable which has a lognormal distribution. Aitchison and Brown (1969) give formulae for the arithmetic average (mean) $\bar{z}$, the median (geometric mean) z_m, and the most probable value (mode) z_p for a variable z which is lognormally distributed. The arithmetic average $\bar{z}$ is

$$\bar{z} = \exp(\mu + \frac{\sigma^2}{2}) \quad . \tag{19}$$

The median z_m is

$$z_m = \exp(\mu) \quad . \tag{20}$$

The most probable value z_p is

$$z_p = \exp(\mu - \sigma^2) \quad . \tag{21}$$

Which of these three quantities is most appropriate for doing assessment calculations? The answer is obtained by investigating the probability of a contamination level not exceeding each of these three quantities. Substituting $\bar{z}$, z_m, and z_p from Eqs. (19) through (21), respectively, for z_u in Eq. (16) the following result is obtained from Eq. (18):

$$P(z \leq \bar{z}) = 1/2\left[1 + \text{erf}\left(\frac{\sigma}{2\sqrt{2}}\right)\right] \quad , \tag{22}$$

$$P(z \leq z_m) = 1/2 \quad , \quad \text{and} \tag{23}$$

$$P(z \leq z_p) = 1/2\left[1 - \text{erf}\left(\frac{\sigma}{\sqrt{2}}\right)\right] \quad . \tag{24}$$

Thus the probability of not exceeding contamination levels associated with $\bar{z}$ and z_p is dependent entirely on the standard deviation σ.

As the data become more precise and σ decreases toward zero, all three probabilities approach 50%. As σ becomes large the probability of not exceeding the average contamination level approaches 100%. The median (geometric mean) contamination level always gives a 50% probability of not being exceeded. The maximum probability of not exceeding the most probable value is 50%. If σ is very large, then the probability of

not exceeding z_p approaches zero. If a probability greater than 50% of not exceeding a contamination limit is desired then the median and most probable values are inappropriate choices. In fact, depending on the magnitude of σ, a 99% probability may be achievable only for some contamination considerably greater than the mean (Shaeffer and Hoffman, 1979).

In many applications the contamination z is proportional to a concentration χ of the contaminant in a given environmental media. Therefore, $\bar{z}$, z_m, and z_p are also dependent on χ. Consequently, a maximum permissible concentration χ_M is associated with the contamination limit. This χ_M is different for $\bar{z}$, z_m, z_p, and $z_{99\%}$ (the value for which there is a 99% probability of not being exceeded). The relationship between these maximum allowable concentrations associated with z_p, z_m, $\bar{z}$, and $z_{99\%}$ may be shown to be, respectively,

$$\chi_p : \chi_m : \chi_{avg} : \chi_{99\%} = 1 : e^{-\sigma^2} : e^{-3\sigma^2/2} : e^{-(2.326+\sigma)\sigma} . \qquad (25)$$

Thus there can be a large difference between χ_p and $\chi_{99\%}$, depending on σ.

The results of an analysis by Shaeffer and Hoffman (1979), using the above method, show that there is considerable imprecision in the values of parameters used to predict doses to an individual's thyroid resulting from the ingestion of milk contaminated following releases of modecular iodine, $^{131}I_2$, from nuclear power facilities. Consequently, the values of dose predicted with a simple multiplicative chain model exhibit considerable spread. This analysis indicates that if short-term (hours or days) averaged values for the pasture-cow-milk pathway parameters are used to calculate annual doses, then the maximum allowable air concentration associated with the most probable dose must be reduced by a factor of 19 to obtain a 99% probability of not exceeding a dose limit. On the other hand, the use of long-term (weeks or months) averaged values implies that the maximum allowable air concentration associated with the most probable dose must be reduced by a factor of 10 to obtain a 99% probability of not exceeding a dose limit, such as the EPA dose limit of 75 millirems per year to an individual's thyroid (U.S.

EPA, 1976). However, because of the quality of the data available for the analysis, these numbers should be treated with caution. A fuller delineation of this point is given by Shaeffer and Hoffman (1979).

VALIDATION

The validation of models and computer codes has been a subject of considerable discussion in the literature. The issue revolves around the question of whether there is any meaningfulness in an attempt to validate a model. The problem stems from the fact that a model is an abstraction and a mathematical representation of reality. As such a model can never completely duplicate reality under all conditions. In fact some models predict quantities which are not measurable and therefore excluded from possible validation. Also, some models are not meant to be validated and are designed for the purpose of guiding research. However, many models are designed for predictive purposes and often may be compared against historical data. Some of these models are intended to more fully understand fundamental processes, and others exist to predict the consequences of a postulated set of actions.

Mankin et al. (1975), in their treatment of the validity of ecological models from the point of view of set theory, define all models as being invalid but consider a model to be "useful" if it faithfully predicts some of the behavior of the system being modeled. House (1974), in a discussion of models used for forecasting, contends that validation of social science models is an impossibility because of Heisenberg's Uncertainty Principle, which implies that a complete picture of reality is unobtainable. Such an argument is an excursion into the realm of epistemology. Caswell (1977) makes a distinction between corroboration and validation. The former term is applied to models designed to increase understanding of the laws of nature. In this regard Caswell raises the question as to how one can infer the truth of a universal statement from observations of particular phenomena. He maintains that a scientist's efforts should be directed toward invalidating theories rather than the contrary. Caswell also maintains that the important issue to be settled with predictive models is the range of conditions over which they give

meaningful results. Validity, a term applied by Caswell only to predictive models, is a matter of degree and depends on the intended use of the model and the validity criteria imposed by the model user.

The users of models designed for assessment purposes are probably aligned with Caswell's thinking. Modeling for the sake of enhancing the understanding of nature has a need for invalidity in order to allow room for additional fundamental research (or funds) to discover new truths and guide further experimentation. Environmental assessment modeling has a need for validity, in order to provide confidence in the predictions of the impact of technology on man and his environment.

Only those routines or formulae which predict results that are measurable and are not dependent upon inputs unsupported by data can be validated. In other words, predicted results will have meaning only if they depend on experimentally measured parameters. The degree of the meaningfulness may be checked only by comparing the predicted quantities with experimental data. Also, the validity must be determined only for those circumstances for which the model is intended, as stated or implied in the model or computer code documentation. If a model or any of its subcomponents cannot be validated due to a lack of data then the model examination procedure must be relied upon to make a judgment as to the reasonableness of the assumptions and the structure of the model.

Validity Types

The kind of validity concept to be applied to a given parameter will depend upon the amount of data available. Three types of validity are defined here: statistical, deviative, and qualitative validity. These concepts are applicable to single output quantities. If the output is statistical in nature, the accuracy and precision of the predicted and measured quantities should be similar. Conceivably, model predictions could yield a statistical average which is acceptably close in value to the average of the data against which the predictions are being compared and yet exhibit a variance which is unacceptably larger than the variance of the comparison data. The converse situation is

also conceivable. Consequently, two statistical tests are needed, that is, one for the average value and one for the variance.

Various statistical procedures (Zar, 1974) are available for testing whether the accuracy or precision of two different distributions are in agreement. If application of these tests leads to the conclusion that the two distributions are equivalent, then we say that the model is statistically valid. For certain quantities (e.g., air or water concentration of pollutants) of interest to environmental assessments, the predicted quantity x is required to be greater than the measured quantity y, that is, $x > y$. This is typically considered to be a conservative approach. The statistical tests given by Zar can be adjusted to allow for the desired conservatism. The acceptable degree of discrepancy between predicted and measured quantities must be determined by the decision maker.

If sufficient data are unavailable to characterize the output as statistical, a test of deviative validity might be applied. The deviation of a predicted value from a measured value might be evaluated in terms of a deviation coefficient defined as

$$D = \frac{x - y}{x} . \tag{26}$$

A model is said to possess deviative validity if the model passes the test $D \leq \varepsilon_d$, where ε_d is a deviative validity criterion supplied by subjective judgment, such as that provided by a decision maker concerned with compliance to a regulatory statute.

More frequently, however, the concept of qualitative validity is applied. This simply means that the agreement (or disagreement) between model predictions and observed data is rated on a qualitative scale. For example, such a scale might consist of the validity levels: excellent, good, fair, poor, and unacceptable. Such a refined scale may not be possible in reality. Again, the satisfactory degree of validity must be determined by the decision maker. From this discussion comes the conclusion that validity is a matter of degree, depends on the information available, and is subject to the requirements established by a decision maker.

There may be a requirement to select the best model among many which are designed to predict the same quantities under similar situations. In this case the models or codes might be ranked according to their degree of validity. This ranking may be referred to as "comparative validity." In other words, one model or code may appear to be more valid than another. This kind of validity does not indicate whether either model is sufficiently valid. Goodall (1972) provides an excellent discussion of model comparison.

The appropriate test of validity depends upon the output, which in turn depends on the input. The output from a computer code will be statistical in nature if it depends on at least one statistical input parameter. Therefore, a statistical test of validity may be applied to an output quantity only if it is a function of some statistical parameter. If the output is nonstatistical, the deviative and qualitative tests of validity may be applied. If comparison data are nonexistent or if a quantity of interest is not one of the outputs of a model, a validity test cannot be performed.

Global Validity

The validity tests mentioned above are applicable to single output variables. However, there may be occasions when the overall judgment of validity of the code may be desirable. For example, a model may have several output quantities, such as concentration and temperature, as a function of distance from the release point. In any given situation one of the quantities might pass a test of validity and the other might not. If all the outputs of a model pass a validity test for a given set of circumstances, the model is considered to be globally valid. If the model is not globally valid, each output quantity must be separately identified as being valid or invalid.

Validity Consistency

"Validity consistency" refers to the variation of validity among calculations having different input and/or comparison data sets. A

model may be valid under one set of conditions but invalid under another. Here, only the range of conditions for which the model is supposedly applicable, as determined from the model or code documentation, is of concern. This concept is similar to that of model "adequacy" developed by Mankin et al. (1975).

The author is aware of at least one model which gave an excellent fit to some observed data and therefore was deemed valid. However, an alternative test performed under different conditions would undoubtedly have proven the model to be invalid because of the inclusion of some incorrect physical assumptions. The model proved to be valid for one set of conditions only because the incorrectness of the physical assumptions was not important in that situation. Thus, a model proven to be valid for one set of input parameters may be invalid for another. There may exist in the model some unreasonable or refuted theories which have no effect in one predictive situation but which surface in another. Hopefully, some of these problems might be uncovered in the model examination process.

Application of Validity Tests and Concepts

Frequently, environmental assessment models are developed without any validation attempts being made. This is particularly true of large systems models which include biotic as well as abiotic processes and which are often used for the development of environmental impact statements. These models may never be subject to validation because of the lack of understanding on how to perform the appropriate experiments. The veracity of the predictions of these large systems models is judged primarily by intuition and the reasonableness of these predictions. For those models which have been validated to some extent, very few have been subjected to a large number of validation experiments performed under different environmental conditions, that is, very few models have been tested for validity consistency. In addition, many models can only be partially validated because of the limited data available. Examples are the environmental transport models of Eraslan et al. (1977) and the ground-water transport models of Duguid and Reeves (1976) and Reeves and Duguid (1975). Only portions of these models can presently be validated.

The model of Jackson and Yotsukura (1977) for predicting thermal loading of natural streams was found not to be globally valid, due to the fact that the transport (convective-diffusion) aspect of the model gave satisfactory agreement with field data but the decay (surface dissipation) of waste heat did not. Additionally, the Jackson-Yotsukura two-dimensional model for excess temperature was tested for validity consistency and found to give satisfactory agreement with data obtained from several different streams. However, some cases were found for which the agreement was not satisfactory. Yotsukura and Sayre (1976) tested for validity consistency their steady state two-dimensional model for solute concentration in a meandering nonuniform natural channel and obtained excellent agreement between model predictions and measurements made in several streams.

A simple multiplicative model for atmospheric transport (Hanna, 1971; Hanna, 1973) has been applied to several different environmental situations (Hanna, 1975; Koch and Fisher, 1976; Hanna and Gifford, 1977). Miller et al. (1976) have applied a combination of a deviation coefficient, similar to Eq. (2), and a qualitative validity scale to the validation of an ecological system model used to predict pollutant transfer in an aquatic ecosystem. The authors contend that model validation and sensitivity analyses should be an integral part of the model development process. These parallel activities facilitate in guiding an experimental program.

Nappo (1974) presented a methodology for validating air pollution models. This methodology involves the calculation of temporaliy and spatially averaged linear correlation coefficients. Nappo also uses time-averaged and space-averaged ratios of predicted concentrations to observed concentrations of pollution in air.

The correlation coefficient ρ is a useful concept but is not sufficient for determining the accuracy of a model, which Nappo realizes. This fact may be seen from the definition of the correlation coefficient:

$$\rho = \frac{N\Sigma xy - \Sigma x \Sigma y}{[N\Sigma x^2 - (\Sigma x)^2]^{1/2} [N\Sigma y^2 - (\Sigma y)^2]^{1/2}} , \qquad (27)$$

where x is the predicted quantity, y is the measured quantity, N is the total number of predictions (or measurements), and Σ represents the sum over all predictions and/or measurements. If all the data y_i are subjected to a linear transformation of the form $y_i' = ay_i + b$, where a is > 0 and b is any real number, then the correlation coefficient remains unchanged. If a is < 0, the magnitude of r stays the same but the value of r changes sign. These facts simply illustrate that two different sets of model predictions can be linearly transformed with respect to each other and still have the same value for the correlation coefficient. Thus model predictions may be off by a factor of 100 (or any other number) and still produce a statistically significant correlation with the data. The correlation coefficient is therefore a measure of trend, but not of accuracy. Consequently, another test is needed for determining accuracy. Nappo supplies such a test in the form of the averaged ratios of predicted to observed air concentrations mentioned above. However, such ratios may be statistically biased if the data come from a censored distribution of values (Shaeffer and Little, 1978).

Nappo fails to mention that the linear correlation coefficient is derived from Gaussian statistics. This is particularly important in view of the fact that air pollution data are often lognormally distributed (Gifford, 1972). If such a situation is anticipated, the model predictions and the comparison data should all be log-transformed before the correlation coefficient is calculated (Miller, 1978). Additionally, nonparametric correlation coefficients such as Spearman's may be more appropriate than the linear correlation coefficient when the data are suspected of obeying neither normal nor lognormal statistics.

Alternative methods to those of Nappo are such statistical tests as the paired-sample t tests (Zar, 1974). The two-tailed hypothesis could be used to test whether there is any significant difference between model predictions and the observed data. The one-tailed hypothesis could be useful in ascertaining whether model predictions are higher or lower than the values of the comparison data. If the data are suspected of being nonnormal (even if they are log-transformed), then perhaps the Wilcoxon paired-sample test, a nonparametric analogue to the paired-sample t test, would be appropriate. Generally, the performance of more

than one type of statistical test is advisable to ensure that errors of Type I or Type II (Zar, 1974) are not committed. Many other statistical tests not mentioned here are available. In addition, there is no substitute for visually inspecting the data rather than blindly going through a lot of numerical calculations.

CODE COMPARISON

The existence of several computer codes that perform similar tasks may require an evaluation of the relative merits of these codes to enable the user to select which ones might be most appropriate for use in a given analysis. This evaluation will involve comparing the outputs of different codes with each other, including actual values and the general character of any time-dependent behavior. In addition, a comparison should be made of the differences in validity among computer codes. If any of the model outputs are statistical, the differences in the confidence intervals of the predicted quantities must be ascertained. A comparison must also be made of the input data requirements, programming needs (i.e., debugging and algorithm improvement), computer requirements (e.g., special compilers, macros, overlays, double precision, etc.), sufficiency of documentation, and the sufficiency and readability of the printed output. Finally, the time and cost required for execution of each of the codes must be compared. Hopefully, a combination of a knowledge of the structure of the models and codes and the results of the sensitivity studies will reveal the reasons for the differences among the codes and will identify needed programming and modeling improvements.

MODEL SELECTION

Models chosen for environmental assessment purposes should be ones which rely upon available data. Many models include detailed physical, chemical, or biological mechanisms for which there are no data. The usefulness of such models is subject to question. Often certain complexities may be omitted such that the model depends only on available data. A model which lacks input data cannot be validated in its entirety.

A model should also be validated not just for one situation but for many (i.e., a model should be tested for validity consistency). If more than one such model is available, one may have to determine the trade-offs of choosing one model over another. Perhaps more than one model should be selected. Many of these decisions involve scientific and engineering judgment. The simplest model which can be acceptably validated is deemed more suitable than a more complex model. This suggests that an attempt should be made to validate the simplest model first. If this attempt fails then the validity of the models representing the next higher level of complexity should be examined, providing that sufficient data exist to permit such an attempt. The validation procedure should be applied to models of progressively increasing complexity until one is found to satisfy the validity criteria. However, failure to validate a simple model does not guarantee that a more complex model can be validated.

CONCLUSIONS

A generalized conceptual approach to the evaluation of environmental assessment models has been presented. This approach consists of six major tasks: model examination, algorithm examination, data evaluation, sensitivity analyses, validation studies, and code comparison studies. Each of these tasks is helpful in determining the capabilities of a model. Data evaluation is perhaps the most neglected step. Often the experimentalist does not report all the errors associated with a given measurement and/or the user will place more weight on the accuracy of a model prediction than is warranted by the data on which that prediction depends. The quality of a model prediction is no better than the quality of the data available for the input parameters to that model. Consequently, environmental assessors should attempt to ascertain the quality of the data used in assessment models. Sensitivity analysis is helpful in determining whether the uncertainty in an input parameter has any effect on the variation of a model prediction.

Models are often developed without any thought given to the performance of validation experiments from which the accuracy of model predictions might be determined. Although validation studies cannot

even be attempted on some models, such as the ecological systems models which include biotic and abiotic processes, the literature reveals that too little effort has been devoted to validating models for which such studies are possible. Validation is an important process whereby confidence in the predictive capability of a model might be enhanced. In the event that validation experiments are not possible then an attempt should be made to quantify the uncertainties in the model predictions due to the error associated with each of the input parameters. A methodology was presented above for performing this kind of analysis on multiplicative chain models having parameter values which are statistically independent and lognormally distributed.

Several different ways of viewing the validity of a model have been presented and validity has been seen to be a matter of degree, dependent on the information available and subject to the requirements established by the decision maker. In addition, the validity of a model should be determined by comparing model predictions to data obtained from several different environmental situations. The process of selecting suitable models for environmental assessment purposes should give weight to those models which have a minimum need for unknown parameter values. Finally, the simplest model which can be acceptably validated is deemed most suitable for environmental assessment purposes.

REFERENCES

Abramowitz, M. and Stegun, I. A. (Editors), 1972. Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables. Dover Publications Inc., New York, 1046 pp.

Aitchison, J. and Brown, J. A. C., 1969. The Lognormal Distribution. Cambridge at the University Press, New York, 176 pp.

Aoyama, I., Yamamoto, M. and Inoue, Y., 1977. Evaluation of the radioactive wastes disposal into the deep ocean. Health Phys., 33: 227-240.

Apt, K. E., 1976. Applicability of the Weibull distribution function to atmospheric radioactivity data. Atmos. Environ., 10: 777-781.

Atchley, W. R., Gaskins, C. T. and Anderson, D., 1976. Statistical properties of ratios. I. empirical results. Syst. Zool., 25: 137-148.

Baker, D. A., Hoenes, G. R. and Soldat, J. K., 1976. FOOD: an interactive code to calculate internal radiation doses from contaminated food products. Battelle Pacific Northwest Laboratories, Richland, Wash., BNWL-SA-5523.

Baker, D. A., 1977. User guide for computer program FOOD. Battelle Pacific Northwest Laboratories, Richland, Wash., BNWL-2209, 39 pp.

Bevington, P. R., 1969. Data Reduction and Error Analysis for the Physical Sciences. McGraw-Hill Book Company, New York, 336 pp.

Bohac, R. L., Derrick, W. R. and Sosebee, J. B., 1974. Sensitivity of the Gaussian plume model. Atmos. Environ., 8: 291-293.

Booth, R. S. and Kaye, S. V., 1971. A preliminary systems analysis model of radioactivity transfer to man from deposition in a terrestrial environment. Oak Ridge National Laboratory, Oak Ridge, Tenn., ORNL/TM-3135, 21 pp.

Booth, R. S., Kaye, S. V. and Rohwer, P. S., 1971. Systems analysis methodology for predicting dose to man from a radioactively contaminated terrestrial environment. In: Proc. Third National Symposium on Radioecology, Oak Ridge, Tenn., USAEC CONF-710501-P2 (NTIS, Springfield, Va.), pp. 887-893.

Caswell, H., 1977. The validation problem. In: B. C. Patten (Editor), Systems Analysis and Simulation in Ecology, vol. IV, Academic Press, New York, pp. 313-325.

Conover, W. J., 1971. Practical Nonparametric Statistics. John Wiley and Sons, Inc., New York, 462 pp.

Duguid, J. O. and Reeves, M., 1976. Material transport through porous media: a finite-element Galerkin model. Oak Ridge National Laboratory, Oak Ridge, Tenn., ORNL-4928, 201 pp.

Eberhardt, L. . and Gilbert, R. O., 1973. Gamma and lognormal distributions as models in studying food-chain kinetics. Battelle Pacific Northwest Laboratories, Richland, Wash., BNWL-1747, 95 pp.

Ellett, W. H. and Brownell, G. L., 1964. The time analysis and frequency distribution of cesium-137 fall-out in muscle samples. In: Assessment of Radioactivity in Man, vol. II. International Atomic Energy Agency, Vienna, pp. 155-166.

Eraslan, A. H., Akin, F. J., Barton, J. M., Bledsoe, J. L., Cross, K. E., Diament, H., Fields, D. E., Fischer, S. K., Harris, J. L., Hetrick, D. M., Holdeman, J. T., Kim, K. K., Lietzke, M. H., Park, J. E., Patterson, M. R., Sharp, R. D. and Thomas, B., Jr., 1977. Development of a unified transport approach for the assessment of power-plant impact. Oak Ridge National Laboratory, Oak Ridge, Tenn., ORNL/NUREG/TM-89, 47 pp.

Espey, W. H., Jr., Gloyna, E. F. and Taylor, R. D., 1978. Hydrologic transport of radionuclides. In: Proc. Workshop on the Evaluation of Models Used for the Environmental Assessment of Radionuclide Releases, Gatlinburg, Tenn., DOE CONF-77-09-01 (NTIS, Springfield, Va.), pp. 33-72.

Gardner, R. H., Mankin, J. B. and Shugart, H. H., Jr., 1976. The COMEX computer code documentation and description. Oak Ridge National Laboratory, Oak Ridge, Tenn., EDFB/IBP-76/4.

Garten, C. T., Jr., 1978. Terrestrial food chain transport of transuranic elements. In: C. W. Miller (Compiler), The Evaluation of Models Used for the Assessment of Radionuclide Releases to the Environment. Oak Ridge National Laboratory, Oak Ridge, Tenn., ORNL-5382, 150 pp.

Gifford, F. A., 1972. The form of the frequency distribution of air pollution concentrations. In: Proc. Symp. on Statistical Aspects of Air Quality Data, Chapel Hill, N.C., pp. 173-181.

Gifford, F. A., 1974. Sensitivity of the Gaussian plume model. Atmos. Environ., 8: 870-871.

Goodall, D. W., 1972. Building and testing ecosystem models. In: J. N. R. Jeffers (Editor), Mathematical Models in Ecology. The Symposium of the British Ecological Society, Grange-over-Sands, Lancashire, 23-26 March 1971. Blackwell Scientific Publications, Oxford, pp. 173-194.

Hanna, S. R., 1971. A simple method of calculating dispersion from urban area sources. J. Air Pollut. Control Assoc., 21(12): 774-777.

Hanna, S. R., 1973. Description of the ATDL computer model for dispersion from multiple sources. In: Industrial Air Pollution. Ann Arbor Science, Ann Arbor, Mich., pp. 23-32.

Hanna, S. R., 1975. Urban diffusion problems. In: Lectures on Air Pollution and Environmental Impact Analyses. American Meteorological Society, Boston, Mass., pp. 209-227.

Hanna, S. R. and Gifford, F. A., 1977. Application of the ATDL simple urban dispersion model to Frankfurt, West Germany. In: Proceedings of the NATO/CCMS 8th International Technical Meeting on Air Pollution Modeling and Its Application. 16 pp.

Heinemann, K., Stoeppler, M., Vogt, K. J. and Angeletti, L., 1976. Studies on the deposition and release of iodine on vegetation. Jülich Nuclear Research Center, Jülich, Federal Republic of Germany, Report No. 1287, 58 pp.

Hermann, C. F., 1967. Critique and comment. Validation problems in games and simulations with special reference t: models of international politics. Behav. Sci., 12: 216-231.

Hoffman, F. O., Miller, C. W., Shaeffer, D. L. and Garten, C. T., Jr., 1977a. Computer codes for the assessment of radionuclides released to the environment. Nucl. Saf., 18(3): 343-354.

Hoffman, F. O., Miller, C. W., Shaeffer, D. L., Garten, C. T., Jr., Shor, R. W. and Ensminger, J. T., 1977b. A compilation of documented computer codes applicable to environmental assessment of radioactivity releases. Oak Ridge National Laboratory, Oak Ridge, Tenn., ORNL/TM-5830, 76 pp.

House, P. B., 1974. Diogenes revisited-the search for a valid model. Simulation, 22(10): 117-125.

International Business Machines Corporation (IBM), 1975. Continuous system modeling program III (CSMP III) program reference manual. IBM Corp., Data Processing Division, White Plains, N.Y., SH19-7001-3, 205 pp.

Jackson, A. P. and Yotsukura, N., 1977. Thermal loading of natural streams. U.S. Government Printing Office, Washington, D.C., Geological Survey Professional Paper 991, 39 pp.

Koch, R. C. and Fisher, G. E., 1976. Evaluation of the multiple-source Gaussian plume diffusion model. Geomet, Inc., 50 Monroe St., Rockville, Md., Geomet Report EF-186, 95 pp.

Lange, R., 1973. ADPIC: a three-dimensional computer code for the study of pollutant dispersal and deposition under complex conditions. Lawrence Livermore Laboratory, University of California, Livermore, Calif., UCRL-51462, 60 pp.

Lee, R. L., Gresho, P. M. and Sani, R. L., 1976. A comparative study of certain finite-element and finite-difference methods in advection-diffusion simulations. Lawrence Livermore Laboratory, University of California, Livermore, Calif., UCRL-77539, 6 pp.

Long, P. E. and Pepper, D. W., 1976. A comparison of six numerical schemes for calculating the advection of atmospheric pollution. In: Third Symposium on Atmospheric Turbulence, Diffusion and Air Quality, ERDA CONF-761003-11 (NTIS, Springfield, Va.), pp. 181-187.

Mankin, J. B., O'Neill, R. V., Shugart, H. H. and Rust, B. W., 1975. The importance of validation in ecosystem analysis. In: G. S. Innis (Editor), New Directions in the Analysis of Ecological Systems, Part I. Simulation Council, Inc., LaJolla, Calif., pp. 63-71.

McKay, M. D. and Bruckner, L. A., 1976. On the statistical analysis of an output of a large computer code as a function of the inputs. In: W. L. Nicholson and J. L. Harris (Editors), Proc. First ERDA Statist. Symp., Los Alamos, N.M. Battelle Pacific Northwest Laboratories, Richland, Wash., BNWL-1986, pp. 142-150.

McKay, M. D., Conover, W. J. and Whiteman, D. E., 1976. Report on the application of statistical techniques to the analysis of computer codes. Los Alamos Scientific Laboratory, Los Alamos, N.M., LA-6479-MS, 15 pp.

Miller, C. W., 1978. An application of the ATDL simple dispersion model. J. Air Pollut. Control Assoc., in press.

Miller, C. W., Hoffman, F. O. and Shaeffer, D. L., 1978. The importance of variations in the deposition velocity assumed for the assessment of airborne radionuclide releases. Health Phys., in press.

Miller, D. R., Butler, G. and Bramall, L., 1976. Validation of ecological system models. J. Environ. Manage., 4: 383-401.

Mitchell, A. R., 1977. The Finite Element Method in Partial Differential Equations. John Wiley and Sons, New York, 246 pp.

Moore, R. E., 1977. The AIRDOS-II computer code for estimating radiation dose to man from airborne radionuclides in areas surrounding nuclear facilities. Oak Ridge National Laboratory, Oak Ridge, Tenn., ORNL-5245, 146 pp.

Nappo, C. J., Jr., 1974. A method for evaluating the accuracy of air pollution prediction models. In: Proc. Symp. Atmospheric Diffusion and Air Pollution, Santa Barbara, Calif., pp. 325-329.

O'Neill, R. V., 1971. Error analysis of ecological models. In: D. J. Nelson (Editor), Radionuclides in Ecosystems. Proceedings of the Third National Symposium on Radioecology. U.S. Atomic Energy Commission, Oak Ridge National Laboratory, Oak Ridge, Tenn., CONF-710501.

Onishi, Y., 1977a. Finite element models for sediment and contaminant transport in surface waters. Transport of sediments and radionuclides in the Clinch River. Battelle Pacific Northwest Laboratories, Richland, Wash., BNWL-2227, 103 pp.

Onishi, Y., 1977b. Mathematical simulation of sediment and radionuclide transport in the Columbia River. Battelle Pacific Northwest Laboratories, Richland, Wash., BNWL-2228, 89 pp.

Pinder, G. F. and Gray, W. G., 1977. Finite Element Simulation in Surface and Subsurface Hydrology. Academic Press, Inc., New York, 310 pp.

Pinder, J. E., III and Smith, M. H., 1975. Frequency distributions of radiocesium concentrations in soil and biota. In: F. G. Howell, J. B. Gentry and M. H. Smith (Editors), ERDA CONF-740513 (NTIS, Springfield, Va.), pp. 107-125.

Polis, M. P. and Goodson, R. E., 1976. Parameter identification in distributed systems: a synthesizing overview. Proc. Inst. Electr. Electron. Eng., 64(1): 45-61.

Reeves, M. and Duguid, J. O., 1975. Water movement through saturated-unsaturated porous media: a finite-element Galerkin model. Oak Ridge National Laboratory, Oak Ridge, Tenn., ORNL-4927, 236 pp.

Romney, E. M., Wallace, A., Gilbert, R. O. and Kinnear, J. T., 1976. $^{239,240}Pu$ and ^{241}Am contamination of vegetation in aged fall-out areas. In: Transuranium Elements in the Environment. International Atomic Energy Agency, Vienna, pp. 479-491.

Schulz, R. K., Tompkins, G. A., Leventhal, L. and Babcock, K. L., 1976. Uptake of plutonium and americium by barley from two contaminated Nevada test site soils. J. Environ. Qual., 5:406-410.

Schwartz, F. W., 1977. Macroscopic dispersion in porous media: The controlling factors. Water Resour. Res., 13(4): 743-752.

Shaeffer, D. L. and Little, C. A., 1978. Unpublished.

Shaeffer, D. L. and Hoffman, F. O., 1979. Uncertainties in radiological assessments - a statistical analysis of radioiodine transport via the pasture-cow-milk pathway. Nucl. Technol., in press.

Siegel, S., 1956. Nonparametric Statistics for the Behavior Sciences. McGraw-Hill Book Company, New York, 312 pp.

Soldat, J. K., Robinson, N. M. and Baker, D. A., 1974. Models and computer codes for evaluating environmental radiation doses. Battelle Pacific Northwest Laboratories, Richland, Wash., BNWL-1754, 165 pp.

Speer, D. R. and Waite, D. A., 1975. Statistical distributions as applied to environmental surveillance data. In: Proc. Third Environmental Protection Conference, Chicago, ERDA CONF-750967 (NTIS, Springfield, Va.), pp. 117-143.

Staley, G. B., Turi, G. P. and Schreiber, D. L., 1977. Radionuclide migration from low-level waste: a generic overview. In: Proc. Symp. Management of Low-Level Radioactive Waste, Atlanta, Ga.

Tiwari, J. L. and Hobbie, J. E., 1976a. Random differential equations as models of ecosystems: Monte Carlo simulation approach. Math. Biosci., 28: 25-44.

Tiwari, J. L. and Hobbie, J. E., 1976b. Random differential equations as models of ecosystems-II. Initial conditions and parameter specifications in terms of maximum entropy distributions. Math. Biosci., 31: 37-53.

Tomovic, R., 1963. Sensitivity Analysis of Dynamic Systems. McGraw-Hill Book Company, Inc., New York, 142 pp.

Toy, A. J. and Lindeken, C. L., 1975. The implications of sampling from a log-normal population. In: Proc. Third Environmental Protection Conference, Chicago. ERDA CONF-750967 (NTIS, Springfield, Va.), pp. 144-167.

U.S. Environmental Protection Agency (U.S. EPA), 1976. 40 CFR 190 Environmental Radiation Protection Requirements for Normal Operations of Activities in the Uranium Fuel Cycle, Vol. I. Washington, D.C., 229 pp.

Vanderploeg, H. A., Parzyck, D. C., Wilcox, W. H., Kercher, J. R. and Kaye, S. V., 1975. Bioaccumulation factors for radionuclides in freshwater biota. Oak Ridge National Laboratory, Oak Ridge, Tenn., ORNL-5002, 222 pp.

Vinokur, M., 1976. The relation of finite element and finite difference methods. National Aeronautics and Space Administration, Washington, D.C., NASA CR-2764, 52 pp.

Vithayathil, F. J. and Hirsch, H. A., 1975. Sensitivity analysis of a simulation model of an agricultural ecosystem-some preliminary results. In: Proc. of Conference on Simulation and Modeling, Pittsburgh, Penn.

Weber, A. H., 1976. Atmospheric dispersion parameters in Gaussian plume modeling, part I, review of current systems and possible future developments. U.S. Environmental Protection Agency, Environmental Sciences Research Laboratory, Research Triangle Park, N.C., EPA-600/4-76-030a.

Yotsukura, N. and Sayre, W. W., 1976. Transverse mixing in natural channels. Water Resour. Res., 12(4): 695-704.

Zar, J. H., 1974. Biostatistical Analysis. Prentice-Hall, Inc., Englewood Cliffs, N.J., 620 pp.