Structure, Use, and Validation of the IEUBK Model

Marlin H. Mickle

Department of Electrical Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania

The potential impact of the effects of lead in children is a major concern. Although measurements of lead concentration can be made in a geographic area, it is difficult to predict the effects of this exposure that involve complicated biologic functions. Dynamic mathematical models that can be simulated on a computer provide one method of analysis to facilitate the prediction process. The integrated exposure uptake biokinetic (IEUBK) model is a dynamic mathematical model that has been used for execution on a digital computer. This paper is concerned with the general difficulties in validating a dynamic model of this type. A number of the general pitfalls of validating a model of this type are presented. The illustrations are of a general nature not requiring an understanding of the physiologic effects of lead on children. The concept of validating a model by comparing results to historical data is discussed. A comparison is made with traditional modeling efforts having this form of dynamic model. Also included are general mathematical concepts illustrating potential difficulties with intuitive analyses in calibrating a dynamic model. — Environ Health Perspect 106(Suppl 6):1531–1534 (1998). http://ehpnet1.inehs.nih.gov/docs/1998/Suppl6/1531-1534mickle/abstract.html

Key words: biokinetic, dynamic model, sensitivity, simulation, stability, validation

Overview

This paper is directed at the integrated exposure uptake biokinetic (IEUBK) model primarily in terms of a) structure of the model, b) use as a dynamic predictor, and c) validation.

The references provided in the strategy (1) on model verification, validation, confirmation, and general use are appropriate for the considerations of this monograph. Thus, this paper will not consider the general concepts of modeling and simulation, but rather the aspects of the IEUBK model that can be discussed with respect to form and dynamic and numerical considerations.

Models of this type have been popular for more than 25 years and have been valuable as pedagogic tools. The models may be either linear or nonlinear, but the approach is essentially the same. Some have been successful in predicting results within a limited range of coverage. This is a valid method for solving a difficult problem.

The IEUBK model is typical of the choice of individual contributing factors in an area of concern. These individual factors are modeled with scientific studies to verify a particular result. The individual factors are then connected by an intuitive approach that is quite accurate as to contributions and changes in the state of certain variables. However, the intuitive connection and the calibration of the interactive effects are different in making the model a tool with any accuracy of prediction. This technique was popularized for socioeconomic systems by Forrester in Urban Dynamics (2) and by Meadows et al. in The Limits to Growth (3).

Once the individual contributing factors have been modeled as a function of time, a basis for updating at a designated time interval is programmed into the computer model to allow time to evolve in a predicting mode. This is the most difficult part of the modeling process to understand and to validate. Validation over time usually means choosing a range of years for which the history is known, starting the model with some previous initial conditions and watching the predictions as they evolve according to the previously recorded factual history. If these predictions agree, one assumes that the model must be correct. This possibly is the most significant pitfall of such an endeavor.

The easiest way to describe the difficulty with such models is to state that the sum of the correct parts is not a true whole. It is also at this point that scientific knowledge of the field in which the model is to be used may be of little use. The problems are centered in the mathematical analysis for the calibration and construction of the interconnection and the mathematical basis for the time evolution. The time evolution is a matter of numerical integration and analysis of the dynamics that have been implicitly defined by the data-gathering process and the choice of constants of combination in the interconnection phase of the model.

Independent of the scientific discipline for which the model is being used, the interconnection is similar to an attempt to describe or solve a partial differential equation with the analysis of one of an infinite number of possible solution planes in some solution space. As an example, consider the function of 

\[ f(x,y) \]

in Figure 1.

In Figure 1, one of the 

\[ f(x,y) \]

is determined analytically and verified for the value of \( x_1 \) with the specific \( y \).

At the outset, it may or may not be known that \( f \) is dependent on both \( x \) and \( y \). If the \( x,y \) dependence is known, then the form and parameters of the slopes of the figure are known. Frequently, it is conjectured that a relationship exists, but it does not appear to be significant.

The different slopes illustrated in Figure 1 become major contributors to the results after the variable \( y \) is coupled, usually by an indirect relationship. The ranges must be known and the model variables

---

This paper is based on a presentation at the Workshop on Model Validation Concepts and Their Application to Lead Models held 21-23 October 1996 in Chapel Hill, North Carolina. Manuscript received at EHP 16 January 1998; accepted 28 April 1998.

Address correspondence to M.H. Mickle, Department of Electrical Engineering, University of Pittsburgh, 348 Benedum Hall, Pittsburgh, PA 15261. Telephone: (412) 624-9682. Fax: (412) 724-8003. E-mail: mickle@ee.pitt.edu

Abbreviations used: IEUBK, integrated exposure uptake biokinetic model; TSD, U.S. Environmental Protection Agency’s Technical Support Document.
must be limited to those ranges. Otherwise, the mathematical model may continue to use the relationship for all x and y. The model is valid only if the coupled variables are recognized and the relationship of the slopes (derivatives) is tested against established data. These data are usually not available. The builder then asserts intuitive arguments as to why it does not matter or why it should work.

In some cases, such a variable limitation is obvious. However, the more subtle inflections can be the most significant to the result. This is due to the intuitive constants that multiply derivatives.

In addition, the verification of a relationship is usually measured with respect to the entire domain and range of a function. However, in many applications, it may only be the boundary that actually becomes the dominant value in a solution. This requires a specific value, not an intuitive curve. This is, in fact, always the case in a linear programming solution.

**Sensitivity Analysis**

Validation of the model must be accompanied by a sensitivity analysis of the structure, in particular, the constants of combination. The data supporting the most sensitive parts of the model must be the most accurate. Finding the most sensitive data components of the model can only be accomplished after the model has been constructed. Otherwise, there is no need for a computer model.

Sensitivities to relationships that may be approximate curve fits to a straight line or an exponential curve would not normally be of particular concern in validating a model. However, certain graphical relationships in the references of (7) have additional inflection points with changes in derivatives that may not be supported by any dynamic data and can render the system unstable. Thus, the entire model could be invalid. Another example is in defining a slope for a relationship of range of values identified in a scatter diagram. The inclusion of the range of values included with the boundaries may be perfectly valid, but the resulting slope suggests a derivative that may be a sensitive relationship.

There is no way of knowing whether a particular range is unique for describing the indicated relationships without examining those aspects that influence the slope or derivative. If the model is sensitive to this relationship, then considerably more effort is required on a scientific basis to support the choice of the appropriate slope(s).

Certain relationships may cover two or more parts of the domain/range of a function. The two are then combined in a form that provides a continuous relationship over the entire domain. If the simulation hovers about this connection point, the results are meaningless. Functions such as splines are used to overcome this problem.

**Dynamics of the Model**

Perhaps the most difficult area to understand in the validation of a model and its use as a dynamic predictor is the dynamic behavior of the processes themselves. The difficulty comes in the time evolution [ref. (4), pp A-11–A-12, eqns B-6a–B-6i]. Although these are the established data points because of a particular data-gathering process, they do not necessarily represent the points at which the calculation should be done. In fact, if they are known to represent the process, then there is likely no need for a computer model.

The interrelationships of the structure, sensitivities, and dynamics are a more complicated problem than is typically thought. The interrelationships are often counterintuitive (5). In economics, recursive relationships similar to those of the IEUBK model were determined on the basis of quarterly data inputs, i.e., 3-month intervals. Although the relationships seemed correct intuitively, and the models and results were updated by extensive data on a quarterly basis, they were mathematically unstable after many years of usage. One difficulty was that they were most often used for predicting on a quarterly basis, which was a very short time evaluation from the historical results that were included. This is also demonstrated with many complicated socioeconomic models. An excellent data predictor is the previous data point (6).

As an example, curves that fit static data but show undocumented slopes and inflection points can be viewed as well-documented results for a static analysis for a particular set of conditions. However, the dynamic prediction is based not on the fit of the curve but on the derivatives of the fitted curve throughout the range of the independent variable which appear to be unduly accentuated in the fitted curve.

**Description of the Model**

Consider the current IEUBK model. The primary description of the model occurs in the U.S. Environmental Protection Agency *Technical Support Document* (TSD) (4). These relationships represent a number of the specific aspects of the individual areas of concern. The IEUBK model is first conceptualized in the TSD (4), which describes the structure of the system. Equations of the B-6 series, Appendix B, of the TSD describe the state variables for the structure of the IEUBK model (4). The output equation is described by equations in the B-6 and B-9 series. The combination of the state variables and the structure provides the description of a model that is described by a differential equation.

The output is the rate of accumulation of lead in the blood over time (7). This output is then integrated to obtain the accumulation. All discharges of lead are assumed to be incorporated in the model. Thus, the integral represents all accumulated lead from the model.

After the system has been put in this form, it is a relatively simple procedure to perform a variety of analyses to study sensitivity, stability, etc. However, none of these aspects are discussed in the model description. The results of such analyses may be satisfactory for model usage, but they have not been validated.

In particular is the question of the result of a zero (0) input test. This would be essential in any validation process.

The system described above represents a linear system according to the definitions of linear system theory (8). The coefficients vary with time and represent the application of rate and constants of combination to the connection structure that is described by nonzero inputs. The ability to put the model in this form is of particular advantage because of the many tools available to test the model. Two tests important to the model and the method of simulation are the stability of the system in the linear system sense and the computational stability due to the choice of internal mathematical forms and the method of solution. The solution method involves both the discretization of the time basis and the method of integration. All these factors are independent of the biologic components of the model.

Perhaps the most important analysis is to establish the validity of the constants and conditions applied as part of the calibration of the model in a sensitivity analysis. On viewing the numbers used for calibration, it is highly improbable that small variations in the numbers have a uniform impact on the resulting output. Although unlimited resources would say check and recheck all numbers equally, it is likely that certain of these have more of an
impact than others. The identification of those parameters that have the most impact on the result is a necessary part of any calibration procedure.

Parameter Uncertainty and Roundoff

The performance of the system is assumed to be a continuous function of time with parameters that are known. As the parameters become uncertain, the effects of this uncertainty can be traced through the system response. This can affect both results and stability (8). In addition to the parameters, the method of calculation can introduce errors in the calculations that also affect the system response and stability.

In particular, any programming on a digital computer introduces an automatic discretization that normally introduces a truncation error in any parameter or input. These errors can be analyzed using a floating point standard, typically Institute of Electrical and Electronic Engineers 754-1985 (9). In the case of simulation, these errors can be changed by scaling the time and magnitude of the equations.

These factors may be of little significance, but they cannot be arbitrarily dismissed.

Validation

The validation strategy for the model (1), presents four considerations for validation; (1) scientific foundations of the model structure, (2) adequacy of the parameter estimates, (3) verification, and (4) empirical comparisons.

Item (1) represents multiple levels of inquiry, i.e., (a) a variety of relatively simple input/output relationships, and (b) a basis, method, conservation law, etc., as a means by which the relationships of (a) are interconnected. Each level must be based on scientific principles.

Based on the scope of inquiry and the above expansion of item (1) into items (1a) and (1b), for this author, items (1a) and (2) can only be addressed superficially looking at the sources. The author is not an expert in these areas and can only say that it appears scientific procedure has been followed in establishing the basic model structure and in estimating the parameters, i.e., items (1a) and (2) of (1).

Item (1b) can be more easily addressed in a system theory context, which is primarily an engineering/mathematical problem. The author did not have the computer code for the IEUBK model available at the time of the workshop and cannot comment on item (3).

This paper is fundamentally concerned with the dynamics and interconnection structure (differential equations and their coupling) of the dynamic IEUBK model.

The dynamic IEUBK (4) model component considered is the part of the model that exists between the front end, the uptake component (4), and the probability distribution component (4) of the total IEUBK model (4). As such, it has a single input, UPTAKE (in μg/unit time), and a single output, PBBLOODEND, in μg(lead)/dl (blood volume).

The single input/output relationship is classical in system theory, but may present some difficulty in validation in the sense of item (4) [empirical comparisons (1)] because of the nature of item (1b). These two variables may not be observable in the sense of system theory (8). Neither of these variables can be accessed explicitly with the current IEUBK digital computer simulation model (4).

If the model were totally intrinsic, e.g., force = mass × acceleration, such a discussion would not be warranted. However, it is a necessary validation consideration because the three components of this particular model (4) can be separated with only a single variable coupling the first and second component; likewise, the second and third components are so coupled, it is a necessary validation consideration. This input/output relationship must also be validated in the sense of validation item (4), in order to validate the entire model.

From the material included in the references available to the author, it appears that no validation of this single input/output system has taken place. Considering statistical comparisons (1), with lead sources as inputs to the front end of the model, and statistical comparisons based on the limited observations for final outputs, it is possible to have positive results for validation, with the dynamic portion of the model having little or no effect on the total system. Such a result does not validate the dynamics. In summary, satisfying input/output data for the overall model may be a necessary condition for model validation. However, it is not a sufficient condition for the differential equation (dynamic) portion of the model.

Thus, the approach of this paper addresses items (1a) and (2) of the validation strategy (1) as structural components without biologic considerations.

The author has taken part in a wide range of modeling efforts (10-13) where the models have had the form of the IEUBK model while satisfying items (1), (2), and (3) of the validation strategy (1). One of these efforts involved the prediction of the levels of persons in various occupations based on social mobility and the accompanying social mobility matrix. The researchers were informed of the difficulty in finding a model that would satisfy the input/output relationships that would incorporate known parameters and support the raw data collected over a number of decades. Using the existing data and calibration parameters, a model was developed that fit the data within all established tolerances for all outputs. However, sociologists would not accept the model because of a structure that implied (postobservation) that the reason why married couples would choose to have children was a good economic outlook. The model incorporated the proper mathematical structure and satisfied all data points but was not considered valid.

Any past data validation satisfies the conditions for many (or all) past scenarios. However, to be useful, a model must be predictive on scenarios that have not necessarily been seen in the past. This is the point at which the structure (1b) must be verifiable to place any validity on future results. Otherwise the model can only be useful if every possible scenario has previously taken place or has been entered and tested as an input. Ensuring all possible scenarios have been taken into account is normally only possible for relatively simple discrete models.

Whether the model is correct cannot be fully demonstrated by simply matching input/output data of known tests. This is necessary but not sufficient. Satisfying item (4) at least requires satisfying item (1b). The first stage of the model that produces uptake can be structurally verified more easily because of the algebra. However, the dynamic portion includes differential equations and coupling that cannot be so easily verified as valid structure and interconnection.

The point of this experience is that validation simply begins with predicting the correct past results as in item (4). This is a single point in a much more complex validation procedure, i.e., both items (1a) and (1b) must be validated. For example, sensitivity analysis of the IEUBK model would suggest certain relationships and parameters that have a major impact on the output. This narrows the scope of inquiry at the point in time to check a reasonable number of relationships as
opposed to all of them, which might be prohibitively complicated. This also suggests a deeper meaning to the adequacy of parameter estimates. The model may be robust with respect to many parameters and sensitive to a few. This suggests a multilevel analysis in which item (2) in the validation strategy (J) is a starting point. The same is true for item (1) in finding those structural relationships that are the most sensitive.

All of the above verifications of the model suggest that the data required for validation is much more extensive than the overall input/output data if the model is to be used in a predictive mode where a previous set of parameter conditions has not been experienced. This is necessary for any model that is not empirical. Empirical models are based on input/output data. However, the IEUBK model under consideration by its construction is not empirical.

REFERENCES AND NOTES

1. U.S. EPA. Validation Strategy for the Integrated Exposure Uptake Biokinetic Model for Lead in Children. EPA 540/R-94-039. Washington: U.S. Environmental Protection Agency, 1994.
2. Forrester JW. Urban Dynamics. Cambridge, MA: MIT Press, 1969.
3. Meadows DH, Meadows DL, Randers J. The Limits to Growth: A Report for the Club of Rome’s Project on the Predicament of Mankind. New York: Universe Books, 1972.
4. U.S. Environmental Protection Agency. Technical Support Document: Parameters and Equations Used in the Integrated Exposure Uptake Biokinetic Model for Lead in Children (v 0.99.d). EPA 540/R-94-040. Washington: U.S. Environmental Protection Agency, 1994;7-22.
5. Forrester JW. Counterintuitive nature of social systems. Technol Rev 73:53-68 (1971).
6. Vogt WG, Mickle MH. The Pitt Dynamic Input/Output Model. University of Pittsburgh, Based on the United States Input-Output Economy, 1947, 1958, 1963 and 1967 Tables.

Pittsburgh, PA: University of Pittsburgh, 1977.
7. Marcus AH, Cohen J. Modeling the blood lead-soil lead relationship. In: Lead in Soil Issues and Guidelines (Davies BE, Wixson BG, eds). Northwood, UK: Science Reviews Limited, 1988.
8. Zadeh LA, Desoer CA. Linear System Theory. New York: McGraw-Hill, 1963.
9. Institute of Electrical and Electronic Engineers. A proposed standard for floating-point arithmetic. Computer 14(5):51–62 (1981).
10. Vogt WG, Mickle MH, Aldermeshian H. A dynamic Leontief model for a productive system. Proc IEEE 63:438–443 (1975).
11. Mickle MH, Vogt WG, Trygar TA. Optimization and interaction of an economic and sociological model. IEEE Trans Systems Man Cybernetics 5:167–175 (1974).
12. Mickle MH, Vogt WG. On the dynamics of a class of optimum routing problems. J Franklin Inst 300:7–23 (1975).
13. Mickle MH, Vogt WG. A dynamic sociological system model. Int J Systems Sci 7:1181–1200 (1976).