THE ROUTINE FITTING OF KINETIC DATA TO MODELS:

A MATHEMATICAL FORMALISM FOR DIGITAL COMPUTERS

MONES BERMAN, EZRA SHAHN, *and* MARJORY F. WEISS

From the National Institute of Arthritis and Metabolic Diseases, National Institutes of Health, Bethesda. Mr. Shahn's present address is the Johnson Foundation for Medical Physics, Philadelphia.

ABSTRACT A mathematical formalism is presented for use with digital computers to permit the routine fitting of data to physical and mathematical models. Given a set of data, the mathematical equations describing a model, initial conditions for an experiment, and initial estimates for the values of model parameters, the computer program automatically proceeds to obtain a least squares fit of the data by an iterative adjustment of the values of the parameters. When the experimental measures are linear combinations of functions, the linear coefficients for a least squares fit may also be calculated. The values of both the parameters of the model and the coefficients for the sum of functions may be unknown independent variables, unknown dependent variables, or known constants. In the case of dependence, only linear dependencies are provided for in routine use. The computer program includes a number of subroutines, each one of which performs a special task. This permits flexibility in choosing various types of solutions and procedures. One subroutine, for example, handles linear differential equations, another, special non-linear functions, etc. The use of analytic or numerical solutions of equations is possible.

INTRODUCTION

Physical and mathematical models provide a useful technique for the study of biological systems. Although no formalism has been developed as yet for the building of models, one may distinguish several stages in this process: (*a*) choice is made for type of model, (*b*) degree of complexity of model is defined, (*c*) values for the parameters of the model are calculated, (*d*) judgment is made whether model is compatible with the data, and (*e*) model is revised when it is inconsistent with the data.

This paper deals chiefly with the third of the above outlined stages, namely, the derivation of values for model parameters to fit the data. This can be a considerable

task even for a relatively simple model, and only through the use of high speed computers has it become possible to treat more complex models. In an accompanying paper (5) some of the other stages of model building are discussed for a special class of models known as linear compartmental systems.

Originally, the methods described in this paper were developed for the solution of linear compartmental systems, such as are encountered with isotope tracer experiments (2–4, 9). They have been extended since to include a variety of linear and non-linear systems, and also serve as a link in a broader formalism for model building which is under development.

The procedures described are sufficiently flexible to accept fragmentary or "soft" data, the type frequently encountered in biological experiments. They also permit the pooling of data from several experiments and other sources into a single mathematical framework so that a solution may be obtained that is at once compatible with all the information available on the system.

The computer program developed for this was written in FORTRAN (1) and has been compiled[1] for routine use on an IBM 7090 computer[2] having a 32,000 word storage capacity. It includes many subroutines which permit various options on the types of problems and methods of solution. At present it is possible to treat simultaneously twenty-five separate functions that may arise as responses of a system having up to fifty-five parameters of which a maximum of twenty-five may be variable.

In principle, the program can be adapted to computers having a smaller storage capacity than the one employed here. This, however, would require a fair amount of reprogramming and may also result in a reduction in the complexity of the systems that could be treated. As may be apparent from the accompanying paper (5), the latter could be a serious limitation in the development of models.

The statistical procedures employed are based mostly on reference (7). A more complete discussion of the principles has recently been presented by Box (11), who also describes a computer program having a number of features similar to the ones described here.

When using the computer program, the data and model are entered in a certain format and the type of solution required is specified. A number of solutions (linear differential equations, sums of exponentials, etc.) are available for routine use, and others may be added as special subroutines when desired. The types of problems or models that can be treated are limited to those that can be described and solved mathematically. The solutions may be either analytical or numerical.

---

[1] Binary program decks can be made available to potential users. Detailed instructions for the preparation of data for automatic processing, and a description of the computer printout are available with the program deck.

[2] We wish to take this opportunity to thank the National Bureau of Standards for the use of their computer, the computer personnel for their cooperation, and Mr. A. Beam for many assists.

SYSTEMS TREATED

The computer program is applicable to physical or mathematical models which can be described by a discrete number of parameters x_i and for which a number of response functions

$$f_j(t) \equiv f_j(x_1, \cdots, x_m, t) \qquad (j = 1, 2, \cdots, n) \qquad [1]$$

to a set of initial conditions may be specified.

In the case of linear compartmental systems, for example, the function f_j is specified by a set of differential equations (2):

$$\frac{df_j(t)}{dt} = \sum_{i=1}^{n} \lambda_{ji} f_i(t) \qquad (j = 1, 2, \cdots, n) \qquad [2]$$

and the λ_{ji} are the parameters of the system equivalent to the x_i in equation [1].

The experimentally measured quantities, $Q_k(t)$, for such a system are usually a linear combination of the functions f_j:

$$Q_k(t) = \sum_{j=1}^{n} \sigma_{kj} f_j(t) \qquad (k = 1, 2, \cdots, l) \qquad [3]$$

where the σ_{kj} are time independent coefficients either known or unknown. For the special case where the measured quantity is proportional to the function f_k, this reduces to

$$Q_k(t) = \sigma_{kk} f_k(t).$$

Given a set of observed quantities $Q_k(t)$, together with the set of equations describing the model, and a set of initial conditions, the program is designed to search for the values of the parameters x_i and the coefficients σ_{kj} which will yield a least squares solution for the observed quantities. This is accomplished by an iterative procedure requiring initial estimates for the independently variable x_i only.

In general, the x_i and σ_{kj} may be independent variables, dependent variables, or fixed. When the x_i and σ_{kj} are independent variables, their range of variation can be specified; when they are dependent variables, only linear dependence relations may be specified routinely, but special dependence relations may be added when desired.

SOLUTION OF SYSTEM EQUATIONS

Given a model, initial conditions, and initial estimates of the values of the parameters x_i, it is first necessary to determine the functions $f_j(t)$ of [1]. This is accomplished in the computer program by using an analytical or numerical procedure depending on the kinds of functions and on the methods available for their determination. A separate subroutine is employed for each kind of function and the proper subroutine is specified by a code number which forms part of the input of each problem. Values of $f_j(t)$ corresponding to each experimental datum are obtained.

Some of the subroutines used routinely in the program include:

(*a*) Linear differential equations: solved numerically using a 4th order Runge-Kutta method (6).

(*b*) Sum of exponentials: solved analytically.

(*c*) Radiation survival of mixed cells populations: solved analytically.

(*d*) Non-linear binding problem (special): solved numerically using 4th order Runge-Kutta method.

(*e*) Sum of Gaussians: solved analytically.

Other subroutines may be added as required.

When the measured quantities $Q_i(t)$ are linear combinations of functions, [3] applies:

$$Q_i(t) = \sum_{j=1}^{n} \sigma_{ij} f_j(t) \qquad (i = 1, 2, \cdots, l) \qquad [4]$$

Once values of the $f_j(t)$ are available, the problem reduces to finding a set of coefficients σ_{ij} that will yield a best fit to the data.

Since the data correspond to measurements at specific times t_k, [4] may be indexed for each k:

$$Q_{ik} = \sum_{j=1}^{n} \sigma_{ij} f_{jk} \qquad (i = 1, 2, \cdots, l) \qquad [5]$$

If all the data are arranged in a linear array and identified by successive values of k, the subscript i may be dropped and [5] may be redefined in terms of new subscripts

$$Q_k = \sum_{j=1}^{h} \sigma_j f_{jk} \qquad (k = 1, 2, \cdots, m) \qquad [6]$$

where m is the number of experimental observations and h is the total number of the original σ_{ij}. One may solve for the unknown σ_j in terms of the Q_k and the known σ_j by rearrangement of [6], yielding the set of *equations of condition* [7]:

$$\sum_{j=1}^{r} \sigma_j f_{jk} = Q_k - \sum_{j=r+1}^{h} \sigma_j' f_{jk} \qquad (k = 1, 2, \cdots, m) \qquad [7]$$

where the σ_i' represent the known σ_j, and r is the number of unknown σ_j.

The σ_j can be determined by linear regression analysis (7). Using matrix notation the set of *normal* equations generated from the equations of condition is

$$(f^T w f)(\sigma) = (f^T w Q) \qquad [8]$$

where σ is a column matrix of the σ_j, f is an $r \times m$ matrix of the f_{jk}, f^T is the transpose of f, Q is a column matrix of the elements

$$\left(Q_k - \sum_{j=r+1}^{h} \sigma_j f_{jk}\right),$$

and w is a diagonal matrix in which the element w_{kk} represents the relative statistical weight of the k^{th} observation.

The least squares solution for σ follows from [8]:

$$\sigma = (f^T w f)^{-1}(f^T w Q) \qquad [9]$$

The matrix $(f^T w f)^{-1}$ multiplied by the weighted sum of squares of the fitted data and divided by the number of degrees of freedom in the data is the variance-covariance matrix for the unknown σ_j, conditional on the specified values of the x_i. The square root of the k^{th} diagonal elements is the standard error for the k^{th} variable. The correlation coefficient between the i^{th} and j^{th} variable is given by the square root of the ratio of the i,j^{th} element to the product of the i^{th} and j^{th} diagonal elements.

The variance-covariance matrix of the σ_j is modified when the x_i are also variable.

When the number of equations of condition are inadequate to permit a solution of the unknown σ_{ij}, the resulting normal equations are singular or ill conditioned (near singular) and a unique solution for all the σ_{ij} is not possible. The addition of more equations of condition may not improve the situation if they are "nearly" dependent on the already available ones. In such cases it is necessary to introduce additional assumptions about the σ_{ij} to permit a solution.

If some σ_{ij} are linearly dependent on x_k and/or other σ_{lm}, a "dependence" subroutine in the program substitutes for the dependent variables and solves the set of equations in terms of the independent variables only. Certain non-linear dependencies may also be treated if the dependence subroutine is properly modified.

The uncertainties for the dependent variables may be obtained from the variance-covariance matrix of the independent variables. If the dependent variable u is expressed as a linear combination of independent variables (x_i);

$$u = l_1x_1 + l_2x_2 + \cdots, + l_nx_n \qquad [10]$$

the variance of u is (7)

$$\text{VAR } u = l^T A l$$

where A is the variance-covariance matrix of the x_i and l is a column matrix of the l_i.

CORRECTIONS FOR INITIAL ESTIMATES

The solution of the functions $f_i(t)$ and the linear coefficients σ_{ji} permits the calculation of a theoretical value $F_i(t) = \Sigma\, \sigma_{ji} f_i(t)$ corresponding to each measured quantity $Q_i(t)$. The solution corresponding to a least squares fit, $F_i^\circ(t)$, may be related to the calculated function $F_i(t)$ using a Taylor expansion

$$F_i^0(t) = F_i(t) + \sum_i \frac{\partial F_i(t)}{\partial x_i}(\delta x_i) + \tfrac{1}{2}\sum_{i,k}\frac{\partial^2 F_i(t)}{\partial x_i\, \partial x_k}(\delta x_i)(\delta x_k) + \cdots, \qquad [11]$$

where δx_j is the difference between the given value x_j and its value for a least squares solution.

Approximation of equation [11] with only the first order term yields

$$\sum_j \frac{\partial F_i(t)}{\partial x_j} \delta x_j = F_i^0(t) - F_i(t). \quad [12]$$

The best available estimate for $F_i^0(t)$ is $Q_i(t)$ and its substitution in [12] for each experimental value of i and t produces a set of equations of condition for the variables δx_j

$$\sum_j \frac{\partial F_i(t)}{\partial x_j} \delta x_j = Q_i(t) - F_i(t). \quad [13]$$

A set of normal equations may be generated from [13] in a way similar to that used for the σ_{ij} in [8] and a least squares solution for the δx_j obtained. These constitute estimates for the corrections of the variables x_j to approximate a least squares fit of the data.

The solution for the δx_j also yields a variance-covariance matrix which permits an estimation of uncertainties of the x_j and of the variables dependent on the x_j. Such an estimate, however, is only valid when obtained in the neighborhood of a least squares fit, and is subject to re-examination in the case of extensive non-linearities.

When linear dependencies of x_j on other x_i and of σ_{kl} on other σ_{pq} and/or x_i are specified, proper substitutions are made in the calculations so that the partial derivatives include the dependencies.

The coefficients $\partial F_i(t)/\partial x_j$ for each observed i and t are calculated numerically. A small increment Δx_j is introduced for the variable x_j and the new value $\bar{F}_i(t)$ is calculated, from which the $\Delta F_i(t) = \bar{F}_i(t) - F_i(t)$ is easily determined. The coefficient $\partial F_i(t)/\partial x_j$ is then approximated with $\Delta F_i(t)/\Delta x_j$.

There are several reasons for using $\Delta F_i/\Delta x_j$ instead of $\partial F_i(t)/\partial x_j$:

(*a*) For some functions it is not possible to derive $\partial F_i(t)/\partial x_j$ analytically.

(*b*) The method is general and independent of the type of function $F_i(t)$.

(*c*) A decoupling is obtained indirectly between the variables x_j and the σ_{ij}. To calculate a change in ΔF_i for a change in Δx_j a new set of σ_{ij} has to be calculated. Thus, in effect, the σ_{ij} are treated in the calculation of $\Delta F_i(t)/\Delta x_j$ as dependent parameters. This partitions a single interdependent variable space into 2 smaller independent spaces, thereby reducing the possibilities of singular or ill conditioned normal equations.

(*d*) The use of a finite Δx_i may compensate for non-linearities of the function with respect to the variables.

CONVERGENCE

Because the functions $F_i(t)$ are in general non-linear with respect to the variables, a first order approximation is usually inadequate to obtain a least squares solution

in one step, and an iterative procedure must be employed. The newly calculated values of one iteration are used as initial estimates for the following one, and convergence to a least squares solution is a critical aspect of the procedure.

Several factors may be responsible for failure to converge. One is a high degree of non-linearity of the functions with respect to the variables x_i, resulting in poor extrapolation in the calculation of the correction vector. A second factor, inadequacy of the data, may result in failure to resolve the proposed model and lead to either singular or ill conditioned normal equations. In the case of singularity no correction vector can be calculated, and in the case of ill conditioned equations the calculated vector is nearly meaningless, especially in the neighborhood of the least squares fit, since it is very sensitive to the statistical fluctuations of the data. A third factor that may lead to failure to converge to a least squares fit is a poor choice of initial estimates for the values of the parameters. In this case convergence to a local minimum in the sums of squares surface may result.

To deal with the above factors and to accelerate convergence some empirical procedures were introduced into the program. First, every variable is assigned an upper and lower limit for its value and the magnitude of the correction vector is limited not to exceed these limits. When a variable is already at a limit, and the calculated correction is in a direction beyond it, that variable is fixed at the limit and treated as a constant for the remainder of the iteration. A new correction vector is calculated for the remaining variables.

In addition to the limitation imposed by the limits of the variables, a correction vector is also tested to determine whether its magnitude can be optimized to yield a lowest sum of squares. This is accomplished by multiplying the magnitude of the correction vector by a factor k. The value of k is determined from the set of equations

$$k[F_i'(t) - F_i(t)] = [Q_i(t) - F_i(t)] \qquad (i = 1, \cdots, m) \qquad [14]$$

in which $F_i(t)$ is the calculated value for the i^{th} observation *before* the correction was made and $F_i'(t)$ is the calculated value of the same observation after adjustment by the initial correction vector. The experimental value for the i^{th} observation is $Q_i(t)$. When the differences between the calculated and observed values are linearly related to the magnitude of the correction vector, a lowest sum of squares will be obtained for the value of k:

$$k = \frac{\sum_{i=1}^{m} [F_i'(t) - F_i(t)][Q_i(t) - F_i(t)]}{\sum_{i=1}^{m} [F_i'(t) - F_i(t)]^2} \qquad [15]$$

Usually, k is not related linearly to the difference between the calculated and observed values of the data, and a "best" value for it is obtained by iterating [15] several times. Limits have also been incorporated to prevent wild excursions and to

force convergence. The correction vector multiplied by the calculated value of k is accepted as the final set of corrections and the new values for the parameters serve as initial estimates for the next complete iteration.

The use of the factor k speeds up convergence and prevents divergence in the case of ill conditioned normal equations. The magnitude of k may also serve as a measure for the degree of ill conditioning.

When the normal equations are ill conditioned it is sometimes possible (10) to determine mathematically the variables that give rise to the "near dependence" by inserting appropriate dependence assumptions. Such a procedure does not yield a unique solution but does guarantee convergence to a "near" least squares solution.

In the case of both singular and ill conditioned normal equations, when further assumptions about the variables of the system are made to permit a least squares solution, the variance-covariance matrix is conditional on the assumptions made.

The entire convergence procedure is terminated when the rate of decrease in the sum of squares with respect to a previous iteration reaches a preassigned value or when a desired number of iterations has been performed, whichever occurs first.

TEST OF FITTED MODEL

When the final fit of the data is obtained, the initial choice of the model may be re-examined in view of the fit obtained. The following three possibilities may arise.

1. The data are adequate to define the model and the final fit yields a random scatter of the data about the theoretically calculated values. The calculated correction terms tend to zero as the solution converges to a least squares fit and the uncertainties in the variable parameters is relatively small compared to their values. In this case the proposed model is considered compatible with the data.

2. The data are inadequate to permit a definition of the model. This can be recognized by the singularity or the ill-conditioned behavior of the normal equations as discussed in the preceding section. In some cases it may be sufficient to state that the model is indeterminate. When, however, a model solution is still desired, it is necessary to introduce one or more additional constraints to permit a solution. The choice of constraints may follow procedures discussed in the previous section or may be based on criteria similar to those followed in choosing the initial model. After a constraint is imposed the data must be refitted to the new model and the final results re-examined.

3. The data are adequate to define the model but do not have a *random* scatter about the calculated values. The presence of systematic deviations suggests that the *model* is inadequate to fit the data and that it requires additional degrees of freedom. Again, the rationale for extending the model may be the same as discussed earlier, although some modifications may suggest themselves from the nature of the inconsistencies of the fit.

To judge how well a model fits the data a "reference" fit is introduced for com-

parison (8). A reference fit may be obtained by fitting the data to a model having at least one or two more degrees of freedom than the model considered. Such degrees of freedom may be introduced by the release of constraints or the introduction of additional parameters in the model. Several reference models may be considered, each with one more degree of freedom than the previous one, and, when the sum of squares fails to improve significantly with increasing degrees of freedom, or when the degrees of freedom are too great to permit convergence, the reference fit is considered acceptable.

A comparison between the model fit and the reference fit may be used to determine the acceptability of the former. The comparison may be made on the basis of sums of squares or the presence of systematic deviation. The level at which acceptance of a model is set is arbitrary and is not made automatically by the program.

SPECIAL FEATURES

The main features of the methodology have been described in the preceding part of the paper. Special features helpful for calculations of certain models may be generated from these. Such features include the simulation of analog computer operations, use of function generators, solution for input functions and the evaluation of transfer functions of systems. A more detailed description of applications to linear compartmental systems is presented separately (5).

EXAMPLE

The paper deals mainly with the mathematical formalism for the routine fitting of data to models using high speed digital computers. To demonstrate specifically where and how, in the over-all employment of models, the computer program is used, a simple example is given.

Let the experimental data be measures of the concentration (amount per unit volume) of some substance X in the blood as a function of time, as shown in Fig. 1.

If the nature of the system that gave rise to the data is not known, it is necessary to assume one so that a model may be proposed. The rationale in choosing a system is quite empirical and is usually based on previous experience, theoretical considerations, or intuition.

Let us assume that the system for this example may be represented by 2 distinguishable states for the substance X, with transition probabilities between the states as shown in Fig. 2, and that this system may be described by the following set of differential equations:

$$\frac{dq_1}{dt} = -\lambda_{21}q_1 + \lambda_{12}q_2$$

$$\frac{dq_2}{dt} = \lambda_{21}q_1 - \lambda_{12}q_2 - \lambda_{02}q_2$$

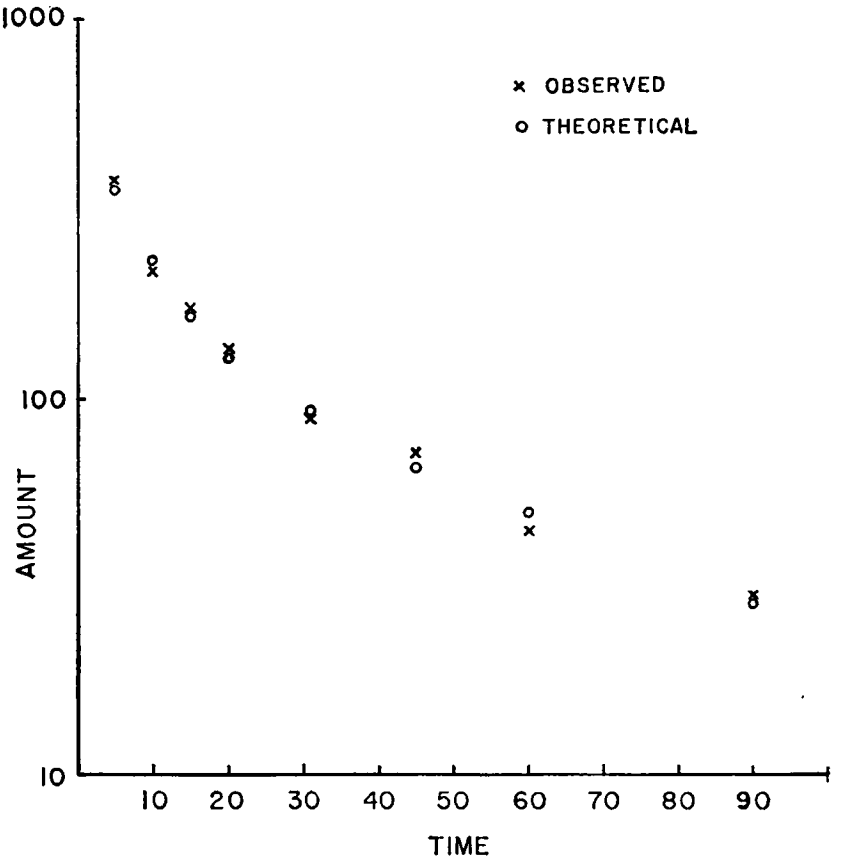

FIGURE 1 Plot of experimental *versus* theoretical values derived from model shown in Fig. 2.

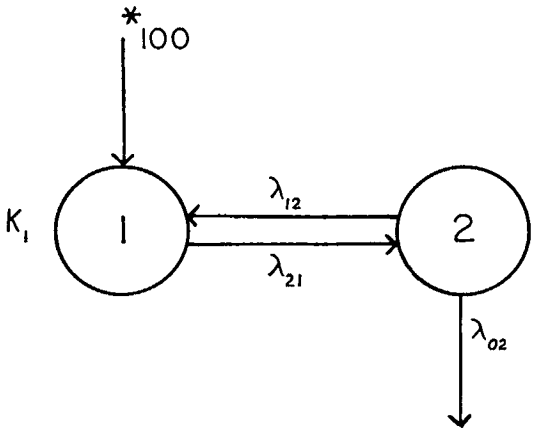

FIGURE 2 Two compartment model used as an example. *100 represents initial conditions of the experiment.

The q_i represent amount of material in compartment i, and λ_{ij} is a transition probability per unit time from the j^{th} to the i^{th} state. λ_{02} represents loss to the outside. State *1* represents the blood and state *2* is unidentified.

Since the data are measures of the concentration of material in state *1*, and since the volume of distribution is unknown, a proportionality constant, k_1, is introduced between the data and q_1. A solution for the model parameters implies the calculation of a set of values for λ_{12}, λ_{21}, λ_{02}, and k_1 that will yield a least squares fit of the data. This is performed completely and automatically on the computer provided the initial conditions of the experiment are specified and initial estimates are provided for the values of λ_{12}, λ_{21}, and λ_{02}. No initial estimates are required for k_1.

To set the problem up for the computer the following information is listed in a prescribed format and placed behind the program deck:

Number of states (compartments) involved: 2
Initial conditions: $q_1(0) = 100$
$q_2(0) = 0$
Type of model involved: linear differential equations (entered by code)

Parameters of model

Parameter		Initial estimates
λ_{12}	=	0.01
λ_{21}	=	0.10
λ_{02}	=	0.05
k_1		

Data

Compartment	Time	Observed amount	Statistical weight
1	5	375	1.0
1	10	219	2.9
1	15	172	4.7
1	20	135	7.8
1	30.1	91	17.0
	etc.		

The program includes a library of various types of models and automatically selects the proper set of equations for the problem in accord with the code specified. When a problem requires a special type of model that is not in the library, a new subroutine has to be written for it and added to the library. Such a subroutine is a small part of the entire program and involves instructions only for the solution of a given set of equations. The available subroutines in the library of the program are usually written with sufficient generality to apply to many models within a class. Thus, for example, the subroutine for the solution of linear differential equations deals with any linear system up to 25 compartments interconnected in any arbi-

trary manner, with a total number of non-zero coefficients not more than 55 and a number of variable parameters less than 25.

The final output of the computer is a printout of a least squares set of values for λ_{12}, λ_{21}, λ_{02} and k_1, their uncertainties, and a comparison between the values predicted by the model and the experimental values:

Computer output

Compartment	Time	Experimental value	Theoretical value
1	5	375.	362.
1	10	219.	233.
1	15	172.	167.
1	20	135.	130.
1	30.1	91.	93.
1	45	73.	67.
1	60	45.	50.
1	90	30.	29.

Final parameter values

$\lambda_{12} = 0.0267 \pm 0.0070$
$\lambda_{21} = 0.115 \pm 0.030$
$\lambda_{02} = 0.0237 \pm 0.0040$
$k_1 = 6.19 \pm 1.10$

In general, the final values obtained must be interpreted in accord with the manner of convergence, as discussed in the text. If the calculated fit is "good," the assumed model is acceptable. If, however, the final solution is not unique (large standard errors) or inconsistent (systematic deviations between calculated and observed values), the assumed model has to be modified by changes in the number of independent parameters or compartments, or by the introduction of a different type of model. The entire procedure has to be repeated for any new model.

Preliminary report presented at the 5th Annual Biophysical Society Meeting, St. Louis.
Received for publication, January 14, 1962.

REFERENCES

1. IBM Reference Manual—709/7090 FORTRAN Programming System, IBM, 1961.
2. BERMAN, M. and SCHOENFELD, R., Invariants in experimental data on linear kinetics and the formulation of models. *J. Appl. Physics*, 1956, **27**, 1361.
3. GARDNER, D. G., GARDNER, J. C., LAUSH, G., and MEINKE, W. W., Method for the analysis of multi-component exponential decay curves. *J. Chem. Physics*, 1959, **31**, 978.
4. PERL, W., A method for curve-fitting by exponential functions, *Internat. J. Appl. Radiation and Isotopes*, 1960, **8**, 211.
5. BERMAN, M., WEISS, M. F., and SHAHN, E., Some formal approaches to the analysis of kinetic data in terms of linear compartmental systems, *Biophysic. J.*, 1962, **2**, 289.
6. LEVY, H., and BAGGOTT, E. A., Numerical Solutions of Differential Equations, New York, Dover Publications, Inc., 1950.

7. WHITTAKER, E., and ROBINSON, G., The Calculus of Observation, London, Blackie and Son Ltd., 1956, Chapter IX.
8. BERMAN, M., Application of differential equations to the study of the thyroid system, *Proc. 4th Berkeley Symp. Mathematical Statistics,* Berkeley, University of California Press, 1961, **7**.
9. BERMAN, M., and SCHOENFELD, R., A note on unique models in tracer kinetics, *Exp. Cell Research,* 1960, **20,** 574.
10. HEAD, J. W., and OULTON, G. M., The solution of "ill-conditioned" linear simultaneous equations, *Aircraft Eng.,* 1958, **30,** 309.
11. BOX, G. E. P., Fitting empirical data, *Ann. New York Acad. Sc.,* 1960, **86,** 792.