BIOMETRICS *27*, 77–102
March 1971

The following paper was presented by H. O. Hartley on August 21, 1970 during the 7th International Biometric Conference held in Hannover, Germany. Bernard G. Greenberg was in the Chair.

STOCHASTIC COMPARTMENTAL ANALYSIS: MODEL AND LEAST SQUARES ESTIMATION FROM TIME SERIES DATA

J. H. MATIS AND H. O. HARTLEY

Institute of Statistics, Texas A&M University, College Station, Texas 77843, U. S. A.

SUMMARY

This paper is concerned with a discrete population of particles in a steady state compartmental system. The system is considered to have m compartments and the transitions are stochastic in nature. Such systems may be used to model a variety of problems; one application is sketched in this work.

Efficient estimation of these transition rate parameters requires the associated distribution theory. This paper advances the distribution theory considerably by providing a compact analytic solution to the compartmental problem.

Often in practice, individual compartments are inaccessible for observation and instead time series data are available only on the passage of material to the system exterior. The covariance kernel of such observations is derived in this paper and utilized for 'efficient' parameter estimation. The recommended estimation procedure is demonstrated on both simulated and biological data, and in the process the merits of the procedure are clarified.

1. INTRODUCTION

Recently a new branch of biomathematical modelling called compartmental analysis has been developed. The concept of compartmental analysis assumes that a system may be divided into homogeneous components, or 'compartments.' Various characteristics of the system are determined by observing the movement of tracer material.

Usually the theory is applied to describe the movement of a population of tracer molecules, e.g., the flow of iron molecules within sheep (Carter *et al.*

[1964]). Possibly since the individual molecules are infinitesimal in size, nearly all the previous literature has made the implicit assumption of a deterministic flow pattern. Papers by Zilversmit *et al.* [1943] and Sheppard and Householder [1951] are principal founders of the deterministic compartmental theory, and more recently Sheppard [1962], Rescigno and Segre [1966], and Whipple and Hart [1963] provide comprehensive reviews of the theory and partial bibliographies of its application. Estimation in the deterministic model utilizes the ordinary least squares technique, and recently 'simultaneous equation' estimation for timewise uncorrelated error has been introduced to compartmental models in a series of papers by Beauchamp and Cornell [1966; 1968; 1969].

On the other hand, stochastic compartmental analysis, which assumes probabilistic behavior of the tracer particles, has been slow to develop. Bartholomay [1958] was an early pioneer of the stochastic assumption, and solved the one-compartment model. Cornfield *et al.* [1960], after reviewing some questions of estimation in the deterministic model, arrive at the conclusion that the stochastic model is more 'realistic' and should be investigated. Their treatment of the stochastic aspects, however, is brief and admittedly 'preliminary'; their assumption that observations are independent over time is 'unsatisfactory' from a practical standpoint. Several subsequent papers have solved for the 1st and 2nd moments of special 2- or 3-compartment models; see e.g. Bernard *et al.* [1965] and Matis and Carter [1970]. But the probability distribution theory and its implementation into an estimation procedure, the two fundamental problems of stochastic compartmental analysis, still remain to be solved.

In a sense, a compartmental model can claim only to represent an 'approximate' theoretical background for the biological observed phenomena, in that it employs an abstraction of 'transfer' of particles without specifying a detailed causative theory responsible for the transfer mechanisms. It is also because of this approximate feature that a stochastic compartmental theory is more realistic in that a detailed causative mechanism, lacking in the deterministic model, is supplied by a stochastic model. This paper treats the case of a finite tracer population in a compartmental system with probabilistic flow. The joint generating function for the m random variables specifying the number of tracer particles in each compartment is derived. Inversion of the generating function gives the joint distribution of the m random variables at any particular point in time; from whence the covariance kernel for the total number of particles in the m-compartment system at different times is derived. The paper also recommends a procedure to incorporate the covariance kernel in the nonlinear least squares estimation of the parameters from time-series data.

2. AN APPLICATION

As is apparent in the above mentioned reviews, compartmental analysis finds application in many diverse areas of biomedical science. Heretofore the applications have been predominately physiological or pharmacokinetic and

it is not as well known that compartmental models have been proposed under other terminology in many other areas of endeavor; see e.g. the sociological model of Herbst [1963], the public health model of Fix and Neyman [1951], and the anatomical model of Blaxter *et al.* [1956]. This paper will illustrate the theory as applied to the last of these models; the application to other areas is immediate.

More specifically, the application motivating this paper is twofold: (1) to develop a stochastic model describing the passage of a given material through the gastrointestinal tract of ruminants, e.g. cattle and sheep, and (2) to estimate the 'turnover-rate' parameters of the given material from fecal output of its tracer. Animal nutritionists identify a great variety of materials for which the above modelling is useful; one particular substance of interest, for example, consists of indigestible, plastic beads used as roughage substitute.

Previous biological evidence suggests that compartmental analysis is applicable to the above problem. The gastrointestinal tract may be conceptualized as a series of 'vats' (see e.g. Hungate [1966] chapter V) and indeed Blaxter *et al.* [1956], by assuming deterministic behavior, have obtained 'good' fits of experimental data to a very restricted 4-compartment model. They suggest the compartments are identified as in Figure 1.

The theory subsequently developed will incorporate stochastic behavior into any configuration of m compartments. All associated turnover rates may then be estimated.

3. SOLUTION OF m-COMPARTMENT STOCHASTIC MODEL

3.1. *Derivation of partial differential equation*

Let a general m-compartment system[1], where each compartment is connected to each other and to the system exterior, be given. It follows that a general system has m^2 parameters. Figure 2 represents a general $m = 2$ compartment system with $m^2 = 4$ parameters. Let $N_i(0)$ be the known number of labelled units introduced into compartment i at time 0 and let $N_i(t)$ be a random variable specifying the number of units in compartment i at time t. Let b_{ji} be the transition intensity or 'turnover rate' from compartment i to compartment j, where b_{oi} represents an excretion from compartment i. Then, by definition, $b_{ji}\ \Delta t$ is the probability that a particular unit migrates from compartment i to compartment j in the time interval Δt.

Two assumptions of (steady-state) compartment analysis are introduced: (1) the b_{ji} are independent of time and (2) each of the $N_i(t)$ units acts in-

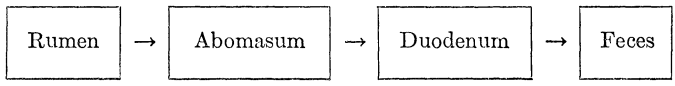

FIGURE 1

[1] A simple example of the following general theory is discussed in section 5.

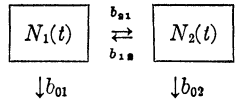

FIGURE 2

dependently. We now derive the fundamental differential equation for the joint distribution of stochastic variables, $N_i(t)$. Although this could be directly derived for differential time, dt, we prefer a direct derivation by starting from a finite increment, Δt, and letting Δt approach 0.

Given the $N_i(t)$ independent units in compartment i at time t, the conditional probability that a single unit moves from i to j in Δt is determined by the binomial law to be

$$N_i(t)(b_{ji}\ \Delta t)(1 - b_{ji}\ \Delta t)^{N_i(t)-1},$$

which may be expanded in terms of Δt to

$$N_i(t)b_{ji}\ \Delta t + o(\Delta t).$$

Similarly the probability of more than one migration in Δt is $o(\Delta t)$.

There are $2m - 1$ possible ways for the number in compartment i, $N_i(t)$, to change in Δt; compartment i may gain a unit from any of the $m - 1$ other compartments, or it may lose a unit either to another compartment or to the system exterior. Though these events are not independent, clearly the probability of two or more events occurring in Δt is again $o(\Delta t)$.

Let a simple event be a change in the number of at least one compartment. Also let

$$p_{k_1,k_2,\cdots,k_m}[t; N_1(t), \cdots, N_m(t)] \quad k_i = 0, \pm 1$$

denote the conditional probability that given $N_1(t), N_2(t), \cdots, N_m(t)$ units in the compartments at time t, the event {each compartment i gains (or loses) k_i units respectively} occurs in the time interval $(t, t + \Delta t)$. Then to first order magnitude of Δt, the above considerations determine m^2 different events with probabilities:

$$\begin{aligned}
p_{-1,0,0\cdots0} &= N_1(t)b_{01}\ \Delta t \\
p_{-1,1,0\cdots0} &= N_1(t)b_{21}\ \Delta t \\
&\vdots \\
p_{-1,0,0\cdots1} &= N_1(t)b_{m1}\ \Delta t \\
&\vdots \\
p_{0,0,0\cdots-1} &= N_m(t)b_{0m}\ \Delta t \\
p_{1,0,0\cdots-1} &= N_m(t)b_{1m}\ \Delta t \\
&\vdots \\
p_{0,0,0\cdots1,-1} &= N_m(t)b_{m-1,m}\ \Delta t
\end{aligned} \tag{1}$$

Note that each event is associated with one of the m^2 $b_{j,i}$ parameters of a general m-compartment system.

The above probabilities (unit jumps in a continuous parameter, t, for integral states, $N_i(t)$) define an m-dimensional birth-death process. By standard techniques (see e.g. Bartlett [1966] or Bailey [1964]), one may first write the Kolmogorov forward equations and then a linear partial differential equation for the generating function. Let $\pi_{n_1,n_2,\cdots,n_m}(t)$ be the probability that $N_i(t) = n_i$, $i = 1, \cdots, m$. Then it follows that

$$\begin{aligned}\pi_{n_1,n_2,\cdots,n_m}(t + \Delta t) = {} & \pi_{n_1+1,n_2,\cdots,n_m}(t)p_{-1,0,0,\cdots,0}(t) \\ & + \pi_{n_1+1,n_2-1,\cdots,n_m}(t)p_{-1,1,0,\cdots,0}(t) + \cdots \\ & + \pi_{n_1+1,n_2,\cdots,n_m-1}(t)p_{-1,0,0,\cdots,1}(t) + \cdots \\ & + \pi_{n_1,n_2,\cdots,n_m+1}(t)p_{0,0,0,\cdots,-1}(t) \\ & + \pi_{n_1-1,n_2,\cdots,n_m+1}(t)p_{1,0,0,\cdots,-1}(t) + \cdots \\ & + \pi_{n_1,n_2,\cdots,n_m-1-1,n_m+1}(t)p_{0,0,0,\cdots,1,-1}(t) \\ & + \pi_{n_1,n_2,\cdots,n_m}(t)[1 - p_{-1,0,0,\cdots,0}(t) \\ & - p_{-1,1,0,\cdots,0}(t) - \cdots - p_{-1,0,0,\cdots,1}(t) \\ & - \cdots - p_{0,0,0,\cdots,-1}(t) - p_{1,0,0,\cdots,-1}(t) \\ & - \cdots - p_{0,0,0,\cdots,1,-1}(t)]. \end{aligned} \tag{2}$$

From equations (1) and (2) the forward equation is readily seen to be

$$\begin{aligned}\frac{d\pi_{n_1,n_2,\cdots,n_m}(t)}{dt} = {} & (n_1 + 1)[\pi_{n_1+1,n_2,\cdots,n_m}(t)b_{01} + \pi_{n_1+1,n_2-1,\cdots,n_m}(t)b_{21} \\ & + \cdots + \pi_{n_1+1,n_2,\cdots,n_m-1}(t)b_{m1}] + \cdots \\ & + (n_m + 1)[\pi_{n_1,n_2,\cdots,n_m+1}(t)b_{0m} + \pi_{n_1-1,n_2,\cdots,n_m+1}(t)b_{1m} \\ & + \cdots + \pi_{n_1,n_2,\cdots,n_m-1-1,n_m+1}(t)b_{m-1,m}(t)] \\ & - \pi_{n_1,n_2,\cdots,n_m}(t) \sum_{i=1}^{m} n_i \sum_{\substack{j=0 \\ j\neq i}}^{m} b_{j,i} . \end{aligned} \tag{3}$$

Since the moment generating function (m.g.f.) is defined as

$$M(\theta_1, \theta_2, \cdots, \theta_m, t) = \sum_{(n_1,n_2,\cdots,n_m)} \pi_{n_1,n_2,\cdots,n_m}(t) \prod_{i=1}^{m} e^{\theta_i n_i},$$

where the θ_i are real, a partial differential equation for the m.g.f. may be found by multiplying both sides of (3) by $\prod_{i=1}^{m} e^{\theta_i n_i}$ and then summing over all possible values of the m-tuple $(n_1, n_2, \cdots, n_m)$. Using such relations as

$$\begin{aligned}\sum_{(n_1,n_2,\cdots,n_m)} n_1\pi_{n_1,n_2,\cdots,n_m}(t) \prod_{i=1}^{m} e^{\theta_i n_i} \\ = \sum_{(n_1^*,n_2^*,\cdots,n_m^*)} n_1^*\pi_{n_1^*,n_2^*,\cdots,n_m^*}(t) \prod_{i=1}^{m} e^{\theta_i n_i},\end{aligned}$$

where $n_1^* = n_1 + 1$, $n_2^* = n_2 - 1$, and $n_j^* = n_j$ for $3 \leq j \leq m$; and then multiplying throughout by M^{-1} to transform the m.g.f. to the cumulant generating function (c.g.f.), say $K(\theta_1 , \theta_2 , \cdots , \theta_m , t)$, one has the relation

$$\frac{\partial K(\theta_1 , \theta_2 , \cdots , \theta_m , t)}{\partial t} = \sum_{i=1}^{m} \left\{ (e^{-\theta_i} - 1) b_{oi} + \sum_{\substack{j=1 \\ j \neq i}}^{m} (e^{-\theta_i + \theta_j} - 1) b_{ji} \right\} \frac{\partial K(\theta_1 , \theta_2 , \cdots , \theta_m , t)}{\partial \theta_i} . \quad (4)$$

Inasmuch as at time $t = 0$ there are $N_i(0) = N_i$ known units in each compartment i, an initial condition is

$$K(\theta_1 , \theta_2 , \cdots , \theta_m , 0) = \sum_{i=1}^{m} N_i \theta_i . \quad (5)$$

The form of equations (4) and (5) is well known; indeed, Bailey [1964] has formalized rules of thumb through which the equations could be written directly from the probability intensity parameters, b_{ji} , and the initial unit counts.

3.2. *Solution of partial differential equation*

The first moments of the compartments, $\mu_i(t) = E[N_i(t)]$, are particularly interesting. A system of equations describing the $\mu_i(t)$ may be derived by (1) expanding the c.g.f. in terms of its cumulants and expanding the exponential terms in powers of the θ_i , (2) performing the appropriate differentiations on both sides of the equation, and finally (3) equating coefficients of θ_1 through θ_m . Let $\mathbf{M}^T(t)$ be the m-vector of expected values, $[\mu_1(t), \cdots , \mu_m(t)]$, and define an $m \times m$ matrix $\mathbf{B} = (b_{ji})$ such that b_{ji} , for $i \neq j$, are the previously introduced transition probabilities and

$$b_{ii} = - \sum_{\substack{j=0 \\ j \neq i}}^{m} b_{ji} .$$

Note that b_{ii} is a linear combination of all rates leaving compartment i. With these definitions, the system of equations derived from (4) to describe the first moments may be written

$$\mathbf{M}'(t) = \mathbf{B}\mathbf{M}(t), \quad (6)$$

where $\mathbf{M}'(t)$ designates the derivative of $\mathbf{M}(t)$. Equation (6), however, is identical to the deterministic equations of a general m-compartment system (see Sheppard [1962] p. 30). Thus by appealing to uniqueness properties, the following proposition is proven, which though not surprising, is nonetheless important subsequently:

Proposition 1. The expected value, $\mu_i(t)$, of the number of units in each compartment i of a stochastic m-compartment system is identically equal to its deterministic solution.

Matis and Carter [1970] have solved for the second moments of a 2-

compartment system by the same technique of equating second order coefficients. However, the second (and higher) moments of an $m > 2$ compartment system are difficult to solve by the same method; the linear system describing the second moments involves $\frac{1}{2}m(m + 1)$ equations. A simpler and more revealing method involves solving for the c.g.f. directly from equations (4) and (5). Of course, another compelling argument for an explicit solution of the c.g.f. rests in its ability to define uniquely the associated probability distribution.

The following definitions are essential for the c.g.f. Let $\mathfrak{B}$ be the negative transpose of the matrix $\mathbf{B}$ in (6), i.e. $\mathfrak{B} = -\mathbf{B}^T$. Let the latent roots of $\mathfrak{B}$ be α_i , $i = 1, \cdots, m$; and assume that the complex numbers α_i are distinct. Corresponding to each α_i , a latent m-vector, say $\mathbf{F}^i$, may be found where for subsequent convenience the first element is either 0 or is standardized to 1. Let $\mathbf{F} = (f_{ij})$ be the matrix of these latent vectors, i.e.,

$$\mathbf{F} = [\mathbf{F}^1, \mathbf{F}^2, \cdots, \mathbf{F}^m] = \begin{bmatrix} f_{11} & f_{12} & \cdots & f_{1m} \\ f_{21} & f_{22} & \cdots & f_{2m} \\ \vdots & \vdots & & \vdots \\ f_{m1} & f_{m2} & \cdots & f_{mm} \end{bmatrix}.$$

The determinant of $\mathbf{F}$ is denoted $|F|$, and the cofactor of f_{ij} by F_{ij} . Another function of the turnover rates, $p_{ki}(t\ ; b_{01}\ , b_{21}\ , \cdots, b_{mm})$, or in abbreviated form $p_{ki}(t)$, is now defined by

$$p_{ki}(t) = \frac{1}{|F|} \sum_{j=1}^{m} f_{ij} F_{kj} e^{-\alpha_j t} \tag{7}$$

and it may be shown to be a real number. Further, the m-variate c.g.f. which solves equations (4) and (5) is

$$K(\theta_1\ , \theta_2\ , \cdots, \theta_m\ , t) = \sum_{i=1}^{m} N_i(0) \ln \left[1 + \sum_{k=1}^{m} (e^{\theta_k} - 1) p_{ki}(t) \right]. \tag{8}$$

Whilst the differential equation (4) is readily available, we are not aware that solution (8) has been given before in the literature of compartmental analysis, and it is this that contributes to the theory[2]. The result is readily verified for the special case $m = 2$ by substitution into (4).

3.3. *Resultant probability distribution*

The remainder of this section will explore the properties of the generating function. The jth marginal c.g.f. is immediately

$$K(0, 0, \cdots, \theta_j\ , \cdots, 0, t) = \sum_{i=1}^{m} N_i \ln [1 + (e^{\theta_j} - 1) p_{ji}(t)]. \tag{9}$$

[2] It has since come to our notice that our problem is related to the so-called 'illness-death' process treated by Chiang [1968] who uses probability generating functions which he derives as solutions of a differential equation treated in his Chapters 4–7. Equation (8) is derivable from his published work.

The mean and variance of $N_v(t)$, say $\mu_v(t)$ and $\sigma_v^2(t)$, are found by expanding and differentiating (9) to be, for $v = 1, \cdots, m$,

$$\begin{aligned} \mu_v(t) &= \sum_{i=1}^{m} N_i p_{vi}(t), \\ \sigma_v^2(t) &= \sum_{i=1}^{m} N_i p_{vi}(t)[1 - p_{vi}(t)]. \end{aligned} \tag{10}$$

Consider the case where only the uth compartment is labelled; i.e., $N_i = N_i(0) = 0$ for all $i \neq u$. Then as an immediate consequence of the non-negativity of the variance, one may show that $0 \leq p_{vu}(t) \leq 1$ for all t. Since u and v are arbitrary, the following inequalities are established:

$$0 \leq p_{v,u}(t) \leq 1 \quad \text{for all} \quad u, v, \quad \text{and} \quad t. \tag{11}$$

Consider also the random variable $N_T(t) = \sum_{i=1}^{m} N_i(t)$. By properties of expectations, its expected value, $\mu_T(t)$, is

$$\mu_T(t) = \sum_{i=1}^{m} N_i \sum_{v=1}^{m} p_{vi}(t). \tag{12}$$

Since the total system cannot gain units in addition to the initial dose, clearly $0 \leq N_T(t) \leq N_T(0)$. If only the uth compartment is labelled, it follows that

$$0 \leq N_T(t) \leq N_u(0)$$

from whence

$$0 \leq \mu_T(t) = N_u(0) \sum_{v=1}^{m} p_{vu}(t) \leq N_u(0).$$

Thus since $\sum_{v=1}^{m} p_{vu}(t) \leq 1$ holds for all t and arbitrary u, the following is true:

$$\sum_{v=1}^{m} p_{vu}(t) \leq 1 \quad \text{for all} \quad u \quad \text{and} \quad t. \tag{13}$$

A fundamental result may now be established:

Proposition 2. Let the m-vector $\mathbf{\Delta}(t)$ be defined by $\mathbf{\Delta}^T(t) = [N_1(t), N_2(t), \cdots, N_m(t)]$. Also let $\mathbf{\Gamma}_i(t)$, where $\mathbf{\Gamma}_i^T(t) = [\gamma_{1i}(t), \gamma_{2i}(t), \cdots, \gamma_{mi}(t)]$ for $i = 1, 2, \cdots, m$, be distributed as a multinomial with parameters $N_i(0)$, $p_{1i}(t)$, $p_{2i}(t)$, $\cdots$, $p_{mi}(t)$; i.e.

Prob $[\gamma_{1i}(t), \gamma_{2i}(t), \cdots, \gamma_{mi}(t)]$

$$= \frac{N_i(0)! \prod_{j=1}^{m} p_{ji}^{\gamma_{ji}} \left[1 - \sum_{j=1}^{m} p_{ji}\right]^{N_i(0) - \sum_{j=1}^{m} \gamma_{ji}}}{\prod_{j=1}^{m} \gamma_{ji}! \left[N_i(0) - \sum_{j=1}^{m} \gamma_{ji}\right]!}.$$

Then $\mathbf{\Delta}(t)$ is distributed as the sum of the m independent $\mathbf{\Gamma}_i(t)$, i.e.,

$$\mathbf{\Delta}(t) = \sum_{i=1}^{m} \mathbf{\Gamma}_i(t).$$

A verification of this claim is now sketched. The $p_{ji}(t)$ functions in equation (8) meet the restrictions of multinomial parameters by virtue of inequalities (11) and (13). Moreover each of the m terms in the c.g.f. (8) has the form of a multinomial c.g.f. Hence by appeal to uniqueness properties of the c.g.f., the vector $\mathbf{\Gamma}_i(t)$ corresponding to each term is distributed as a multinomial with the above parameters. The independence of the $\mathbf{\Gamma}_i(t)$ results from the additivity of the c.g.f.'s.

A physical interpretation may be attached to the $\mathbf{\Gamma}_i(t)$ vectors as follows. The $\gamma_{ji}(t)$ random variables characterize at time t the dispersion throughout the compartments (j) of the $N_i(0)$ units which originated in compartment i. The $\mathbf{\Gamma}_i(t)$ vector is intuitively a multinomial but the $p_{ji}(t)$ parameters are usually involved functions of the b_{ji} turnover rates. Logically, the behavior of the $N_i(0)$ units is independent of the, say, $N_{i'}(0)$ units originating in compartment $i' \neq i$, and indeed Proposition 2 establishes this assertion. Hence in a two-way layout of the $\gamma_{ji}(t)$ elements, as in Figure 3, the rows are independent.

		Compartment at time t			Vector
		1 ⋯ j ⋯ m			Notation
Compartment at time 0	1	$\gamma_{11}(t)$ ⋯	$\gamma_{j1}(t)$ ⋯	$\gamma_{m1}(t)$	$\mathbf{\Gamma}_1(t)$
	⋮	⋮	⋮	⋮	⋮
	i	$\gamma_{1i}(t)$ ⋯	$\gamma_{ji}(t)$ ⋯	$\gamma_{mi}(t)$	$\mathbf{\Gamma}_i(t)$
	⋮	⋮	⋮	⋮	⋮
	m	$\gamma_{1m}(t)$ ⋯	$\gamma_{jm}(t)$ ⋯	$\gamma_{mm}(t)$	$\mathbf{\Gamma}_m(t)$
	Total	$N_1(t)$	$N_j(t)$	$N_m(t)$	$\mathbf{\Delta}(t)$

FIGURE 3

However, the total (over all m origins) number of units in compartment j, $N_j(t)$, is the jth marginal of $\mathbf{\Delta}(t)$ and is not independent of the other marginals; hence the columns of Figure 3 are dependent. If u and v are different compartments, the covariance of γ_{ui} and γ_{vi} is determined by properties of the multinomial distribution to be

$$\operatorname{cov}[\gamma_{ui}(t), \gamma_{vi}(t)] = -N_i(0)p_{vi}(t)p_{ui}(t).$$

Similarly, since $N_j(t) = \sum_{i=1}^{m} \gamma_{ji}$ and by virtue of the independence of the $\mathbf{\Gamma}_i(t)$, it follows that

$$\operatorname{cov}[N_u(t), N_v(t)] = -\sum_{i=1}^{m} N_i(0)p_{vi}(t)p_{ui}(t).$$

It is also of interest here to recall a well-known result of a one-compartment model, or a pure death process. It has been established that the number of

individuals remaining in the process follows a binomial distribution. Note that Proposition 2 generalizes this previous result in a natural manner, i.e., the $N_i(t)$ for given t can be regarded as a mixture of multinomial distributions. Indeed $N_T(t_1) - N_T(t_2)$ for various t intervals may also be regarded as a mixture of multinomial distributions, where the ith component results from the $N_i(0)$ units placed in the ith compartment. The present theory will not utilize this latter mixture property.

4. LEAST SQUARES ESTIMATION OF THE PARAMETERS b_{ji}.

Proposition 2 contains the stochastic model of the general m-compartment system; hence it provides a complete basis for the least squares estimation of the (at most) m^2 transition probability rates, or b_{ji}, from output data. We will now spell out estimation in the special case where in our experience data are most readily available, i.e., where data are available only on the total number of units in the system at various times, denoted previously as $N_T(t_i)$ for $i = 1, \cdots, z$. Beauchamp and Cornell [1966; 1968; 1969] have considered a generalization where time-series data is obtained at specified times from several compartments simultaneously. The present approach differs from theirs in two regards: (1) it allows for the correlation of observations over time and (2) by recognizing that the variance-covariance matrix, $\mathbf{\Sigma}$, is a known mathematical function of the parameters, it introduces a two-stage estimation procedure based on that fact. Both above features are direct consequences of Proposition 2 rather than being an assumed error structure and hence, we feel, they are imperative to implement.

The present approach also differs fundamentally from previous practice by estimating the b_{ji} parameters directly. It has been customary in the past first to estimate the exponents and coefficients of a sum of exponentials model, and then to transform them by the invariant theory of Berman and Schoenfeld [1956]. Since it is our experience, however, that often many b_{ji} parameters are known a priori to be 0, the present approach reduces the number of parameters immediately and thus enhances the efficiency without resort to the algebraically difficult alternative of restricted least squares (see e.g. Goldberger [1964] p. 256).

4.1. *Compartmental analysis regression function*

Equation (12) above gives the expected value, $\mu_T(t_i)$, for the number of units remaining in a pulsed compartmental system at time t. With the additional definition

$$a_s(t;b_{01}, b_{21}, \cdots, b_{mm}) = a_s(t) = \sum_{k=1}^{m} p_{ks}(t), \tag{14}$$

equation (12) simplifies to

$$\mu_T(t) = \sum_{i=1}^{m} N_i(0)a_i(t). \tag{15}$$

The following definitions are now necessary:

(1) $\mathbf{Y}^T = [y_1 , y_2 , \cdots , y_z]$ is the z-vector of observations $N_T(t_1), N_T(t_2), \cdots , N_T(t_z)$,

(2) $\mathbf{\Omega}^T = [\omega_1 , \omega_2 , \cdots , \omega_{m^2}]$ is the m^2-parameter vector of the rates $b_{01} , b_{21} , b_{31} , \cdots , b_{mm}$, respectively,

(3) $\mathbf{g}^T(t, \mathbf{\Omega}) = [g(t_1), g(t_2), \cdots , g(t_z)]$ is the z-vector of expected values $\mu_T(t_1), \mu_T(t_2), \cdots , \mu_T(t_z)$, and

(4) $\boldsymbol{\varepsilon}^T = [\epsilon_1 , \epsilon_2 , \cdots , \epsilon_z]$ is the z-vector of error variables.

The 'regression function' may now be written

$$\mathbf{Y} = \mathbf{g}(t; \mathbf{\Omega}) + \boldsymbol{\varepsilon}, \tag{16}$$

where it is apparent by construction that $E(\boldsymbol{\varepsilon}) = \mathbf{0}$. Hence the $z \times z$ covariance matrix is $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T)$ which is denoted subsequently by $\mathbf{\Sigma} = (\sigma_{ij})$.

Reference back to equations (7) and (12) indicates that the parameters, ω_i , enter into equation (16) nonlinearly. Thus following a well-known nonlinear estimation procedure, $\mathbf{g}(t; \mathbf{\Omega})$ will be expanded by a first order Taylor series about some initial estimate of the parameter vector, say ${}_o\mathbf{\Omega}$, and the modified Gauss-Newton technique employed to iterate for $\mathbf{\Omega}$ (see Hartley [1961]). With the above linearization, equation (16) is approximately

$$\mathbf{Y} - \mathbf{g}(t, {}_o\mathbf{\Omega}) = {}_o\mathbf{G}\,{}_o\mathbf{D} + \boldsymbol{\varepsilon}, \tag{17}$$

where

(1) ${}_o\mathbf{G} = ({}_o g_{ij}) = [\partial \mathbf{g}(t_i , \mathbf{\Omega})]/(\partial \omega_j) \,|_{\Omega = {}_o\Omega}$ is the $z \times m^2$ matrix of first partials of the elements $\mathbf{g}(t_i , \mathbf{\Omega})$, $i = 1, \cdots , z$, with respect to the parameters ω_j , $j = 1, \cdots , m^2$, and

(2) ${}_o\mathbf{D}^T = [{}_o d_1 , {}_o d_2 , \cdots , {}_o d_{m^2}]$ is the m^2-vector of differences $\omega_i - {}_o\omega_i$, Since equation (17) is linear in the ${}_o d_i$, the best linear unbiased estimates (BLUE) of the ${}_o d_i$ are found by minimizing the generalized sum of squared deviations, $\boldsymbol{\varepsilon}^T\mathbf{\Sigma}^{-1}\boldsymbol{\varepsilon}$.

Estimation in the deterministic model proceeds exactly as above, inasmuch as by Proposition 1, the expected values, $\mu_T(t_i)$, identically equal the deterministic models. In the absence of stochastic considerations, however, the observations are assumed independent and usually the (measurement) errors are assigned equal magnitude; hence $\mathbf{\Sigma} = \sigma^2\mathbf{I}$, where $\mathbf{I}$ is the identity matrix. Thus the modified Gauss-Newton technique minimizes the sum of squares by using the Gauss-Markov theorem internally for BLUE estimates.

Yet, as previously hypothesized by Cornfield *et al.* [1960], it will be demonstrated presently that the individual stochastic disturbances, ϵ_i , are neither homoscedastic nor independent. In stochastic compartment analysis, then, 'efficient' estimation requires the modified *G-N* technique to minimize the generalized sum of squares by using the Aitken generalized least squares theorem (see e.g. Goldberger [1964] p. 233). By the Aitken theorem, the BLUE of ${}_o\mathbf{D}$, say ${}_o\tilde{\mathbf{D}}$, is given by

$${}_o\tilde{\mathbf{D}} = [{}_o\mathbf{G}^T\mathbf{\Sigma}^{-1}{}_o\mathbf{G}]^{-1}{}_o\mathbf{G}^T\mathbf{\Sigma}^{-1}[\mathbf{Y} - \mathbf{g}(t, {}_o\mathbf{\Omega})]. \tag{18}$$

The balance of this section will consider first the derivation of the matrix

Σ and then suggest a method of incorporating it into the estimation procedure.

4.2. *Covariance kernel of observations*

The diagonal elements of Σ may be derived utilizing the corollary in section 3. Note that

$$\sigma_{aa} = \text{var } [N_T(t_a)] = \text{var } \left[\sum_{k=1}^{m} N_k(t_a) \right]$$

$$= \sum_{k=1}^{m} \text{var } [N_k(t_a)] + \sum_{k \neq l}^{m} \sum \text{cov } [N_k(t_a), N_l(t_a)].$$

From the corollary and by the properties of independent multinomials, it follows that

$$\text{var } [N_k(t_a)] = \sum_{s=1}^{m} N_s p_{ks}(t_a)[1 - p_{ks}(t_a)]$$

and (19)

$$\text{cov } [N_k(t_a), N_l(t_a)] = -\sum_{s=1}^{m} N_s p_{ks}(t_a) p_{ls}(t_a).$$

Substituting equations (19) into the above, and then simplifying, one has

$$\sigma_{aa} = \sum_{k=1}^{m} \sum_{s=1}^{m} N_s p_{ks}[1 - p_{ks}(t_a)] - \sum_{k \neq l}^{m} \sum \sum_{s=1}^{m} N_s p_{ks}(t_a) p_{ls}(t_a)$$

$$= \sum_{s=1}^{m} N_s \sum_{k=1}^{m} p_{ks}(t_a) \left[1 - \sum_{r=1}^{m} p_{rs}(t_a) \right]$$

$$= \sum_{s=1}^{m} N_s a_s(t_a)[1 - a_s(t_a)], \tag{20}$$

where $a_s(t_a)$ was defined previously by equation (14).

The derivation of σ_{ab}, the covariance of the process at two different times, t_a and $t_b = t_a + \Delta t$, $\Delta t > 0$, requires the identity

$$a_i(t_b) = \sum_{r=1}^{m} p_{ri}(t_a) a_r(\Delta t) \tag{21}$$

and also the Markov property. The general covariance element is found to be

$$\sigma_{ab} = \sum_{s=1}^{m} N_s a_s(t_b)[1 - a_s(t_a)]. \tag{22}$$

Proofs of identity (21) and claim (22) are lengthy and thus deferred to the Appendix. Equations (20) and (21) may be combined into the following result: *Proposition 3.* Let $\sigma_{ab} = \text{cov } [N_T(t_a), N_T(t_b)]$ be the covariance kernel of the process describing the total number of beads in the system at times t_b and t_a such that $t_b \geq t_a$. Then

$$\sigma_{ab} = \sum_{s=1}^{m} N_s a_s(t_b)[1 - a_s(t_a)]. \tag{23}$$

The covariance matrix, $\mathbf{\Sigma}$, is thus specified by equation (23) for observations at any arbitrary time points.

4.3. *Recommended estimation procedure*

If the model (16) were linear in the parameters $\mathbf{\Omega}$ and if the covariance matrix $\mathbf{\Sigma}$ were completely specified, the Aitken least squares theorem would provide BLUE estimates of the b_{ji} parameters. The stochastic compartmental problem, however, does not satisfy either condition. Instead, it is proposed that the b_{ji} parameters be found iteratively according to the scheme below; such estimates will be 'efficient' in the sense that the heteroscedasticity and interdependence of the observations is considered.

The logic of the following procedure is two-fold. One iterative loop starts by initially assuming a spherical covariance matrix, $\mathbf{\Sigma} = \mathbf{I}$, to estimate the parameters. These estimates may be used to improve the estimate of $\mathbf{\Sigma}$, which in turn may be used for improved parameter estimation by an Aitken method. This loop is repeated until the parameter estimates converge. The problem of an incompletely specified $\mathbf{\Sigma}$ matrix is thus tackled.

The nonlinearity of the regression model may be handled by inducing an iteration process within every cycle of the above loop. This second iterative procedure, as outlined in section 4.1, consists of the modified Gauss-Newton algorithm extended to minimize the generalized sum of squares of a specific $\mathbf{\Sigma}$ matrix.

In outline form, let ${}_{i}\mathbf{\Sigma}$ be the ith estimate of $\mathbf{\Sigma}$ with ${}_{o}\mathbf{\Sigma} = \mathbf{I}$. Similarly let ${}_{i}\mathbf{\Omega}$ be the ith estimate of $\mathbf{\Omega}$. The estimation procedure then consists of the following steps:

(1) Holding ${}_{o}\mathbf{\Sigma}$ fixed, iterate for the parameter estimates ${}_{1}\mathbf{\Omega}$ by the modified Gauss-Newton algorithm.

(2) Substitute the ${}_{1}\mathbf{\Omega}({}_{k}\mathbf{\Omega})$ estimates into the matrix β and

(a) find its latent roots, α_i , and vectors, $\mathbf{F}^i$,

(b) using (a), find the $p_{ji}(t)$ and $a_i(t)$ parameters from equations (7) and (13), respectively, and

(c) using (b), find the new estimated covariance matrix ${}_{1}\mathbf{\Sigma}({}_{k}\mathbf{\Sigma})$ according to (23).

(3) Iterate for new parameter estimates ${}_{2}\mathbf{\Omega}({}_{k+1}\mathbf{\Omega})$ using ${}_{1}\mathbf{\Sigma}({}_{k}\mathbf{\Sigma})$ in formula (18) with ${}_{1}\mathbf{\Omega}({}_{k}\mathbf{\Omega})$ as the initial values.

(4) Repeat steps 2 and 3 obtaining ${}_{i}\mathbf{\Sigma}$ and ${}_{i}\mathbf{\Omega}$ estimates successively until the process converges.

The convergence properties of this two-stage procedure will not be investigated in the present work. However, in practice, as also illustrated by the examples in section 4, convergence has always been attained to numerical satisfaction.

5. EXAMPLES OF ESTIMATION PROCEDURE

In this section the above estimation procedure is illustrated with two examples. One is a simulation with known parameter values and the other consists of data from the application described in Section 2.

5.1. *Example of simulated data*

Consider first simulated data from the compartmental system represented by Figure 4. The data were generated by choosing parameter values $b_{21} =$

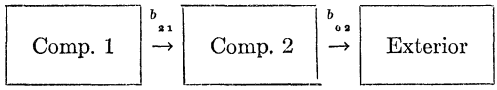

FIGURE 4

0.125 and $b_{02} = 0.250$, and initializing $N_1(0) = 4000$ and $N_2(0) = 0$. Table 1 contains a realization of the above stochastic process at 40 time points.

The matrix β corresponding to the above system is

$$\beta = \begin{bmatrix} b_{21} & -b_{21} \\ 0 & b_{02} \end{bmatrix}.$$

TABLE 1

DATA OF EXAMPLE 1

t	$N_T(t)$	t	$N_T(t)$	t	$N_T(t)$	t	$N_T(t)$	t	$N_T(t)$
1	3949	9	2175	17	917	25	335	33	127
2	3799	10	1968	18	819	26	288	34	112
3	3618	11	1771	19	729	27	258	35	102
4	3399	12	1591	20	643	28	232	36	92
5	3147	13	1448	21	562	29	219	37	81
6	2883	14	1303	22	509	30	185	38	70
7	2653	15	1147	23	441	31	161	39	62
8	2418	16	1012	24	386	32	147	40	55

and simple matrix algebra reveals that

$$\alpha_1 = b_{21}, \qquad \alpha_2 = b_{02},$$

$$f_{11} = f_{12} = F_{22} = -F_{21} = 1, \qquad f_{21} = F_{12} = 0,$$

$$f_{22} = F_{11} = |F| = (b_{21} - b_{02})/b_{21}.$$

Substituting the above into (7) and (14), it follows that

$$a(t) = (b_{02} - b_{21})^{-1}(b_{02}e^{-b_{21}t} - b_{21}e^{-b_{02}t}),$$

from whence the mean and covariance kernel of the random variable $N_T(t)$ are given by

$$\mu_T(t_a) = 4000a(t_a), \quad \sigma_{ab} = 4000a(t_b)[1 - a(t_a)].$$

Assuming now that the parameters are unknown, we judiciously select initial parameter estimates, ${}_0b_{21} = 0.125$ and ${}_0b_{02} = 0.250$, and iterate for the least squares estimates (step 1 of the estimation procedure). The derived estimates,

$_1b_{21}$ and $_1b_{02}$, are then substituted into the covariance matrix (step 2) from whence subsequent estimates, $_2b_{21}$ and $_2b_{02}$, are again obtained by Gauss-Newton iteration (step 3). The procedure is repeated until these estimates converge.

Table 2 summarizes the results of the estimation procedure. Note the

TABLE 2

PARAMETER ESTIMATION OF EXAMPLE 1

Iteration	$_ib_{21}$ ± estimated std. dev.	$_ib_{02}$ ± estimated std. dev.	s^2
0	0.12500	0.25000	
1	0.12547 ± .00092	0.24454 ± .00312	82.016
2	0.12561 ± .00527	0.24419 ± .01851	1.063
3	0.12561 ± .00527	0.24419 ± .01851	1.063

very rapid convergence of the $_ib$ estimates. The procedure also provides an indicator of goodness of fit. Assuming the model to be true, the random variable s^2 is approximately distributed as $\bar{\chi}^2/n$; hence $s^2 = 1.063$ indicates an acceptable fit.

Theoretically, the convergence is independent of the starting values for well-behaved surfaces. In the present simulation, the following 7 alternative sets of initial values were also used for the first iteration: (.2500, .5000), (.0625, .5000), (.0625, .1250), (.0312, .0625), (.0312, 1.000), (.5000, 1.000), and (.5000, .0625). It is gratifying that the estimation cycle of Table 2 was reproduced in each case.

Another noteworthy fact is the difference in the standard deviations. As previously observed, the Aitken estimates are BLUE for a linear model with known covariance matrix. In the present simulation, with the parameters and hence the covariance matrix determined, the standard deviations of the Aitken estimates are given by

$$\mathbf{\Sigma}_{\Omega\Omega} = [\mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}]^{-1}$$

to be $\sigma_{b_{21}} = 0.00488$ and $\sigma_{b_{02}} = 0.01806$. As expected, any other unbiased estimates have a greater variance; in particular the variability of the ordinary least squares (OLS) estimates, $\tilde{\Omega}$, is calculated from

$$\mathbf{\Sigma}_{\tilde{\Omega}\tilde{\Omega}} = [\mathbf{G}^T\mathbf{G}]^{-1}\mathbf{G}^T\mathbf{\Sigma}\mathbf{G}[\mathbf{G}^T\mathbf{G}]^{-1}$$

to be $\sigma_{\tilde{b}_{21}} = 0.00528$ and $\sigma_{\tilde{b}_{02}} = 0.02407$. Note from Table 2 that the recommended estimation procedure estimates the standard deviations of the parameters to be $S_{b_{21}} = 0.00527$ and $S_{b_{02}} = 0.01851$ which are close to the above $\sigma_{b_{21}}$ and $\sigma_{b_{02}}$. However the OLS estimates of the standard deviations, by failing to recognize the interdependence of the observations, use the improper law

$$\mathbf{\Sigma}_{\tilde{\Omega}\tilde{\Omega}} = \sigma^2[\mathbf{G}^T\mathbf{G}]^{-1}$$

and thereby seriously underestimate the variability. Iteration 1 of Table 2 gives these improper OLS estimates as $S_{\hat{b}_{21}} = 0.00092$ and $S_{\hat{b}_{02}} = 0.00312$. In summary, experimenters who use ordinary least squares estimation in stochastic compartmental problems are led to believe such estimates are exceptionally significant when in fact such estimates may be shown inferior to those of the recommended iterative estimation.

5.2. *Animal science application*

As a second example, Table 3 contains data on the passage of beads

TABLE 3

DATA ON BEAD RETENTION IN SHEEP 148

t	$N_T(t)$	t	$N_T(t)$	t	$N_T(t)$	t	$N_T(t)$	t	$N_T(t)$	t	$N_T(t)$
0.25	3989	2.75	3826	5.25	3589	7.75	3374	10.50	3011	14.00	2774
0.50	3970	3.00	3813	5.50	3570	8.00	3347	10.75	2976	14.25	2766
0.75	3954	3.25	3795	5.75	3563	8.25	3318	11.25	2941	14.50	2758
1.00	3935	3.50	3778	6.00	3556	8.50	3250	11.50	2930	14.75	2744
1.25	3931	3.75	3761	6.25	3542	8.75	3228	12.25	2911	15.00	2736
1.50	3905	4.00	3705	6.50	3531	9.00	3202	12.50	2891	15.25	2727
1.75	3888	4.25	3662	6.75	3503	9.25	3179	12.75	2866	15.50	2714
2.00	3872	4.50	3629	7.00	3473	9.75	3127	13.00	2839	16.00	2701
2.25	3864	4.75	3622	7.25	3450	10.00	3105	13.25	2821	16.25	2696
2.50	3832	5.00	3599	7.50	3391	10.25	3062	13.75	2796		

through the gastrointestinal tract of a sheep. Similar experiments have been conducted by Blaxter *et al.* [1956] whose findings are well received among animal scientists and, indeed, constitute the state-of-the-art in the above-mentioned modelling. At time $t' = 0$, 4000 indigestible plastic beads were placed into the rumen of the sheep. The sheep were fed every 6 hours and their feces were also collected then and analyzed for bead passage. The transformed argument t of Table 2 represents the argument t' in days less a 4-day fixed transit time or 'time delay,' i.e., $t = t' - 4$.

TABLE 4

PARAMETER ESTIMATION OF SHEEP 148

Iteration	${}_ib_{21} \pm$ estimated std. dev.	${}_ib_{02} \pm$ estimated std. dev.	s^2
1	.0290 ± .0005	0.6580 ± 0.0654	
2	.0218 ± .0015	5.5656 ± 1.8699	1.523
3	.0239 ± .0018	2.1988 ± 0.9871	1.708
4	.0231 ± .0017	3.6266 ± 1.6228	1.615
5	.0235 ± .0017	2.8129 ± 1.2981	1.636
6	.0234 ± .0017	3.1611 ± 1.4416	1.623

5.2.1. *Two sequential compartment model*

Assuming initially the model of Figure 4 for the data, the mean value function and compartmental covariance kernel, say $\mathbf{\Sigma}_c$, are given above in section 5.1. The complete covariance matrix of $N_T(t)$, however, includes two other kernels in addition to the compartmental kernel; one is due to the 'end-period' error recognized by Blaxter *et al.* [1956], and the other is due to some unfortunate mastication of the beads by the sheep. Subsequent experimentation will be designed to practically eliminate both of these latter errors, hence their form is not presented at this time. For the present data, then, the complete covariance matrix, $\mathbf{\Sigma}_T$, is the sum of the three components which are assumed independent, i.e.,

$$\mathbf{\Sigma}_T = \mathbf{\Sigma}_c + \mathbf{\Sigma}_e + \mathbf{\Sigma}_m .$$

The estimation procedure of Result 3 is now employed using the matrix $\mathbf{\Sigma}_T$ in the place of the previous $\mathbf{\Sigma}$.

Table 4 lists the cycles of the procedure. The fit is poor ($s^2 = 1.6$) but it is within reason for biological data. Inasmuch as current methodology uses ordinary least squares estimates, the fact that the final estimates differ considerably from the OLS estimates is noteworthy. The OLS estimates are 0.0290 and 0.6580 while the terminal estimates of the above procedure are approximately (by extrapolation) 0.0234 and 3.07; in another light one parameter estimate decreased by 19% and the other increased by an incredible 370%. In the event one used the compartmental covariance kernel $\mathbf{\Sigma}_c$ alone, the final parameter estimates $\hat{b}_{21} = 0.0244$ and $\hat{b}_{02} = 2.552$ are close to the above terminal estimates but again far apart from the OLS estimates.

Also, as in the simulated data example, the estimated standard deviations are deceptively low in OLS estimation. The coefficients of variation for the parameters in OLS are 0.017 and 0.099 compared to 0.073 and 0.456 in the recommended procedure.

5.2.2. *Other models*

Animal scientists proposed two other compartmental models which were tested on the sheep data. The first was a three sequential compartment model, where a third compartment was added to Figure 3, and the other was a model of two compartments in equilibrium, where $b_{02} = 0$ in Figure 2. Only the results are presented below; the algebraic detail may be derived from the general formulations of sections 3 and 4.

The three-compartment model was fitted to the data of Table 3 by ordinary least squares with resulting parameter estimates

$$\hat{b}_{21} = 0.0294 \pm 0.0005, \hat{b}_{32} = 0.6265 \pm 0.0673, \hat{b}_{03} = 16{,}384. \pm 3.401 \times 10^7,$$

and with no appreciable reduction in the error mean square. Clearly the astronomical turnover of the third compartment indicates the physical absence of such a compartment; the model was thus rejected in favor of the previous two-compartment system.

The other alternative model was likewise fitted by ordinary least squares with parameter estimates $\hat{b}_{01} = 0.0093$, $\hat{b}_{21} = 7.54$, and $\hat{b}_{12} = 14.55$, and with an increase in the error mean square. The magnitude of the crossflow was judged excessive, and this model was also discarded.

In summary, we too are led to accept the two sequential compartment model proposed by Blaxter. It seems clear, however, that the stochastic considerations contribute (1) improved parameter estimates, (2) more realistic error estimates, and (3) a goodness-of-fit test.

ACKNOWLEDGMENTS

This particular application of compartmental analysis described in section 2 was formulated during discussions with Dr. W. C. Ellis, Department of Animal Science, Texas A&M University. We are indebted to him for his interest and guidance of the biological aspects of the paper. We are also indebted to our colleague, Dr. W. B. Smith, for many stimulating discussions about the problem, and to NASA for partial support under grant NASA NGR 44-001-097.

REFERENCES

Bailey, N. T. J. [1964]. *The Elements of Stochastic Processes with Applications to the Natural Sciences.* Wiley, New York.

Bartholomay, A. F. [1958]. Stochastic models for chemical reactions: I. Theory of the unimolecular reaction process. *Bull. Math. Biophys. 20*, 176–90.

Bartlett, M. S. [1966]. *An Introduction to Stochastic Processes with Special Reference to Methods and Applications.* 2nd Edn. University Press, Cambridge.

Beauchamp, J. J. and Cornell, R. G. [1966]. Simultaneous nonlinear estimation. *Technometrics 8*, 319–26.

Beauchamp, J. J. and Cornell, R. G. [1968]. Simultaneous estimation by partial totals for compartmental models. *J. Amer. Statist. Ass. 63*, 573–83.

Beauchamp, J. J. and Cornell, R. G. [1969]. Spearman simultaneous estimation for a compartmental model. *Technometrics 11*, 551–60.

Berman, M. and Schoenfeld, R. [1956]. Invariants in experimental data on linear kinetics and the formation of models. *J. Appl. Phys. 27*, 1361–70.

Bernard, S. R. Shenton, L. R., and Uppuluri, V. R. R. [1965]. Stochastic models for the distribution of radioactive material in a connected system of compartments. *Proc. 5th Berkeley Symp. Math. Stat. Prob. IV*, 481–510.

Blaxter, K. L., Graham, N. M., and Wainman, F. W. [1956]. Some observations on the digestibility of food by sheep, and on related problems. *Brit. J. Nutr. 10*, 69–91.

Carter, M. W., Matrone, G., and Mendenhall, W. [1964]. Estimation of the life span of red blood cells. *J. Gen. Physiol. 47*, 851–58.

Chiang, C. L. [1968]. *Introduction to Stochastic Processes in Biostatistics.* Wiley, New York.

Cornfield, J., Steinfeld, J., and Greenhouse, S. W. [1960]. Models for the interpretation of experiments using tracer compounds. *Biometrics 16*, 212–34.

Fix, E. and Neyman, J. [1951]. A simple stochastic model of recovery, relapse, death, and loss of patients. *Human Biology 23*, 205–41.

Goldberger, A. S. [1964]. *Econometric Theory.* Wiley, New York.

Hartley, H. O. [1961]. The modified Gauss-Newton method for the fitting of nonlinear regression functions by least squares. *Technometrics 3*, 269–80.

Herbst, P. G. [1963]. Organizational commitment: a decision process model. *Acta Sociologica 7*, 34–45.

Hungate, R. E. [1966]. *The Rumen and its Microbes.* Academic Press, New York.

Matis, J. H. and Carter, M. W. [1970]. Multi-compartmental analysis in steady state as a stochastic process. Accepted by *Acta Biotheoretica.*

Rescigno, A. and Segre, G. [1965]. *Drug and Tracer Kinetics.* Blaisdell, Waltham, Mass.

Sheppard, C. W. [1962]. *Basic Principles of the Tracer Method.* Wiley, New York.

Sheppard, C. W. and Householder, A. S. [1951]. The mathematical bases of the interpretation of tracer experiments in closed steady state systems. *J. Appl. Phys. 22,* 510–20.

Whipple, H. E. and Hart, H. E. (Eds.) [1963]. Multi-compartment analysis of tracer experiments. *Ann. N. Y. Acad. of Sciences 108,* 1–338.

Zilversmit, D. G., Entenman, C., and Fishler, M. C. [1943]. On the calculation of turnover time and turnover rate for experiments involving the use of labelling agents. *J. Gen. Physiol. 26,* 325–31.

APPENDIX

(a) *Proof.* $a_i(t_b) = \sum_{r=1}^{m} p_{ri}(t_a)a_r(\Delta t)$ for $t_b = t_a + \Delta t$, $\Delta t > 0$

By definition (14)

$$\sum_{r=1}^{m} p_{ri}(t_a)a_r(\Delta t) = \sum_{r=1}^{m} p_{ri}(t_a) \sum_{k=1}^{m} p_{kr}(\Delta t).$$

Using definition (7)

$$= \frac{1}{|F|^2} \sum_{r=1}^{m} \sum_{s=1}^{m} f_{is}F_{rs}e^{-\alpha_s t_a} \sum_{k=1}^{m} \sum_{j=1}^{m} f_{rj}F_{kj}e^{-\alpha_j \Delta t}$$

$$= \frac{1}{|F|^2} \sum_{k=1}^{m} \sum_{s=1}^{m} \sum_{j=1}^{m} f_{is}F_{kj}e^{-\alpha_s t_a}e^{-\alpha_j \Delta t} \sum_{r=1}^{m} f_{rj}F_{rs} .$$

Using a fundamental result of matrix theory,

$$\sum_{r=1}^{m} f_{rj}F_{rs} = \begin{cases} |F| & \text{for } j = s \\ 0 & \text{for } j \neq s, \end{cases}$$

the above may be written

$$\sum_{r=1}^{m} p_{ri}(t_a)a_r(\Delta t) = \frac{1}{|F|} \sum_{k=1}^{m} \sum_{s=1}^{m} f_{is}F_{ks}e^{-\alpha_s t_b}.$$

The right hand side reduces to $a_i(t_b)$ by successively reapplying definitions (7) and (14), thus establishing the desired identity.

(b) *Proof.* $\sigma_{ab} = \sum_{s=1}^{m} N_s a_s(t_b)[1 - a_s(t_a)]$

The following definitions are helpful:

$E_{1,2,\cdots,m}$ = the expectation operator with respect to (the random variables) $z_1 , z_2 , \cdots , z_m ,$

$V_{1,2,\cdots,m}$ = the variance operator with respect to $z_1 , z_2 , \cdots , z_m ,$

$E_{m+1,\cdots,n|1,2,\cdots,m}$ = the conditional expectation operator with respect to $z_{m+1} , \cdots , z_n$ given the values of $z_1 , z_2 , \cdots , z_m .$

Let variables $N_i(t_a)$, $i = 1, \cdots , m$, be designated $z_1 , \cdots , z_m ,$ and $N_i(t_b)$, $i = 1, \cdots , m$, designated $z_{m+1} , \cdots , z_{2m} .$ Since $N_T(t_a)$ is a function only of the first m variables, appeal to fundamental properties of expectations

establishes the relation below:

$$E_{1,2,\cdots,m,m+1,\cdots,2m}\{N_T(t_a)\cdot N_T(t_b)\}$$
$$E_{1,2,\cdots,}\{N_T(t_a)E_{m+1,\cdots,2m|1,2,\cdots,m}[N_T(t_b)]\}.$$

By virtue of the Markov property inherent in the system, the conditional expectation on the right is equivalent to starting the process with $N_1(t_a), \cdots, N_m(t_a)$ units in the m compartments, and taking its expectation at time Δt. Formula (15) permits the right hand side to be rewritten as

$$E_{1,2,\cdots,m}\left\{\sum_{i=1}^{m} N_i(t_a)\sum_{i=1}^{m} N_i(t_a)a_i(\Delta t)\right\}.$$

Utilizing expectation properties again, the above simplifies to

$$\sum_{i=1}^{m} a_i(\Delta t)V_i[N_i(t_a)] + \sum_{i>j}^{m}\sum [a_i(\Delta t) + a_j(\Delta t)]\ \text{cov}\ [N_i(t_a), N_j(t_a)]$$
$$+ \sum_{i=1}^{m} a_i(\Delta t)\{E_i[N_i(t_a)]\}^2 + \sum_{i>j}^{m}\sum [a_i(\Delta t) + a_j(\Delta t)]E_i[N_i(t_a)]E_j[N_j(t_a)]$$

Substituting in equations (10) and (19), one arrives at the following expression:

$$\sum_{i=1}^{m} a_i(\Delta t)\sum_{s=1}^{m} N_s p_{is}(t_a)[1 - p_{is}(t_a)]$$
$$- \sum_{i>j}^{m}\sum [a_i(\Delta t) + a_j(\Delta t)]\sum_{s=1}^{m} N_s p_{is}(t_a)p_{js}(t_a) + \sum_{i=1}^{m} a_i(\Delta t)\left\{\sum_{s=1}^{m} N_s p_{is}(t_a)\right\}^2$$
$$+ \sum_{i>j}^{m}\sum [a_i(\Delta t) + a_j(\Delta t)]\left\{\sum_{s=1}^{m} N_s p_{is}(t_a)\right\}\left\{\sum_{s=1}^{m} N_s p_{js}(t_a)\right\}.$$

Combining the first two terms and the last two, the above becomes

$$\sum_{s=1}^{m} N_s\left[1 - \sum_{i=1}^{m} p_{is}(t_a)\right]\left[\sum_{i=1}^{m} a_i(\Delta t)p_{is}(t_a)\right]$$
$$+ \sum_{i=1}^{m} a_i(\Delta t)\left[\sum_{s=1}^{m}\sum_{r=1}^{m} N_s N_r p_{is}(t_a)p_{ir}(t_a) + \sum_{s=1}^{m}\sum_{r=1}^{m}\sum_{j\neq i}^{m} N_s N_r p_{is}(t_a)p_{jr}(t_a)\right].$$

Simplifying the second term, one has

$$\sum_{s=1}^{m} N_s\left[1 - \sum_{i=1}^{m} p_{is}(t_a)\right]\left[\sum_{i=1}^{m} a_i(\Delta t)p_{is}(t_a)\right]$$
$$+ \left[\sum_{i=1}^{m} N_i \sum_{s=1}^{m} a_i(\Delta t)p_{is}(t_a)\right]\left[\sum_{r=1}^{m} N_r \sum_{j=1}^{m} p_{jr}(t_a)\right].$$

Applying identity (21), the above reduces to

$$\sum_{s=1}^{m} N_s a_s(t_b)[1 - a_s(t_a)] + \sum_{s=1}^{m} N_s a_s(t_b)\sum_{r=1}^{m} N_r a_r(t_a).$$

Using definition (15), and recalling the left hand side, it follows that

$$E[N_T(t_a) \cdot N_T(t_b)] = \sum_{s=1}^{m} N_s a_s(t_b)[1 - a_s(t_a)] + E[N_T(t_a)]E[N_T(t_b)]$$

from whence, by definition,

$$\sigma_{ab} = \text{cov}\,[N_T(t_a), N_T(t_b)] = \sum_{s=1}^{m} N_s a_s(t_b)[1 - a_s(t_a)].$$

Received March 1970, *Revised September* 1970

DISCUSSION

M. E. Wise (University of Leiden):

I have become an active opponent of models based on homogeneous compartments for interpreting physiological tracer data, which lead to linear kinetics. I would assume that a tracer mixes quickly and homogeneously with blood, but not with soft tissue, muscle, or bone or other organs. This is a minority view-point,[1] but there is now some uneasiness among many users of these models, and this shows up again and again in the 500 page book of the latest symposium (at Oak Ridge, edited by Bergner and Lushbaugh [1967]).

Matis and Hartley have produced a method that works well with simulated data, but they too are obviously having difficulties with the excretion data on sheep. However, they have reasonable biological grounds for compartmentalising the digestive system in an unusual model that does not involve other parts of the body. Two compartments A and B are assumed (on the left and right, respectively, in their Figure 2) and there are two unknown transition rates, from A to B and from B out of the system. The reverse transition B to A is assumed not to exist. The compartments are in series, the tracer is moving through the system, and the model predicts and fits the distribution of the total time it spends there. This consists of two parts, the time interval between entry and the A to B transition, or the *first* A to B transition if there are reverse ones, and the interval between this event and departure from the system.

Nearly all observed clearance curves, such as these, can be interpreted as distributions of time intervals. These are between the tracer's entering and leaving different parts of the system, for example plasma (blood), soft tissues, and bone when the tracer is injected radioactive calcium. If these parts are homogeneous compartments, these intervals (or they may be called passage or sojourn times) will be distributed as negative exponentials. In Matis and Hartley's case they are actually binomial, and closer to exponentials the larger the number of beads (which correspond to the tracer). But we can explain why so many clearance curves fit negative powers of time as well or better, and more simply, than they fit negative exponentials if these time interval distributions take a different form. They have to be like those for a drifting cloud of Brownian particles moving from A to B according to Einstein's law, but through a *mixed* medium, i.e., with varying rates of spreading out (of diffusion in the classical case) compared with the rate of drift (Wise *et al.* [1968]). This mixed random walk in series distribution has the form

$$y(t) = At^{-W} \exp\left[-\varphi\left(\frac{t}{\mu} + \frac{\mu}{t}\right)\right]. \qquad (1)$$

When $W = \frac{3}{2}$ this becomes the well-known distribution of first passage times. For mixed media, $W > \frac{3}{2}$, and various kinds of superposition, as can occur if the tracer has left the part being observed and returned to it many times, can yield almost any smaller negative power W down to about 0.1.

[1] A group at Kings College Hospital, London, have been criticising these analyses, almost alone, for a number of years (Anderson *et al.* [1963]). See also Neuman *et al.* [1968].

The beads in the sheep's gut may or may not undergo this kind of mixed random walk in series. But there is one obvious consequence for Matis and Hartley's general model and procedure when there are several compartments in series, say 4. Then the observed distribution of total time in the system will consist of the sum of just 4 random variables; these will be more or less of the same type, and positively skew. Surely then this distribution will approach a limiting type—something like a log normal one—or at any rate one with fewer parameters than are in the model. This could lead to trouble, just as in any situation where there are too many unknown parameters for the amount of numerical data available.

I would like, however, to support strongly another general feature of their method. The output is a cumulative distribution. This is fitted with success by least squares, as the simulation shows, with major corrections for the correlations between successive observations. This could be very valuable. For example, observed total radioactivity in the body is subject to Poisson fluctuations which will be independent, whilst random uncorrelated fluctuations in the excretion rate produce correlated ones in the observed output.

In reanalyses of published clearance curves (Wise [1971]) the results from cumulative data were much less consistent than those from direct measurements of blood or urine activity. It looks as if the Matis and Hartley type of correction was not made but should have been. The example in Figure 1 shows how wildly in error an uncorrected least squares fit to this kind of cumulative data can be.

The corresponding distribution of time to excretion in the sheep is obviously of interest. This is obtained at once by differencing Table 3, which yields the numbers of beads excreted in successive 6-hour periods. This does not yield a negative power of time, but it does reveal a near-daily rhythm in the excretion pattern. It is surprising that the biologists concerned seem not to have been interested in this.

Figure 2 shows the ratios of numbers of beads excreted in successive 6-hour periods, plotted on a log scale. Clearly the rate of excretion very often increases rapidly from the first to the second quarter of the day, in fact there were 10 increases and one decrease from the 12 first quarters: and there is usually an increase from the third quarter too. Hence there is very often a 24-hour rhythm—not quite always—and often a 12-hour rhythm: possibly this is related to feeding every 6 hours.

Incidentally there is a new journal that is devoted to biological rhythms and cycles of all kinds (Tromp [1970]).

When this daily rhythm has been sorted out, the methods of this paper should certainly be adequate and efficient for assessing the model for a sheep's gut. I think there is not as yet convincing evidence, from the data, for or against it.

Finally I would like to mention one or two more general issues. In interpreting physiological tracer data, and much else in biometry, a good method of analysis certainly yields a better answer than a less good one. But the choice of model makes far more difference, whether you are concerned with this or only with an empirical answer. And so far biometricians have tended to leave this choice to biologists (or physiologists, clinicians, bio- or medical physicists, etc.) even when it is a mathematical model. One realises that methodology has to be a full time occupation for many. But there should be more work on both clean and dirty data (cf. Finney's plea at this congress) to sort out the best mathematical description; and even this should be kept clearly separate from any underlying biological or biophysical model.

Until this is done more often, quantitative biology in its widely different forms (anything involving tracers, biological cycles, much of neurophysiology, etc. etc.) will by and large stay separate from biometry[2] and still further away will be "Ein neues goldenes Zeitalter . . in dem die verschiedensten Fachdisziplinen friedlich unter dem grossen Dach der Biometric Society zusammenleben" (Schneider [1971])[3]. Is this a consequence of the

[2] See also Skellam [1969].

[3] Summarizing some reactions at our Society's foundation.

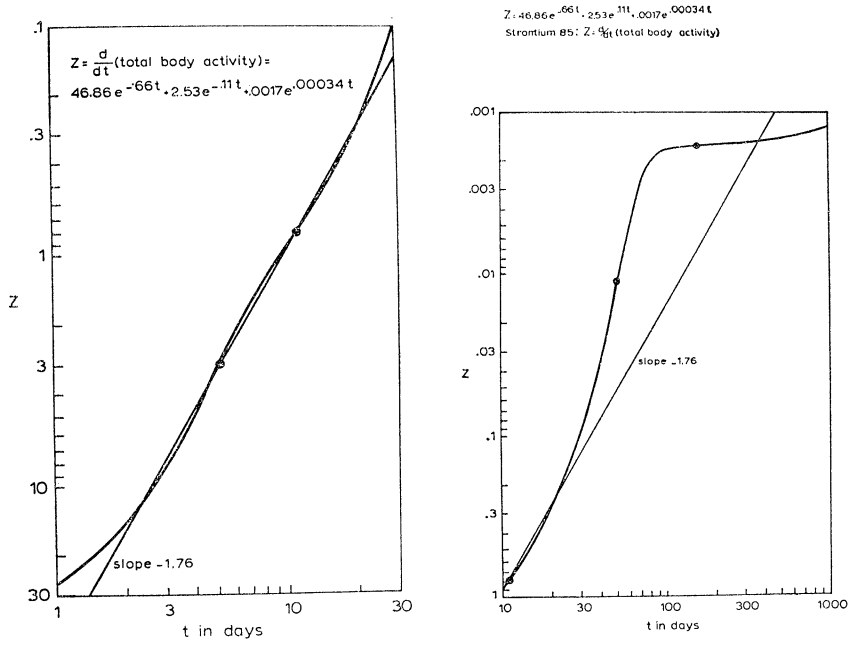

FIGURE 1

REPLOT OF A COMPARTMENTAL FIT TO A STRONTIUM RETENTION CURVE FOR A PATIENT (SARGENT *et al.* [1965] TABLE 2, 6TH LINE FROM THE BOTTOM). THE FITTED SUM OF 3 EXPONENTIALS IS DIFFERENTIATED TO GIVE THE RATE OF EXCRETION Z, AND LOG Z IS PLOTTED AGAINST LOG t. THE NEARLY STRAIGHT CURVE IN (A) GIVES CLEAR EVIDENCE THAT Z IS PROPORTIONAL TO A POWER OF TIME, t^{-W}, UP TO AT LEAST 20 DAYS. CURVE (B) SHOWS THE CONTINUATION OF Z AND WIDE AND WILD DEVIATIONS FROM THE STRAIGHT LINE, I.E., AT 50 DAYS THE FITTED EXCRETION RATE IS 5 TIMES TOO SMALL IF THE POWER FUNCTION CONTINUES TO FIT. IT DOES SO CONTINUE ACCORDING TO MANY SETS OF DIRECT OBSERVATIONS (SEE E.G. J. H. MARSHALL IN BERGNER AND LUSHBAUGH [1967] P. 451–68)

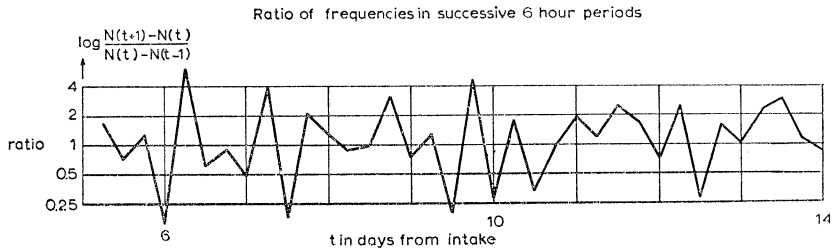

FIGURE 2

RATIOS OF NUMBERS OF BEADS EXECRETED IN SUCCESSIVE 6-HOUR PERIODS, PLOTTED AGAINST THE TIME AT THE MIDDLE OF THE 12 HOURS. CALCULATED FROM MATIS AND HARTLEY'S TABLE 3. BEYOND 14 DAYS THE PERIODS BETWEEN COLLECTIONS WERE SOMETIMES LONGER THAN 6 HOURS. THE DAILY RHYTHM IS BECOMING LESS REGULAR AFTER 10 DAYS, BUT DOES NOT DISAPPEAR.

very success of our subject in providing a service for research workers in fields like agriculture where little depended on the underlying models or these could be kept very simple?

REFERENCES

Anderson, J., Osborn, S. B., Tomlinson, R. W. S., and Weinbren, I. [1963]. Some applications of power law analysis to radiosotope studies in man. *Phys. Med. Biol. 8*, 287–95.

Bergner, E. E. and Lushbaugh, C. C., Eds. [1967]. Compartments, pools, and spaces in medical physiology. U. S. Atomic Energy Commission, Symposium series 11.

Neuman, W. F., Terepka, A. R., Canas, F., and Tippet, J. T., [1968]. The cycling concept of exchange in bone. *Calc. Tiss. Res. 2*, 262–70.

Sargent, T., Linfoot, J. A., and Isaac, Elsa C. [1965]. Whole body counting of ^{47}Ca and ^{85}Sr. Report UCRL 16246. Lawrence Radiation Lab., Berkeley, California.

Schneider, B. [1971]. Presidential address. Biometrie in den 70er Jahren. *Biometrics 27*, 264–7.

Skellam, J. G. [1969]. Models, inference, and strategy. *Biometrics 25*, 457–75 (in particular p. 464, middle paragraph).

Tromp, S. W., Ed. [1970]. *J. Interdisciplinary Cycle Res.*, Swets and Zeitlinger N. V., Amsterdam.

Wise, M. E. [1971]. The evidence against compartments (Abstract). *Biometrics 27*, 262.

Wise, M. E., Osborn, S. B., Anderson, J., and Tomlinson, R. W. S. [1968]. A stochastic model for turnover of radiocalcium based on the observed power laws. *Math. Biosci. 2*, 199–224.

See also the reply to a letter on 'Calcium kinetics: the philosophy and practice of science' by J. S. Beck and A. Rescigno [1970]. *Phys. Med. Biol. 15*, 567–8, and the abstracts in: *Brit. J. Radiology* [1968] *41*, 953–4.

The authors replied briefly at the meeting and subsequently more fully in writing as follows:

Lest anyone feel that the present paper advocates the indiscriminate adoption of the compartmental model, we welcome the comments of Dr. Wise as a word of caution. His comments also raise several pertinent issues which call for brief remarks:—

May we first note that the intent of the paper was to provide a stochastic version of compartmental analysis which is already widely used in a deterministic context. However, we *do* of course take the responsibility for 'justifying' the model in the situations where we apply it. We are gratified that Dr. Wise seems to 'support strongly' the *methods* proposed for the model. At this time we would like to make a brief case also for the relevancy of the compartmental *model* itself in regard to clearance data.

The classical compartmental model assumes 'the random appearance and disappearance of molecules' within the compartments (Zilversmit *et al.* [1943]). Hence, as pointed out, one must assume instant mixing and also the inability of the system to distinguish between 'old' and 'newly'-introduced units. Both conditions are roughly satisfied with indigestible plastic beads. However, they are not true in many other cases of passage data as, for example, in the passage of hay particles which undergo physical alteration in the rumen. In such cases the transfer probability coefficients, or b_{ij}, are functions of the age of the unit.

Suppose, as suggested by Wise, that such age dependency induces a gamma distribution of lifetimes rather than the exponential implied in the above assumptions. This gamma random variable is readily incorporated into the compartmental model by the artifice of a sequence of irreversible subcompartments with identical internal flowrates whilst a transfer to other compartments only occurs from the last subcompartment. Since the internal rates are known to be identical, no new parameters are introduced but instead the corresponding eigenvalues of the β matrix are equal. According to well-known theory, the solution of the differential equations introduces powers of t, and equation (7) in the paper

above is generalized to the form

$$p_{ki}(t) = \sum_{i=1}^{m} e^{-\alpha_i t} \sum_{l=1}^{\lambda_i} c_{kijl} t^{l-1}.$$

This solution not only 'explains' the success of the 'power law' approximation[1] but adds 1) the advantage of the biological concept of the compartmental model, 2) the interpretation of the coefficients as known functions of the unknown turnover rates, and 3) a simple yet complete stochastic derivation. The details of the theory together with simulations and supportive examples are contained in Matis [1970]. Could it be that many examples which purport to discredit the compartmental model are merely cases of compartmental systems with age dependency, i.e. eigenvalue degeneracy? Of course, as Dr. Wise observes, the power law does not fit well to the present data; indeed we did not expect to find age dependency in bead transfers.

Quite aside from these modelling considerations, we also wholeheartedly concur with the plea for more analysis of the basic data. Biological models are, or at least should be, evolved by a repeated see-saw process of monitoring against experimental evidence. Starting from a comparatively simple model a comparison with experimental evidence will normally point to modifications of the original model which in turn call for new experimental evidence to provide a decision between alternative model modifications and so on. Dr. Wise rightly criticizes numerous instances in the literature when a compartmental model was used as a starting point and its failure to explain experimental data was never heeded. In the present case, the first comparison with experimental data showed only a moderately good fit but at the same time provided pointers for reasonable modifications as, for example, the 'end-period' factor. However, the data did not appear to us to call for the complete rejection of the compartmental model.

Dr. Wise suggests, in light of the above considerations, that a diurnal rhythm should have been incorporated by us. Should his assumption prove true through further experimentation, it could be incorporated as a final stage of the compartmental model and should improve the goodness-of-fit statistic *if* such a phenomenon *is* present. However, we do not believe that the present data warrant the use of a diurnal rhythm model in addition to the end-period error which we refer to in the paper. That reference (and indeed more extensive experience) indicates a strong negative correlation between consecutive fecal collections and gives the natural reason for this end-period phenomenon. Note that the phenomenon is vividly illustrated by the zig-zag effect in Dr. Wise's Figure 2. It is likely that most of the 'diurnal effect' noticed by him is a reflection of the end-period factor which fact can be demonstrated in the following manner:

Table 1 is constructed by first calculating all differences $N(t + \frac{1}{4}) - N(t)$ (which Wise calls $N(t + 1) - N(t)$). The trends in the adjacent differences, or 6-hour collections, are then observed for each diurnal quarter. Note that a slight trend in the pattern of increases vs. decreases is induced by the negative correlation between successive collections. If one now takes the ratio of those successive collections, i.e. $[N(t + \frac{1}{4}) - N(t)]/[N(t) - N(t - \frac{1}{4})]$ the negative correlation is amplified even more and one would expect the original end-period

TABLE 1

TRENDS IN DIFFERENCES, $N(t + \frac{1}{4}) - N(t)$, BY QUARTERS

	1st quarter	2nd quarter	3rd quarter	4th quarter
Decrease	5	9	3	8
Increase	8	3	8	4
No Change	0	1	1	1

[1] At least for positive powers; Dr. Wise seems to prefer negative powers.

TABLE 2

TRENDS IN RATIOS, $\frac{N(t+\frac{1}{4}) - N(t)}{N(t) - N(t-\frac{1}{4})}$, BY QUARTERS

	1st quarter	2nd quarter	3rd quarter	4th quarter
Decrease	7	5	9	1
Increase	4	7	3	10
No Change	1	0	0	0

factor to be even more strongly evidenced. Table 2 lists the trends among these consecutive ratios. In light of the increased correlation, the observation of 10 increases to 1 decrease in a particular quarter is not surprising.

ADDITIONAL REFERENCE

Matis, J. H. [1970]. An example of age dependency in compartmental models. *NASA Tech. Report #4*, Institute of Statistics, Texas A&M University.