

Regression Analysis of Compartmental Models

T. L. Lai

Columbia University, New York, NY 10027

Accepted: June 24, 1985

Herein we study the problem of assessing, on the basis of noisy and incomplete observations, how much information there is in the data for model identification in compartmental systems. The underlying concept is that of an "information distance" between competing models, and estimation of this distance on the basis of the given data is discussed. Useful reduction of the dimensionality of the corresponding least squares problem is accomplished by regarding the decay rate constants as primary parameters of interest and the other parameters of the model as nuisance parameters. Estimation of the decay rate function is also discussed.

Key words: compartmental models; decay rate constants; separable least squares problems; system identification.

1. Introduction

Compartmental models are widely used in many fields of science and engineering—pharmacokinetics, biochemistry, physiology, radioactive isotopes and tracers, to name a few. The basic equations of a typical (linear) compartmental model with k compartments are

$$dx_i/dt = c_{i0} + \sum_{j=1, j \neq i}^k (c_{ij}x_j - c_{ji}x_i) - c_{0i}x_i, \quad (1)$$

$0 \leq t < \infty$, $i = 1, \dots, k$; where c_{ij} are nonnegative constants (called the "turnover rate constants"), compartment 0 denotes the environment, and $x_i = x_i(t)$ is the amount of material in compartment i at time t . Letting $f_{i0} = c_{i0}$ and $f_{ij} = c_{ij}x_j$ for $j \neq 0$, f_{ij} represents the mass flow rate to compartment i from compartment j . Under certain assumptions (cf. [3], [12])¹, integration of differential eq (1) leads to

$$x_i(t) = \beta_i + \sum_{j=1}^k \alpha_{ij} e^{-\lambda_j t}, \quad (2)$$

About the Author, Paper: T. L. Lai is with the Department of Statistics at Columbia. The work he describes was supported by the National Science Foundation and the National Institutes of Health.

$i = 1, \dots, k$; where $\lambda_1, \dots, \lambda_k$ are positive constants (called the "decay rate constants") depending only on the turnover rate constants c_{ij} , and β_i and α_{ij} are non-negative constants.

Noting that it is often not possible to have accurate and complete measurements from all the compartments, Berman and Schoenfeld [2] considered the degrees of freedom in choosing a compartmental model compatible with the data when measurements are incomplete. In particular, when one can only measure aggregate input-output characteristics so that one observes only $x(t) = \sum_{i=1}^k x_i(t)$, then one can only identify the rate constants $\lambda_1, \dots, \lambda_k$ of the model (2) from the equation

$$x(t) = B + \sum_{j=1}^k A_j e^{-\lambda_j t}, \quad (3)$$

where $B = \sum_{i=1}^k \beta_i$ and $A_j = \sum_{i=1}^k \alpha_{ij}$. In pharmacokinetic applications, Wagner [10] has shown that many basic quantities of interest can be expressed in terms of the parameters of the reduced model (3), the use of which he recommends in lieu of the full model (2) whose specification usually leads to ambiguities in these applications because of noisy and incomplete measurements.

The difficulties in model formulation and identification are compounded when the measurements are not only incomplete but also are subject to error. This leads to the question of assessing, on the basis of noisy

¹ Figures in brackets indicate literature references.

and incomplete observations, how much information there is in the data for model identification. If the amount of information is found to be inadequate for meaningful specification of the full model, then it may be more useful to work with a reduced model or directly with certain scientifically important characteristics (functions of the parameters) of the model.

Since both eqs (2) (for the individual compartments) and (3) (for the whole system) express the response as a polyexponential function of time t , the statistical problems in the present context are basically those of parameter estimation for polyexponential regression models. We discuss herein 1) the information content in the data from these regression models, and 2) estimation of model parameters and certain functions thereof.

2. Noisy Data and Information Content

Consider the polyexponential regression model

$$y_i = \beta + \sum_{j=1}^k \alpha_j \exp(-\lambda_j t_i) + \epsilon_i, \quad i = 1, \dots, n, \quad (4)$$

where the ϵ_i represent random errors. The errors ϵ_i are usually assumed to be independent random variables with zero means. Letting $\theta = (\lambda_1, \dots, \lambda_k; \alpha_1, \dots, \alpha_k, \beta)$ and

$$f_\theta(t) = \beta + \alpha_1 e^{-\lambda_1 t} + \dots + \alpha_k e^{-\lambda_k t},$$

common models for $\text{Var}(\epsilon_i)$ are:

- (i) $\text{Var}(\epsilon_i) = \sigma^2$ (constant variance error model),
- (ii) $\text{Var}(\epsilon_i) = f_\theta(t_i) \sigma^2$ (constant coefficient of variation model),
- (iii) $\text{Var}(\epsilon_i) = f_\theta(t_i) \sigma^2$ (Poisson-type error model).

Statistical methods for estimating the unknown parameters of the regression function try to "average out" the random errors in various ways. Assuming that observations are taken at equally spaced times $t_i = (i-1)\Delta$ and that $n = (2k+1)m$, the method of Lanczos [6, p. 273] and Cornell [4] uses the sample means

$$\bar{y}_r = m^{-1} \sum_{i=(r-1)m}^{rm-1} y_i \quad (r = 1, \dots, 2k+1)$$

to estimate the moving averages $\mu_r = m^{-1} \sum_{i=(r-1)m}^{rm-1} f_\theta(t_i)$ of the regression function. Replacing $\mu_r - \mu_{r-1}$ by $\bar{y}_r - \bar{y}_{r-1}$ in Prony's algebraic equation defining $\zeta_i = e^{-\lambda_i m \Delta}$, solution of the algebraic equation then gives the estimate of λ_j . This method therefore tries to average out the random errors by introducing the new parameters μ_r and using the sample means \bar{y}_r to estimate μ_r .

Since the \bar{y}_r are strongly consistent estimates of μ_r , it follows that the Cornell-Lanczos method is consistent, as was established by Cornell [4]. However, Lanczos [6] gave an example to demonstrate "surprising numerical snags which may develop on account of the exceedingly nonorthogonal behavior of exponential functions." The true regression function in Lanczos' example is

$$f_\theta(t) = 0.0951 e^{-t} + 0.8607 e^{-3t} + 1.5576 e^{-5t}. \quad (5a)$$

On the basis of 24 successive decay observations in the time interval $(0, 1.2]$ of the form 2.51, 2.04, ..., 0.07, 0.06, which are accurate up to 1/2 unit of the second decimal, and assuming knowledge of $\beta = 0$ and $k = 3$, the preceding method gives the fitted model

$$f_\theta(t) = 0 + 0.305 e^{-1.53t} + 2.202 e^{-4.45t}. \quad (5b)$$

Although the true parameters are disappointingly far from their estimates, the fitted function (5b) is remarkably close to the true function (5a), and one cannot distinguish between the two models within the errors of the measurements.

An alternative method to estimate the unknown parameters is that of least squares. The estimate $\hat{\theta}$ of θ is the value of γ that minimizes

$$S(\gamma) = \sum_{i=1}^n w_i [y_i - f_\gamma(t_i)]^2, \quad (6)$$

where the w_i are suitably chosen weights. Note that

$$S(\gamma) - S(\theta) = \sum_{i=1}^n w_i [f_\theta(t_i) - f_\gamma(t_i)]^2 + 2 \sum_{i=1}^n w_i [f_\theta(t_i) - f_\gamma(t_i)] \epsilon_i. \quad (7)$$

Moreover, if the weights w_i are so chosen that $w_i \text{Var}(\epsilon_i)$ are bounded, then

$$\sum_{i=1}^n w_i [f_\theta(t_i) - f_\gamma(t_i)] \epsilon_i = o\left(\sum_{i=1}^n w_i [f_\theta(t_i) - f_\gamma(t_i)]^2\right) \text{ w.p. } 1 \quad (8)$$

(with probability 1) as

$$d(\theta, \gamma) = \sum_{i=1}^n w_i [f_\gamma(t_i) - f_\theta(t_i)]^2 \rightarrow \infty. \quad (9)$$

The quantity $d(\theta, \gamma) (= E\{S(\gamma) - S(\theta)\})$ defined in (9) is a measure of the separation (distance) between γ and the parameter vector θ reflected by the data. When the ϵ_i are normal $N(0, 1/w_i)$ random variables, $1/2 d(\theta, \gamma)$ is the Kullback-Leibler information number and the least squares estimate coincides with the maximum

likelihood estimate. Thus, the least squares method averages out the errors ϵ_i in the weighted sums $\sum w_i [f_\theta(t_i) - f_\gamma(t_i)] \epsilon_i / d(\theta, \gamma)$ for all choices of γ under consideration. Consistency of this method under assumption (9) is therefore an immediate consequence of (7) and (8) if there are only finitely many such choices among which one is the true parameter vector, and its extension to infinite parameter spaces involves additional compactness arguments (cf. [7], [11]).

Since $f_\theta(t)$ is linear in the parameters $\beta, \alpha_1, \dots, \alpha_k$, least squares estimates of these "linear" parameters are given by the standard formulas in multiple linear regression theory for every fixed $\lambda = (\lambda_1, \dots, \lambda_k)$. For fixed λ , define

$$S^*(\lambda) = \min_{\beta, \alpha_1, \dots, \alpha_k} \sum_{i=1}^n w_i [y_i - (\beta + \alpha_1 e^{-\lambda_1 t_i} + \dots + \alpha_k e^{-\lambda_k t_i})]^2, \quad (10)$$

and the original least squares problem is reduced to that of minimizing $S^*(\lambda)$ involving only the decay rate constants $\lambda_1, \dots, \lambda_k$. This approach, suggested by Golub and Pereyra [5] and Osborne [8], has the great advantage of reducing the dimensionality of the parameter vector from $2k + 1$ to k . Using this approach to fit a two-compartment model to the Lanczos data, Varah [9] has recently shown that $S^*(\lambda_1, \lambda_2)$ (with equal weights) has ill-conditioned Hessian matrices and is very flat over a broad region containing the minimum.

We now show how a global analysis of the least squares function $S^*(\lambda)$ enables us to assess how much information there is in the data to specify the model. Not only does such analysis provide a relatively stable numerical algorithm for finding the least squares estimates of the model parameters, but it also sheds light on the range of models that are compatible with the data.

To fix the ideas, consider the Poisson-type error model $\text{Var}(\epsilon_i) = f_\theta(t_i) \sigma^2$, in which $f_\theta(t)$ is large, at least during the initial portion of the sampling times, as is often the case in tracer measurements. For large $f_\theta(t_i)$,

$$\begin{aligned} \log y_i &= \log f_\theta(t_i) + \log \{1 + \epsilon_i / f_\theta(t_i)\} \\ &\approx \log f_\theta(t_i) + \eta_i / (f_\theta(t_i))^{\frac{1}{2}}, \end{aligned} \quad (11)$$

where $\eta_i = \epsilon_i / (f_\theta(t_i))^{\frac{1}{2}}$ has mean 0 and constant variance σ^2 . This suggests that $f_\theta(t_i)$ can be estimated with small relative error when $f_\theta(t_i)$ is large. Therefore we introduce "ideal" weights of the form $w_i^* = 1 / \max\{f_\theta(t_i), C\}$, where C is some large constant, and define the actual weights

$$w_i = 1 / \max\{\tilde{f}_\theta(t_i), C\}, \quad (12)$$

where $\tilde{f}_\theta(t)$ denotes some initial estimate of $f_\theta(t)$ such

that $\tilde{f}_\theta(t)$ is proportionally close to $f_\theta(t)$ if $f_\theta(t)$ is large.

With this choice of weights w_i , we define the least squares function $S^*(\lambda)$ by (10) and study its global and local properties by using both discretization and gradient methods. The idea is to partition the k -dimensional parameter space Λ (set of all possible values for λ) into a finite number of subregions. The minimum $S^*(\lambda_D)$ in each subregion D is found by standard gradient-type (such as the Marquardt or Fletcher) algorithms. The minimum of $S^*(\lambda_D)$ over all subregions D then gives the least squares estimate $\hat{\lambda}$ ($S^*(\hat{\lambda}) = \min_{\lambda \in \Lambda} S^*(\lambda)$). Moreover, those values of $S^*(\lambda_D)$ that are proportionally close to $S^*(\hat{\lambda})$ also give a range of models compatible with the data.

Figure 1 illustrates the results of this analysis in the regression model

$$y_i = 100 e^{-t_i} + 1000 e^{-5t_i} + \epsilon_i, \quad i = 1, \dots, n, \quad (13)$$

where $n = 50$, $t_i = (0.01)i$, and the ϵ_i are independent normal random variables with zero means and $\text{Var}(\epsilon_i) = E y_i$. Here $\lambda_1 = 1$, $\lambda_2 = 5$, $\alpha_1 = 100$ and $\alpha_2 = 1000$ are the unknown parameters, and $\beta = 0$ is assumed known. The initial estimates $\tilde{f}_\theta(t_i)$ in the weights (12), where we set $C = 30$, are obtained by using the Cornell-Lanczos estimate of θ . Prior knowledge of the inequality constraints $0 < \lambda_1 \leq 4$ and $4 \leq \lambda_2 \leq 10$ is assumed, and we divide this parameter space into 24 unit squares. In 100 simulation experiments performed, we obtained results similar to those in figure 1. In figure 1 reporting one such simulation, $S^*(\lambda_D)$ is shown inside each unit square D , near the minimizing point λ_D which is represented by a small triangle. The solid triangles denote those λ_D whose $S^*(\lambda_D)$ values lie within 10% of the minimum value, which is underlined in the figure. At the true parameter vector $\lambda = (1, 5)$, $S^*(\lambda) = 51.5$, which differs from the minimum value of 47.1 by about 9%. The curves represent the contour $S^*(\lambda) = 52$, so that S^* lies within 10% of its minimum value in the region between the curves.

The wide range of models compatible with the data in figure 1 is in sharp contrast to figure 2, where in addition to the data during the time interval $[0, 0.5]$ of figure 1 we took 75 additional observations (generated by the model (13)) at equally spaced times in the subsequent period $(0.5, 1.25)$. In figure 2 there is a relatively small range of parameter vectors λ whose $S^*(\lambda)$ values are near the underlined minimum value at the least squares estimate $\hat{\lambda}$, which is remarkably close to true parameter $\lambda = (1, 5)$.

Regarding $S(\gamma) - S(\hat{\theta})$ as an estimate of the "information distance" $d(\theta, \gamma) = \sum_{i=1}^n w_i [f_\theta(t_i) - f_\gamma(t_i)]^2$, we can use it to assess the compatibility of the model f_γ with the data. To illustrate this idea, consider the 14 models repre-

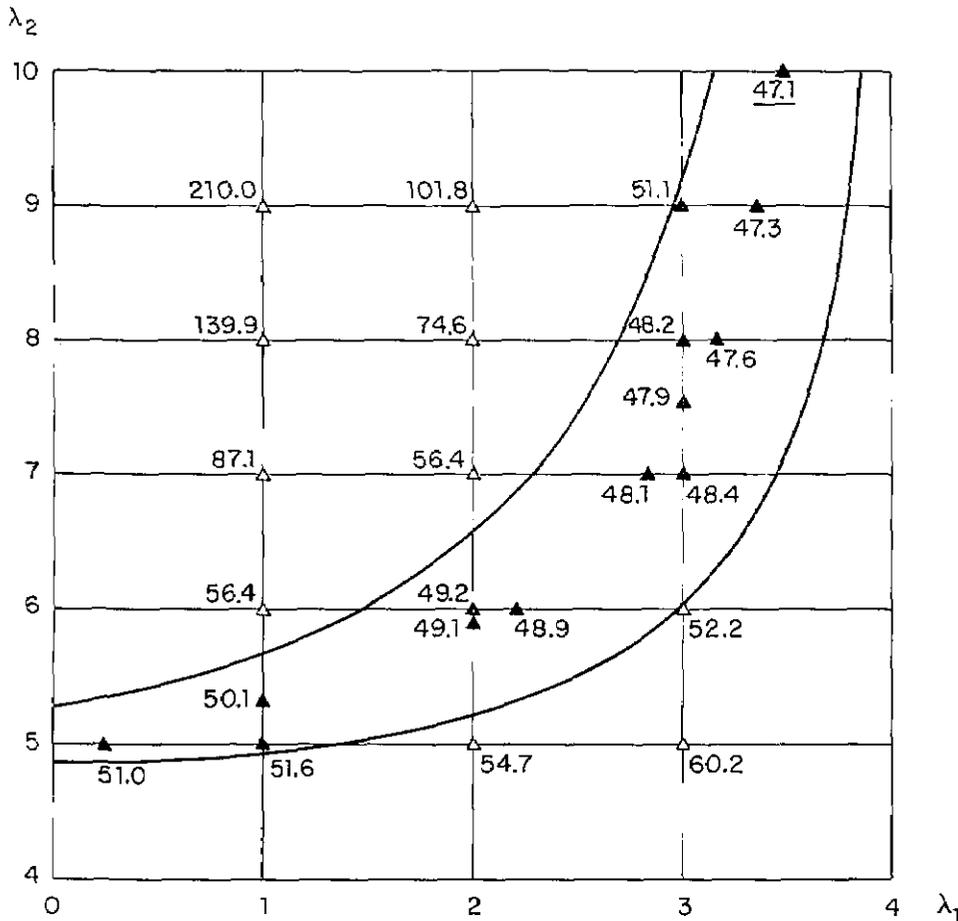


Figure 1.

sented by solid triangles in figure 1, whose S^* values lie 10% of the minimum value of 47.1. We tabulate below $S(\gamma) - S(\hat{\theta})$ for each of these 14 values of γ . In addition, the corresponding information distances $d(\theta, \gamma)$ and $d^*(\theta, \gamma) = \sum [f_{\theta}(t_i) - f_{\gamma}(t_i)]^2 / f_{\theta}(t_i)$ are also tabulated for comparison. Note that $d^*(\theta, \gamma)$ is remarkably close to $d(\theta, \gamma)$; moreover, $d^*(\theta, \gamma)$ is so small (in good agreement with $S(\gamma) - S(\hat{\theta})$) that $f_{\gamma}(t)$ is within 2% of $f_{\theta}(t)$ over the observed time range $0.01 \leq t \leq 0.5$ in each of the 14 cases.

$S(\gamma) - S(\hat{\theta})$	$d(\theta, \gamma)$	$d^*(\theta, \gamma)$	$S(\gamma) - S(\hat{\theta})$	$d(\theta, \gamma)$	$d^*(\theta, \gamma)$
0	0.63	0.63	1.0	2.20	2.24
0.2	3.18	3.25	1.8	1.58	1.61
4.0	1.67	1.69	2.1	2.80	2.84
0.5	2.72	2.78	2.0	2.11	2.14
1.1	5.21	5.28	3.0	1.67	1.69
0.8	3.01	3.06	4.5	0.06	0.06
1.3	1.20	1.23	3.9	0.67	0.63

In the case of normal $N(0, 1/w_i)$ random errors ϵ_i , $\exp\{-1/2(S(\gamma) - S(\hat{\theta}))\}$ is the generalized likelihood ratio for testing $H_0: \theta = \gamma$. We can also construct confidence regions for θ by using contours of the function

$S(\gamma)$, as will be shown elsewhere. In this connection, Bates and Watts [1] recently proposed another useful method involving parameter transformations to improve the standard asymptotic approximations for constructing confidence regions.

The separation of the full parameter vector θ into its linear and nonlinear components is not only of computational interest, but it also has basic statistical implications. Analogous to the preceding paragraph, in the case of independent $N(0, 1/w_i)$ random errors, $\exp\{-1/2(S^*(\lambda) - S^*(\hat{\lambda}))\}$ is the generalized likelihood ratio for testing whether λ is the true vector of decay rate constants, with $\beta, \alpha_1, \dots, \alpha_k$ as the nuisance parameters. This idea can be easily extended to simultaneous equations (2) defining k -compartment models, where we have

$$y_v(t) = \beta_v + \sum_{j=1}^k \alpha_{vj} e^{-\lambda_j t} + \epsilon_v(t), \quad v = 1, \dots, k.$$

We can similarly define

$$S^*(\lambda) = \min_{(\beta_v, \alpha_{vj})_{1 \leq v, j \leq k}} \sum_{t=1}^k \sum_{v=1}^k w_v(t) [y_v(t) - \beta_v - \sum_{j=1}^k \alpha_{vj} e^{-\lambda_j t}]^2,$$

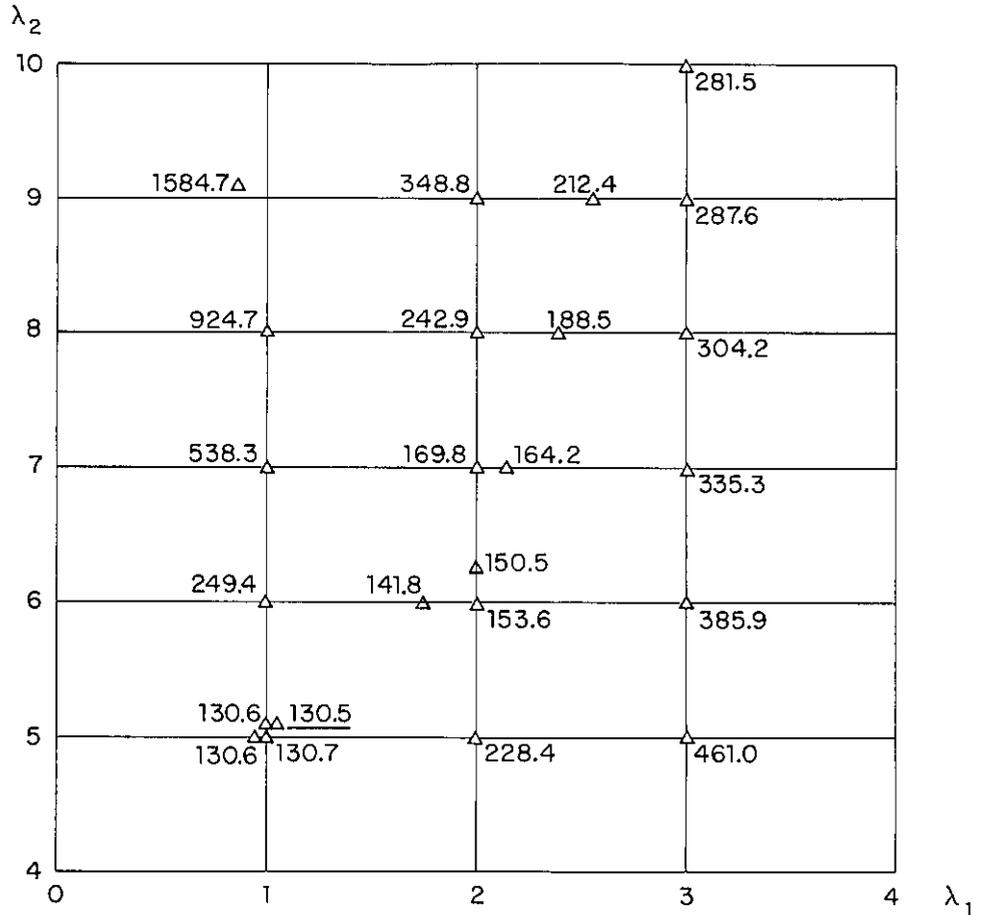


Figure 2.

and interpret $\exp\{-1/2(S^*(\lambda) - S^*(\hat{\lambda}))\}$ as a generalized likelihood ratio for testing whether λ is the true vector of decay rate constants on the basis of data from all k compartments.

3. Decay Rates and Their Estimation

Consider the polyexponential regression model (4). As illustrated in figure 1, one often encounters a wide range of rate parameter vectors $(\lambda_1, \dots, \lambda_k)$ that are compatible with one's data. In such circumstances, since there is not enough information to estimate the individual rate constants, it is more meaningful to consider them in a combined characteristic, such as the fractional rate of decay

$$r(t) = \lim_{\delta \rightarrow 0} \delta^{-1} \{1 - f_\theta(t + \delta)/f_\theta(t)\} = -(d/dt) \log f_\theta(t)$$

at different time points t of interest. Letting $\lambda_0 = 0$ and $\alpha_0 = \beta$, the logarithmic derivative of $f_\theta(t) = \sum_{j=0}^k \alpha_j e^{-\lambda_j t}$ is given by -1 times

$$r(t) = \sum_{i=0}^k p_i(t) \lambda_i, \text{ where } p_i(t) = \alpha_i e^{-\lambda_i t} / \sum_{j=0}^k \alpha_j e^{-\lambda_j t}. \quad (14)$$

Thus, $r(t)$ is a convex combination of the rate constants λ_i , with the proportional size of the i^{th} exponential term as the natural weighting factor for the decay rate λ_i .

To estimate $r(\tau)$ at a particular time point τ within the range of sampling times, we propose to balance "global information" from all sampling times (leading to weighted least squares estimates described in sec. 2) with "local information" from only the sampling times near τ . We start by using the method of section 2 on all the data to find a region Λ_0 of rate parameters $(\lambda_1, \dots, \lambda_k)$ that are compatible with the data. To choose a vector in Λ_0 that will provide the best estimate of $r(\tau)$, we note that the logarithmic derivative $-r(\tau)$ is a "local" quantity involving only sampling times near τ , and it is therefore reasonable to weight the observations not only by their variability but also by how far their sampling times are from τ , putting more weight on sampling times near τ . With this new set of weights $w_i(\tau)$, we calculate the (constrained) least squares estimate $\hat{\theta}(\tau)$ of the parameter vector θ , under the constraints $\lambda \in \Lambda_0$ and $\alpha_i \geq 0 (i = 0, 1, \dots, k)$. Substituting the unknown parameters in eq (14)

by these least squares estimates, we obtain the estimate $\hat{r}(\tau)$ of $r(\tau)$.

A detailed discussion of the procedure sketched above, together with a comparative study of this approach and the popular curve-peeling methods for compartmental analysis (cf. [12]), will be presented elsewhere.

The author thanks Dr. Hung Chen for helpful discussions.

References

- [1] Bates, D. M., and D. G. Watts, Parameter transformations for improved approximate confidence regions in nonlinear least squares *Ann. Statist.* **9**, 1151 (1981).
- [2] Berman, M., and R. Schoenfeld, Invariants in experimental data on linear kinetics and the formulation of models. *J. Appl. Phys.* **27**, 1361 (1956).

- [3] Coddington, E. A., and N. Levinson, *Theory of Ordinary Differential Equations*. McGraw-Hill, New York (1955).
- [4] Cornell, R. G. A method for fitting linear combinations of exponentials. *Biometrics* **18**, 104 (1962).
- [5] Golub, G. H., and V. Pereyra, The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate. *SIAM J. Numer. Anal.* **10**, 413 (1973).
- [6] Lanczos, C., *Applied Analysis*. Prentice Hall, Englewood Cliffs (1956).
- [7] Ljung, L., Consistency of the least squares identification method. *IEEE Trans. Autom. Contr.* **AC-21**, 779 (1976).
- [8] Osborne, M. R., Some special nonlinear least squares problems. *SIAM J. Numer. Anal.* **12**, 571 (1975).
- [9] Varah, J. M., On fitting exponentials by nonlinear least squares. *SIAM J. Sci. Stat. Comput.* **6**, 30 (1985).
- [10] Wagner, J. G., Linear pharmacokinetic equations allowing direct calculation of many needed pharmacokinetic parameters from the coefficients and exponents of polyexponential equations which have been fitted to data. *J. Pharmacokin. Biopharm.* **4**, 443 (1976).
- [11] Wu, C. F., Asymptotic theory of nonlinear least squares estimation. *Ann. Statist.* **9**, 501 (1981).
- [12] Zeirler, K., A critique of compartmental analysis. *Ann. Rev. Biophys. Bioeng.* **10**, 531 (1981).

DISCUSSION

of the T. L. Lai paper,
Regression Analysis of Compartmental Models

T.-H. Peng

Oak Ridge National Laboratory

One of the major tasks of marine geochemists is determining the uptake by the sea of CO_2 derived from the combustion of fossil fuels. Until valid models of the general circulation of the ocean are constructed, this task will have to be done with box models calibrated through use of the distribution of natural radioisotopes and transient tracers.

We need to explore how sensitive the uptake of fossil fuel CO_2 is to the basic design of these models and how the design can be improved by simultaneously fitting the distributions of several tracers. Five different 11-box thermocline circulation models of the temperate North Atlantic were constructed for this purpose.* Anthropogenic tritium, ^3He , and radiocarbon are used as tracers to calibrate these models. The temporal input functions of these tracers differ considerably from one another. So also do the geographic patterns of their inputs and their geochemistries in the sea.

Using the basic equation of the box model [e.g., eq (1) of T.L. Lai's presentation at this conference] and the finite difference method of computation for mass bal-

ance in each box, these thermocline ventilation models with differing circulation patterns were calibrated to yield a tritium distribution similar to that observed during the Geochemical Ocean Section Studies (GEOSECS) survey in 1973. These models were then run for ^3He and bomb-produced ^{14}C . While the models differ significantly in their ability to match the observed ^3He and ^{14}C distributions, these differences are not large enough to clearly single out one model as superior. This insensitivity of tracer to tracer ratio to model design is reflected by the nearly identical uptake of CO_2 by the various models. This result also suggests that the uptake of CO_2 by the sea is limited more by the rates of physical mixing within the sea than by the rate of gas exchange across the sea surface.

* Research of the application of box models for the geochemical modeling of oceans was supported jointly by the National Science Foundation's Ecosystems Studies Program under Interagency Agreement DEB 8115316 and the Carbon Dioxide Research Division, Office of Energy Research, U.S. Department of Energy, under contract DE-AC05-84OR21400 with Martin Marietta Energy Systems, Inc.