

**A HIERARCHICAL MODEL FOR SERIALY
CORRELATED DATA: ANALYSIS OF
BIOLOGICAL RHYTHM DATA**

Joel B. Greenhouse
Robert E. Kass
Department of Statistics
Carnegie Mellon University
Pittsburgh, PA 15213

C.Y. Teresa Lam
Department of Industrial & Operations Engineering
University of Michigan
Ann Arbor, MI 48109-2117

Ruey S. Tsay
Graduate School of Business
University of Chicago
Chicago, IL 60637

Technical Report 90-13

April 1990

**A Hierarchical Model for Serially Correlated Data:
Analysis of Biological Rhythm Data**

Joel B. Greenhouse,¹ Robert E. Kass¹
Teresa Lam,² and Ruey S. Tsay³

¹Department of Statistics
Carnegie Mellon University, Pittsburgh, PA 15213

²Department of Industrial and Operations Engineering
University of Michigan, Ann Arbor, MI 48109

³Graduate School of Business
University of Chicago, Chicago, IL 60637

Technical Report #476

April 25, 1990

Author's Note

Joel B. Greenhouse and Robert E. Kass are both Associate Professors in the Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213. Teresa Lam is an Assistant Professor in the Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI 48109. Ruey S. Tsay is a Professor in the Graduate School of Business, University of Chicago, Chicago, IL 60637. The authors would like to thank David B. Jarrett, M.D., of the Western Psychiatric Institute and Clinic, University of Pittsburgh for many stimulating discussions concerning biological rhythms and for the use of the data set. This work was supported in part by the NIMH CRC Grant No. MH30915 (JBG), by the National Science Foundation under Grants No. MCS-8301831 (REK) and DMS-8902177 (RST), the William S. Fishman Research Fund at the Graduate School of Business, the University of Chicago (RST), and a grant from the John D. and Catherine T. MacArthur Research Network on the Psychobiology of Depression (TL). This work was begun when all four authors were at Carnegie Mellon University.

**A Hierarchical Model for Serially Correlated Data:
Analysis of Biological Rhythm Data**

Abstract

Cortisol is a hormone secreted by the adrenal gland, and, like other physiological measurements, the concentration of cortisol in the blood over time is periodic with an approximate 24 hour cycle. For descriptive purposes, linear regression using a simple sinusoidal curve with a fixed 24 hour period is sometimes an adequate tool for analyzing data from such processes. Inferences based on regression models under the assumption of independent and identically distributed errors, however, can often mislead seriously. We have previously presented a general class of models for fitting a single biological rhythm, with use of higher order harmonic terms of one or more unknown fundamentals and ARMA processes for the errors (Greenhouse, Kass, and Tsay, 1987). Since comparative studies of changes in biological rhythms due to a change in disease state often involve many subjects, it is likely that there exists sufficient inhomogeneity among subjects that a single model would be inadequate and a model that incorporates variation across subjects is necessary. We present an application of the use of a two-stage hierarchical model to study differences in the circadian variation of plasma cortisol concentration between depressed patients and healthy controls. In our harmonic regression model the level, amplitudes, and phases are treated as randomly varying across individuals and the frequency and the ARMA parameters are taken as common to all subjects.

Key Words: ARMA Process; Depression; Harmonic Regression Model; Parametric Empirical Bayes Estimation; Plasma Cortisol.

1 Introduction

Sequential measurements of physiological processes follow patterns that typically repeat themselves approximately every 24 hours (Moore-Ede, et al., 1982). Examples in humans include the sleep-wake cycle, core body temperature, and neuroendocrine activity, such as the secretion of plasma cortisol. These patterns of measurements, called biological or circadian rhythms, are endogenous and have been observed not only in physiological systems but in behavioral activities as well. Circadian rhythms, according to Moore-Ede, et al. (1982), are the outputs of a biological system whose main function is to measure time and that the functioning of this system has important implications for physiological regulation and for health. These rhythms are surprisingly stable, in the sense of being regular and repeatable, and often pass unnoticed until they become disturbed as might occur due to a change in the environment, such as flying across several time zones resulting in "jet lag", or due to a disease state. For example, patients suffering from a major depressive disorder are consistently reported to have disturbances in their rest-activity cycle, disturbances which are often characterized by a phase advance of the sleep-wake cycle (Kupfer and Foster, 1972). Just as jet lag may arise because of an acute disruption of an individual's circadian rhythms (that is, their internal biological clock is "out of sync" with the external clock), it is now widely believed that patients with major affective disorders have disturbances in the regulation of their biological rhythms that play an important role in the clinical course and pathophysiology of their disease (Moore-Ede, et al., 1983; Sack, et al., 1987).

To better understand the role these disturbances play in the pathology and etiology of depression, psychiatrists and others are attempting to characterize changes in the biological rhythms of depressed patients. It is hoped that by observing consistent changes in these rhythms, new strategies for the treatment of affective disorders will be suggested, and new understandings of the mechanisms by which physiological systems are regulated will be obtained. Fundamental to the study of biological rhythms is the need to accurately characterize the rhythm so that changes in a rhythm from a baseline state following an intervention can be made or the assessment of abnormalities in a rhythm between a group of patients and controls can be determined. In this paper we develop statistical models for studying changes in the circadian variation of the hormone cortisol in a group of depressed patients as compared to a group of healthy control subjects. The data described here are part of a larger study conducted at the Western Psychiatric Institute and Clinic.

University of Pittsburgh, under the direction of Dr. David Jarrett.

Cortisol is a hormone secreted by the adrenal gland, and, like other physiological and behavioral measurements, the concentration of cortisol in the blood over time is periodic with an approximate 24 hour cycle. Plasma cortisol concentration reaches a peak at or near the time of waking and then falls off over the course of the day, reaching the lowest level shortly after sleep onset (see Figure 1). In a review of 14 studies of the secretory activity of cortisol, Sack, et al. (1987) found that changes in the circadian rhythm of plasma cortisol levels in depressed patients have not been conclusively established. Although most studies report an increase in the mean level of cortisol concentration in depressed patients, there is disagreement as to whether some features of the cortisol circadian rhythm are advanced. For example, in six relatively carefully controlled studies, three found a phase advance in the cortisol rhythm and three did not. These diverse results with respect to cortisol are typical of problems with studies of biological rhythms in general and are due in part to poor designs (e.g., many studies only observe subjects for 24 hours or less), to the use of inefficient statistical methods, and to the problem of substantial variation among subjects.

In the University of Pittsburgh study, blood samples were obtained from both healthy and depressed subjects every 20 minutes for approximately 48 hours. The purpose of this paper is to develop statistical methods for analyzing periodic data such as these that will allow us to make comparisons between groups of subjects while accounting for intersubject variability. We will analyze the data collected to investigate the effect of depression on the circadian variation of cortisol. In the next section we describe a model for characterizing a single rhythm based on a harmonic regression model with an ARMA process for the errors. We then generalize that model to incorporate a two-stage model with hierarchically structured parameters. At the first stage, the model is assumed to be identical for all individuals and variability among individuals is accounted for by the distribution of the parameters specified at the second stage. This approach is similar to the random-effects model for longitudinal data considered in Laird and Ware (1982). The likelihood function for the parameters is presented in section 2 and issues related to the estimation of parameters and to computation are discussed in section 3. In section 4, we present the analysis of the cortisol data for the depressed patients and controls in the Pittsburgh study. We conclude in section 5 with a discussion of the results of the data analysis and of the methodology. We note that the family of models considered in this article have a wide range of application, such as, to

the investigation of the impact of chlorofluoromethanes on stratospheric ozone (e.g., Reinsel and Tiao, 1987,) and to the analysis of gait, that is, free-speed walking on a level surface as discussed in Olshen, Biden, Wyatt and Sutherland (1989).

2 Model Specification and the Exact Likelihood Function

2.1 First-Stage Model

Statistical models for characterizing biological rhythm data, or “cosinor” analysis (Tong, 1976; Minors and Waterhouse, 1988), have generally consisted of a single sinusoidal regression function, having a *fixed* period of 24 hours, together with independent and identically distributed (i.i.d.) errors. Greenhouse, Kass and Tsay (GKT) (1987) describe a more general class of models for fitting biological rhythm data. This class contains models that include higher-order harmonics of one or more unknown fundamentals and, in addition, have ARMA processes for the errors. Specifically, the GKT model for an observation, Y_t , at time t of a rhythm for a single individual, is of the additive form

$$Y_t = f(t) + Z_t \quad (1)$$

where $f(t)$ denotes the deterministic component of the data and Z_t the stochastic component. In applications to biological rhythm data, trigonometric series are used for $f(t)$. Since the k -th harmonic term of a fundamental frequency ω can be written as

$$R \cdot \cos\{2\pi k(\omega t - \phi)\} = A \cdot \cos(2\pi k\omega t) + B \cdot \sin(2\pi k\omega t)$$

where $A = R \cdot \cos(2\pi k\phi)$ and $B = R \cdot \sin(2\pi k\phi)$, we may write $f(t)$ in (1) as

$$f(t) = \mu + \sum_{k=1}^K \left\{ A_k \cos(2\pi k\omega t) + B_k \sin(2\pi k\omega t) \right\}, \quad (2)$$

where μ is an added constant which we call the level of Y_t , ω is the fundamental frequency, and $k\omega$, ($k = 1, \dots, K$) are the harmonics of ω . Here the unknown parameters are μ , ω , the A_k 's and B_k 's, and in this form the only nonlinear parameter in $f(\cdot)$ is ω .

Further, the stochastic component, Z_t , in (1) is taken to be an autoregressive moving average, ARMA(p,q), process. That is, Z_t satisfies

$$Z_t = \sum_{i=1}^p \phi_i Z_{t-i} - \sum_{j=1}^q \theta_j a_{t-j} + a_t \quad (3)$$

where the φ_i 's and θ_j 's are real-valued parameters, and $\{a_t\}$ is a sequence of independent, identically distributed, Gaussian random variables with mean zero and variance σ^2 . Methods for specifying K , p and q for model (1)-(3) are described in Greenhouse, Kass and Tsay (1987) and illustrated with an example of the circadian variation of core body temperature in a healthy adult.

Model (1)-(3) is useful as a description of a particular circadian rhythm in a single individual. Comparative studies often involve many subjects, however, and it is likely that there exists sufficient intersubject variability that a single model of the form (1)-(3) would be inadequate. In this case, it is necessary to introduce a new model that incorporates variation across subjects of the parameters appearing in model (2).

2.2 Second-Stage Model

To accommodate the intersubject variability of cortisol secretion we generalize the GKT model by considering two-stage hierarchical or "random-effects" models, in which certain parameters in (2) are assumed to follow probability distributions. For generality, we allow subjects to belong to one of g distinct groups, though in the application of section 4 we will have $g = 2$. We use h as the index for group, $h = 1, 2, \dots, g$. We denote the j -th observation for the i -th subject in group h by Y_{ijh} . We let m_h be the number of subjects in the h -th group and let n_{ih} be the number of observations on the i -th subject. The first stage of the hierarchical model is:

$$Y_{ijh} = \mu_{ih} + \sum_{k=1}^K [A_{ih}^k \cdot \cos(2\pi k\omega_h j) + B_{ih}^k \cdot \sin(2\pi k\omega_h j)] + Z_{ijh}, \quad (4)$$

$j = 1, 2, \dots, n_{ih}$, $i = 1, 2, \dots, m_h$, ω_h is the frequency for the h -th group, and A_{ih}^k and B_{ih}^k are the coefficients of the k -th harmonic of ω_h for the i -th subject in the h -th group. In (4) we now use the subscript j in place of t to index the repeated measurements on the i -th subject. When it is clear from the context that we are referring to the h -th group, we will for simplicity denote the total number of observations for the i -th subject by n_i . For the stochastic component of (4), we assume, as in GKT, that $\{Z_{ijh}\}_{j=1}^{n_i}$ is an ARMA(p, q) process, and that $\{Z_{ijh}\}_{j=1}^{n_i}$ are all independent for $i = 1, 2, \dots, m_h$ and $h = 1, 2, \dots, g$.

At the second stage of the hierarchical model, we will assume that the μ_i 's, A_i^k 's and B_i^k 's, are jointly normally distributed, and are distributed independently across subjects. This specification provides a simple manageable description of the variation among individuals with respect to the

mean level, amplitude and phase of the circadian rhythm for the nonlinear harmonic regression model in (2). In this paper, the fundamental frequency ω_h and the ϕ 's, θ 's, and σ of the stochastic component will be treated as common to all subjects, or as "fixed" effects.

It is convenient to write Equation (4) in matrix form.

$$y_{ih} = \mathbf{X}_{ih}\beta_{ih} + z_{ih}$$

where

$$\begin{aligned} y_{ih} &= (Y_{i1h}, Y_{i2h}, \dots, Y_{in_ih})^T \\ z_{ih} &= (Z_{i1h}, Z_{i2h}, \dots, Z_{in_ih})^T \\ \beta_{ih} &= (\mu_{ih}, A_{ih}^1, B_{ih}^1, \dots, A_{ih}^K, B_{ih}^K)^T. \end{aligned}$$

and

$$\mathbf{X}_{ih} = \begin{bmatrix} 1 & \cos(2\pi\omega_h) & \sin(2\pi\omega_h) & \dots & \cos(2\pi K\omega_h) & \sin(2\pi K\omega_h) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \cos(2\pi\omega_h j) & \sin(2\pi\omega_h j) & \dots & \cos(2\pi K\omega_h j) & \sin(2\pi K\omega_h j) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \cos(2\pi\omega_h n_i) & \sin(2\pi\omega_h n_i) & \dots & \cos(2\pi K\omega_h n_i) & \sin(2\pi K\omega_h n_i) \end{bmatrix}$$

with \mathbf{X}_{ih} depending on the frequency ω_h .

Letting λ_h be a vector of the fixed effects, we can express the distributional assumptions for the two-stage hierarchical model as:

$$\begin{aligned} \text{Stage I: } y_{ih} | \mathbf{X}_{ih}, \beta_{ih}, \lambda_h &\sim N(\mathbf{X}_{ih}\beta_{ih}, \Sigma_W(\phi, \theta, \sigma^2)) \\ \text{Stage II: } \beta_{ih} | \lambda_h &\sim N(\bar{\beta}_h, \Sigma_B) \end{aligned} \tag{5}$$

where $\Sigma_W(\phi, \theta, \sigma^2)$ denotes the within-subject component of variance expressed as a function of the parameters of the stochastic component, and

$$\begin{aligned} \bar{\beta}_h &= (\bar{\mu}_h, \bar{A}_h^1, \bar{B}_h^1, \bar{A}_h^2, \bar{B}_h^2, \dots, \bar{A}_h^K, \bar{B}_h^K)^T, \\ \Sigma_B &= \text{DIAG}(\sigma_0^2, \sigma_1^2, \dots, \sigma_{2K}^2). \end{aligned}$$

It will be useful for future reference to write $\lambda_h = (\lambda_c, \lambda_h^*)^T$, where $\lambda_c = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma^2)^T$ are the parameters common to all subjects, and $\lambda_h^* = (\omega_h, \bar{\beta}_h, \Sigma_B)^T$ are the parameters common to all subjects in group h . Finally, let $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_g)$.

2.3 The Exact Likelihood Function

In this section we derive the exact likelihood function for λ . First, we will obtain the joint distribution of \mathbf{y}_h and β_h given λ_h which we express for the h -th group as

$$p(\mathbf{y}_h, \beta_h | \lambda_h) = \prod_{i=1}^{m_h} p(\mathbf{y}_{ih} | \beta_{ih}, \lambda_h) p(\beta_{ih} | \lambda_h). \quad (6)$$

We will then find the marginal distribution of \mathbf{y}_h given λ_h by integrating over β_{ih} , i.e.,

$$L(\lambda_h) \propto \prod_{i=1}^{m_h} \int p(\mathbf{y}_{ih} | \beta_{ih}, \lambda_h) p(\beta_{ih} | \lambda_h) d\beta_{ih}. \quad (7)$$

In the remainder of this section we consider the details for the derivation of the exact likelihood function for λ , $L(\lambda)$, where $L(\lambda) = \prod_{h=1}^g L(\lambda_h)$.

Let $N = \sum_{h=1}^g \sum_{i=1}^{m_h} n_{ih}$ be the total number of observations of all the subjects in all g groups. Using the results in the Appendix for the exact likelihood function of an ARMA(p,q) process, we have

$$p(\mathbf{z}_{ih} | \lambda_c) = (2\pi\sigma^2)^{-n_{ih}/2} |W_{ih}^T W_{ih}|^{-1/2} \exp\left\{-\frac{1}{2\sigma^2} \mathbf{z}_{ih}^T \Psi_{ih}^T \Psi_{ih} \mathbf{z}_{ih}\right\} \quad (8)$$

where W_{ih} and Ψ_{ih} , $i = 1, 2, \dots, m_h$ are the corresponding time series matrices given in the Appendix associated with the n_{ih} observations for the i -th subject in group h . These two matrices depend only on the time series parameter $\lambda_c = (\phi_1, \phi_2, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma^2)$ and n_{ih} .

It follows from (5) and from (8) above, that

$$p(\mathbf{y}_{ih} | \lambda_c, \mathbf{X}_{ih}, \beta_{ih}) = (2\pi\sigma^2)^{-n_{ih}/2} |W_{ih}^T W_{ih}|^{-1/2} \exp\left\{-\frac{1}{2\sigma^2} (\mathbf{y}_{ih} - \mathbf{X}_{ih}\beta_{ih})^T \Psi_{ih}^T \Psi_{ih} (\mathbf{y}_{ih} - \mathbf{X}_{ih}\beta_{ih})\right\}.$$

The joint probability density function of the data $\mathbf{y} = (\mathbf{y}_{11}^T, \dots, \mathbf{y}_{m_{11}}^T, \mathbf{y}_{12}^T, \dots, \mathbf{y}_{m_{gg}}^T)^T$ and $\beta = (\beta_{11}^T, \dots, \beta_{m_{11}}^T, \beta_{12}^T, \dots, \beta_{m_{gg}}^T)^T$ given λ and \mathbf{X} , is

$$p(\mathbf{y}, \beta | \lambda, \mathbf{X}) =$$

$$\begin{aligned}
& (2\pi\sigma^2)^{-N/2} \left[\prod_{h=1}^g \prod_{i=1}^{m_h} |W_{ih}^T W_{ih}|^{-1/2} \right] \exp \left\{ -\frac{1}{2\sigma^2} \sum_{h=1}^g \sum_{i=1}^{m_h} (\mathbf{y}_{ih} - \mathbf{X}_{ih} \boldsymbol{\beta}_{ih})^T \Psi_{ih}^T \Psi_{ih} (\mathbf{y}_{ih} - \mathbf{X}_{ih} \boldsymbol{\beta}_{ih}) \right\} \\
& \times (2\pi)^{-m/2} \left[\prod_{h=1}^g \prod_{i=1}^{m_h} |\Sigma_B|^{-1/2} \right] \exp \left\{ -\frac{1}{2} \sum_{h=1}^g \sum_{i=1}^{m_h} (\boldsymbol{\beta}_{ih} - \bar{\boldsymbol{\beta}}_h)^T \Sigma_B^{-1} (\boldsymbol{\beta}_{ih} - \bar{\boldsymbol{\beta}}_h) \right\}
\end{aligned} \tag{9}$$

where $m = (\sum_{h=1}^g m_h) \times (2K + 1)$. Letting $\hat{\boldsymbol{\beta}}_{ih} = (X_{ih}^T \Psi_{ih}^T \Psi_{ih} X_{ih})^{-1} (X_{ih}^T \Psi_{ih}^T \Psi_{ih} \mathbf{y}_{ih})$, we have

$$\begin{aligned}
& (\mathbf{y}_{ih} - \mathbf{X}_{ih} \boldsymbol{\beta}_{ih})^T \Psi_{ih}^T \Psi_{ih} (\mathbf{y}_{ih} - \mathbf{X}_{ih} \boldsymbol{\beta}_{ih}) = \\
& (\mathbf{y}_{ih} - \mathbf{X}_{ih} \hat{\boldsymbol{\beta}}_{ih})^T \Psi_{ih}^T \Psi_{ih} (\mathbf{y}_{ih} - \mathbf{X}_{ih} \hat{\boldsymbol{\beta}}_{ih}) + (\boldsymbol{\beta}_{ih} - \hat{\boldsymbol{\beta}}_{ih})^T X_{ih}^T \Psi_{ih}^T \Psi_{ih} X_{ih} (\boldsymbol{\beta}_{ih} - \hat{\boldsymbol{\beta}}_{ih}).
\end{aligned} \tag{10}$$

Using the following matrix identity (see, e.g., the Appendix of Chapter 7 of Box and Tiao, 1973),

$$\begin{aligned}
& (\mathbf{x} - \mathbf{a})^T A (\mathbf{x} - \mathbf{a}) + (\mathbf{x} - \mathbf{b})^T B (\mathbf{x} - \mathbf{b}) = \\
& (\mathbf{x} - \mathbf{c})^T (A + B) (\mathbf{x} - \mathbf{c}) + (\mathbf{a} - \mathbf{b})^T A (A + B)^{-1} B (\mathbf{a} - \mathbf{b})
\end{aligned}$$

where $\mathbf{c} = (A + B)^{-1} (A\mathbf{a} + B\mathbf{b})$, we obtain

$$\begin{aligned}
& (\boldsymbol{\beta}_{ih} - \hat{\boldsymbol{\beta}}_{ih})^T \frac{X_{ih}^T \Psi_{ih}^T \Psi_{ih} X_{ih}}{\sigma^2} (\boldsymbol{\beta}_{ih} - \hat{\boldsymbol{\beta}}_{ih}) + (\boldsymbol{\beta}_{ih} - \bar{\boldsymbol{\beta}}_h)^T \Sigma_B^{-1} (\boldsymbol{\beta}_{ih} - \bar{\boldsymbol{\beta}}_h) = \\
& (\boldsymbol{\beta}_{ih} - \mathbf{c}_{ih})^T (A_{ih} + B_{ih}) (\boldsymbol{\beta}_{ih} - \mathbf{c}_{ih}) + (\hat{\boldsymbol{\beta}}_{ih} - \bar{\boldsymbol{\beta}}_h)^T A_{ih} (A_{ih} + B_{ih})^{-1} B_{ih} (\hat{\boldsymbol{\beta}}_{ih} - \bar{\boldsymbol{\beta}}_h)
\end{aligned} \tag{11}$$

where $A_{ih} = \frac{X_{ih}^T \Psi_{ih}^T \Psi_{ih} X_{ih}}{\sigma^2}$, $B_{ih} = \Sigma_B^{-1}$ and $\mathbf{c}_{ih} = (A_{ih} + B_{ih})^{-1} (A_{ih} \hat{\boldsymbol{\beta}}_{ih} + B_{ih} \bar{\boldsymbol{\beta}}_h)$.

Putting (10) and (11) into (9), the joint probability density function of \mathbf{y} and $\boldsymbol{\beta}$ given λ and \mathbf{X} can be written as

$$\begin{aligned}
p(\mathbf{y}, \boldsymbol{\beta} \mid \lambda, \mathbf{X}) &= (2\pi\sigma^2)^{-N/2} (2\pi)^{-m/2} \left(\prod_{h=1}^g \prod_{i=1}^{m_h} |W_{ih}^T W_{ih}|^{-1/2} \right) \left(\prod_{h=1}^g \prod_{i=1}^{m_h} |\Sigma_B|^{-1/2} \right) \\
&\times \exp \left\{ -\frac{1}{2} \sum_{h=1}^g \sum_{i=1}^{m_h} \left[(\mathbf{y}_{ih} - \mathbf{X}_{ih}(\omega_h) \hat{\boldsymbol{\beta}}_{ih})^T \frac{\Psi_{ih}^T \Psi_{ih}}{\sigma^2} (\mathbf{y}_{ih} - \mathbf{X}_{ih}(\omega_h) \hat{\boldsymbol{\beta}}_{ih}) \right. \right. \\
&\quad + (\boldsymbol{\beta}_{ih} - \mathbf{c}_{ih})^T (A_{ih} + B_{ih}) (\boldsymbol{\beta}_{ih} - \mathbf{c}_{ih}) \\
&\quad \left. \left. + (\hat{\boldsymbol{\beta}}_{ih} - \bar{\boldsymbol{\beta}}_h)^T A_{ih} (A_{ih} + B_{ih})^{-1} B_{ih} (\hat{\boldsymbol{\beta}}_{ih} - \bar{\boldsymbol{\beta}}_h) \right] \right\}.
\end{aligned}$$

Integrating out $\boldsymbol{\beta}_{ih}$ yields the marginal distribution of \mathbf{y} given λ and \mathbf{X} ,

$p(\mathbf{y} \mid \lambda, \mathbf{X}) =$

$$(2\pi\sigma^2)^{-N/2} \left[\prod_{h=1}^g \prod_{i=1}^{m_h} (|W_{ih}^T W_{ih}|^{-1/2} |\Sigma_B|^{-1/2} |A_{ih} + B_{ih}|^{-1/2}) \right] \exp \left\{ -\frac{1}{2} \sum_{h=1}^g \sum_{i=1}^{m_h} (S_{1ih} + S_{2ih}) \right\}$$

where

$$S_{1ih} = (\mathbf{y}_{ih} - \mathbf{X}_{ih}(\omega_h)\hat{\boldsymbol{\beta}}_{ih})^T \frac{\Psi_{ih}^T \Psi_{ih}}{\sigma^2} (\mathbf{y}_{ih} - \mathbf{X}_{ih}(\omega_h)\hat{\boldsymbol{\beta}}_{ih})$$

and

$$S_{2ih} = (\hat{\boldsymbol{\beta}}_{ih} - \bar{\boldsymbol{\beta}}_h)^T A_{ih} (A_{ih} + B_{ih})^{-1} B_{ih} (\hat{\boldsymbol{\beta}}_{ih} - \bar{\boldsymbol{\beta}}_h).$$

The loglikelihood function for λ is therefore given by

$$\begin{aligned} L(\lambda) = & \\ & -\frac{N}{2} \log(2\pi\sigma^2) - \frac{1}{2} \sum_{h=1}^j \sum_{i=1}^{m_h} \left(\log(|W_{ih}^T W_{ih}|) + \log|\Sigma_B| + \log(|A_{ih} + B_{ih}|) + S_{1ih} + S_{2ih} \right). \end{aligned} \quad (12)$$

3 Estimation and Computation

3.1 Estimation

We wish to compare depressed and control patients' rhythms by fitting the data using the two-stage hierarchical model described in the previous section. Assuming the values of the parameters were known, the model would enable us to make comparisons in two ways. First, we could judge the magnitude of the discrepancies between the population parameters, $\lambda_{depressed}^*$ and $\lambda_{control}^*$. Second, we could plot the systematic components of (4) for all individuals in the sample. This would help us understand the variability among subjects, as well as the distinctions between the rhythms in the two groups.

Since we do not know the values of the parameters in the model, we will estimate them. The population parameters λ will be estimated by $\hat{\lambda}$ that maximizes the likelihood function in (12). Then, the $\boldsymbol{\beta}_{ih}$'s, will be estimated by their conditional posterior means, conditionally on $\lambda = \hat{\lambda}$. These estimates are the usual "parametric empirical Bayes" (PEB) estimates of the quantities $\boldsymbol{\beta}_{ih}$. According to standard large-sample theory, the assumption that the number of patients in each group is becoming infinite provides asymptotic Normality of the MLE's, with the observed information matrix (i.e., the negative Hessian of the loglikelihood, $-D^2 L(\hat{\lambda})$) furnishing the asymptotic precision matrix. We can then examine the MLE's of parameters of interest, together with their standard errors. We will also be able to produce plots of the individual systematic components, using the PEB estimates of the subject-specific parameters.

We note that both maximum likelihood and parametric empirical Bayes estimation may be justified as providing approximate fully Bayes estimates. That is, inference regions having confidence $1 - \alpha$ based on the asymptotic Normal distribution of the MLE $\hat{\lambda}$ also have posterior probability $1 - \alpha$ based on the limiting Normal posterior distribution of λ . Similarly, the parametric empirical Bayes estimates of β_{ih} differ from fully Bayes posterior expectations by a term that vanishes at a rate inversely proportional to the number of patients (see Kass and Steffey, 1989).

3.2 Computation

To use the two-stage hierarchical model, we must first identify a common first-stage model that can adequately represent the rhythm under study, and then estimate the unknown parameters of the model. GKT describe a procedure for model specification and estimation for the first-stage model. We apply this procedure to each individual and obtain initial estimates for the second-stage model using steps 1 and 2 below, and then obtain final estimates of all the parameters simultaneously using step 3.

1. Tentative model selection for the deterministic process:

Obtain initial estimates of the frequency for each subject in group h either from *a priori* knowledge or from spectral analysis of Y_{ijh} . An initial estimate of ω_h , in (4) can be obtained as the average of the individual frequencies. With this initial estimate of ω_h , we can then obtain initial estimates of β_h and σ^2 using maximum likelihood estimation by treating the stochastic component of (4) as i.i.d. random variables. This step can be done using standard statistical packages, such as BMDP or ISP.

2. Tentative model selection for the stochastic component:

Using the initial estimates of the deterministic component from step 1, form the estimated residuals

$$\hat{Z}_{ijh} = Y_{ijh} - \hat{f}_{ih}(t)$$

where $\hat{f}_{ih}(t)$ is the deterministic component evaluated at the initial values obtained in step 1. A tentative ARMA model for the stochastic component \hat{Z}_{ijh} can be specified for each subject i (for details, see GKT) and then an overall ARMA model can be chosen for \hat{Z}_{ijh} .

The maximum likelihood estimates of the common time series parameters of \hat{Z}_{ijh} can then be obtained.

3. Full model fitting:

By means of maximum likelihood estimation with the combined, tentative model obtained in steps 1 and 2, reestimate simultaneously all the parameters in the full model, i.e., both the deterministic and stochastic components. The likelihood function to be maximized is specified in section 2.3.

To carry out steps 2 and 3, we have written and used two FORTRAN programs. The first program minimizes the residual sum of squares of the time series \hat{Z}_{ijh} in step 2. This allows us to obtain the maximum likelihood estimates of the time series parameters across all of the subjects. The second program simultaneously reestimates all of the parameters in the full model in step 3 and obtains standard errors of the parameters based on the inverse of the observed information matrix (Kass, 1987).

4 Comparative Analysis of the Circadian Variation of Cortisol

As described earlier, we are concerned with studying changes in the circadian variation of the hormone cortisol in a group of depressed patients as compared to a group of healthy control subjects. The protocol for the University of Pittsburgh study required all subjects to enter the hospital for a period of 7 days. Following several adaptation days, blood sampling began on day 4 around 6:00 p.m., to obtain plasma cortisol measurements. Samples were obtained every 20 minutes thereafter for the next approximately 48 hours. The samples were labelled with respect to time, stored, and assayed within 72 hours to obtain the cortisol values. This analysis is based on 14 depressed outpatients (11 females; median age 50, range: 30-63), and 6 normal controls (1 female; median age=28.5, range: 24-54). All the patients satisfied research diagnostic criteria for a major depression and the normal controls were volunteers from the community. The large proportion of females in the depressed group reflects the prevalence of the disease among men and women in the population.

Figure 1 presents the cortisol levels for 4 healthy subjects as an illustration of the pattern of cortisol concentration in the blood over a 2 day period. We see that the general circadian pattern of cortisol concentration is similar in these subjects, but we also see a great deal of individual variation.

Using the iterative methods for model specification described in steps 1 and 2 in section 3.2, we determined that a suitable harmonic regression model for stage one of our two-stage hierarchical model based on these data should consist of, for the deterministic component, the fundamental and the first harmonic ($K=2$), and for the stochastic component, an autoregressive process of order 2. We treat as randomly varying across individuals, i.e., “random effects” in this model, the level, the amplitudes, and the phases as in (4) of section 2.2. The frequency and the autoregressive parameters are taken as common to all subjects, i.e., “fixed effects.” Given the sample size, this model with 20 parameters is the most general with respect to the number of parameters that we could seriously consider.

Tables 1 and 2 give the empirical Bayes or approximate Bayes estimates for β_{ih} , the parameters of the first-stage model for each subject in the depressed group and the control group, respectively. For ease of interpretation, we have transformed the A_i^k 's and the B_i^k 's back to amplitudes and phases (in radians). Table 3 gives the estimates of the second-stage parameters, λ_h^* , for both groups. The estimates of the common AR parameters (standard error in parentheses) are $\phi_1 = 0.792$ (0.0196) and $\phi_2 = -0.140$ (0.0194), and $\sigma = 2.17$. From this table, we see that the differences between the two groups with respect to the estimates of the second-stage parameters, the level, the amplitudes (R 's), the phases (ϕ 's), and the frequency, respectively, are not large compared to their standard errors.

In Figures 2 and 3 we plot the posterior fit for the deterministic component of the model for each individual along with the estimate of the overall cortisol rhythm for each group, respectively. It can be seen from the figures that the estimates of the two group curves for the cortisol rhythm are very similar. We also see that subject 1 in the control group seems to have a higher level of cortisol than the other subjects in that group. We reanalyzed the data without that subject and again found no substantial differences between the groups with respect to the amplitudes, phases, and frequency. However, there is an apparent difference between the groups with respect to level. When subject 1 is excluded, patients have higher average cortisol levels than controls (patients = 7.425, controls = 5.896, $se(\text{diff}) = .688$).

5 Discussion

5.1 Data Analysis

The understanding and analysis of biological rhythms has important implications for health and disease. In this paper we have been concerned with abnormalities in the circadian rhythm of plasma cortisol concentration in depressed patients. We have developed a statistical model for describing the circadian variation of plasma cortisol using a harmonic regression model to fit the underlying circadian rhythm and an ARMA process to pick up fluctuations from it. To accommodate variation across subjects and to compare the cortisol rhythm in depressed patients to controls, we use a two-stage hierarchical model in which certain parameters in the harmonic regression model are assumed to follow probability distributions. The analysis of the cortisol data presented here suggests that differences in the circadian rhythm of plasma cortisol between depressed patients and healthy controls are probably small. In addition, using the parametric empirical Bayes estimates of the subject-specific parameters, we have been able to examine the posterior fit for the deterministic component of the model for each individual; this helped us to identify one subject in the control group who had a higher level of cortisol than the other subjects in the group.

There are three components in the model we have used, the deterministic part of the first stage, the stochastic part of the first stage, and the second stage. In the deterministic component we chose to fit a trigonometric series with a fundamental frequency and one higher harmonic. This harmonic picks up a substantial part of the departure from the model using only the fundamental frequency. However, our empirical fitting procedure did not include a careful attempt to determine the precise nature of this departure. Thus, although the extra harmonic might genuinely represent what is often called an “ultradian rhythm” with a period of about 12 hours, there is nothing in our analysis that provides real evidence of such a claim. Further effort to establish the existence of a 12-hour rhythm would require additional data so that higher-order harmonics could be considered, as well. We also emphasize that our approach has been to model the concentration of cortisol in the blood over time rather than attempting to model the physiological process responsible for the secretion of the hormone. The latter is an interesting and challenging modeling problem addressing different substantive questions than are being asked in the Pittsburgh study.

Although some of our preliminary analysis included diagnostic checking of first-stage assump-

tions. Our choices for all three components were based largely on convenience and we did not perform thorough sensitivity studies. Additional checks designed for two-stage models could be conducted as in, for example, Waternaux, Laird, and Ware (1989). Also it would be possible to use a non-Normal second-stage distribution; this would entail a nontrivial computational burden, though we believe the problem could be solved with current numerical integration techniques.

5.2 Methodology

We have used a general ARMA model for the error structure in (1) to represent the dependence among the repeated measurements made for each subject. With sufficiently many subjects it would be possible to estimate a more general covariance structure. Use of the ARMA model, on the other hand, reduces the characterization of the within-subject covariance to a small number of parameters, and is therefore suitable for studies with relatively small numbers of both subjects and observations per subject.

In a recent article, Chi and Reinsel (1989) consider a linear model with random effects and AR(1) errors for longitudinal data. They derive a score test to detect serial correlation in the within-individual errors for random-effects models that are “conditionally independent” in the sense that conditional on the random effects the observations on a given individual are assumed to be independent. They further use a scoring method for maximum likelihood estimation. However, there are several differences between the present article and that of Chi and Reinsel. First, the model used in this article is non-linear; it estimates the fundamental frequency jointly with all the other parameters and this frequency is often of major interest. Second, the error terms of the present article may assume the general ARMA models; an AR(1) model often cannot capture the periodic pattern showing in the data. Third, this article emphasizes the comparison of a circadian rhythm between depressed patients and normal healthy controls, not on the detection of serial correlation. Finally, Chi and Reinsel allowed for irregularly sampled data whereas only equally spaced observations were considered in the present article. Of course, a combination of the ideas given in the two articles could be useful in application.

An alternative method to compute the exact likelihood function of Z_t is to use a state-space formulation and Kalman filter. Nice properties of the Kalman filter approach include the ease in handling missing observations and the explicit use of a diffuse prior in the case of unit-root

nonstationarity of Z_t . See Jones (1980, 1986), Harvey and Pierce (1984), Ansley and Kohn (1985), and Kohn and Ansley (1986), among others. We used a regression approach in this article for various reasons. First, the regression approach provides a quadratic expression that is handy in the hierarchical setting. Second, recent research shows that by treating missing observations as additive outliers the regression approach can also easily handle the missing value problem (see Ljung 1989, 1990). Third, by adopting the techniques in Wincek and Reinsel (1986), it is not difficult to generalize the regression approach to the case of irregularly spaced observations.

APPENDIX:

Exact likelihood function of an ARMA(p,q) Model

Consider a stationary ARMA(p,q) model

$$Z_t - \sum_{j=1}^p \phi_j Z_{t-j} = a_t - \sum_{j=1}^q \theta_j a_{t-j} \quad (A.1)$$

where the a_t 's are independent normal random variates with mean zero and variance σ^2 . Suppose the data $\mathbf{z} = (Z_1, \dots, Z_n)^T$ are available and our goal is to derive the associated exact likelihood function.

There are two approaches to evaluating the exact likelihood function of a stationary ARMA process. The first approach uses a state-space model representation and the Kalman filter recursion. e.g. Jones (1980). The second approach uses linear regression techniques to obtain estimates of the starting values of the process so that the exact likelihood function can be expressed as a quadratic function of the data. e.g. Box and Jenkins (1976), Hillmer and Tiao (1979), and Abraham and Ledolter (1983). The first approach is useful in computation, especially when missing observations exist. However, since we are interested in obtaining an exact expression of the likelihood of a two-stage hierarchical model, the second approach is used in this article.

Rewrite the ARMA(p,q) model (A.1) more compactly as

$$\phi(B)Z_t = \theta(B)a_t$$

where $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$, $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$, and B is the usual backshift operator such that $B^j Z_t = Z_{t-j}$. Let π_i be the π -weights of the model, that is, π_i is the coefficient of B^i in $\pi(B) = 1 - \sum_{i=1}^{\infty} \pi_i B^i = \phi(B)/\theta(B)$. Let $\mathbf{z}^* = (z_0, \dots, z_{1-p}, a_0, \dots, a_{1-q})^T$ be the vector of unknown starting values of the process Z_t . Furthermore, for $1 \leq k \leq p$, define $\phi_{(k)} = (\phi_k, \dots, \phi_p, 0, \dots, 0)$ be a p -dimensional row vector with the last $k-1$ elements being zero. For $k > p$, $\phi_{(k)}$ is the p -dimensional null row vector. Similarly, define the q -dimensional row vector $\theta_{(k)} = (\theta_k, \dots, \theta_q, 0, \dots, 0)$ for $1 \leq k \leq q$, and $\theta_{(k)} = \mathbf{0}$ for $k > q$. Finally, let $\mathbf{G}_{(k)} = (-\phi_{(k)}, \theta_{(k)})$. Then, simple algebra shows that

$$a_t = Z_t - \sum_{j=1}^{t-1} \pi_j Z_{t-j} + \mathbf{h}_t \mathbf{z}^* \quad (A.2)$$

where $\mathbf{h}_1 = \mathbf{G}_1$ and other \mathbf{h}_t 's are defined recursively by

$$\mathbf{h}_t = \begin{cases} \mathbf{G}_{(t)} + \sum_{j=1}^t \theta_j \mathbf{h}_{t-j} & \text{for } 2 \leq t \leq \max\{p, q\} \\ \sum_{j=1}^q \theta_j \mathbf{h}_{t-j} & \text{for } t > \max\{p, q\} \end{cases} \quad (\text{A.3})$$

with $\theta_j = 0$ if $j > q$. Putting the equation (A.2) in a matrix form, we have

$$\begin{bmatrix} \mathbf{z}^* \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} 0 \\ L \end{bmatrix} \mathbf{z} + \begin{bmatrix} I_{p+q} \\ D \end{bmatrix} \mathbf{z}^* \quad (\text{A.4})$$

where $\mathbf{a} = (a_1, \dots, a_n)^T$, D is an $n \times (p+q)$ matrix with i th row equal to \mathbf{h}_i defined in (A.3), and L is the $n \times n$ lower-triangular matrix of π -weights of Z_t with unity on the diagonal and the i -th row being $(-\pi_{i-1}, \dots, \pi_1, 1, 0, \dots, 0)$. Let $\sigma^2 \Sigma$ be the covariance matrix of \mathbf{z}^* , i.e., $\mathcal{E}(\mathbf{z}^* \mathbf{z}^{*T}) = \sigma^2 \Sigma$. Then one can find a matrix P such that $P \Sigma P^T = I_{p+q}$.

Pre-multiplying Equation (A.4) by the $(n+p+q) \times (n+p+q)$ matrix $M = \text{diag}\{P, I_n\}$, where I_n is the $n \times n$ identity matrix, we obtain

$$\begin{bmatrix} \mathbf{u}^* \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} 0 \\ L \end{bmatrix} \mathbf{z} + \begin{bmatrix} I_{p+q} \\ DP^{-1} \end{bmatrix} \mathbf{u}^* \quad (\text{A.5})$$

where $\mathbf{u}^* = P \mathbf{z}^*$. Let

$$W = \begin{bmatrix} I_{p+q} \\ DP^{-1} \end{bmatrix} \quad \text{and} \quad L^* = \begin{bmatrix} 0 \\ L \end{bmatrix}$$

so that $(\mathbf{u}^{*T} + \mathbf{a}^T)^T = L^* \mathbf{z} + W \mathbf{u}^*$. Now, it is important to note that

- (1) $\mathcal{E}(\mathbf{u}^* \mathbf{u}^{*T}) = \mathcal{E}(P \mathbf{z}^* \mathbf{z}^{*T} P^T) = P \mathcal{E}(\mathbf{z}^* \mathbf{z}^{*T}) P^T = P \Sigma P^T \sigma^2 = I_{p+q} \sigma^2$, and that
- (2) \mathbf{u}^* involves only $a_{1-q}, \dots, a_0, z_{1-p}, \dots, z_0$, which is independent of \mathbf{a} .

Thus the elements of $(\mathbf{u}^{*T}, \mathbf{a}^T)^T$ are independent normal random variables with mean zero and variance σ^2 . The joint probability density function of $(\mathbf{u}^{*T}, \mathbf{a}^T)^T$ is given by

$$p(\mathbf{u}^{*T}, \mathbf{a}^T | \lambda_c) = (2\pi\sigma^2)^{-(n+p+q)/2} \exp\left\{-\frac{1}{2\sigma^2}\{\mathbf{u}^{*T}\mathbf{u}^* + \mathbf{a}^T\mathbf{a}\}\right\}$$

where $\lambda_c = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma^2)^T$ is the vector of common unknown parameters. Since the Jacobian of the transformation of (A.5) is unity, we have that

$$p(\mathbf{z}^T, \mathbf{u}^{*T} | \lambda_c) = (2\pi\sigma^2)^{(n+p+q)/2} \exp\left\{-\frac{1}{2\sigma^2}S(\lambda_c, \mathbf{u}^*)\right\} \quad (\text{A.6})$$

where

$$S(\lambda_c, \mathbf{u}^*) = (L^*\mathbf{z} + W\mathbf{u}^*)^T(L^*\mathbf{z} + W\mathbf{u}^*)$$

is a sum of squares that depends on, among others, \mathbf{u}^* , which in turn is a function of the unknown starting values \mathbf{z}^* . Now it can be shown that

$$S(\lambda_c, \mathbf{u}^*) = S(\lambda_c) + (\mathbf{u}^* - \hat{\mathbf{u}}^*)^T Z^T Z (\mathbf{u}^* - \hat{\mathbf{u}}^*)$$

where $\hat{\mathbf{u}}^* = -(W^T W)^{-1} W^T L^* \mathbf{z}$ and

$$S(\lambda_c) = S(\lambda_c, \hat{\mathbf{u}}^*) = (L^*\mathbf{z} + W\hat{\mathbf{u}}^*)^T(L^*\mathbf{z} + W\hat{\mathbf{u}}^*)$$

Equation (A.6) then becomes

$$p(\mathbf{z}, \mathbf{u}^* | \lambda_c) = (2\pi\sigma^2)^{-(n+p+q)/2} \exp\left\{\frac{-1}{2\sigma^2}\{S(\lambda_c) + (\mathbf{u}^* - \hat{\mathbf{u}}^*)^T W^T W (\mathbf{u}^* - \hat{\mathbf{u}}^*)\}\right\}, \quad (\text{A.7})$$

the left-hand side of which can be written as

$$p(\mathbf{z}, \mathbf{u}^* | \lambda_c) = p(\mathbf{z} | \lambda_c) p(\mathbf{u}^* | \mathbf{z}, \lambda_c).$$

Integrating Equation (A.7) over \mathbf{u}^* , we obtain

$$p(\mathbf{z} | \lambda_c) = (2\pi\sigma^2)^{-n/2} |W^T W|^{-1/2} \exp\left\{\frac{-1}{2\sigma^2}S(\lambda_c)\right\} \quad (\text{A.8})$$

which is the exact likelihood function for an ARMA(p,q) process.

Furthermore, by

$$\begin{aligned} L^*\mathbf{z} + W\hat{\mathbf{u}}^* &= L^*\mathbf{z} - W(W^T W)^{-1}W^T L^*\mathbf{z} \\ &= (I_{p+q+n} - W(W^T W)^{-1}W^T)L^*\mathbf{z}, \end{aligned}$$

and letting

$$I_{p+q+n} - W(W^T W)^{-1} W^T = \begin{bmatrix} D_1 & | & \\ & & \Lambda \\ D_2 & | & \end{bmatrix},$$

where Λ is an $(p+q+n) \times n$ matrix, we have that

$$L^* \mathbf{z} + W \hat{\mathbf{u}}^* = \Lambda L \mathbf{z} = \Psi \mathbf{z}$$

and

$$p(\mathbf{z} | \lambda_c) = (2\pi\sigma^2)^{-n/2} |W^T W|^{-1/2} \exp\left\{\frac{-1}{2\sigma^2} \mathbf{z}^T \Psi^T \Psi \mathbf{z}\right\}.$$

which is the result used in Section 2.

References

- Abraham. B. and Ledolter. J. (1983), *Statistical Methods for Forecasting*. New York: Wiley.
- Ansley. C. F. and Kohn. R. (1985), "Estimation, Filtering and Smoothing in State Space Models with Incompletely Specified Initial Conditions," *Annals of Statistics*, **13**, 1286-1316.
- Box. G.E.P.. and Jenkins. G.M. (1976), *Time Series Analysis: Forecasting and Control*, 2nd Edition. San Francisco: Holden-Day.
- Box. G.E.P. and Tiao. G.C. (1973), *Bayesian Inference in Statistical Analysis*. London: Addison-Wesley.
- Chi. E. M.. and Reinsel. G. C. (1989), "Models for Longitudinal Data with Random Effects and AR(1) Errors." *Journal of the American Statistical Association*, **84**, 452-459.
- Harvey. A. C. and Pierce. R. G. (1984), "Estimating Missing Observations in Economic Time Series." *Journal of the American Statistical Association*, **79**, 125-131.
- Hillmer. S. C.. and Tiao. G. C. (1979), "Likelihood Function of Stationary Multiple Autoregressive Moving Average Models." *Journal of the American Statistical Association*. **74**, 652-660.
- Greenhouse. J.B.. Kass, R.E.. and Tsay, R.S. (1987), "Fitting Nonlinear Models with ARMA Errors to Biological Rhythm Data," *Statistics in Medicine*, **6**:167-183.
- Jones. R. H. (1980), "Maximum Likelihood Fitting of ARMA Models to Time Series With Missing Observations." *Technometrics*, **22**, 389-395.
- Jones, R. H. (1986). "Random Effects and the Kalman Filter," in *Proceedings of the Business and Economic Statistics Section. American Statistical Association*, pp. 69-75.
- Kass. R.E. (1987), "Computing Observed Information by Finite Differences." *Communications in Statistics: Simulation and Computation*, **B16**:587-599.
- Kass. R.E. and Steffey, D. (1989), "Approximate Bayesian Inference in Conditionally Independent Hierarchical Models (Parametric Empirical Bayes Models)," *Journal of the American Statistical Association*. **84**:717-726.

- Kohn, R., and Ansley, C. F. (1986), "Estimation, Prediction, and Interpolation for ARIMA Models With Missing Values." *Journal of the American Statistical Association*, **81**, 751-761.
- Kupfer, D.J., and Foster, F.G. (1972), "Interval between Onset of Sleep and Rapid-Eye-Movement Sleep as an Indicator of Depression." *Lancet*, 2:684-686.
- Laird, N.M. and Ware, J.H. (1982), "Random-Effects Models for Longitudinal Data," *Biometrics*, **38**:963-974.
- Ljung, G. M. (1989). "A Note on the Estimation of Missing Values in Time Series." *Communications in Statistics. Series B*, **18**, 459-465.
- Ljung, G. M. (1990). "Outlier Detection in Time Series," Working Paper, Statistics Center, Massachusetts Institute of Technology.
- Minors, D.S. and Waterhouse, J.M., (1988), "Mathematical and Statistical Analysis of Circadian Rhythms." *Psychoneuroendocrinology*, **13**:443-464.
- Moore-Ede, M.C., Czeisler, C.A., Richardson, G.S. (1983), "Circadian Timekeeping in Health and Disease: Part 2. Clinical Implications of Circadian Rhythmicity," *The New England Journal of Medicine*, **9**:530-536.
- Moore-Ede, M.C., Sulzman, F.M., and Fuller, C.A. (1982), *The Clocks That Time Us: Physiology of the Circadian Timing System*, Cambridge: Harvard University Press.
- Olshen, R.A., Biden, E.N., Wyatt, M.P., Sutherland, D.H. (1989), "Gait Analysis and the Bootstrap," *The Annals of Statistics*, **17**:1419-1440.
- Reinsel, G. C., and Tiao, G. C. (1987), "Impact of Chlorofluoromethanes on Stratospheric Ozone." *Journal of the American Statistical Association*, **82**, 20-30.
- Sack, D.A., Rosenthal, N.E., Parry, B.L., Wehr, T.A. (1987), "Biological Rhythms in Psychiatry." in *Psychopharmacology: The Third Generation of Progress*, edited by H. Y. Meltzer. New York: Raven Press.
- Tong, Y.L. (1976), "Parameter Estimation in Studying Circadian Rhythms," *Biometrics*, **32**:85-94.

Waterman, C., Laird, N.M., and Ware, J.H. (1989), "Methods for Analysis of Longitudinal Data: Blood-Lead Concentrations and Cognitive Development." *Journal of the American Statistical Association*. 84:33-41.

Winckel, M. A., and Reinsel, G. C. (1986), "An Exact Maximum Likelihood Estimation Procedure for Regression-ARMA Time Series Models With Possibly Nonconsecutive Data." *Journal of the Royal Statistical Society, Ser. B*, 48, 303-313.

Table 1: Posterior Estimates of the Parameters of the First-Stage Model
for Each Subject in the Depressed Group

i	μ_{i1}	R_{i1}^1	θ_{i1}^1	R_{i1}^2	θ_{i1}^2
1	5.75073	2.762761	2.313958	1.526266	-0.021732
2	7.15767	4.795436	2.135773	1.669860	-0.019927
3	7.60592	4.189216	2.298618	1.715918	-0.019342
4	7.37679	3.847971	2.284967	1.710920	-0.019436
5	6.07798	4.633564	2.363536	1.703300	-0.019489
6	7.07930	3.961659	2.204823	1.619559	-0.020524
7	9.53174	4.425398	2.285647	1.556337	-0.021368
8	8.68716	4.016897	2.340257	1.575313	-0.021084
9	6.79398	4.271711	2.255000	1.689855	-0.019683
10	9.65074	4.132880	2.265791	1.673365	-0.019837
11	8.67777	4.845589	2.313336	1.640384	-0.202470
12	7.78879	4.586333	2.320656	1.722901	-0.019279
13	6.26715	4.010485	2.162387	1.691691	-0.029659
14	5.52294	4.119850	2.329102	1.626388	-0.020408

Table 2- Posterior Estimates of the Parameters of the First-Stage Model
for Each Subject in the Control Group

i	μ_{i2}	R_{i2}^1	θ_{i2}^1	R_{i2}^2	θ_{i2}^2
1	11.41896	5.085662	2.565209	2.577948	-0.016294
2	4.941418	3.490282	2.461203	2.293906	-0.018283
3	5.357593	3.735369	2.386339	2.374498	-0.017650
4	7.561059	4.597357	2.669817	2.331090	-0.017967
5	5.508472	4.388673	2.625529	2.298040	-0.018226
6	6.551861	3.941468	2.467842	2.247713	-0.018660

Table 3: Estimates of the Parameters of the Second-Stage Model for Each Group

	Patients	Controls	Difference	SE(Diff)
$\bar{\mu}$	7.425	6.890	0.536	0.865
\bar{R}^1	4.177	4.186	-0.009	0.489
$\bar{\theta}^1$	2.276	2.539	-0.263	0.155
\bar{R}^2	1.652	2.357	-0.705	0.378
$\bar{\theta}^2$	-0.020	-0.0018	-0.0002	0.284
ω	.01402	.01378	0.0002	0.0003

Figure Titles

Figure 1(a-d). Examples of the circadian pattern of cortisol secretion for an approximate 48 hour period in four healthy control subjects. The smooth curve is the posterior fit based on the two-stage hierarchical model (see text). Measurements are obtained three times an hour. The units of time are indexed by sequential measurement number.

Figure 2. Posterior fit of the deterministic component of the model for each subject in the depressed group (—) and the estimate of the overall population curve from the second-stage of the hierarchical model for the depressed group (●—●).

Figure 3. Posterior fit of the deterministic component of the model for each subject in the control group (—) and the estimate of the overall population curve from the second-stage of the hierarchical model for the control group (●—●).

Figure 1a.

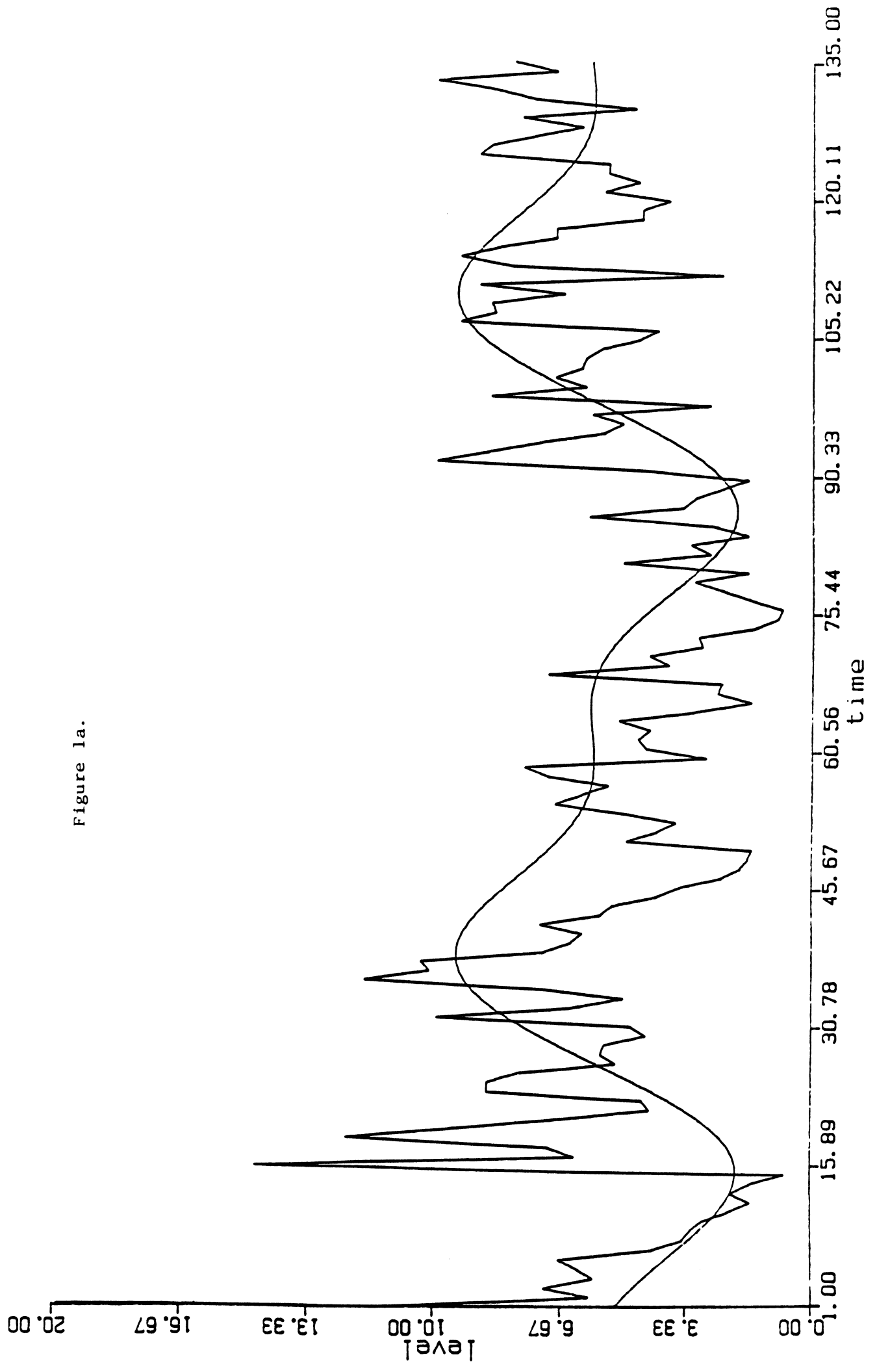


Figure 1b.

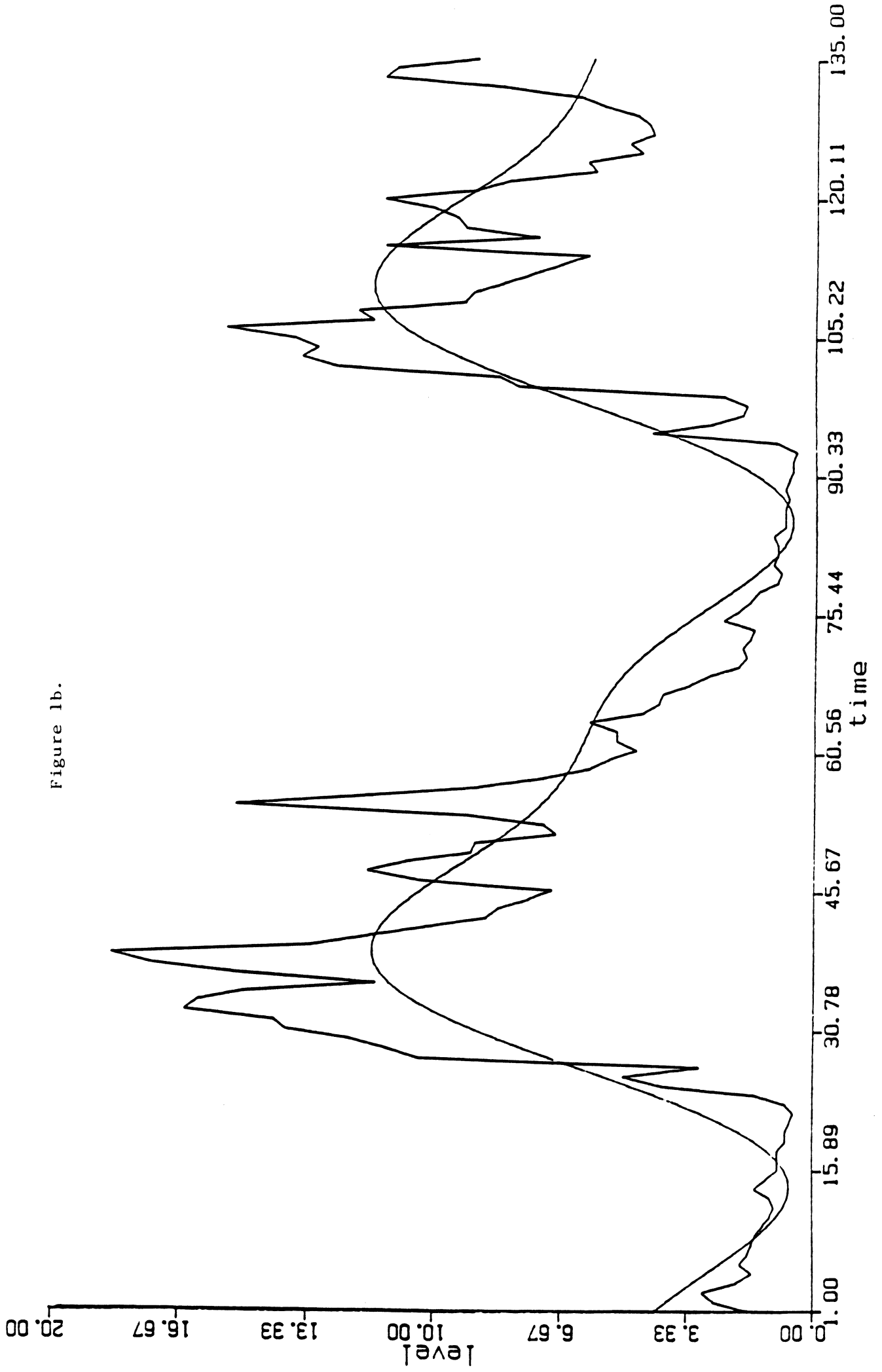


Figure 1c.

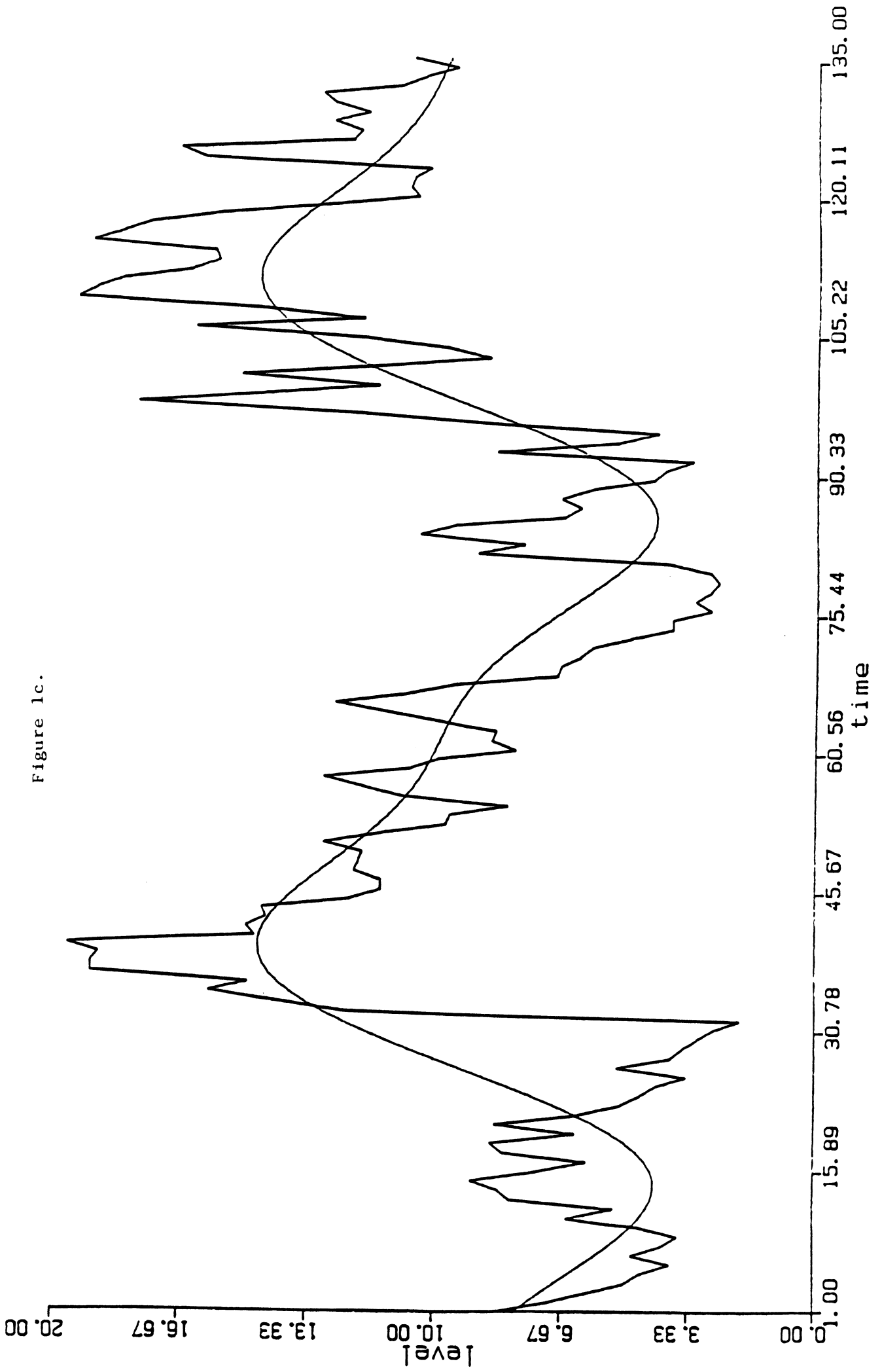


Figure 1d.

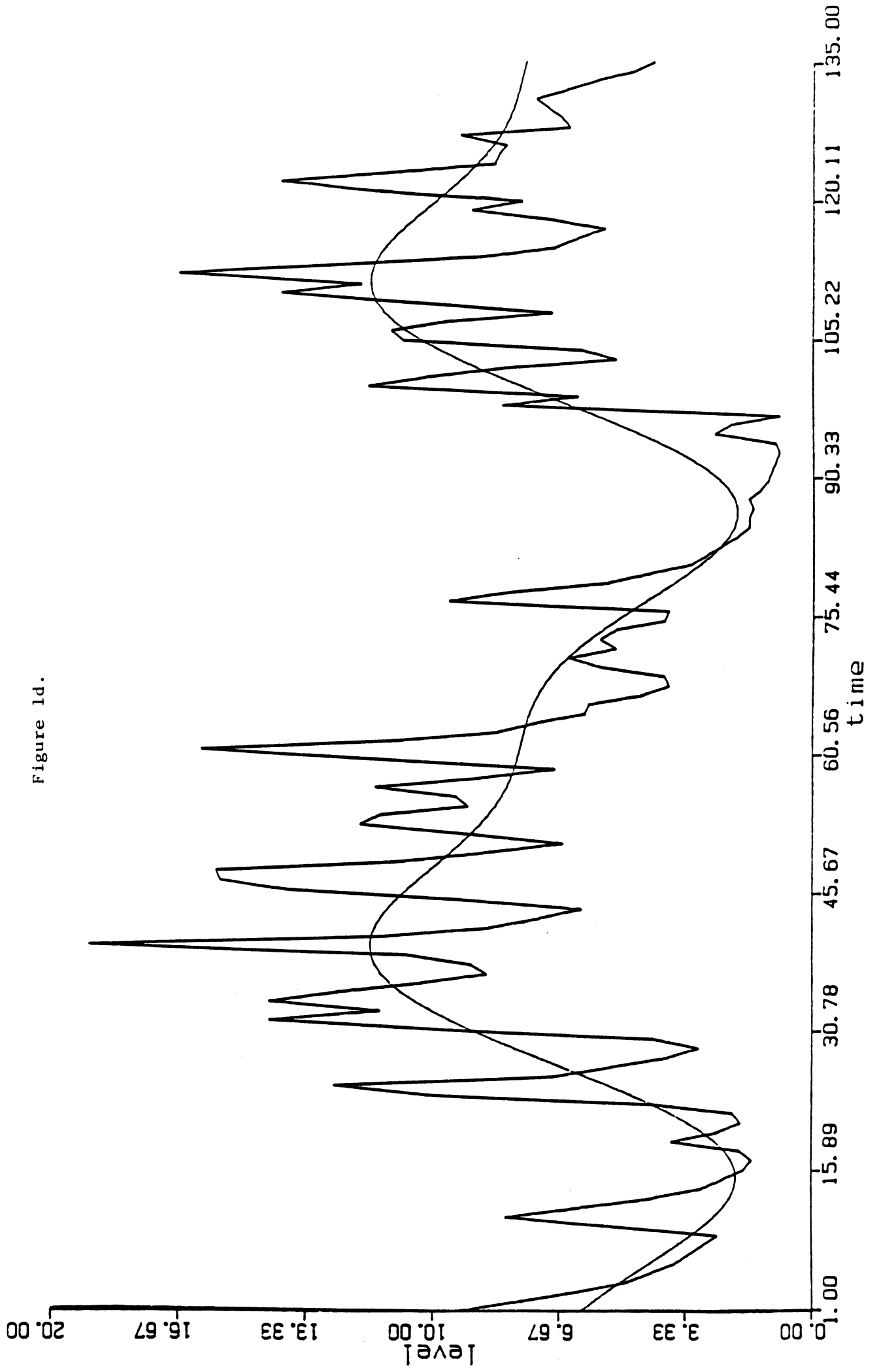


Figure 2

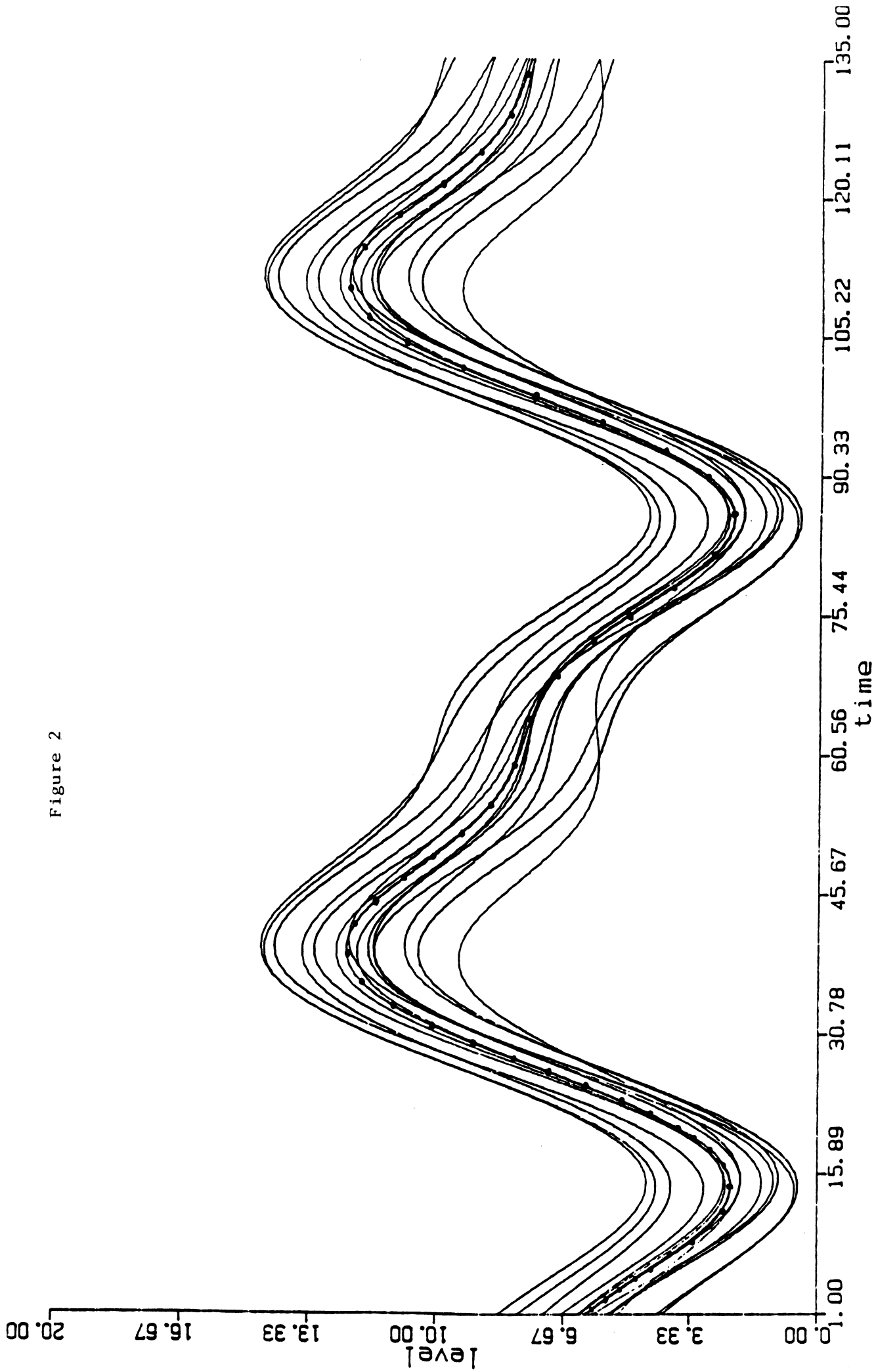
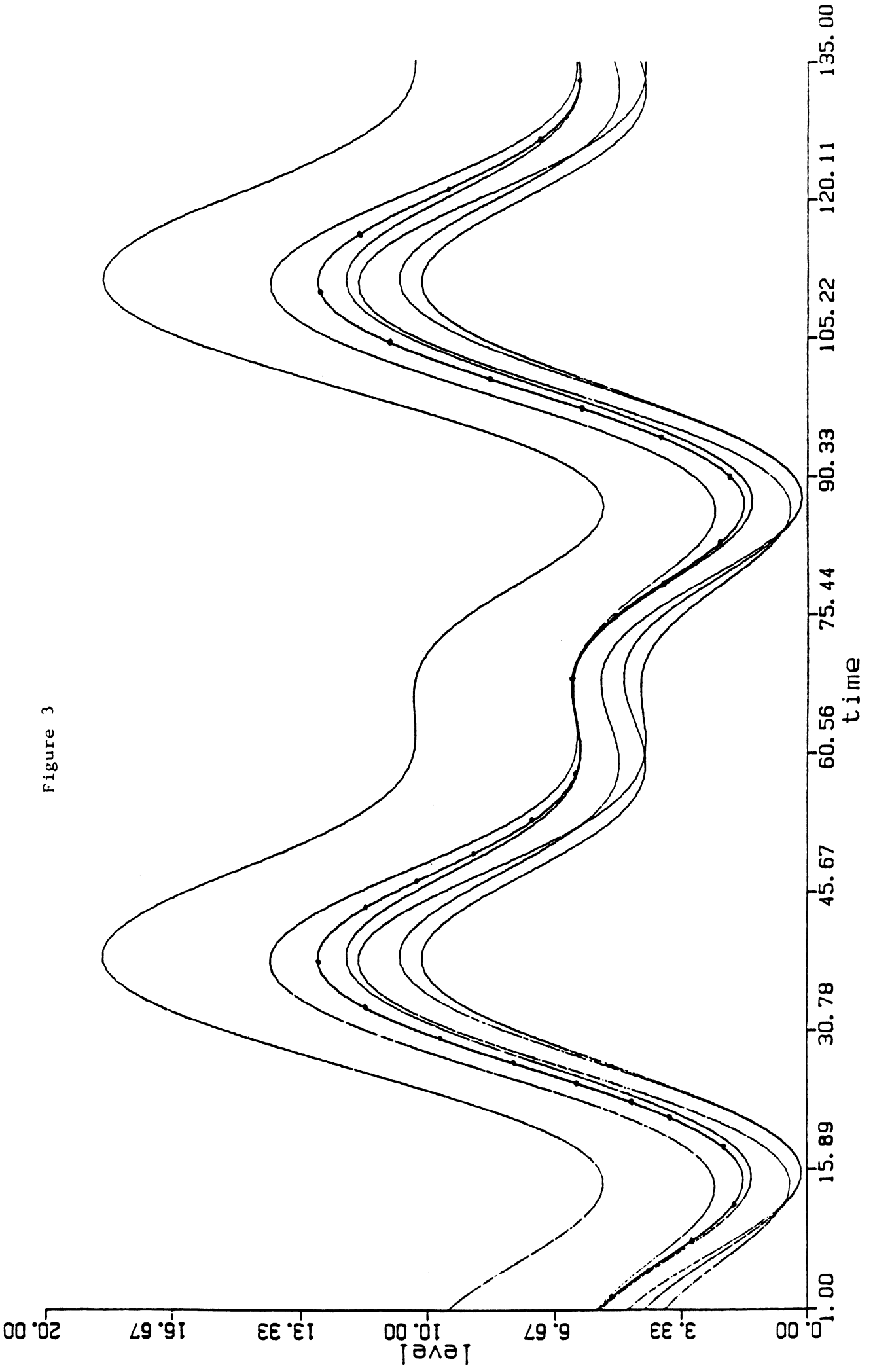


Figure 3



UNIVERSITY OF MICHIGAN



3 9015 04732 7542