Aust. N. Z. J. Stat. **45**(3), 2003, 257–270

ON MODELLING DATA FROM DEGRADATION SAMPLE PATHS OVER TIME

TSUNG I. LIN[1]* AND JACK C. LEE[2]

National Chiao Tung University

Summary

This paper is mainly concerned with modelling data from degradation sample paths over time. It uses a general growth curve model with Box–Cox transformation, random effects and ARMA(p, q) dependence to analyse a set of such data. A maximum likelihood estimation procedure for the proposed model is derived and future values are predicted, based on the best linear unbiased prediction. The paper compares the proposed model with a nonlinear degradation model from a prediction point of view. Forecasts of failure times with various data lengths in the sample are also compared.

Key words: ARMA(p, q) dependence; Box–Cox transformation; ECME; maximum likelihood estimation; semi-variogram.

1. Introduction

In this paper, we are concerned with modelling degradation data such as measurements of the growth of fatigue cracks. In general, engineers need to produce units of material with acceptable reliability and at an acceptable life-cycle cost. Most material accumulates irreversible damage during its life, which leads to failure. The cumulative damage reduces the reliability of the material as time increases. To maintain an acceptable reliability in the unit, inspections and repairs must be made, which increase life-cycle costs. Thus, cumulative damage plays a very important role in the design of the unit.

An example of cumulative damage is the fatigue crack growth data of Bogdanoff & Kozin (1985 p. 242) as plotted in Figure 1(a). The crack lengths represent the growth of cracks in metal for 21 notched test units under several constant load amplitude duty cycles. The crack lengths of each unit were recorded at every 0.01 million cycles over a period of 0.12 million cycles. Initial crack length was 0.9 inches. Testing was stopped if the crack length exceeded 1.60 inches, defined as a 'failure', or at 0.12 million cycles. The failure times for the 21 test units are given by Lu & Meeker (1993 Table 2).

For degradation data such as the fatigue crack growth data described above, it is very important to develop a model capable of predicting the fatigue crack growth and, most important of all, predicting the time to failure. Then engineers can order repair or replacement before the failure actually occurs. Once the unit or material has failed, it is too late to repair, and failure could cause heavy physical and/or economic losses.

Received July 2002; revised December 2002; accepted March 2003.

* Author to whom correspondence should be addressed.

[1] Institute of Statistics, National Chiao Tung University, Hsinchu 30050, Taiwan.
e-mail: tilin.st87g@nctu.edu.tw

[2] Institute of Statistics and Graduate Institute of Finance, National Chiao Tung University, Hsinchu 30050, Taiwan.

Acknowledgments. The authors thank two referees and the editors for their constructive suggestions that led to significant improvement of the paper.

This type of data is quite typical in studies such as accelerated life testing, because the product usually takes a long time to wear out. One important characteristic of the observations obtained in degradation studies is that they are measurements of several units, and each unit is measured over time. The measurements on a single unit are not independent because they are time-series in nature. If there are only a few measurements on the unit, say fewer than 20, then the dependence may be too hard to estimate. Fortunately, such data are usually obtained for several similar and independent units. Also, the linearity of growth function can be enhanced by the well known Box–Cox transformation (Box & Cox, 1964), as seen in Figure 1(b). These phenomena occur in many studies including technology substitutions as reported by Keramidas & Lee (1990). This paper predicts that a general growth curve model having ARMA(p, q) dependence coupled with the Box–Cox transformation can be applied to degradation data. A model is proposed and compared, in terms of its prediction accuracy and failure time prediction, with the degradation model of Lu & Meeker (1993) using the fatigue crack data of Bogdanoff & Kozin (1985). The failure time is the time to grow a crack from 0.90 inches to the critical crack length of 1.60 inches. A credible prediction of failure time is important, particularly to engineers.

Applying the Paris–Erdogan law (Paris & Erdogan, 1963), the degradation model for failure time proposed by Lu & Meeker (1993) is

$$Y_{ij} = \eta(t_{ij}; \boldsymbol{b}_i) + \varepsilon_{ij}\,, \tag{1}$$

where $t_{ij} = j \times 10^4$ cycles and $\eta(t_{ij}; \boldsymbol{b}_i) = -1/b_{i2} \log(1 - 0.90^{b_{i1}} b_{i1} b_{i2} t_{ij})$ $(i = 1, \ldots, N,$ $j = 1, \ldots, N_i)$, in which $\boldsymbol{b}_i = (b_{i1}, b_{i2})$ denotes a vector of the ith unit random-effects parameters, representing individual unit characteristics. We also have $a_i(t_{ij}) = 0.90 \exp(Y_{ij})$, where $a_i(t_{ij})$ denotes the crack length of the ith unit at time t_{ij}. The ε_{ij} and the $\boldsymbol{b}_i$ are all independent with $\varepsilon_{ij} \overset{\mathrm{d}}{=} \mathrm{N}(0, \sigma_\varepsilon^2)$ and $\boldsymbol{b}_i \overset{\mathrm{d}}{=} \mathrm{N}(\boldsymbol{\tau}, \boldsymbol{\Sigma}_b)$. The model specified by (1) is called the Lu and Meeker model (LMM) throughout the paper.

The LMM is essentially a nonlinear growth curve model with random coefficients. Lu & Meeker (1993) provide a two-stage method for estimating the random-effect parameters assuming no autocorrelation since each unit is relatively short in length. To use this estimating procedure in practical situations, the number of measurement times for each test unit, N_i, should be sufficiently large.

In Section 2, we present the proposed model. Section 3 is devoted to the prediction of future values. Section 4 illustrates the application of the model to the fatigue crack data. Section 5 gives some concluding remarks. The Appendix gives the derivation of a Hessian matrix for the asymptotic standard errors of maximum likelihood estimates (MLEs).

2. General growth curve model

2.1. Model specification

The growth curve model (GCM) has been considered by many authors since Potthoff & Roy (1964). Laird & Ware (1982) considered the random effects model with white noise errors. Jennrich & Schluchter (1986) discussed various types of patterned covariance structures, including random effects models and the AR(1) model. Chi & Reinsel (1989) considered the MLEs for the model with both random effects and AR(1) errors by the score method. Rochon (1992) presented a fixed-effects model for analysing repeated measures experiments having ARMA(p, q) covariance structures with time heteroscedasticity.

Some transformations on the observations could enhance the justification of assumptions such as normality of the distribution or linearity of the growth curve function. Keramidas & Lee (1990) showed tremendous improvement in predictive accuracy using the Box–Cox transformation models for technology substitutions. This is primarily because the linearity assumption for the growth function can be enhanced significantly with the Box–Cox transformation, and the observations can be given a proper dependence structure. Assumptions of normality and constancy of variance are of relatively minor importance in improving predictive accuracy. In this article, we extend the growth curve model to incorporate the Box–Cox transformation, random effects and ARMA(p, q) errors dependence.

The model considered here is

$$\boldsymbol{Y}_i^{(\lambda)} = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i \qquad (i = 1, 2, \ldots, N)\,,$$

where $\boldsymbol{Y}_i$ denotes an $N_i \times 1$ vector of measurements, $\boldsymbol{\beta}$ denotes an unknown $m_1 \times 1$ vector of fixed effects regression coefficients, $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ denote known design matrices, $\boldsymbol{b}_i$ denotes an $m_2 \times 1$ vector of random effects distributed as $\mathrm{N}(\boldsymbol{0}, \sigma^2\boldsymbol{\Gamma})$ and $\boldsymbol{\varepsilon}_i$ denotes an independent $N_i \times 1$ vector of within-individual errors whose components are assumed to follow the ARMA(p, q) model. That is,

$$\varepsilon_{it} = \sum_{j=1}^{p} \phi_j \varepsilon_{i,t-j} - \sum_{j=1}^{q} \theta_j a_{i,t-j} + a_{it}\,,$$

where $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iN_i})$ and $\{a_{it}\}$ denotes a series of shocks or white noise, which are identically and independently distributed as $\mathrm{N}(0, \sigma_a^2)$. In our study, we assume the observations for each unit are made at equally spaced intervals in time. Following Box, Jenkins & Reinsel (1994 p. 53), we write $\phi(B)\varepsilon_{it} = \theta(B)a_{it}$, where B is the backshift operator such that $B^k\varepsilon_{it} = \varepsilon_{i,t-k}$, $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ and $\theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$ are polynomials of B of degrees p and q, respectively. For the process to be stationary and invertible so that there will be a unique model corresponding to the likelihood function, the roots of $\phi(B)$ and $\theta(B)$ must lie outside the unit circle, which constrains the parameter vectors $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)$ to lie in regions $\mathbb{C}_p$ and $\mathbb{C}_q$, respectively.

Let $\sigma^2\boldsymbol{C}_i$ denote the covariance matrix of $\boldsymbol{\varepsilon}_i$, where $\boldsymbol{C}_i = [\rho_{|g-h|}]$, where $g, h = 1, \ldots, N_i$, and $\sigma^2 = (1 + \theta_1^2 + \cdots + \theta_q^2)\sigma_a^2/(1 - \phi_1\rho_1 - \cdots - \phi_p\rho_p)$. The ρ_j are implicit functions of $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$.

The Box–Cox transformation is defined as:

$$Y_{ij}^{(\lambda)} = \begin{cases} \dfrac{(Y_{ij} + \nu)^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0\,, \\ \log(Y_{ij} + \nu) & \text{if } \lambda = 0\,, \end{cases}$$

where Y_{ij} is the jth component of $\boldsymbol{Y}_i$, ν is a known constant such that $Y_{ij} + \nu > 0$ for all i, j, and λ is an unknown parameter. Thus the covariance matrix of $\boldsymbol{Y}_i^{(\lambda)}$ can be written as

$$\boldsymbol{\Sigma}_i = \sigma^2(\boldsymbol{Z}_i\boldsymbol{\Gamma}\boldsymbol{Z}_i^{\mathrm{T}} + \boldsymbol{C}_i) = \sigma^2\boldsymbol{\Lambda}_i(\boldsymbol{\Gamma}, \boldsymbol{\phi}, \boldsymbol{\theta})\,.$$

2.2. Reparameterization

To facilitate the estimating procedure and ensure admissibility of $(\boldsymbol{\phi}, \boldsymbol{\theta})$, we need to reparameterize the model. Barndorff-Nielsen & Schou (1973) proposed the following one-to-one and onto transformation which reparameterizes $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)$ in terms of the partial autocorrelations $\boldsymbol{\gamma}_\phi = (\gamma_{\phi_1}, \ldots, \gamma_{\phi_p})$ for the AR(p) process:

$$\phi_k^{(k)} = \gamma_{\phi_k}, \qquad \phi_i^{(k)} = \phi_i^{(k-1)} - \gamma_{\phi_k}\phi_{k-i}^{(k-1)} \qquad (i = 1, 2, \ldots, k-1), \tag{2}$$

and the condition that $\boldsymbol{\phi} \in \mathbb{C}_p$ becomes $|\gamma_{\phi_k}| \leq 1,\ k = 1, \ldots, p$.

For the MA(q) process, the reparameterization scheme is identical to the AR(p) process, with ϕ_k in (2) replaced by θ_k, as noted by Monahan (1984). Thus, we let $\boldsymbol{\gamma}_\theta = (\gamma_{\theta_1}, \ldots, \gamma_{\theta_q})$ and the condition that $\boldsymbol{\theta} \in \mathbb{C}_q$ becomes $|\gamma_{\theta_k}| \leq 1,\ k = 1, \ldots, q$. For the general ARMA($p, q$) process, a reparameterization of $(\boldsymbol{\phi}, \boldsymbol{\theta})$ in terms of $(\boldsymbol{\gamma}_\phi, \boldsymbol{\gamma}_\theta)$ can be obtained by applying (2) on both $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$.

Because GCM contains the Box–Cox transformed parameter λ, the MLEs could be computed by extending the Newton–Raphson method of Lindstrom & Bates (1988). We do not provide those expressions here. The Newton–Raphson iterations are quite unstable when the dimension of parameter space is high or initial values of parameters are far from optimum. Instead, the ECME algorithm (Liu & Rubin, 1994), is computationally easier and more stable than the Newton–Raphson method.

2.3. Using ECME algorithm for MLEs

The EM algorithm (Dempster, Laird & Rubin, 1977) is a popular iterative algorithm for maximum likelihood estimation in models with incomplete data. The EM algorithm has the advantage that each iteration is easy to compute and the initial iterations approach the optimum quite quickly. Laird & Ware (1982) described the EM algorithm as a method of calculating MLEs assuming $\mathrm{D}(\boldsymbol{\varepsilon}_i) = \sigma^2 \boldsymbol{I}_{N_i}$. Jennrich & Schluchter (1986) used the hybrid EM scoring algorithm in fitting arbitrary structural models for the within-subject covariance. For GCM, the M-step in the EM is difficult to implement. Meng & Rubin (1993) suggested that it can be replaced by a sequence of constrained maximization (CM) steps. This simple extension of the EM algorithm is called the ECM algorithm. A further extension of EM algorithm is the ECME algorithm (Liu & Rubin, 1994). This algorithm replaces each CM-step of ECM with a CM-step that maximizes either the constrained Q function (CMQ-step) or the corresponding constrained actual likelihood function (CML-step). Here the Q function is obtained as the expectation of the complete-data log-likelihood function given the observed data and the current estimates of the parameters.

The GCM can alternatively be written as:

$$\begin{bmatrix} \boldsymbol{Y}_i^{(\lambda)} \\ \boldsymbol{b}_i \end{bmatrix} \overset{\mathrm{d}}{=} \mathrm{N}\left(\begin{bmatrix} \boldsymbol{X}_i\boldsymbol{\beta} \\ \boldsymbol{0} \end{bmatrix}, \quad \sigma^2 \begin{bmatrix} \boldsymbol{\Lambda}_i & \boldsymbol{Z}_i\boldsymbol{\Gamma} \\ \boldsymbol{\Gamma}\boldsymbol{Z}_i^{\top} & \boldsymbol{\Gamma} \end{bmatrix} \right) \qquad (i = 1, \ldots, N).$$

Following Dempster *et al.* (1977), the random effects $\boldsymbol{b}_i$ are treated as missing data. We have the following ECME algorithm:

E-step. Using the current estimate of $\boldsymbol{\alpha}$, calculate, for $i = 1, \ldots, N$,

$$\hat{\boldsymbol{b}}_i = \mathrm{E}(\boldsymbol{b}_i \mid \boldsymbol{Y}_i^{(\lambda)}, \hat{\boldsymbol{\alpha}}) = \hat{\boldsymbol{\Gamma}} \boldsymbol{Z}_i^{\top} \hat{\boldsymbol{\Lambda}}_i^{-1} (\boldsymbol{Y}_i^{(\hat{\lambda})} - \boldsymbol{X}_i \hat{\boldsymbol{\beta}}),$$

$$\hat{\boldsymbol{S}}_b = \mathrm{E}\Big(\sum_{i=1}^{N} \boldsymbol{b}_i \boldsymbol{b}_i^{\top} \mid \boldsymbol{Y}^{(\lambda)}, \hat{\boldsymbol{\alpha}} \Big) = \sum_{i=1}^{N} \big(\mathrm{D}(\boldsymbol{b}_i \mid \boldsymbol{Y}_i^{(\lambda)}, \hat{\boldsymbol{\alpha}}) + \hat{\boldsymbol{b}}_i \hat{\boldsymbol{b}}_i^{\top} \big),$$

$$\hat{\boldsymbol{S}}_e = \mathrm{E}\Big(\sum_{i=1}^{N} \boldsymbol{e}_i^{\top} \boldsymbol{C}_i^{-1} \boldsymbol{e}_i \mid \boldsymbol{Y}^{(\lambda)}, \hat{\boldsymbol{\alpha}} \Big) = \sum_{i=1}^{N} \hat{\boldsymbol{e}}_i^{\top} \hat{\boldsymbol{C}}_i^{-1} \hat{\boldsymbol{e}}_i + \mathrm{tr}\Big(\sum_{i=1}^{N} \boldsymbol{Z}_i \mathrm{D}(\boldsymbol{b}_i \mid \boldsymbol{Y}_i^{(\lambda)}, \hat{\boldsymbol{\alpha}}) \boldsymbol{Z}_i^{\top} \hat{\boldsymbol{C}}_i \Big),$$

where $\boldsymbol{e}_i = \boldsymbol{Y}_i^{(\lambda)} - \boldsymbol{X}_i \boldsymbol{\beta} - \boldsymbol{Z}_i \boldsymbol{b}_i$, $\hat{\boldsymbol{e}}_i = \boldsymbol{Y}_i^{(\hat{\lambda})} - \boldsymbol{X}_i \hat{\boldsymbol{\beta}} - \boldsymbol{Z}_i \hat{\boldsymbol{b}}_i$ and $\mathrm{D}(\boldsymbol{b}_i \mid \boldsymbol{Y}_i^{(\lambda)}, \hat{\boldsymbol{\alpha}}) = \hat{\sigma}^2 \boldsymbol{G}_i$, where $\boldsymbol{G}_i = (\boldsymbol{Z}_i^{\top} \hat{\boldsymbol{C}}_i^{-1} \boldsymbol{Z}_i + \hat{\boldsymbol{\Gamma}}^{-1})^{-1}$.

CM-step 1. Update $\hat{\sigma}^2$ by maximizing the expected complete-data likelihood:

$$\hat{\sigma}^2 = \frac{1}{n + N m_2} \big(\hat{\boldsymbol{S}}_e + \mathrm{tr}(\hat{\boldsymbol{S}}_b \hat{\boldsymbol{\Gamma}}^{-1}) \big), \quad \text{where } n = \textstyle\sum_{i=1}^{N} N_i .$$

CM-step 2. Fix the updated $\sigma^2 = \hat{\sigma}^2$, and update $\hat{\boldsymbol{\Gamma}}$ by maximizing the constrained expected complete-data likelihood:

$$\hat{\boldsymbol{\Gamma}} = \frac{1}{N \hat{\sigma}^2} \hat{\boldsymbol{S}}_b .$$

CM-step 3. Update $\hat{\boldsymbol{\beta}}$ by maximizing the expected complete-data likelihood:

$$\hat{\boldsymbol{\beta}} = \Big(\sum_{i=1}^{N} \boldsymbol{X}_i^{\top} \hat{\boldsymbol{C}}_i^{-1} \boldsymbol{X}_i \Big)^{-1} \sum_{i=1}^{N} \boldsymbol{X}_i^{\top} \hat{\boldsymbol{C}}_i^{-1} (\boldsymbol{Y}_i^{(\hat{\lambda})} - \boldsymbol{Z}_i \hat{\boldsymbol{b}}_i) .$$

CML-step 4. Update $\hat{\lambda}$, $\hat{\boldsymbol{\gamma}}_{\phi}$ and $\hat{\boldsymbol{\gamma}}_{\theta}$ by maximizing the following constrained actual log-likelihood:

$$-\frac{1}{2} \sum_{i=1}^{N} \log |\boldsymbol{Z}_i \hat{\boldsymbol{\Gamma}} \boldsymbol{Z}_i^{\top} + \boldsymbol{C}_i| + (\lambda - 1) \sum_{i=1}^{N} \sum_{j=1}^{N_i} \log |Y_{ij} + \nu|$$

$$-\frac{1}{2\hat{\sigma}^2} \sum_{i=1}^{N} (\boldsymbol{Y}_i^{(\lambda)} - \boldsymbol{X}_i \hat{\boldsymbol{\beta}})^{\top} (\boldsymbol{Z}_i \hat{\boldsymbol{\Gamma}} \boldsymbol{Z}_i^{\top} + \boldsymbol{C}_i)^{-1} (\boldsymbol{Y}_i^{(\lambda)} - \boldsymbol{X}_i \hat{\boldsymbol{\beta}}) ,$$

where $\boldsymbol{C}_i$ is an implicit function of $\boldsymbol{\gamma}_{\phi}$ and $\boldsymbol{\gamma}_{\theta}$. The CML step can be done by a straightforward function-minimization algorithm such as S-PLUS (nlminb function) with bounded constraint of $(-1, 1)$ on the γ_{ϕ_i} and the γ_{θ_i}. Once the MLEs of $\boldsymbol{\gamma}_{\phi}$ and $\boldsymbol{\gamma}_{\theta}$ are obtained, we can obtain the MLEs of $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ by (2).

The above ECME algorithm is easy to implement but might be slow to converge in some situations. In the case that the convergence is too slow, one can first perform a moderate number of iterations of ECME and then switch to Newton–Raphson iterations.

3. Prediction of future values

Consider the prediction of $\boldsymbol{y}_j$, a set of future k-dimensional measurements of $\boldsymbol{Y}_j$, given the observed repeated measurements $\boldsymbol{Y} = (\boldsymbol{Y}_{(j)}, \boldsymbol{Y}_j)$; $\boldsymbol{Y}_{(j)} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{j-1}, \boldsymbol{Y}_{j+1}, \ldots, \boldsymbol{Y}_N)$.

This is a time-series prediction that is of practical interest for many types of growth curve data. Because the ARMA(p, q) dependence structure considered in this paper is covariance stationary, the functional form of the dependence between $\boldsymbol{y}_j^{(\lambda)}$ and $\boldsymbol{Y}_j^{(\lambda)}$ is known.

Suppose that $\boldsymbol{y}_j^{(\lambda)} = \boldsymbol{x}_j\boldsymbol{\beta} + \boldsymbol{z}_j\boldsymbol{b}_i + \boldsymbol{\varepsilon}_j^*$ satisfies the GCM assumptions, where $\boldsymbol{x}_j$ and $\boldsymbol{z}_j$ denote the design matrices corresponding to $\boldsymbol{y}_j^{(\lambda)}$; $\boldsymbol{x}_j$ is $k\times m_1$ and $\boldsymbol{z}_j$ is $k\times m_2$. Then $\mathrm{E}((\boldsymbol{Y}_j^{(\lambda)}, \boldsymbol{y}_j^{(\lambda)})) = [\boldsymbol{X}_j \ \boldsymbol{x}_j]\boldsymbol{\beta}$ and $\mathrm{D}((\boldsymbol{Y}_j^{(\lambda)}, \boldsymbol{y}_j^{(\lambda)})) = \sigma^2(\boldsymbol{Z}_j^*\boldsymbol{\Gamma}\boldsymbol{Z}_j^{*\top} + \boldsymbol{C}_j^*) = \sigma^2\boldsymbol{\Omega}$, where $\boldsymbol{Z}_j^* = [\boldsymbol{Z}_j \ \boldsymbol{z}_j]$ and $\boldsymbol{C}_j^* = [\rho_{|g-h|}]$ $(g, h = 1, \ldots, N_j + k)$ and $\boldsymbol{\Omega}$ is correspondingly partitioned as

$$\boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{bmatrix}.$$

Furthermore, the autocorrelations ρ_j can be obtained by the following recursive relations

$$\rho_h = \begin{cases} \phi_1\rho_{h-1} + \cdots + \phi_p\rho_{h-p} - \theta_1\psi_{h-1} - \cdots - \theta_q\psi_{h-q} & h = 1, \ldots, q, \\ \phi_1\rho_{h-1} + \cdots + \phi_p\rho_{h-p} & h = q+1, \ q+2, \ldots, \end{cases}$$

where $\psi_h = \mathrm{E}(\varepsilon_{t-h}a_t)/\operatorname{var}(\varepsilon_{t-h})$.

Conditional on the variance components in $\boldsymbol{\alpha}$ and λ,

$$\tilde{\boldsymbol{y}}_j^{(\lambda)} = \boldsymbol{x}_j\tilde{\boldsymbol{\beta}} + \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}(\boldsymbol{Y}_j^{(\lambda)} - \boldsymbol{X}_j\tilde{\boldsymbol{\beta}}) \tag{3}$$

is the best linear unbiased predictor (BLUP) of $\boldsymbol{y}_j^{(\lambda)}$, where

$$\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\boldsymbol{\Gamma}, \boldsymbol{\phi}, \boldsymbol{\theta}) = \left(\sum_{i=1}^{N} \boldsymbol{X}_i^\top\boldsymbol{\Lambda}_i^{-1}\boldsymbol{X}_i\right)^{-1} \sum_{i=1}^{N} \boldsymbol{X}_i^\top\boldsymbol{\Lambda}_i^{-1}\boldsymbol{Y}_i^{(\lambda)}.$$

After some algebraic manipulations, (3) can be rewritten as

$$\tilde{\boldsymbol{y}}_j^{(\lambda)} = \boldsymbol{x}_j\tilde{\boldsymbol{\beta}} + \boldsymbol{z}_j\tilde{\boldsymbol{b}}_j + \boldsymbol{V}_{j*}\boldsymbol{C}_j^{-1}(\boldsymbol{Y}_j^{(\lambda)} - \boldsymbol{X}_j\tilde{\boldsymbol{\beta}} - \boldsymbol{Z}_j\tilde{\boldsymbol{b}}_j),$$

where $\tilde{\boldsymbol{b}}_i = \boldsymbol{\Gamma}^{-1}\boldsymbol{Z}_i^\top\boldsymbol{\Lambda}_i^{-1}(\boldsymbol{Y}_i^{(\lambda)} - \boldsymbol{X}_i\tilde{\boldsymbol{\beta}})$ is the BLUP of $\boldsymbol{b}_i$ and $\boldsymbol{V}_{j*} = \mathrm{D}(\boldsymbol{\varepsilon}_j, \boldsymbol{\varepsilon}_j^*)$.

Based on the BLUP of $\boldsymbol{y}_j^{(\lambda)}$ with all the parameters replaced by their MLEs, we can predict the k-dimensional future value of the jth measurement, $\boldsymbol{y}_j$, as follows:

$$\hat{\boldsymbol{y}}_j = \begin{cases} \left(\boldsymbol{1} + \hat{\lambda}\left(\boldsymbol{x}_j\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\Omega}}_{21}\hat{\boldsymbol{\Omega}}_{11}^{-1}(\boldsymbol{Y}_j^{(\hat{\lambda})} - \boldsymbol{X}_j\hat{\boldsymbol{\beta}})\right)\right)^{1/\hat{\lambda}} - \nu\boldsymbol{1} & \text{where } \hat{\lambda} \neq 0, \\ \exp\left(\boldsymbol{x}_j\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\Omega}}_{21}\hat{\boldsymbol{\Omega}}_{11}^{-1}(\boldsymbol{Y}_j^{(\hat{\lambda})} - \boldsymbol{X}_j\hat{\boldsymbol{\beta}})\right) - \nu\boldsymbol{1} & \text{where } \hat{\lambda} = 0, \end{cases} \tag{4}$$

where $\boldsymbol{1} = (1, \ldots, 1)$ is a $k\times 1$ vector and $\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\hat{\boldsymbol{\Gamma}}, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})$. This type of prediction has been considered by Rao (1987) and Lee (1988), among others. In (4), we use the convention that $\boldsymbol{b}^a = (b_1^a, \ldots, b_k^a)$ and $\exp(\boldsymbol{b}) = (e^{b_1}, \ldots, e^{b_k})$. When $k = 1$, an approximate $1-\alpha$ predictive interval for y_j is given by $\hat{y}_j \pm \xi_\alpha\hat{\sigma}_{y_j}$, where ξ_α denotes the $(1 - \frac{1}{2}\alpha)$-quantile of the standard normal distribution and $\sigma_{y_j}^2 = \sigma^2(h'(\mu_{2\cdot1}))(\boldsymbol{\Omega}_{22\cdot1} + \boldsymbol{H}_j\boldsymbol{W}^{-1}\boldsymbol{H}_j^\top)$ is the variance of the forecast error, where $\boldsymbol{H}_j = \boldsymbol{x}_j - \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{X}_j$, $\boldsymbol{W} = \sum_{i=1}^{N}\boldsymbol{X}_i^\top\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{X}_i$, $\mu_{2\cdot1} = \boldsymbol{x}_j\hat{\boldsymbol{\beta}} + \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}(\boldsymbol{Y}_j - \boldsymbol{X}_j\hat{\boldsymbol{\beta}})$, $h'(u) = (1+\lambda u)^{(1-\lambda)/\lambda}$ and $\boldsymbol{\Omega}_{22\cdot1} = \boldsymbol{\Omega}_{22} - \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\Omega}_{12}$.

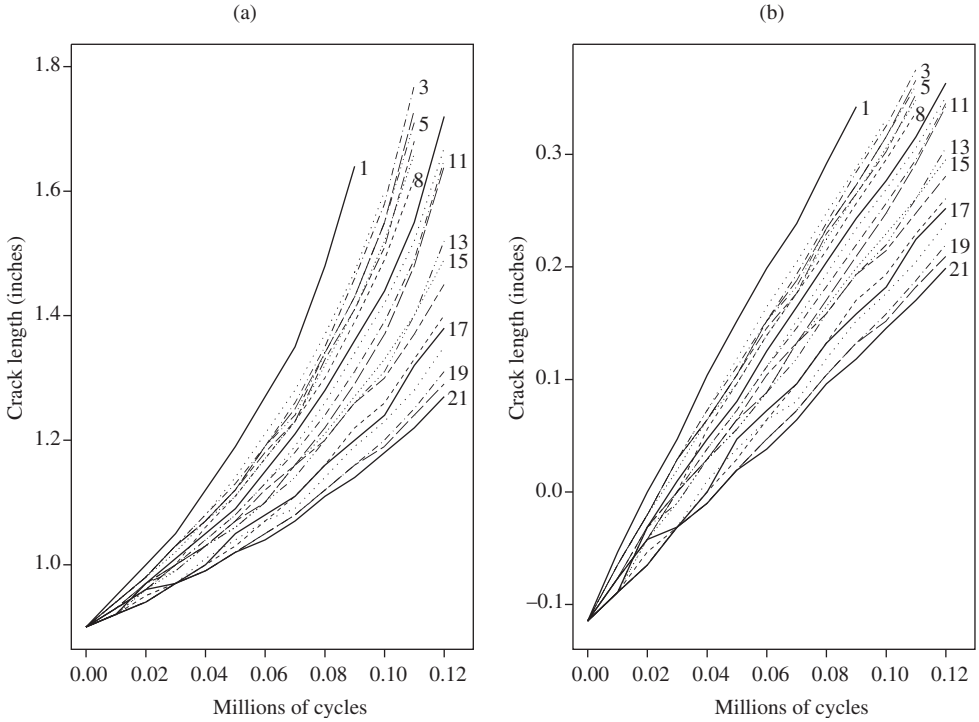

Figure 1. Fatigue crack growth data over 21 units: (a) original data; (b) data after Box–Cox transformation with $\lambda = -1.59$

For the LMM, the prediction of a future value of any lead time will be similar, because the measurements within a given unit are assumed independent. Here we consider the prediction of y_{ik}, for $i = 1, \ldots, N$, a future value corresponding to time t_{N_i+k} for the ith unit.

$$\hat{y}^*_{ik} = \eta(t_{N_i+k};\ \hat{b}_{i1},\ \hat{b}_{i2}) = -\frac{1}{\hat{b}_{i1}} \log(1 - 0.90^{\hat{b}_{i2}}\, \hat{b}_{i1}\, \hat{b}_{i2}\, t_{N_i+k}),$$

where $\hat{y}^*_{ik} = \log(\hat{y}_{ik}/0.90)$, so that the prediction of y_{ik} is given by

$$\hat{y}_{ik} = 0.90 e^{\hat{y}^*_{ik}}. \tag{5}$$

4. Application to fatigue crack growth data

In this section, we apply the results developed in Sections 2 and 3 to fatigue crack growth data of Bogdanoff & Kozin (1985). The data are unbalanced because the numbers of observations for the 21 units are not all equal. After applying the suitable Box–Cox transformation with $\lambda = -1.59$, which is the MLE, the transformed data are plotted in Figure 1(b). We found the growth function for the mean response to be linear and that all observations could be considered in the same group. The growth function also exhibits random effects on the slope. Thus the design matrices are $\boldsymbol{X}_i = [\mathbf{1}\ \ \boldsymbol{k}_i]$ and $\boldsymbol{Z}_i = \boldsymbol{k}_i$, where $\boldsymbol{k}_i = (1, 2, \ldots, N_i)$, for $i = 1, 2, \ldots, 21$.

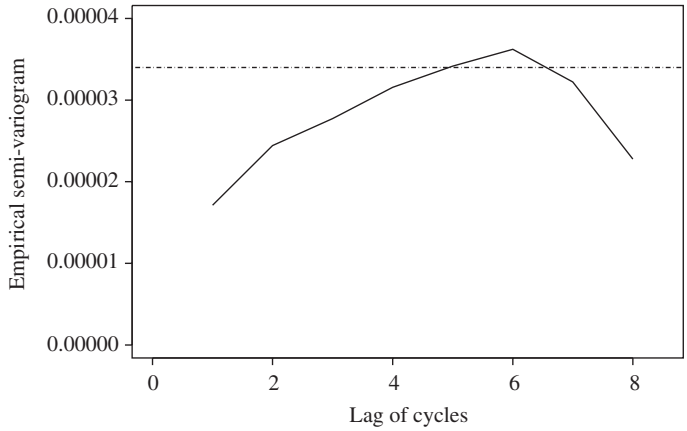

Figure 2. Empirical semi-variogram of subject-specific residuals for the fatigue crack growth data. The horizontal line is the estimated within errors variance ($\hat{\sigma}^2$).

TABLE 1

MLEs of selected ARMA models with the standard errors in parentheses for the fatigue crack growth data from Bogdanoff & Kozin (1985)

Model	$\hat{\Gamma}$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\lambda}$
	0.94	0.52	–	–	–	–	–	−1.59
AR(1)	(0.063)	(0.044)						(0.039)
	0.88	0.48	0.13	–	–	–	–	−1.57
AR(2)	(0.059)	(0.055)	(0.069)					(0.059)
	0.87	0.48	0.12	0.03	–	–	–	−1.58
AR(3)	(0.059)	(0.056)	(0.076)	(0.074)				(0.040)
	1.09	–	–	–	−0.39	–	–	−1.59
MA(1)	(0.074)				(0.057)			(0.033)
	1.04	–	–	–	−0.43	−0.27	–	−1.60
MA(2)	(0.070)				(0.056)	(0.065)		(0.033)
	0.97	–	–	–	−0.46	−0.29	−0.20	−1.59
MA(3)	(0.066)				(0.055)	(0.065)	(0.069)	(0.039)
	0.87	0.71	–	–	0.22	–	–	−1.58
ARMA(1,1)	(0.059)	(0.083)			(0.103)			(0.040)
	0.87	0.67	–	–	0.19	−0.04	–	−1.58
ARMA(1,2)	(0.059)	(0.137)			(0.102)	(0.110)		(0.040)

To check for serial correlation, the empirical semi-variogram plot of the residuals is a very useful graphical display for exploring the degree of association within units. The approach was first discussed by Diggle (1988) and later extended by Verbeke, Lesaffre & Brant (1998). We start by fitting the GCM with white noise errors,

$$\boldsymbol{Y}_i^{(\lambda)} = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i b_i + \boldsymbol{\varepsilon}_i\,, \qquad b_i \stackrel{\mathrm{d}}{=} \mathrm{N}(0, \sigma^2\Gamma)\,, \qquad \boldsymbol{\varepsilon}_i \stackrel{\mathrm{d}}{=} \mathrm{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I})\,. \tag{6}$$

Using the ECME algorithm to obtain the MLEs for the model (6), the fitted values are $\hat{\boldsymbol{\beta}} = (-0.150, 0.033)$, $\hat{\sigma}^2 = 3.31\times10^{-5}$, $\hat{\Gamma} = 1.112$ and $\hat{\lambda} = -1.583$.

We follow Morrell *et al.* (1995) and examine the subject-specific residuals, that is $\boldsymbol{r}_i = \boldsymbol{Y}_i^{(\hat{\lambda})} - \boldsymbol{X}_i\hat{\boldsymbol{\beta}} - \boldsymbol{Z}_i\hat{\boldsymbol{b}}$, in which $\hat{\boldsymbol{b}}_i = \hat{\Gamma}\boldsymbol{Z}_i^{\top}\hat{\boldsymbol{\Lambda}}_i^{-1}(\boldsymbol{Y}_i^{(\hat{\lambda})} - \boldsymbol{X}_i\hat{\boldsymbol{\beta}})$ is the empirical Bayes estimate for the random effects $\boldsymbol{b}_i$ (Laird & Ware, 1982), and $\boldsymbol{r}_i = (r_{i1}, \ldots, r_{iN_i})$. We computed the semi-variogram $d^2_{ii,jk} = \frac{1}{2}(r_{ij} - r_{ik})^2$ for all distinct pairs of observations within each unit. Figure 2 is the plot of the empirical semi-variogram for the subject-specific residuals, using

TABLE 2

Comparison of parameter estimates and forecast accuracy in terms of MARD($\times 10^{-2}$) for the GCM with various ARMA(p, q) models and LMM, starting with a sample of size 5 for each unit

				Step ahead					Step ahead	
T	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$	One	Two	$\hat{\phi}_1$	$\hat{\theta}_1$	$\hat{\theta}_2$	One	Two
				AR(1)					ARMA(1,1)	
6	0.24	–	–	0.635	–	0.52	0.24	–	0.542	–
7	0.32	–	–	0.556	0.946	0.42	0.10	–	0.551	0.693
8	0.36	–	–	0.681	0.829	0.69	0.32	–	0.627	0.799
9	0.40	–	–	0.891	0.652	0.81	0.43	–	0.939	0.723
10	0.34	–	–	0.559	1.485	0.62	0.28	–	0.584	1.538
11	0.39	–	–	0.622	0.878	0.60	0.21	–	0.607	0.838
12	0.43	–	–	1.123	1.232	0.73	0.32	–	1.085	1.196
Average				0.724	1.004				0.705	0.965
				AR(2)					ARMA(1,2)	
6	0.26	0.12	–	0.628	–	–0.01	–0.29	–0.26	0.622	–
7	0.32	0.06	–	0.548	0.820	–0.36	–0.69	–0.30	0.565	0.743
8	0.36	0.15	–	0.633	0.797	0.54	0.17	–0.08	0.636	0.845
9	0.39	0.17	–	0.915	0.736	0.82	0.44	0.01	0.940	0.728
10	0.33	0.11	–	0.558	1.531	0.01	0.33	0.20	0.558	1.569
11	0.38	0.10	–	0.608	0.846	0.13	–0.24	–0.20	0.628	0.881
12	0.41	0.15	–	1.084	1.188	0.75	0.33	0.01	1.088	1.166
Average				0.711	0.986				0.720	0.989
				AR(3)					LMM	
6	0.23	0.10	–0.25	0.613	–	–	–	–	0.870	–
7	0.27	0.02	–0.21	0.602	0.780	–	–	–	0.661	2.150
8	0.36	0.15	–0.03	0.645	0.928	–	–	–	0.704	1.341
9	0.39	0.16	0.07	0.937	0.719	–	–	–	1.035	1.465
10	0.33	0.11	–0.05	0.569	1.560	–	–	–	0.638	1.687
11	0.37	0.11	–0.06	0.617	0.858	–	–	–	0.833	1.234
12	0.41	0.14	0.02	1.081	1.165	–	–	–	1.127	1.804
Average				0.723	1.002				0.838	1.614

a mean for each time point. The plot shows that the mean of the semi-variogram exhibits an increasing trend up to the estimated process variance (horizontal line) indicating some serial correlations among the residuals. However, it decreases at higher lags because the estimates are obtained by using progressively fewer observations. We therefore explore some possible dependence of the data through the ARMA(p, q) models as it is a rich family for describing dependence among observations.

Table 1 displays the MLEs of the variance components and λ with their asymptotic standard errors for the selected ARMA(p, q) models when the entire dataset is used. The ARMA (p, q) models are chosen with $p + q \leq 3$ for parsimony. In the table, we see that the parameters ϕ_3 for AR(3) and θ_2 for ARMA(1,2) cannot be distinguished from 0, suggesting that the two models can be reduced to AR(2) and ARMA(1,1), respectively. The MLEs do not converge for ARMA(2,1).

We now compare the prediction abilities of the ARMA models and LMM. We adopt the predictive sample reuse, or leave-one-out, method as advocated by Geisser (1975). In the empirical comparison of forecast accuracy among the various models, we start with the first five measurements for each unit as the sample, so that we have stable parameter estimation, especially for the LMM. We sequentially fit GCM with various ARMA(p, q) models and LMM, then compute the forecasts (up to two-step-ahead) by applying the MLEs in (4) and (5). In

TABLE 3

Forecast of failure time for the proposed GCM and LMM using the first six observations of each unit as the sample. The tabulated numbers represented the deviation of the prediction from the observed failure time.

Path	Failure time		Models					
	Million cycles	Order of observations	AR(1)	AR(2)	AR(3)	ARMA(1,1)	ARMA(1,2)	LMM
1	0.088	10	+1	+1	+1	+1	+1	+1
2	0.100	11	+1	+1	+1	+1	+1	+1
3	0.101	12	0	0	0	0	0	+1
4	0.103	12	0	0	0	0	0	+3
5	0.103	12	0	0	0	0	0	+3
6	0.106	12	+1	0	0	0	0	+3
7	0.106	12	+1	+1	+1	+1	+1	+4
8	0.109	12	+1	+1	+1	+1	+1	−1
9	0.113	13	+1	+1	+1	0	+1	+1
10	0.115	13	+1	+1	+1	+1	+1	−1
11	0.118	13	+1	+1	+1	+1	+1	+1
12	0.118	13	+1	+2	+1	+1	+1	+8
13	0.129	14	+1	+1	+1	+1	+1	+3
14	0.133	15	0	0	0	0	0	+1
15	0.138	15	0	0	0	0	0	0
16	0.144	16	+1	+1	+1	+1	+1	+1
17	0.146	16	0	0	0	0	0	−2
18	0.151	17	0	0	0	0	0	−6
19	0.160	17	+1	+1	+1	+1	+1	0
20	0.167	18	0	0	0	0	0	−1
21	0.170	18	0	0	0	0	0	−1
Sum of absolute deviations:			12	12	11	10	11	43

assessing the relative merits of the models considered in this paper, we compare the actual observed values with the prediction of $\boldsymbol{Y}_{j,N_j+k}$ for $k = 1, 2,$ given the past observations. Let T denote the time of the point being forecast. We compare the actual observed values with one-step-ahead and two-step-ahead forecasts of crack length using the most recent $T-1$ and $T-2$ observations for each time series. For comparing the forecast accuracy, we assess the relative merits of the predictors from their mean absolute relative deviations (MARD). The best model yields the smallest MARD. Let $\hat{y}_j$ be the forecast for the measurement y_j. This discrepancy measurement is defined as

$$\text{MARD} = \frac{1}{N}\sum_{j=1}^{N}\frac{|\hat{y}_j - y_j|}{y_j}.$$

We fit the GCM with various ARMA(p, q) models and LMM and list the results of MARD from $T = 6$ to 12 in Table 2. The data are incomplete when $T > 10$ and we omit the result for $T = 13$ since over one-third of the units are missing at that time. Since MA models perform worse than the AR and ARMA models, they are excluded from comparison. Table 2 shows that the best model is ARMA(1,1), followed by AR(2), ARMA(1,2), AR(3), AR(1) and LMM, for both the one-step-ahead and the two-step-ahead predictions. Thus, the selected ARMA(p, q) models are superior to AR(1) and are much better than LMM. We conclude that ARMA(1,1) and AR(2) are the two best models for this dataset.

In Table 3, we compare the prediction of failure times among the various models. The second column in Table 3 gives the failure time in millions of cycles for each unit, and the

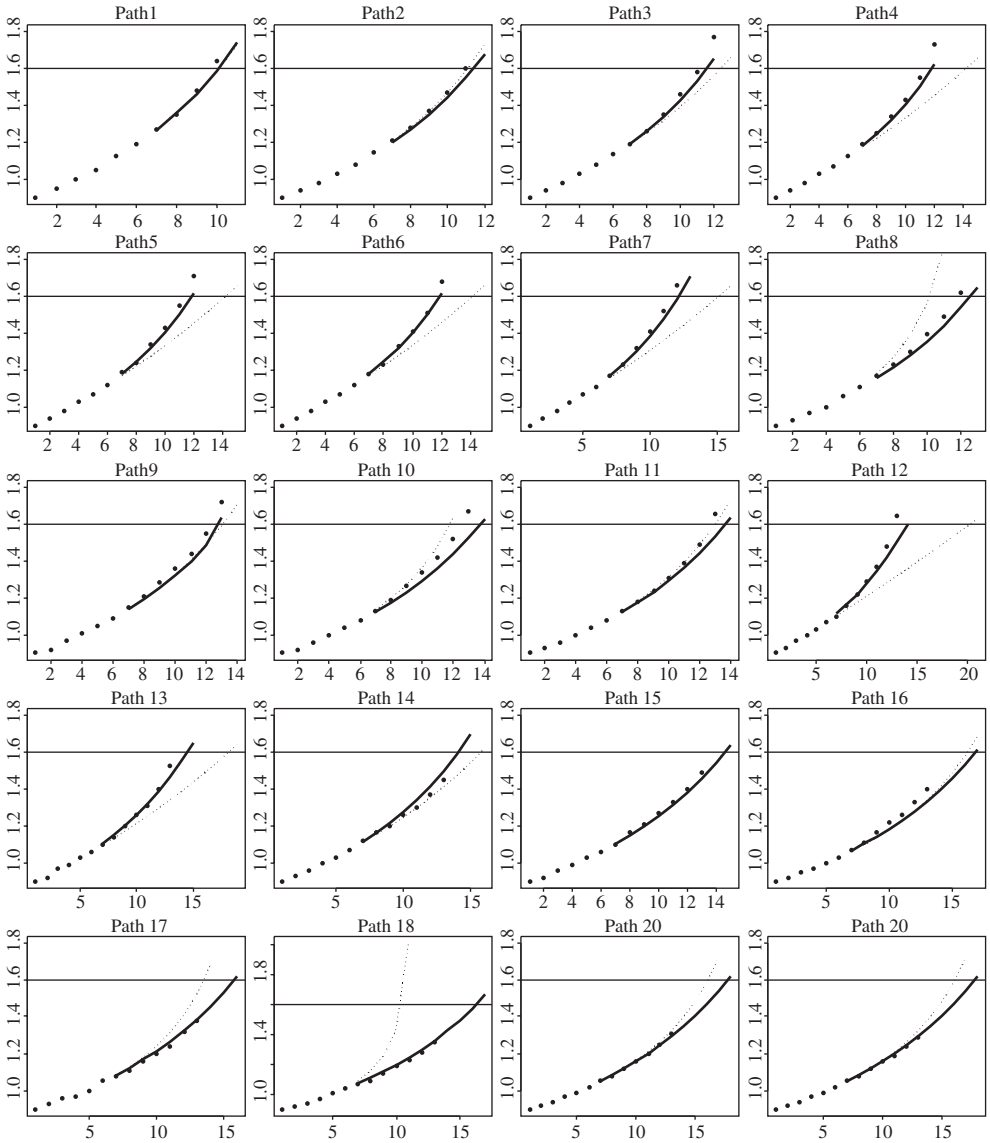

Figure 3. Forecast of failure time against million cycles for the first twenty units: the original data (•); the forecast values of GCM–ARMA(1,1) (solid line) and the forecast values of LMM (dotted line). The failure crack length is 1.6 inches, indicated by the horizontal line.

corresponding order for the failure times is given in column 3. A better crack length prediction model should provide more accurate predictions of failure times because there is a one-to-one correspondence between the crack length and the failure time. In comparing the prediction accuracy of failure times among the various models, we use the sum of absolute deviations between the predictions and the actual measurements as the measure of precision, for the first six observations of each unit (Table 3).

Figure 3 shows the crack growth length forecasts for GCM with ARMA(1,1) dependence and LMM for the first six observations of each unit. The GCM performs much better than

LMM for forecasting both the crack length and the failure time. The failure time prediction results are similar using the first five or the first seven observations of each unit as the sample.

5. Concluding remarks

From the results presented in Section 4 we know that if there are only a few measurements on each unit, it may be too hard to estimate the autocorrelation. Also, the data in each unit are time-series in nature and hence are not independent. Therefore, we can use the general growth curve model with ARMA(p, q) covariance structures to analyse this kind of data, using measurements from similar units to get better prediction results. The advantage of our modelling for this type of data is evident in the comparisons of forecast accuracy in future values and in failure times.

As remarked in Rochon (1992), ARMA(p, q) covariance structures are worth considering and may have better performance than AR(1) dependence in many applications. For the modelling of degradation data, with appropriate ARMA(p, q) covariance structure and coupled with random effects and the Box–Cox transformation, our modelling approach makes the prediction results quite appealing.

Appendix

In this appendix, we derive the information matrix which is useful for obtaining the standard errors of the MLEs. Let $\boldsymbol{\omega} = (\boldsymbol{\eta}, \boldsymbol{\phi}, \boldsymbol{\theta})$, $\boldsymbol{\eta}$ involve the distinct elements of $\boldsymbol{\Gamma}$. Then the log-likelihood of $\boldsymbol{\alpha} = (\boldsymbol{\beta}, \sigma^2, \boldsymbol{\omega}, \lambda)$, omitting the constant term for the data $\boldsymbol{Y}$, is:

$$\ell(\boldsymbol{\alpha}) = -\tfrac{1}{2} n \log(\sigma^2) - \frac{1}{2} \sum_{i=1}^{N} \log \left| \boldsymbol{\Lambda}_i(\boldsymbol{\Gamma}, \boldsymbol{\phi}, \boldsymbol{\theta}) \right| - \frac{1}{2\sigma^2} \sum_{i=1}^{N} \operatorname{tr} \left(\boldsymbol{e}_i \boldsymbol{e}_i^{\top} \boldsymbol{\Lambda}_i^{-1}(\boldsymbol{\Gamma}, \boldsymbol{\phi}, \boldsymbol{\theta}) \right) + (\lambda - 1) \sum_{i=1}^{N} \sum_{j=1}^{N_i} \log |Y_{ij} + \nu| ,$$

where $\boldsymbol{e}_i = \boldsymbol{Y}_i^{(\lambda)} - \boldsymbol{X}_i \boldsymbol{\beta}$ and $\operatorname{tr}(\boldsymbol{A})$ denotes the trace of $\boldsymbol{A}$.

The Hessian matrix, $\mathbf{H}(\boldsymbol{\alpha}) = -\partial^2 \ell / \partial \boldsymbol{\alpha} \boldsymbol{\alpha}^{\top}$, has the following form:

$$\mathbf{H}(\boldsymbol{\alpha}) = \begin{bmatrix} \mathbf{H}_{\beta} & \mathbf{H}_{\beta\sigma^2} & \mathbf{H}_{\beta\omega} & \mathbf{H}_{\beta\lambda} \\ \mathbf{H}_{\sigma^2\beta} & \mathbf{H}_{\sigma^2} & \mathbf{H}_{\sigma^2\omega} & \mathbf{H}_{\sigma^2\lambda} \\ \mathbf{H}_{\omega\beta} & \mathbf{H}_{\omega\sigma^2} & \mathbf{H}_{\omega} & \mathbf{H}_{\omega\lambda} \\ \mathbf{H}_{\lambda\beta} & \mathbf{H}_{\lambda\sigma^2} & \mathbf{H}_{\lambda\omega} & \mathbf{H}_{\lambda} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial^2 \ell}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^{\top}} & \dfrac{\partial^2 \ell}{\partial \boldsymbol{\beta} \partial \sigma^2} & \dfrac{\partial^2 \ell}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^{\top}} & \dfrac{\partial^2 \ell}{\partial \boldsymbol{\beta} \partial \lambda} \\ \dfrac{\partial^2 \ell}{\partial \sigma^2 \partial \boldsymbol{\beta}^{\top}} & \dfrac{\partial^2 \ell}{\partial (\sigma^2)^2} & \dfrac{\partial^2 \ell}{\partial \sigma^2 \partial \boldsymbol{\omega}^{\top}} & \dfrac{\partial^2 \ell}{\partial \sigma^2 \partial \lambda} \\ \dfrac{\partial^2 \ell}{\partial \boldsymbol{\omega} \partial \boldsymbol{\beta}^{\top}} & \dfrac{\partial^2 \ell}{\partial \boldsymbol{\omega} \partial \sigma^2} & \dfrac{\partial^2 \ell}{\partial \boldsymbol{\omega} \partial \boldsymbol{\omega}^{\top}} & \dfrac{\partial^2 \ell}{\partial \boldsymbol{\omega} \partial \lambda} \\ \dfrac{\partial^2 \ell}{\partial \lambda \partial \boldsymbol{\beta}^{\top}} & \dfrac{\partial^2 \ell}{\partial \lambda \partial \sigma^2} & \dfrac{\partial^2 \ell}{\partial \lambda \partial \boldsymbol{\omega}^{\top}} & \dfrac{\partial^2 \ell}{\partial \lambda^2} \end{bmatrix} .$$

Letting $s = \dim(\boldsymbol{\eta}) + \dim(\boldsymbol{\phi}) + \dim(\boldsymbol{\theta}) = \frac{1}{2} m_2 (m_2 + 1) + p + q$ denote the dimension of $\boldsymbol{\omega}$, the expressions for the elements of $\mathbf{H}(\boldsymbol{\alpha})$ are:

$$\mathbf{H}_{\beta} = \frac{-1}{\sigma^2} \sum_{i=1}^{N} \boldsymbol{X}_i^{\top} \boldsymbol{\Lambda}_i^{-1} \boldsymbol{X}_i ,$$

$$\mathbf{H}_{\beta\sigma^2} = \frac{-1}{\sigma^4}\sum_{i=1}^{N} \boldsymbol{X}_i^{\top}\boldsymbol{\Lambda}_i^{-1}\boldsymbol{e}_i\,,$$

$$[\mathbf{H}_{\beta\omega}]_{jk} = \frac{-1}{\sigma^2}\sum_{i=1}^{N} X_{ij}^{\top}\boldsymbol{\Lambda}_i^{-1}\dot{\boldsymbol{\Lambda}}_{ik}\boldsymbol{\Lambda}_i^{-1}\boldsymbol{e}_i \qquad (j,k=1,\ldots,s)\,,$$

$$\mathbf{H}_{\beta\lambda} = \frac{1}{\sigma^2}\sum_{i=1}^{N} \boldsymbol{X}_i^{\top}\boldsymbol{\Lambda}_i^{-1}\boldsymbol{g}_{1i}\,,$$

$$\mathbf{H}_{\sigma^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{N}\operatorname{tr}\left(\boldsymbol{e}_i\boldsymbol{e}_i^{\top}\boldsymbol{\Lambda}_i^{-1}\right),$$

$$[\mathbf{H}_{\sigma^2\omega}]_k = \frac{1}{2\sigma^4}\sum_{i=1}^{N}\operatorname{tr}\left(\boldsymbol{\Lambda}_i^{-1}(\boldsymbol{e}_i\boldsymbol{e}_i^{\top}-\boldsymbol{\Lambda}_i)\boldsymbol{\Lambda}_i^{-1}\dot{\boldsymbol{\Lambda}}_{ik}\right) \qquad (k=1,\ldots,s)\,,$$

$$\mathbf{H}_{\sigma^2\lambda} = \frac{1}{\sigma^4}\sum_{i=1}^{N}\operatorname{tr}(\boldsymbol{e}_i\boldsymbol{g}_{1i}^{\top}\boldsymbol{\Lambda}_i^{-1})\,,$$

$$[\mathbf{H}_{\omega}]_{jk} = \frac{1}{2\sigma^2}\sum_{i=1}^{N}\operatorname{tr}\left(\boldsymbol{\Lambda}_i^{-1}\dot{\boldsymbol{\Lambda}}_{ij}\boldsymbol{\Lambda}_i^{-1}(2\boldsymbol{e}_i\boldsymbol{e}_i^{\top}-\boldsymbol{\Lambda}_i)\boldsymbol{\Lambda}_i^{-1}\dot{\boldsymbol{\Lambda}}_{ik}\right)$$
$$-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\operatorname{tr}\left(\boldsymbol{\Lambda}_i^{-1}(\boldsymbol{e}_i\boldsymbol{e}_i^{\top}-\boldsymbol{\Lambda}_i)\boldsymbol{\Lambda}_i^{-1}\ddot{\boldsymbol{\Lambda}}_{i,jk}\right) \qquad (j,k=1,\ldots,s)\,,$$

$$[\mathbf{H}_{\omega\lambda}]_k = \frac{-1}{\sigma^2}\sum_{i=1}^{N}\operatorname{tr}(\boldsymbol{e}_i\boldsymbol{g}_i^{\top}\boldsymbol{\Lambda}_i^{-1}\ddot{\boldsymbol{\Lambda}}_{ik}\boldsymbol{\Lambda}_i^{-1}) \qquad (k=1,\ldots,s)\,,$$

$$\mathbf{H}_{\lambda} = \frac{-1}{\sigma^2}\sum_{i=1}^{N}\operatorname{tr}(\boldsymbol{e}_i\boldsymbol{g}_{2i}^{\top}\boldsymbol{\Lambda}_i^{-1}+\boldsymbol{g}_{1i}\boldsymbol{g}_{1i}^{\top}\boldsymbol{\Lambda}_i^{-1})\,,$$

where X_{ij} is the jth column of $\boldsymbol{X}_i$, $\boldsymbol{g}_{1i}=\partial\boldsymbol{e}_i/\partial\lambda$, $\boldsymbol{g}_{2i}=\partial^2\boldsymbol{e}_i/\partial\lambda^2$, $\dot{\boldsymbol{\Lambda}}_{ik}=\partial\boldsymbol{\Lambda}_i/\partial\omega_{ik}$, and $\ddot{\boldsymbol{\Lambda}}_{i,jk}=\partial^2\boldsymbol{\Lambda}_i/\partial\omega_{ij}\partial\omega_{ik}$.

It follows from classical likelihood theory that, when the assumed model is correct, the MLE $\hat{\boldsymbol{\alpha}}$ has an asymptotic multivariate normal distribution with mean vector $\boldsymbol{\alpha}$ and covariance matrix $-\mathbf{H}^{-1}(\hat{\boldsymbol{\alpha}})$.

References

Barndorff-Nielsen, O.E. & Schou, G. (1973). On the reparameterization of autoregressive models by partial autocorrelations. *J. Multivariate Anal.* **3**, 408–419.

Bogdanoff, J.L. & Kozin, F. (1985). *Probabilistic Models of Cumulative Damage.* New York: Wiley.

Box, G.E.P. & Cox, D.R. (1964). An analysis of transformations (with discussion). *J. Roy. Statist. Soc. Ser. B* **26**, 211–252.

Box, G.E.P., Jenkins, G.M. & Reinsel, G.C. (1994). *Time Series Analysis Forecasting and Control.* 3rd edn. San Francisco: Holden–Day.

Chi, E.M. & Reinsel, G.C. (1989). Models for longitudinal data with random effects and AR(1) errors. *J. Amer. Statist. Assoc.* **84**, 452–459.

Dempster, A.P., Laird, N.M. & Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39**, 1–38.

DIGGLE, P.J. (1988). An approach to the analysis of repeated measurement. *Biometrics* **44**, 959–971.

GEISSER, S. (1975). The predictive sample reuse method with applications. *J. Amer. Statist. Assoc.* **70**, 320–328.

JENNRICH, R.I. & SCHLUCHTER, M.D. (1986). Unbalanced repeated–measures models with structured covariance matrices. *Biometrics* **42**, 805–820.

KERAMIDAS, E.M. & LEE, J.C. (1990). Forecasting technological substitutions with concurrent short time series. *J. Amer. Statist. Assoc.* **85**, 625–632.

LAIRD, N.M. & WARE, J.H. (1982). Random effects models for longitudinal data. *Biometrics* **38**, 963–974.

LEE, J.C. (1988). Prediction and estimation of growth curve with special covariance structures. *J. Amer. Statist. Assoc.* **83**, 432–440.

LINDSTROM, M.J. & BATES, D.M. (1988). Newton–Raphson and EM algorithms for linear mixed-effects models for repeated measures data. *J. Amer. Statist. Assoc.* **83**, 432–440.

LIU, C. & RUBIN, D.B. (1994). The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence. *Biometrika* **81**, 633–648.

LU, C.J. & MEEKER, W.Q. (1993). Using degradation measures to estimate a time-to-failure distribution. *Technometrics* **35**, 161–174.

MENG, X.L. & RUBIN, D.B. (1993). Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika* **80**, 267–278.

MONAHAN, J.F. (1984). A note on enforcing stationarity in autoregressive moving average models. *Biometrika* **71**, 403–404.

MORRELL, C.H., PEARSON, J.D., CARTER, H.B. & BRANT, L.J. (1995). Estimating unknown transition times using a piecewise non-linear mixed-effects model in men with prostate cancer. *J. Amer. Statist. Assoc.* **90**, 45–53.

PARIS, P.C. & ERDOGAN, F. (1963). A critical analysis of crack propagation laws. *J. Basic Engineering* **85**, 528–534.

POTTHOFF, R.F. & ROY, S.N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika* **51**, 313–326.

RAO, C.R. (1987). Prediction of future observations in growth curve model. *Statis. Sci.* **2**, 434–471.

ROCHON, J. (1992). ARMA covariance structures with time heteroscedasticity for repeated measures experiments. *J. Amer. Statist. Assoc.* **87**, 777–784.

VERBEKE, G., LESAFFRE, E. & BRANT, L.J. (1998). The detection of residual serial correlation in linear mixed models. *Statist. Med.* **17**, 1391–1402.