Generalized Linear Mixed Models for Reliability Analysis of Multi-Copy Repairable Systems

Furong Tan, Zhibin Jiang, Suk Joo Bae

Abstract — The power law process (PLP) is usually applied to the failure data from a single repairable system. When a system has a number of copies for analysis, the usual approach is to pool the data from all copies under the assumption that each copy is modeled by a same PLP. In the real world, however, it may be more reasonable to assume heterogeneity among all system copies. Therefore, in this paper, a new model based on the generalized linear mixed models (GLMM) is proposed for analyzing the failure data from multi-copy repairable systems. In the context of GLMM, the underlying model for each system copy is assumed to be a PLP at Stage 1, and parameters vary among copies at Stage 2. This GLMM-based model can make inferences about both the population and each system copy. A modified Anderson-Darling test is applied to the goodness-of-fit test of the model and a numerical application shows the effectiveness of the model.

Index Terms — multi-copy repairable system, reliability analysis, generalized linear mixed models, the power law process, maximum likelihood estimation, empirical Bayes estimate

ACRONYMS

CI confidence interval
GLIM generalized linear models
GLMM Generalized linear mixed models
HPP homogeneous Poisson process
i.i.d. independent and identically distributed
IUD in-service unplanned derated
MCF mean cumulative function
ML maximum likelihood
MLE maximum likelihood estimate
NHPP non-homogeneous Poisson process
PLP power law process
SE standard error
TAAF test, analyze and fix
s- implies: statistically

F. Tan is with the Dept. of Industrial Engineering and Management, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200030, China (e-mail: frtan@sjtu.edu.cn).
Z. Jiang is with the Dept. of Industrial Engineering and Management, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200030, China (e-mail: zbjiang@sjtu.edu.cn).
S.J. Bae is with the Department of Industrial Engineering, Hanyang University, Seoul, Korea (sjbae@hanyang.ac.kr).

2 The singular and plural of an acronym are always spelled the same.
NOTATION

$A'$, $a'$, transpose of a matrix $A$ or a vector $a$

$\hat{\xi}$, point estimate of unknown parameter (or parameter vector) $\xi$

$\text{Se}(\hat{\xi})$, standard error of the estimate $\hat{\xi}$

$A_i$, $t \times q$ known design matrix for subject $i$ (or copy $i$)

$b_i$, $v \times 1$ vector of random effects for subject $i$

$B_i$, $t \times v$ known matrix for copy $i$, elements of which are usually 0's and 1's

$D_i^2$, modified Anderson-Darling statistic for copy $i$

$D^2$, overall Anderson-Darling statistic

$g$, link function of GLMM or GLIM

$g^{-1}$, inverse function of $g$

$M_i$, $n_i - 1$

$\bar{N}$, overall mean failure number

$n$, number of copies

$n_i$, number of failures for copy $i$, $i = 1, \ldots, n$

$t_{ij}$, failure time of the $j$th failure of the $i$th copy, $j = 1, \ldots, n_i$

$\hat{U}_{ij}$, sequence of order statistics from uniform distribution for copy $i$, $j = 1, \ldots, n_i$

$w_i$, weight of copy $i$

$x_{ij}$, $q \times 1$ vector of regressors for the $j$th measurement of the $i$th subject

$X_i$, $n_i \times q$ design matrix for copy $i$

$X_i^*$, $n_i \times t$ design matrix for copy $i$

$y_{ij}$, the $j$th measurement (or failure) of the $i$th subject

$y_i$, $n_i \times 1$ vector of response variable for copy $i$

$z_{i,j}$, $1 - \alpha$ percentile of the standard normal distribution

$z_{ij}$, $v \times 1$ vector for the $j$th measurement of the $i$th subject

$Z_i$, $n_i \times v$ design matrix for copy $i$

$s$-significance level

$\beta$, $q \times 1$ vector of parameters, the fixed effects

$\eta$, symbol of the linear predictor

$\beta_i$, $t \times 1$ parameter vector for copy $i$

$\gamma, \theta$, shape, scale parameter of PLP

$\gamma_i, \theta_i$, shape, scale parameter of PLP for copy $i$

$\Theta, \Theta_1, \Theta_2$, full, fixed and random parameter space, respectively
\[
\lambda(t) \quad \text{intensity function of an NHPP}
\]
\[
\mu(t) \quad \text{mean value function, mean number of failures up to } t
\]
\[
\mu_{ij} \quad \text{mean value of the } j \text{th measurement (or failure) of the } i \text{th subject, i.e., } E(y_{ij})
\]
\[
\mu_i \quad n \times 1 \text{ vector of } \mu_{ij} \text{'s}
\]
\[
\Sigma \quad \text{variance-covariance matrix of } b_i
\]
\[
\sigma_{11}, \sigma_{12}, \sigma_{22} \quad \text{elements of } \Sigma
\]
\[
\phi \quad \text{dispersion parameter}
\]
\[
D_v(0, \Sigma) \quad v \text{-dimension multivariate distribution with mean } 0 \text{ and variance-covariance matrix } \Sigma
\]
\[
E(y_{ij} | b_i) \quad \text{conditional mean of } y_{ij} \text{ given the random effects } b_i
\]
\[
f(b_i) \quad \text{distribution of the random effects } b_i
\]
\[
f(y_i | b_i) \quad \text{probability of the response variable vector } y_i, \text{ conditional on } b_i
\]
\[
h(\mu_{ij}) \quad \text{a variance function}
\]
\[
H(\mu_i) \quad n \times n \text{ diagonal variance matrix for subject } i \text{ with the } j \text{th diagonal element equal to } h(\mu_{ij})
\]
\[
I(\hat{\Theta}_i) \quad \text{observed Fisher information}
\]
\[
I(\Theta_i) \quad \text{observed information matrix, Hessian matrix}
\]
\[
L(\gamma, \theta) \quad \text{single-system likelihood function}
\]
\[
L(y_i | \Theta_i) \quad \text{marginal density of } y_i
\]
\[
l(\Theta_i) \quad \text{marginal log-likelihood from the sample of } n \text{ copies}
\]
\[
N_v(0, \Sigma) \quad v \text{-dimension multivariate normal distribution with mean } 0 \text{ and variance-covariance matrix } \Sigma
\]

1. Introduction

The most commonly used stochastic process for modeling reliability growth of complex repairable systems is the non-homogeneous Poisson process (NHPP). Within the class of NHPP models, the Power Law Process (PLP) is most commonly discussed in the literature. The PLP, which is proposed by Crow [1] based on the empirical formulation of Duane [2], is an NHPP model with the intensity function:

\[
\lambda(t) = \frac{\gamma}{\theta} \left(\frac{t}{\theta}\right)^{\gamma-1}, \quad \gamma, \theta > 0; t \geq 0.
\] (1.1)

where \(\gamma\) and \(\theta\) are the shape and scale parameter, respectively. The corresponding mean cumulative number of failures over \((0, t]\) is \(\mu(t) = \left(\frac{t}{\theta}\right)^{\gamma}\).

Usually, NHPP models, like the PLP model (1.1), are applied to the data from a single repairable system [3], including some new models [4]. When a system has a
number of copies for analysis and the system reliability may be of interest in some applications (e.g., to estimate the system reliability at the end of each stage of a TAAF program [5], or to predict a new system copy), the usual approach is to pool the data from all copies under the assumption that each copy is modeled by an HPP or NHPP (i.e., every system copy has the same intensity function $\lambda(t)$). This approach is actually to assume homogeneity among all system copies. In the real world, however, when a system has a number of copies for analysis, there usually are known differences among copies which may affect the trend of their reliability. The copies of a same type of system may be operated on different platforms, in different positions on a given platform, with different stresses, operators, maintenance men, etc. [3]. Therefore it may be more reasonable to assume heterogeneity among all system copies.

Some researchers have considered this problem, and assume that the differences are reflected by the varying scale parameters, while the shape parameters are the same. If we treat all unknown parameters as random variables, we may have an alternative approach to account for the copy-to-copy variance. Mixed-effects models are then introduced to the reliability analysis of a repairable system with multiple copies.

Mixed-effects models, also called random-effects models, are a regression type of analysis which provides a common baseline for all subjects and enables the analyst not only to describe the trend over time within each subject, but also to describe the variation among different subjects. In the formulation of mixed-effects models, the probability distribution for the multiple observations has the same form for each subject, but the parameters of that distribution vary over subjects [6, 7].

Mixed-effects models provide a powerful and useful tool for analyzing repeated-measures data that arise in different areas of investigation, such as economics and pharmacokinetics. They can easily handle unbalanced repeated-measures data and allow explicit modeling and analysis of between- and within-subject variation. Because of these properties, mixed-effects models are widely used in medical studies [8-12].

However, in the field of reliability, the application of mixed-effects models seems not to be emphasized. In this paper, we attempt to use mixed-effects models to the reliability analysis of multi-copy repairable systems.

For analyzing the reliability of a multi-copy repairable system, it may be reasonable to assume that the underlying model for each system copy is a PLP at Stage 1, and parameters vary among copies at Stage 2. Under this assumption, i.e., the failure processes of all system copies are assumed to be Poisson processes, the generalized linear mixed models (GLMM) are typically used for analyzing failure data from the multi-copy repairable system.

GLMM are the extension of generalized linear models (GLIM). The GLIM is an important class of univariate nonlinear mixed-effects models introduced by Nelder and Wedderburn [13]. It extends the Gaussian-based linear model to models on the exponential family of distributions and allows the existence of a monotonic differentiable function $g$ that relates the mean $\mu$ of response variable $y$ to a set of regressors, $x = (x_1, ..., x_q)$, through a linear predictor $x'\beta$, denoted as $\eta$, where $\beta$ is a $q \times 1$ vector of parameters, and usually a fixed parameter vector [14-16]. This can be expressed as

$$g(\mu) = x'\beta = \eta.$$  (1.2)
GLMM extend GLIM by the inclusion of random effects in the predictor [17]. For binary, count and categorical data, GLMM are testified to be a useful tool [13-18].

In this paper, we will consider a GLMM-based approach for analyzing the failure data from a multi-copy repairable system. The primary purpose of this study is to introduce a new method to the reliability analysis of multi-copy systems. We will use this method to analyze the real data from a type of generating unit with 8 copies. Since the baseline model for all system copies is assumed to be the PLP, the assumptions of “ignoring repair times” and “minimal repair” are made throughout the rest of the paper.

This paper is organized as follows. In Section 2, after a brief statement of the GLMM, a unified approach to use GLMM to analyze the reliability of a multi-copy repairable system is described. Section 3 briefly discusses the estimation of parameters, including the fixed parameters and random parameters. In Section 4, a modified Anderson-Darling statistic is used for the goodness-of-fit test of the GLMM-based model. Section 5 is a numerical application of this method to the failure data from 8 generating units. In Section 6, the PLP model is applied to these data sets to make a comparison between the GLMM and the usual method. Section 7 gives a conclusion of this analysis and some suggestions for further investigations.

2. GLMM for Multi-copy Repairable Systems

2.1 The General Form of GLMM

Assume there are \( i = 1, \ldots, n \) subjects and \( j = 1, \ldots, n_i \) repeated observations within each subject. Denote \( y_{ij} \) as the \( j \)th measurement of the \( i \)th subject and the \( n_i \times 1 \) vector \( y_i = (y_{i1}, \ldots, y_{in_i})' \) contains all the measurements of the \( i \)th subject. The observations within subject \( i \) are assumed to be not independent of each other, while the observations among subjects are assumed to be independent. Based on the GLIM (1.2), a \( v \times 1 \) vector of random effects, \( b_i \), is added to the linear predictor to account for the correlation of the data within subject \( i \) [17]. The resulting model is a GLMM expressed by

\[
g(\mu_{ij}) = \eta_{ij} = x'_i \beta + z'_i b_i,
\]

where \( \mu_{ij} = E(y_{ij} \mid b_i) \) is the conditional mean of \( y_{ij} \) given the random effects \( b_i \); \( x_i \) is a \( v \times 1 \) vector of regressors with fixed effects \( \beta \) for the \( j \)th measurement of the \( i \)th subject; \( z_{ij} \) is a \( v \times 1 \) vector. The random effects vector \( b_i \) (not varying with \( j \) ) is usually assumed to follow a \( v \)-dimension multivariate normal distribution with mean \( 0 \) and variance-covariance matrix \( \Sigma \). \( b_i \)'s are assumed as independent and identically distributed random vectors [16]. Moreover, the model (2.1) can be more general by relaxing the assumptions on the \( b_i \). In particular, \( b_i \) can be assumed to follow any \( v \)-dimension distribution [14].

Writing \( \mu_i \) as the vector of \( \mu_{ij} \)'s, i.e., \( \mu_i = [\mu_{i1} \cdots \mu_{in_i}]' \), and thinking of \( g(\bullet) \) as a function that applies element by element, the GLMM of (2.1) is then written as
\[ g(\mu_i) = X_i\beta + Z_i b_i, \]
\[ b_i \sim i.i.d. D_i(0, \Sigma), \]
\[ i = 1, \ldots, n. \]

where \( X_i \) is a \( n_i \times q \) design matrix, the \( j \)th row of which is \( x_i' \); \( Z_i \) is a \( n_i \times v \) design matrix, the \( j \)th row of which is \( z_i' \); and \( D_i(0, \Sigma) \) denotes a \( v \)-dimensional distribution with mean 0 and variance-covariance \( \Sigma \).

The model (2.2) can be re-written in terms of the subject-specific GLMM of Zeger, Liang and Albert [19]. Let \( X_i\beta + Z_i b_i = X_i'\beta_i \) and \( \beta_i = A_i \beta + B_i b_i \), where \( X_i' \) is a \( n_i \times t \) known design matrix for subject \( i \), \( \beta_i \) is a \( t \times 1 \) parameter vector for subject \( i \), \( A_i \) is a \( q \times t \) known design matrix, and \( B_i \) is a \( t \times v \) known matrix which is usually an incidence matrix of 0’s and 1’s formed in such a manner as to enable one to select which of the parameters are random [14]. Expressed in vector notation, the two-stage GLMM [14] is given by

\[ E(y_i \mid b_i) = g^{-1}(X_i'\beta_i), \quad \text{Var}(y_i \mid b_i) = \phi H_i(\mu_i), \]
\[ \beta_i = A_i \beta + B_i b_i, \]
\[ b_i \sim i.i.d. D_i(0, \Sigma), \]
\[ i = 1, \ldots, n. \]

where \( g^{-1}(X_i'\beta_i) \) is the \( n_i \times 1 \) vector of means \( \mu_i \) expressed in terms of the inverse of the link function \( g \); \( \phi \) is the dispersion parameter (for Poisson distribution, \( \phi = 1 \)); \( H_i(\mu_i) \) is the \( n_i \times n_i \) diagonal variance matrix with the \( j \)th diagonal element equal to \( h(\mu_j) \), where \( h(\mu_j) \) is a variance function, and the variance of \( y_i \), conditional on \( b_i \), equal to \( \phi h(\mu_j) \).

2.2 GLMM for Multi-copy Repairable Systems

Suppose a repairable system has \( n \) copies. Let \( t_{ij} \) denote the failure time of \( j \)th failure of the \( i \)th copy, and \( y_{ij} \) be the cumulative number of failure in \((0, t_{ij}] \) \( i = 1, \ldots, n \); \( j = 1, \ldots, n_i \), where \( n_i \) is the total number of failures of copy \( i \). If the failure process of each copy is assumed to be a PLP (1.1), then \( y_{ij} \) is Poisson-distributed with mean \( \mu_{ij} \):

\[ P_i(y_{ij} \mid \mu_{ij}) = \frac{\mu_{ij}^{y_{ij}} e^{-\mu_{ij}}}{y_{ij}!}, \]

where \( \mu_{ij} = E(y_{ij}) = \left(\frac{t_{ij}}{\theta_i}\right)^{\gamma_i}, \gamma_i \) and \( \theta_i \) are the shape and scale parameters for copy \( i \), respectively. The variance of \( y_{ij} \) is also \( \left(\frac{t_{ij}}{\theta_i}\right)^{\gamma_i} \).

In the GLMM context, the Poisson regression model utilizes the logarithm link, namely

\[ g(\mu_{ij}) = \log(\mu_{ij}) = \gamma_i \left[ \log(t_{ij}) - \log(\theta_i) \right] = \eta_{ij}. \]

If we re-parameterize the formulation (2.5) and let
\( \beta_1 = \gamma_i, \quad \beta_{12} = -\gamma_i \log(\theta_i), \quad x_{ij} = \log(t_{ij}), \) \hfill (2.6)

then (2.4) has the form as
\( \eta_{ij} = x_{ij} \beta_1 + \beta_{12} . \) \hfill (2.7)

where \( \beta_1 \) and \( \beta_{12} \) are, respectively, the slope and intercept parameters for copy \( i \).

Let \( \beta_i = [\beta_{i1} \quad \beta_{i2}]', \quad x_{ij} = [x_{ij} \quad 1]' \), then we have
\( \eta_{ij} = x_{ij}' \beta_i . \) \hfill (2.8)

Here we consider that both \( \beta_i \) and \( \beta_{12} \) vary from copy to copy and they have means \( \beta_1 \) and \( \beta_{2} \), respectively. Let
\( \beta_1 = \beta_1 + b_1, \quad \beta_{12} = \beta_{2} + b_{12}, \) \hfill (2.9)

where \( \beta = [\beta_1 \quad \beta_{2}]' \) is a vector of population parameters, the fixed effects of this model; \( b_i = [b_{1i} \quad b_{12}]' \) is the independent vector of random effects for copy \( i \), and it is assumed to follow a two-dimension normal distribution with mean \( 0 \) and variance-covariance matrix \( \Sigma \), i.e., \( b_i \sim i.i.d. N_2(0, \Sigma) \). The variance-covariance matrix has the form
\[
\Sigma = \begin{bmatrix}
\sigma_{11}^2 & \sigma_{12} \\
\sigma_{12} & \sigma_{22}^2
\end{bmatrix},
\]
where the covariance \( \sigma_{12} \) represents the correlation of the random effects.

Upon substituting (2.9) into (2.8), the generalized linear mixed model for multi-copy repairable systems under the assumption of PLP is obtained as:
\( \eta_{ij} = x_{ij}' \beta + z_{ij}' b_i, \quad b_i \sim i.i.d. N_2(0, \Sigma), \quad i = 1,...,n; \quad j = 1,...,ni, \) \hfill (2.10)

where \( z_{ij}' = [x_{ij} \quad 1] \).

Let \( \eta_i = [\eta_{i1} \ldots \eta_{ini}]' \), the model (2.10) can be expressed in vector notation as
\[
\eta_i = g(\mu_i) = X_i \beta + Z_i b_i, \quad b_i \sim i.i.d. N_2(0, \Sigma), \quad i = 1,...,n, \) \hfill (2.11)

where \( X_i = [x_{i1} \ldots x_{ini}]' \), \( Z_i = [z_{i1} \ldots z_{ini}]' \).

Furthermore, let \( y_i = [y_{i1} \ldots y_{ini}]' \), the model (2.11) can be re-written in terms of the form of (2.3) as
\[
E(y_i | b_i) = \exp(X_i' \beta_i), \quad Var(y_i | b_i) = \exp(X_i' \beta_i), \quad \beta_i = A \beta + B b_i, \quad b_i \sim i.i.d. N_2(0, \Sigma), \) \hfill (2.12)
by setting
\[
X_i^* = \begin{bmatrix}
    x_{i1}^* & \cdots & x_{in_i}^*
  \end{bmatrix}' = \begin{bmatrix}
    x_{i1} & 1 \\
    \vdots & \vdots \\
    x_{in_i} & 1
  \end{bmatrix},
A_i = \begin{bmatrix}
    1 & 0 \\
    0 & 1
  \end{bmatrix},
B_i = \begin{bmatrix}
    1 & 0
  \end{bmatrix}.
\]

3. Estimation of Parameters

For the model (2.11), the full parameter space is
\[
\Theta = (\Theta_1, \Theta_2) = (\beta_1, \beta_2, \sigma^2_{11}, \sigma^2_{12}, \sigma^2_{22}, b_{1,1}, b_{1,2}, \ldots, b_{n,1}, b_{n,2}),
\]
where the fixed parameter is \( \Theta_1 = (\beta, \Sigma) = (\beta_1, \beta_2, \sigma^2_{11}, \sigma^2_{12}, \sigma^2_{22}) \), and the random parameter is \( \Theta_2 = (b_{1,1}, b_{1,2}, \ldots, b_{n,1}, b_{n,2}) \).

In general, the fixed parameter estimation typically involves maximum likelihood (ML) or variants of ML [20], while the random parameter estimation utilizes empirical Bayes methods [17]. Several software packages (i.e., SAS proc NLMIXED, STATA) and independent programs (HLM, EGRET, LIMDEP, GLLAMM, etc.) are available for fitting GLMM and implementing the estimation procedure. Since not all of these programs fit all of the GLMM, and different programs may have different integration methods over the random effects, an appropriate statistical package should be chosen for a particular problem. Here we choose the SAS Proc NLMIXED to fit the model (2.11) and to implement the likelihood-based estimation procedure [21].

A brief introduction to the parameter estimation in SAS Proc NLMIXED is as follows.

3.1 Estimation of the Fixed Parameters

For the model (2.11), although the response variables \( y_{i1}, \ldots, y_{in_i} \) are correlated, the probability of \( y_{ij} \) conditional on \( b_i \) are assumed to be independent of each other, \( j = 1, \ldots, n_i \). This is known as the conditional independence assumption [22, 23]. Then the probability of the response variable vector \( y_i \), conditional on \( b_i \), is expressed as
\[
f(y_i \mid b_i) = \prod_{j=1}^{n_i} P_i(y_{ij} \mid b_i) = \prod_{j=1}^{n_i} \frac{\mu_{ij}^{y_{ij}} \exp(-\mu_{ij})}{y_{ij}!}
\]
where the random effects \( b_i \) enters through the mean \( \mu_{ij} \).

The marginal density of \( y_i \) is
\[
L(y_i \mid \Theta_1) = \int_{b_i} f(y_i \mid b_i) f(b_i) db_i
\]
where \( f(b_i) \) is the distribution of the random effects \( b_i \). In our case, \( f(b_i) \) is a two-dimension normal density function.

The marginal log-likelihood from the sample of \( n \) copies is then obtained as
\[
\]
\[
I(\Theta_i) = \sum_{i=1}^{n} \log L(y_i | \Theta_i) \\
= \sum_{i=1}^{n} \log \int_{b_i} f(y_i | b_i) f(b_i) db_i 
\]  

(3.3)

Maximizing this log-likelihood will yield the MLE of \( \Theta_i \).

In general, there is no closed-form solution for integrating the random effects \( b_i \). Numerical integration techniques are then needed. Different approaches can be used to approximate the integral over the random effects. In the SAS Proc NLMIXED, there are two principal methods, and the default is adaptive Gaussian quadrature described in [24]. Another approximation method is the first-order method in [25, 26], which is used only in the case where the response variable \( y_{ij} \) is normally distributed.

A variety of optimization techniques can be used to carry out the maximization. In the SAS Proc NLMIXED, there are several optimization techniques available for choosing. The second-derivative methods are best for small problems where the Hessian matrix is not expensive to compute; the first-derivative methods are best for medium-sized problems where the objective function and the gradient are much faster to evaluate than the Hessian; the no-derivative method is best for small problems where derivatives are not continuous or are very difficult to compute. The default is the dual-quasi Newton algorithm [21].

The MLE of \( \Theta_i \) along with their approximate standard errors are obtained when the optimization problem converges. The SE of \( \hat{\Theta_i} \) are computed from the second derivative matrix of the likelihood function [21]. In general, the observed Fisher information \( I(\hat{\Theta}_i) \) provides approximately the standard error of \( \hat{\Theta}_i \) [20], where

\[
I(\Theta_i) = -\frac{\partial^2 l(\Theta_i)}{\partial \Theta_i^2},
\]

the Hessian matrix.

Confidence intervals are a useful supplement to the MLE. In the SAS Proc NLMIXED, Wald-type intervals for parameters are given when the optimization method converges. This type of CI is computed by using the Wald-type statistic [20, 27].

Base on the estimates of \( \beta \), denoted as \( \hat{\beta} \), the point estimates of the shape and scale parameters of the population, \( \hat{\gamma} \) and \( \hat{\theta} \), respectively, can be obtained by

\[
\hat{\gamma} = \hat{\beta}_1, \quad \hat{\theta} = \exp(-\hat{\beta}_2 / \hat{\beta}_1).
\]

Approximate standard errors for \( \hat{\beta} \) are computed using the delta method [28]. Wald-type confidence intervals are also computed.

### 3.2 Estimation of Random Effects

Estimation of the random effects is also known as prediction of the random effects. The random effects are usually estimated by using empirical Bayes methods. For the multivariate case, the vector of the empirical Bayes estimate, \( \hat{b}_i \), is given by [17]:

\[
\hat{b}_i = \hat{\beta}_i + \hat{\gamma} \hat{\epsilon}_i
\]
This is the posterior mean of $\mathbf{b}_i$. The standard error vector of $\mathbf{\hat{b}}_i$ is computed by using the delta method [21, 28, 29]. Similarly, Wald-type confidence intervals of the random effects are obtained by using Wald-type statistics [21].

3.3 Estimation of $\mathbf{\beta}_i$, $\mathbf{\gamma}_i$ and $\mathbf{\theta}_i$

For the GLMM-based model (2.11), $\mathbf{\beta}_i$, $\mathbf{\gamma}_i$ and $\mathbf{\theta}_i$ are all treated as random variables. Since we have $\mathbf{\beta}_i = \mathbf{A}_i \mathbf{\beta} + \mathbf{B}_i \mathbf{b}_i$, $\mathbf{\gamma}_i = \mathbf{\beta}_i$, $\mathbf{\theta}_i = \exp(-\mathbf{\beta}_{i2}/\mathbf{\beta}_{i1})$, their estimates are obtained by using empirical Bayes methods. The estimation procedures are similar to that of $\mathbf{b}_i$.

4. Goodness of Fit Test for the GLMM-based Model

Since the GLMM-based model (2.11) is constructed on the basis of PLP, many of the standard goodness-of-fit statistics used in PLP can be adapted to our present setting. Among the GOF tests utilized to test the hypothesis that failure-time data of a repairable system came from a PLP [1, 30-37], the Anderson-Darling test is shown to be one of the most powerful goodness-of-fit tests for PLP [34]. In this paper, a modified Anderson-Darling test is applied to the goodness-of-fit test of the model (2.11).

Let $0 < t_{i1} < \ldots < t_{in_i}$ denote the successive failure times of the $i$th copy, and $n_i$ is the total number of failures of copy $i$, $i = 1, \ldots, n$. For failure-truncated data, the modified Anderson-Darling statistic [30] for copy $i$ is

$$D_i^2 = -\frac{1}{M_i} \left\{ \sum_{j=1}^{M_i} (2j - 1) [\ln \hat{U}_{ij} + \ln (1 - \hat{U}_{M_i+1-j})] \right\} - M_i$$

(4.1)

where $M_i = n_i - 1$; $\hat{U}_{ij} = \left[ \frac{t_{ij}}{t_{jn_i}} \right]^{\hat{\gamma}_i}$, $\hat{\gamma}_i$ is the point estimate of $\gamma_i$.

We define the overall Anderson-Darling statistic as the weighted average

$$D^2 = \sum_{i=1}^{n} w_i D_i^2,$$

(4.2)

where $\sum_{i=1}^{n} w_i = 1$. In particular, we define $w_i = n_i \left/ \sum_{i=1}^{n} n_i \right.$.

To compare the computed value $D^2$ to critical values of the Andersen-Darling statistic (Table 3 in [30]) at an $s$-significance level $\alpha$, we need to specify the average failure number of the system. Here, we define the overall mean failure number as

$$\bar{N} = \left[ \frac{1}{n} \sum_{i=1}^{n} n_i \right]$$

(4.3)
where the symbol \( [x] \) means the smallest integer not smaller than \( x \).

If the computed \( D^2 \) value is less than the corresponding critical value of the Andersen-Darling test, the model (2.11) cannot be rejected at the specified \( s \)-significance level. Thus, the data set could be modeled by the GLMM-based model (2.11).

5. Numerical Application

In this section we use the model (2.11) to analyze the real failure data from 8 steam-turbine generating units. Steam-turbine generating units are typically complex, repairable systems, each of which has a generator, a steam turbine and a boiler. In a fossil-fueled steam turbine, the fuel is burned in a boiler to produce steam. The resulting steam then turns the turbine blades that turn the shaft of the generator to produce electricity. Fossil-fueled steam-turbine generating units range in size from 1 megawatt (MW) to more than 1,000 megawatts. In our application, the data come from eight 300MW steam-turbine generating units operated in several electric utility plants. Since generators, turbines and boilers of these units were manufactured by the same companies respectively, it seems to be reasonable to treat these units as multiple copies of a system. The main purpose of such application is to show the effectiveness of the model (2.11) in estimating the change trend in reliability of both the population and each copy based on the data.

For an operated generating unit, a failure is in general an unexpected outage (UO) or an in-service unplanned derated (IUD). Usually, a UO event is regarded as “one failure”, while an IUD is not “one failure” since the generating unit still produces electricity when an IUD event occurs. The method of handling IUD’s is discussed in [38].

5.1 Failure Data from 8 Generating Units

We collected failure data from eight 300MW generating units over the period from 1988 to 1994. Since the starting times (in calendar times) are different, the total operating hours and numbers of failures vary among the copies. Fig. 1 shows the cumulative numbers of failures for 8 copies of 300MW generating unit versus operating hours.

![Fig.1 Cumulative numbers of failures for 8 copies of 300MW generating unit versus operating hours](image-url)
5.2 Estimates of Parameters

Under the assumption of the model (2.11), the estimates of parameters can be obtained by using the SAS Proc NL MIXED. The ML estimates, standard errors and 95% confidence intervals of the fixed parameters are presented in Table 5.1, while the estimates of the random effects are listed in Table 5.2.

**TABLE 5.1**
RESULTING MLE, SE, 95% CI OF THE FIXED PARAMETERS

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>MLE</th>
<th>SE</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_1 )</td>
<td>0.7298</td>
<td>0.04115</td>
<td>[0.6291, 0.8305]</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>-3.2683</td>
<td>0.4449</td>
<td>[-4.3568, -2.1797]</td>
</tr>
<tr>
<td>( \sigma_{21} )</td>
<td>0.01193</td>
<td>0.00463</td>
<td>[0.000599, 0.02326]</td>
</tr>
<tr>
<td>( \sigma_{12} )</td>
<td>-0.1293</td>
<td>0.04723</td>
<td>[-0.2449, -0.0137]</td>
</tr>
<tr>
<td>( \sigma_2 )</td>
<td>1.4538</td>
<td>0.5859</td>
<td>[0.0196, 2.888]</td>
</tr>
</tbody>
</table>

**TABLE 5.2**
RESULTING ESTIMATE, SE, 95% CI OF THE RANDOM EFFECTS

<table>
<thead>
<tr>
<th>Copy ( i )</th>
<th>Random Effects ( h_{1i} ) Estimates</th>
<th>SE</th>
<th>95% CI</th>
<th>Random Effects ( h_{2i} ) Estimates</th>
<th>SE</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1201</td>
<td>0.04774</td>
<td>[-0.00325, 0.2369]</td>
<td>-1.4467</td>
<td>0.5095</td>
<td>[-2.6934, -0.2000]</td>
</tr>
<tr>
<td>2</td>
<td>-0.2123</td>
<td>0.04366</td>
<td>[-0.3191, -0.1054]</td>
<td>2.2653</td>
<td>0.4664</td>
<td>[1.1241, 3.4065]</td>
</tr>
<tr>
<td>3</td>
<td>-0.0899</td>
<td>0.04646</td>
<td>[-0.2036, 0.0238]</td>
<td>0.8309</td>
<td>0.4912</td>
<td>[-0.3711, 2.0328]</td>
</tr>
<tr>
<td>4</td>
<td>0.1165</td>
<td>0.05324</td>
<td>[-0.0138, 0.2468]</td>
<td>-1.3801</td>
<td>0.5523</td>
<td>[-2.7315, -0.0286]</td>
</tr>
<tr>
<td>5</td>
<td>0.0424</td>
<td>0.05459</td>
<td>[-0.0912, 0.1760]</td>
<td>-0.7932</td>
<td>0.5634</td>
<td>[-2.1718, 0.5854]</td>
</tr>
<tr>
<td>6</td>
<td>0.0180</td>
<td>0.05118</td>
<td>[-0.1072, 0.1433]</td>
<td>0.0882</td>
<td>0.5247</td>
<td>[-1.1957, 1.3721]</td>
</tr>
<tr>
<td>7</td>
<td>0.0346</td>
<td>0.05084</td>
<td>[-0.0898, 0.1590]</td>
<td>-0.1038</td>
<td>0.5210</td>
<td>[-1.3786, 1.1710]</td>
</tr>
<tr>
<td>8</td>
<td>-0.0339</td>
<td>0.06397</td>
<td>[-0.1904, 0.1226]</td>
<td>0.5860</td>
<td>0.5841</td>
<td>[-0.8433, 2.0153]</td>
</tr>
</tbody>
</table>

Based on the results of Table 5.1 and 5.2, the estimates of the slope and intercept parameters for 8 copies are listed in Table 5.3, and the estimates of the shape and scale parameters for both the population and 8 copies are summarized in Table 5.4.

**TABLE 5.3**
RESULTING ESTIMATE, SE, 95% CI OF THE SLOPE AND INTERCEPT FOR 8 COPIES

<table>
<thead>
<tr>
<th>Copy ( i )</th>
<th>( \beta_{1i} ) Estimates</th>
<th>SE</th>
<th>95% CI</th>
<th>( \beta_{2i} ) Estimates</th>
<th>SE</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.8499</td>
<td>0.0287</td>
<td>[0.7798, 0.9200]</td>
<td>-4.7150</td>
<td>0.2932</td>
<td>[-5.4325, -3.9975]</td>
</tr>
<tr>
<td>2</td>
<td>0.5176</td>
<td>0.0168</td>
<td>[0.4764, 0.5587]</td>
<td>-1.0030</td>
<td>0.1610</td>
<td>[-1.3970, -0.6090]</td>
</tr>
<tr>
<td>3</td>
<td>0.6399</td>
<td>0.0250</td>
<td>[0.5788, 0.7010]</td>
<td>-2.4374</td>
<td>0.2411</td>
<td>[-3.0274, -1.8474]</td>
</tr>
<tr>
<td>4</td>
<td>0.8463</td>
<td>0.0399</td>
<td>[0.7485, 0.9441]</td>
<td>-4.6483</td>
<td>0.3866</td>
<td>[-5.5942, -3.7024]</td>
</tr>
<tr>
<td>5</td>
<td>0.7722</td>
<td>0.0410</td>
<td>[0.6719, 0.8725]</td>
<td>-4.0615</td>
<td>0.3967</td>
<td>[-5.0322, -3.0907]</td>
</tr>
<tr>
<td>6</td>
<td>0.7478</td>
<td>0.0373</td>
<td>[0.6566, 0.8391]</td>
<td>-3.1801</td>
<td>0.3376</td>
<td>[-4.0061, -2.3540]</td>
</tr>
<tr>
<td>7</td>
<td>0.7644</td>
<td>0.0366</td>
<td>[0.6748, 0.8541]</td>
<td>-3.3721</td>
<td>0.3297</td>
<td>[-4.1787, -2.5654]</td>
</tr>
<tr>
<td>8</td>
<td>0.6959</td>
<td>0.0583</td>
<td>[0.5532, 0.8387]</td>
<td>-2.6823</td>
<td>0.4523</td>
<td>[-3.7891, -1.5755]</td>
</tr>
</tbody>
</table>
TABLE 5.4
RESULTING ESTIMATE, SE, 95% CI OF THE SHAPE PARAMETERS AND SCALE PARAMETERS FOR BOTH THE POPULATION AND 8 COPIES

<table>
<thead>
<tr>
<th>Copy</th>
<th>Shape Parameters $\gamma_i$</th>
<th>Scale Parameters $\theta_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimates</td>
<td>SE</td>
</tr>
<tr>
<td>1</td>
<td>0.8499</td>
<td>0.0287</td>
</tr>
<tr>
<td>2</td>
<td>0.5176</td>
<td>0.0168</td>
</tr>
<tr>
<td>3</td>
<td>0.6399</td>
<td>0.0250</td>
</tr>
<tr>
<td>4</td>
<td>0.8463</td>
<td>0.0399</td>
</tr>
<tr>
<td>5</td>
<td>0.7722</td>
<td>0.0410</td>
</tr>
<tr>
<td>6</td>
<td>0.7478</td>
<td>0.0373</td>
</tr>
<tr>
<td>7</td>
<td>0.7644</td>
<td>0.0366</td>
</tr>
<tr>
<td>8</td>
<td>0.6959</td>
<td>0.0583</td>
</tr>
</tbody>
</table>

The Population

<table>
<thead>
<tr>
<th>Estimates</th>
<th>SE</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7298</td>
<td>0.0412</td>
<td>[0.6291, 0.8305]</td>
</tr>
<tr>
<td>88.0782</td>
<td>32.1657</td>
<td>[9.3715, 166.78]</td>
</tr>
</tbody>
</table>

From Table 5.1, the value of $\hat{\sigma}_{12}$ shows that $b_{11}$ and $b_{12}$ are correlated. This means that the shape and scale parameters for copy $i$ are correlated. From Table 5.4, we can see that the estimates of the shape and scale parameters vary from copy to copy. The same results are obtained in Section 6 when using PLP to 8 data sets separately. In this situation, if we ignore the copy-to-copy variance and pool data from these copies to make inferences about the population, the statistical results may be not reasonable.

From Table 5.4, we also see that the estimate of the shape parameter of the population, $\hat{\gamma}$, is less than 1, indicating that this type of generating unit improves over operating time. For Copy 1 and 4, the estimates of the shape parameters are close to 1, indicating that the intensity function, $\lambda(t)$, of these two copies are nearly constant (reduced to an HPP model); while the estimate of the shape parameter for Copy 2 is very less than 1, indicating that there has been a discernible trend over operating time for this copy. For other five copies, estimates of the shape parameters are greater than 0.6 and less than 0.8, indicating that significant improvements in reliability over time exist, but smaller than copy 2. These statistical results are consistent with the plots in Fig.1.

For a future copy of this system, a point prediction for the future number of failures in an interval $[t_a, t_b]$ is

$$\int_{t_a}^{t_b} \left(\frac{t}{88.0782}\right)^{0.7298} dt = \left(\frac{1}{88.0782}\right)^{0.7298} \left( t_b^{0.7298} - t_a^{0.7298} \right)$$

5.3 Goodness-of-fit Test

Table 5.5 summarizes the calculated values of the Anderson-Darling statistic on the basis of Table 5.4. According to Table 3 in [30], the critical value of the Anderson-Darling test at 5% $\alpha$-significance level is 1.33, therefore the hypothesis that the model (2.11) can be used to estimate the change trend in reliability of both the population and all copies cannot be rejected.
TABLE 5.5
CALCULATED VALUES OF The ANDERSON-DARLING STATISTIC UNDER THE MODEL
(2.11)

<table>
<thead>
<tr>
<th>Copy</th>
<th>( n_i )</th>
<th>( D_i^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>87</td>
<td>0.955</td>
</tr>
<tr>
<td>2</td>
<td>83</td>
<td>0.933</td>
</tr>
<tr>
<td>3</td>
<td>67</td>
<td>0.800</td>
</tr>
<tr>
<td>4</td>
<td>55</td>
<td>1.386</td>
</tr>
<tr>
<td>5</td>
<td>48</td>
<td>1.224</td>
</tr>
<tr>
<td>6</td>
<td>70</td>
<td>2.082</td>
</tr>
<tr>
<td>7</td>
<td>54</td>
<td>0.963</td>
</tr>
<tr>
<td>8</td>
<td>22</td>
<td>0.916</td>
</tr>
</tbody>
</table>

The weighted average

\[ N = 61 \]
\[ D^2 = 1.157 \]

6. Comparison with Other Methods

To study the effectiveness of the model (2.11), comparison is made with other commonly used methods. The PLP model (1.1) is often used to analyze the data from a single system, here we fit the 8 failure data sets to PLP separately.

Let \( 0 < t_1 < ... < t_{n_i} \) denote the successive failure times of a generating unit. For the PLP model (1.1), the single-system likelihood for exact failure times is

\[
L(\gamma, \theta) = \left( \frac{\gamma}{\theta} \right)^{n_i} \left( \prod_{j=1}^{n_i} t_j^{\gamma-1} \right) \exp \left\{ -\left( \frac{t_{n_i}}{\theta} \right)^{\gamma} \right\}
\]

The ML estimates, \( \hat{\gamma} \) and \( \hat{\theta} \), are found to be:

\[
\hat{\gamma} = \frac{n_i}{\sum_{j=1}^{n_i-1} \ln(t_{n_i}/t_j)} , \quad \hat{\theta} = t_{n_i}/n_i^{1/\hat{\gamma}}.
\]

Standard errors of \( \hat{\gamma} \) and \( \hat{\theta} \) are computed by using the delta method [29].

An asymptotic 100(1 - \( \alpha \))% pointwise confidence interval for \( \gamma \) is given by

\[
[\hat{\gamma} - z_{\alpha/2} \text{se}(\hat{\gamma}), \hat{\gamma} + z_{\alpha/2} \text{se}(\hat{\gamma})],
\]

where \( z_{\alpha/2} \) is the \( (1 - \alpha / 2) \) percentile of the standard normal distribution.

For \( \theta \), the lower bound of the Wald-type confidence interval may be negative, so we use an asymptotic 100(1 - \( \alpha \))% confidence interval with always positive bounds ([39])
\[
\hat{\theta} \exp \left( - z_{\alpha/2} \frac{\ln n}{\sqrt{n}} \right) \tilde{\theta} \exp \left( z_{\alpha/2} \frac{\ln n}{\sqrt{n}} \right)
\]

The most part of the method described in this section can be implemented by using S-plus Life Data Analysis (SPLIDA). The results are summarized in Table 6.1. Fig. 2 presents the plots of nonparametric MCF (the solid lines) and fitted PLP MCF (the dotted lines). Fig. 2 shows that the PLP provides a good fit to the 8 failure data sets up.

**TABLE 6.1**

RESULTS OF FITTING THE DATA FROM 8 COPIES TO THE PLP MODEL (1.1) SEPARATELY

<table>
<thead>
<tr>
<th>Copy</th>
<th>( \gamma )</th>
<th>SE</th>
<th>95% CI</th>
<th>( \theta )</th>
<th>SE</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.9123</td>
<td>0.09837</td>
<td>[0.7195, 1.105]</td>
<td>354.399</td>
<td>191.242</td>
<td>[126.694, 991.358]</td>
</tr>
<tr>
<td>2</td>
<td>0.5365</td>
<td>0.05924</td>
<td>[0.4204, 0.6526]</td>
<td>10.278</td>
<td>9.560</td>
<td>[1.747, 60.459]</td>
</tr>
<tr>
<td>3</td>
<td>0.6419</td>
<td>0.07901</td>
<td>[0.487, 0.7968]</td>
<td>51.892</td>
<td>42.862</td>
<td>[10.812, 249.054]</td>
</tr>
<tr>
<td>4</td>
<td>0.8211</td>
<td>0.1117</td>
<td>[0.6021, 1.04]</td>
<td>240.009</td>
<td>163.573</td>
<td>[66.078, 871.758]</td>
</tr>
<tr>
<td>5</td>
<td>0.7392</td>
<td>0.1078</td>
<td>[0.5279, 0.9505]</td>
<td>182.508</td>
<td>143.259</td>
<td>[41.481, 802.999]</td>
</tr>
<tr>
<td>6</td>
<td>0.7657</td>
<td>0.09286</td>
<td>[0.5837, 0.9477]</td>
<td>90.612</td>
<td>62.227</td>
<td>[24.699, 332.423]</td>
</tr>
<tr>
<td>7</td>
<td>0.8112</td>
<td>0.1114</td>
<td>[0.5928, 1.03]</td>
<td>119.515</td>
<td>82.854</td>
<td>[32.197, 443.642]</td>
</tr>
<tr>
<td>8</td>
<td>0.7534</td>
<td>0.1644</td>
<td>[0.4312, 1.076]</td>
<td>75.964</td>
<td>70.507</td>
<td>[13.678, 421.874]</td>
</tr>
</tbody>
</table>

Fig. 2 Cumulative failure numbers for 8 generating units versus operating hours with fitted PLP models

Comparing Table 5.4 and Table 6.1 shows that the point estimates of \( \gamma_i \) and \( \theta_i \) under the model (2.11) are similar to those under the model (1.1), while the 95% confidence intervals are narrower under the model (2.11).

The modified Anderson-Darling statistic described in [29] is applied to these data to implement the goodness-of-fit test. The computed values of all copies along with their corresponding critical values at the 5% significance level are listed in Table 6.2. Since the computed values are less than their corresponding critical values, the hypothesis that the 8 failure data sets are separately from PLP cannot be rejected. All data sets can be
modeled by PLP. This also helps us safeguard the GLMM-based model (2.11) against the baseline model misspecification.

<table>
<thead>
<tr>
<th>Copy (i)</th>
<th>(D^2)</th>
<th>Critical Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.760</td>
<td>1.32</td>
</tr>
<tr>
<td>2</td>
<td>1.110</td>
<td>1.32</td>
</tr>
<tr>
<td>3</td>
<td>0.816</td>
<td>1.31</td>
</tr>
<tr>
<td>4</td>
<td>1.223</td>
<td>1.33</td>
</tr>
<tr>
<td>5</td>
<td>0.993</td>
<td>1.33</td>
</tr>
<tr>
<td>6</td>
<td>1.061</td>
<td>1.33</td>
</tr>
<tr>
<td>7</td>
<td>1.179</td>
<td>1.33</td>
</tr>
<tr>
<td>8</td>
<td>1.086</td>
<td>1.33</td>
</tr>
</tbody>
</table>

7. Conclusion

In this paper, we propose a GLMM-based model to analyze the failure data from multi-copy repairable systems. The main purpose of this study is to introduce a new method to the analysis of repairable systems reliability. This method can make inferences about both the population and each system copy when accounting for copy-to-copy variance. The modified Anderson-Darling test is adapted to the goodness-of-fit test of the model.

More analysis is need in this area. For example, during each stage of a TAAF program, one or more copies of the developed system are put on test in order to estimate the system reliability. In many cases there are only few failures on any one copy during the limited observation intervals. In this case, the GLMM-based model (2.11) may be needed adapting to the sparse data. In addition, when the individual point processes are not Poisson, or the intensity function is not the Power law (e.g. the NHPP with log-linear intensity function), the analysis may have different results. Moreover, we have assumed a multivariate normal distribution for random effects in this paper. If we relax this assumption, for instance to extend the normality assumption to incorporate mixtures of normal components, the model may be more general for applications.

ACKNOWLEDGEMENT

The authors would like to thank the anonymous referees for their helpful comments and suggestions that have led to a substantial improvement of the paper.
REFERENCES


Authors

**Furong Tan** is associate professor in Dept. of Industrial Engineering and Management, School of Mechanical Engineering at the Shanghai Jiao Tong University and visiting scholar in Dept. of Electrical and Computer Engineering at the University of Tennessee in 2005. She received her M.S. and B.S. degrees in Electrical Engineering from the Shanghai Jiao Tong University. She mainly focuses on the reliability analysis of repairable systems.

**Dr. Zhibin Jiang** is professor and head of Department of Industrial Engineering & Management, at the Shanghai Jiao Tong University. He received his Ph.D in Manufacturing Engineering & Engineering Management from City University of Hong Kong in 1999. He has been active in research in the area of advanced manufacturing system modeling, simulation and control. He is a senior member of IEEE and IIE.

**Dr. Suk Joo Bae** is assistant professor in Department of Industrial Engineering, Hanyang University, Seoul, Korea. He received his B.S. and M.S. degrees in Industrial Engineering at the Hanyang University, Seoul, in 1994 and 1996 respectively, and his Ph.D. in the School of Industrial and Systems Engineering at the Georgia Institute of Technology in 2003. He worked as a reliability engineer at the SDI, Korea from 1996 to 1999 and post-doctoral research associate in the Department of Electrical and Computer Engineering at the University of Tennessee from 2004 to 2005. His research interests are centered on reliability evaluation of light displays and nano-devices via accelerated life and degradation testing, statistical robust parameter design, and process control for large-volume on-line processing data. He is a member of INFORMS, ASA, and IMS.