Editor's Note: This article and accompanying discussion were presented orally at the *Technometrics* Invited Paper Session at the Sixth Annual Spring Research Conference, held in Minneapolis, Minnesota, June 2–4, 1999. This conference is cosponsored by the American Statistical Association and the Institute of Mathematical Statistics.

Estimating Fatigue Curves With the Random Fatigue-Limit Model

Francis G. PASCUAL

Department of Mathematics and
Program in Statistics
Washington State University
Pullman, WA 99164-3113
(*jpascual@math.wsu.edu*)

William Q. MEEKER

Department of Statistics and
Center for Nondestructive Evaluation
Iowa State University
Ames, IA 50011
(*wqmeeker@iastate.edu*)

In a fatigue-limit model, units tested below the fatigue limit (also known as the threshold stress) theoretically will never fail. This article uses a random fatigue-limit model to describe (a) the dependence of fatigue life on the stress level, (b) the variation in fatigue life, and (c) the unit-to-unit variation in the fatigue limit. We fit the model to actual fatigue datasets by maximum likelihood methods and study the fits under different distributional assumptions. Small quantiles of the life distribution are often of interest to designers. Lower confidence bounds based on likelihood ratio methods are obtained for such quantiles. To assess the fits of the model, we construct diagnostic plots and perform goodness-of-fit tests and residual analyses.

KEY WORDS: Akaike information criterion; Fatigue data; Maximum likelihood methods; Probability (P-P) plots; Random fatigue limit; Right censoring.

1. INTRODUCTION

1.1 Background

The relationship between fatigue life of metal, ceramic, and composite materials and applied stress is an important input to design-for-reliability processes. This article suggests a practical model to describe the relationship between fatigue life and applied stress and provides and illustrates corresponding data-analysis methods. This work is motivated by the need to develop and present quantitative fatigue-life information used in the design of jet engines.

Fatigue data are often presented in the form of a median S-N curve, a log-log plot of cyclic stress or strain s versus the median fatigue life N, which is expressed in cycles to failure. An extension of this concept is the p-quantile S-N curves, also called S-N-P curves, a generalization that relates the p quantile of fatigue life to the applied stress or strain. Thus, each curve represents a constant probability of failure p, as a function of s. We shall use the .05 and .95-quantile S-N curves to illustrate the variability of fatigue life. Unless otherwise specified, the S-N curve in the literature generally refers to the median curve. We shall use the S-N curve as such.

Fatigue data on ferrous and titanium alloys indicate that experimental units tested below a particular stress level are unlikely to fail. This limiting stress level is called the "fatigue limit" or "endurance limit." The S-N curve for these materials exhibits a strong curvature and an asymptotic behavior near the fatigue limit. Most nonferrous metals such as aluminum, copper, and magnesium appear not to have a fatigue limit. S-N curves for these materials gradually drop off and never become horizontal. Failure will occur eventually if units are tested or are in service long enough.

In the literature, there is no clear agreement on the meaning of the terms "fatigue limit" and "endurance limit." For nonferrous materials, it is common practice to define the "fatigue strength" to be the stress level below which failure will not occur before an arbitrary large number of cycles (e.g., 10^7 or 10^8 cycles). Collins (1993) and Dieter (1976) defined fatigue strength as such and used the term "fatigue limit" to imply infinite life. On the other hand, to represent fatigue strength at a prescribed long but finite life, Nelson (1990) and Colangelo and Heiser (1974) used the term "endurance limit," whereas others use the term "fatigue limit." In this article, we will use the term "fatigue limit" to mean the stress below which an infinite or a specified large number of cycles can be sustained, whichever case is appropriate.

The existence of infinite-life fatigue limits is still a subject of debate. Proponents of fatigue limits argue that the fatigue limit can be interpreted as the minimum force or stress required to cause crack propagation. The randomness in the fatigue limit is due in part to the location, orientation, size, and number of cracks, which are random themselves. The range of stress that will not result in crack propagation is important to the classic safe-life design approach that is based on high-cycle fatigue and infinite life. Information on the fatigue limit can be used to design components that are intended to last indefinitely.

On the other hand, others suggest that all units under cyclic stress will eventually fail if they are cycled long enough. Because it is not feasible to run fatigue tests indefinitely, it is difficult, if not impossible, to obtain evidence to support the existence (or nonexistence) of fatigue limits.

Nonetheless even when fatigue limits do not really exist, the random fatigue-limit model presented here provides a useful empirical model to describe the life-stress curvature and nonconstant variability that are typically observed in fatigue data. Moreover, for most materials, empirical S-N curves are generally relatively flat in the high-cycle or long-life range. In these cases, as with other applications of empirical modeling, it is, of course, dangerous to extrapolate outside of the range of the data.

1.2 Related Work

Hirose (1993) used maximum likelihood (ML) methods to estimate the fatigue limit and the mean life of polyethylene terephthalate films (used in electrical insulation) at the service stress. He fitted a Weibull inverse power relationship that includes a fixed fatigue limit parameter. Nelson (1984) studied the fatigue life of a nickel-base superalloy and fitted fatigue curves with nonconstant standard deviation to data with censored observations using ML methods. He fitted a quadratic relationship to describe the curvature in the plot of (log) fatigue life versus (log) pseudo-stress. Shen, Wirsching, and Cashman (1996a) reviewed previous work on statistical models that characterize fatigue strength and describe trends in fatigue data. They compared the fits of such models to several fatigue datasets. Pascual and Meeker (1997) presented a model with a fatigue limit parameter and nonconstant standard deviation of log fatigue life to describe curvature and nonconstant variance in stress-life relationships. They fit the model to the nickel-base superalloy data studied by Nelson (1984). They also studied the effect that test length has on precision of estimates by analyzing simulated datasets based on their model. Nelson (1990, pp. 93–95) suggested modeling the fatigue limit as a random parameter; that is, test specimens have different fatigue limits according to some distribution called the "strength distribution." Dieter (1976) discussed the statistical (random) nature fatigue limits and mentioned that, in heat-treated alloy of forging steel, 95% of fatigue limits fall between 40,000 and 52,000 psi. Symonds (1996) gave typical approximate fatigue limits for different types of metal in reversed bending tests. Klesnil and Lukáš (1992, pp. 178–180) discussed the influence of grain size on the fatigue limit. Collins (1993, chap. 7) listed factors such as material composition, grain size and direction, heat treatment, and surface conditions that affect the S-N-P curves.

Little (1974) discussed the use of the up-and-down method to estimate the median fatigue limit with extreme value distributions based on ML and minimum chi-squared methods. The up-and-down method tests specimens in sequence at equally spaced stress levels for a specified large number of cycles (e.g., 10^7 cycles). A specimen is tested at the next lower (higher) stress level if the previous test produces a failure (right-censored observation). Little (1990) presented a modified up-and-down test that uses a minimum variance strategy to choose the next stress level. The up-and-down method is an efficient and effective way of estimating the median fatigue limit. It is not used to estimate the stress-life relationship because the fatigue data are analyzed as quantal response data in which the main concern is whether or not a specimen on test has failed.

1.3 Overview

In Section 2, we discuss a statistical model for fatigue life that includes a random fatigue limit. This model describes and provides motivation for a fatigue life distribution that has the standard deviation as a function of stress. Section 3 describes ML methods. ML methods allow for censoring, which is common in fatigue testing, particularly at low levels of stress. Sections 4 and 5 illustrate the application of the model to actual fatigue data. We show how to compute ML estimates of parameters and fatigue-life distribution quantiles. Estimates of these quantiles are important inputs to product design processes and, thus, of most interest to engineers. Lower confidence bounds, based on likelihood ratio methods, are computed for these quantiles. We assess the fit of the model to the data by constructing diagnostic plots and performing goodness-of-fit tests and residual analyses. Section 6 outlines possible areas for further research.

2. THE RANDOM FATIGUE–LIMIT MODEL

There are two main considerations in modeling the relationship between the applied stress and fatigue life. First, often the standard deviation of fatigue life decreases as the applied stress increases. Second, curvature in fatigue curves suggests the inclusion of a fatigue limit in the statistical model for fatigue life. The random fatigue-limit model describes both of these characteristics.

Let Y be the fatigue life and s the stress level. We model Y as

$$\log(Y) = \beta_0 + \beta_1 \log(s - \gamma) + \varepsilon, \qquad s > \gamma,$$

where β_0 and β_1 are fatigue curve coefficients, γ is the fatigue limit of the specimen, ε is the error term, and log denotes natural logarithm. Let $V = \log(\gamma)$, and suppose that V has probability density function (pdf)

$$f_V(v; \mu_\gamma, \sigma_\gamma) = \frac{1}{\sigma_\gamma} \phi_V \left(\frac{v - \mu_\gamma}{\sigma_\gamma} \right)$$

with location and scale parameters μ_γ and σ_γ, respectively. $\phi_V(\cdot)$ is either the standardized smallest extreme value (sev) or normal pdf.

Let $x = \log(s)$ and $W = \log(Y)$. Assume that, conditioned on a fixed value of $V < x, W|V$ has pdf

$$f_{W|V}(w; \beta_0, \beta_1, \sigma, x, v) = \frac{1}{\sigma} \phi_{W|V} \left(\frac{w - [\beta_0 + \beta_1 \log(\exp(x) - \exp(v))]}{\sigma} \right)$$

with location parameter $\beta_0 + \beta_1 \log(\exp(x) - \exp(v))$ and scale parameter σ. $\phi_{W|V}(\cdot)$ is either the standardized sev or

normal pdf. The marginal pdf of W is given by

$$f_W(w; x, \boldsymbol{\theta}) = \int_{-\infty}^{x} \frac{1}{\sigma\sigma_\gamma}\phi_{W|V}\left(\frac{w-\mu(x,v,\boldsymbol{\theta})}{\sigma}\right)\phi_V\left(\frac{v-\mu_\gamma}{\sigma_\gamma}\right)dv,$$

where $\boldsymbol{\theta} = (\beta_0, \beta_1, \sigma, \mu_\gamma, \sigma_\gamma)$ and $\mu(x, v, \boldsymbol{\theta}) = \beta_0 + \beta_1 \log[\exp(x) - \exp(v)]$. The marginal cumulative distribution function (cdf) of W is given by

$$F_W(w; x, \boldsymbol{\theta}) = \int_{-\infty}^{x} \frac{1}{\sigma_\gamma}\Phi_{W|V}\left(\frac{w-\mu(x,v,\boldsymbol{\theta})}{\sigma}\right)\phi_V\left(\frac{v-\mu_\gamma}{\sigma_\gamma}\right)dv,$$

where $\Phi_{W|V}(\cdot)$ is the cdf of $W|V$. We will refer to this statistical model as the random fatigue-limit model. There are no closed forms for the density and distribution functions of W. They are, however, easy to evaluate numerically.

We will see that this model has the properties that one usually sees in fatigue-limit data. In particular, the model adequately describes curvature in the stress-life relationship and the increase in variability in log fatigue life at low stress/strain levels.

3. MAXIMUM LIKELIHOOD ESTIMATION

We use ML methods to estimate the parameters of the random fatigue-limit model. Statistical theory suggests that ML estimators, in general, have favorable asymptotic (large-sample) properties. For "large" sample sizes and under certain conditions on the fatigue life distribution, the distribution of ML estimators is approximately multivariate normal with mean vector equal to the vector of true values being estimated and standard deviations no larger than that of any other competing estimators. See Nelson (1990, chap. 5) for an in-depth discussion of ML estimation with censored data.

Let $y_p(s)$ be the p quantile of the life distribution at stress level s. We obtain ML estimates of $y_p(s)$ for $p = .05, .50$, and .95. We compute approximate likelihood-ratio-based lower confidence bounds for the .01 and .05 quantiles of the life distribution. Ostrouchov and Meeker (1988) used Monte Carlo simulations to compare the accuracy of confidence intervals based on likelihood ratio and those based on asymptotic normal theory for interval-censored Weibull and lognormal data. Vander Wiel and Meeker (1990) did a similar study for a simple accelerated life-test model. Both articles concluded that likelihood confidence intervals have coverage probabilities generally closer to nominal confidence levels than those of normal approximation intervals even in small to moderate size samples.

3.1 Parametric Likelihood

For the random fatigue-limit model defined previously with sample data $w_1 = \log(y_1), \ldots, w_n = \log(y_n)$ at log stress levels $x_1, \ldots, x_n$, respectively, the likelihood is

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} [f_W(w_i; x_i, \boldsymbol{\theta})]^{\delta_i}[1 - F_W(w_i; x_i, \boldsymbol{\theta})]^{1-\delta_i},$$

where

$$\delta_i = \begin{cases} 1 & \text{if } w_i \text{ is a failure} \\ 0 & \text{if } w_i \text{ is a censored observation.} \end{cases}$$

The function $L(\boldsymbol{\theta})$ can be interpreted as being approximately proportional to the probability of observing $y_1, \ldots, y_n$ for a given set of parameters $\boldsymbol{\theta}$. Generally, it is easier to work with the log-likelihood function

$$\mathcal{L}(\boldsymbol{\theta}) = \log[L(\boldsymbol{\theta})] = \sum_{i=1}^{n} \mathcal{L}_i(\boldsymbol{\theta}),$$

where

$$\mathcal{L}_i(\boldsymbol{\theta}) = \delta_i \log[f_W(w_i; x_i, \boldsymbol{\theta})] + (1 - \delta_i)\log[1 - F_W(w_i; x_i, \boldsymbol{\theta})]$$

is the contribution of the ith observation. The ML estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is the set of parameter values that maximizes $L(\boldsymbol{\theta})$ or $\mathcal{L}(\boldsymbol{\theta})$.

3.2 Profile Likelihoods and Likelihood-Ratio-Based Confidence Regions

We use the profile likelihood to compute approximate confidence intervals for quantities or vectors of interest. These intervals are based on inverting a likelihood ratio test. Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ be a partition of $\boldsymbol{\theta}$, where $\boldsymbol{\theta}_1$ is a vector of k quantities of interest. Let $\hat{\boldsymbol{\theta}}$ denote the ML estimate of $\boldsymbol{\theta}$. The profile likelihood for $\boldsymbol{\theta}_1$ is defined by

$$R(\boldsymbol{\theta}_1) = \max_{\boldsymbol{\theta}_2}\left[\frac{L(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{L(\hat{\boldsymbol{\theta}})}\right].$$

A large value (close to 1) of $R(\boldsymbol{\theta}_1)$ indicates that the observed data are highly probable for that value of $\boldsymbol{\theta}_1$, relative to the ML estimate. On the other hand, a small value (close to 0) of $R(\boldsymbol{\theta}_1)$ indicates that the observed data are relatively unlikely for the given value of $\boldsymbol{\theta}_1$. When $k = 1$, plotting $R(\boldsymbol{\theta}_1)$ against different values of $\boldsymbol{\theta}_1$ yields a profile likelihood plot for $\boldsymbol{\theta}_1$.

When evaluated at the true value $\boldsymbol{\theta}_1$, the asymptotic distribution of $-2\log[R(\boldsymbol{\theta}_1)]$ is a chi-squared distribution with k df. As a result, an approximate $100(1-\alpha)\%$ confidence region for $\boldsymbol{\theta}_1$ is given by the set of all $\boldsymbol{\theta}_1$ such that

$$-2\log[R(\boldsymbol{\theta}_1)] \le \chi^2_{(k;1-\alpha)}$$

or, equivalently,

$$R(\boldsymbol{\theta}_1) \ge \exp\left[-\frac{\chi^2_{(k;1-\alpha)}}{2}\right],$$

where $\chi^2_{(k;1-\alpha)}$ is the $(1-\alpha)$ quantile of a chi-squared distribution with k df. If $k = 1$, the preceding inequalities yield confidence intervals. If $k = 2$, the equation

$$R(\boldsymbol{\theta}_1) = \exp\left[-\frac{\chi^2_{(2;1-\alpha)}}{2}\right]$$

defines a constant-likelihood contour line corresponding to a $100(1-\alpha)\%$ joint confidence region.

Confidence intervals based on the approximate normal distribution of studentized ML estimators can also be computed. As mentioned earlier, however, likelihood confidence intervals perform better in the sense that coverage probabilities are closer to nominal confidence levels than those of normal-approximation intervals.

4. LAMINATE PANEL DATA

In this section, we fit the random fatigue-limit model to fatigue data given by Shimokawa and Hamaguchi (1987). The data come from 125 specimens in four-point out-of-plane bending tests of carbon eight-harness-satin/epoxy laminate. Fiber fracture and final specimen fracture occurred simultaneously. Thus, fatigue life is defined to be the number of cycles until specimen fracture. The dataset includes 10 right-censored observations (known as "runouts" in the fatigue literature). Figure 1 shows the data and fitted random fatigue-limit models on a log-log scale with time on the horizontal axis, as is traditional in the fatigue literature. In this figure, "•" and "▷" represent failures and censored observations, respectively.

4.1 Maximum Likelihood Estimates of the Random Fatigue-Limit Model Parameters

We fit the random fatigue-limit model discussed in Section 2 to the data under the sev-sev, normal-normal, sev-normal, and normal-sev combinations for the respective distributions of V and $W|V$. Table 1 gives the ML estimates of the model parameters and the value of the log-likelihood for these estimates. The table includes ML estimates $\hat{y}_{.05}(s)$ of the .05 quantile of fatigue life in units of thousands of cycles at stress levels $s = 270, 280, 300, 340$, and 380 MPa. We compute values of the Akaike information criterion (AIC) statistic for each model to identify which best approximates the true underlying model. The AIC statistic is given by

$$\text{AIC} = -2[\log_{\max} L(\boldsymbol{\theta}) - k],$$

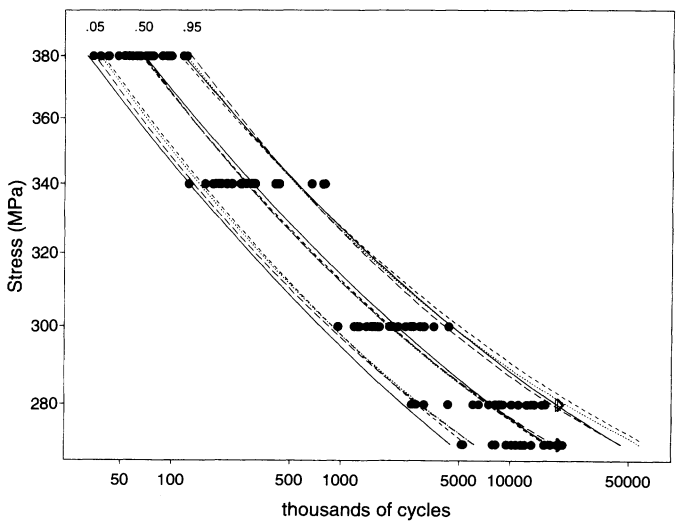

Figure 1. Log-Log S-N Plot for the Laminate Panel Data With ML Estimates of the .05, .50, and .95 Quantiles (•, failure; ▷, censored observation): ———, Sev-Sev; - - - -, Normal-Normal; — — —, Sev-Normal; — — —, Normal-Sev.

Table 1. ML Results for the Laminate Panel Data

		Model			
		Sev-sev	Normal-normal	Sev-normal	Normal-sev
Log-likelihood	log[$L(\theta)$]	−92.706	−86.221	−87.292	−87.603
AIC statistic		195.412	182.442	184.584	185.206
Parameters	β_0	35.575	30.272	33.025	29.435
	β_1	−5.993	−5.100	−5.570	−4.950
	σ	.239	.289	.141	.367
	μ_γ	5.295	5.366	5.323	5.390
	σ_γ	.033	.031	.041	.020
Quantiles	$\hat{y}_{.05}(270)$	4,443.0	6,136.0	5,530.0	6,139.0
	$\hat{y}_{.05}(280)$	2,319.0	2,963.0	2,810.0	2,899.0
	$\hat{y}_{.05}(300)$	751.0	884.0	888.0	840.0
	$\hat{y}_{.05}(340)$	126.0	144.0	150.0	134.0
	$\hat{y}_{.05}(380)$	32.0	38.0	39.0	35.0

where k is the number of model parameters. Smaller values of AIC indicate better fits. See Akaike (1973) for more details. Based on the AIC values, the normal-normal model (V and $W|V$ are both normal) provides the best fit to the data among the four models. On the other hand, the sev-sev model gives the worst fit.

Figure 1 shows curves for the ML estimates of the .05, .50, and .95 quantiles of fatigue life under the different distribution combinations. When compared with the other combinations, the sev-sev combination consistently yields lower estimates of the .05 quantile and higher estimates of the .50 quantile of fatigue life. For the .95 quantile, all combinations yield similar estimates at the intermediate stress levels; the sev-normal combination yields higher estimates at the extreme stress levels. Comparing the fitted curves in these plots to the actual data suggests that the normal-normal and normal-sev models fit the data well.

Table 2 gives approximate confidence intervals for the normal-normal model parameters based on large-sample asymptotics and likelihood ratio methods for the laminate panel data. It also gives the asymptotic standard errors and coefficients of variation (the standard error as a percentage of the estimate) of the estimators. Note that the lower endpoint of the likelihood ratio confidence interval for σ is 0. Both confidence intervals for σ_γ exclude 0. Thus, there is evidence suggesting the random nature of the fatigue limit. If the normal-normal model is appropriate, the estimated (log) fatigue-limit distribution is normal with mean 5.366 and standard deviation .031 log (MPa).

Figure 2 gives the profile likelihood plots for the model parameters for the laminate panel data. The plots indicate the approximate 95% likelihood-ratio-based confidence intervals for the parameters. These plots suggest that the likelihood has a unique maximum.

Viewing ML estimators as random variables, it is sometimes important or interesting to consider the propensity of these estimators to vary together. Issues involved in studying the correlation among parameter estimators are similar to issues of multicollinearity that arise in linear regression analysis. Two parameters with high correlation cannot be estimated independently and interpretation can be difficult. In severe cases, it may be that certain parameters are not

Table 2. Confidence Intervals for the Parameters of the Normal-Normal Model for the Laminate Panel Data

Parameter	Normal approximation		Coefficient of variation (%)	Likelihood ratio confidence interval
	Confidence interval	Standard error		
β_0	(21.814, 38.731)	4.316	14.26	(23.809, 42.691)
β_1	(−6.600, −3.600)	.765	−15.01	(−7.230, −3.927)
σ	(.125, .454)	.084	29.03	(.000, .435)
μ_γ	(5.233, 5.499)	.068	1.26	(5.151, 5.462)
σ_γ	(.016, .047)	.008	25.80	(.017, .053)

identifiable and numerical problems (i.e., difficulty in finding the maximum of the likelihood) will arise. An estimate of the correlation matrix is often an output or an optional output of ML estimation. These correlation estimates are functions of the curvature of the likelihood, evaluated at its maximum (when one exists), as quantified by partial derivatives of the likelihood, evaluated at the ML estimates.

For a given model, the correlations among parameter estimators depend on the parameterization of the model and on the available data. As suggested by Ross (1990), some of numerical difficulties can be addressed by using a "stable" parameterization. Using a stable parameterization, if one is available, will also suggest what aspects of the model are easily identifiable. As with multicollinearity, identifiability can, in some circumstances, also be avoided by collecting appropriate data.

Table 3 gives the estimates of the correlations between parameter estimators for the laminate panel data. In this case, the estimators $\hat{\beta}_0, \hat{\beta}_1$, and $\hat{\mu}_\gamma$ are highly correlated with each other, due in large part to the fact that these three parameters all strongly affected the position of the S-N curve. The high correlations could also be attributed to the lack of curvature in the S-N curve. A substantial amount of curvature is needed to precisely estimate these parameters. The high (negative) correlation between estimators $\hat{\sigma}$ and $\hat{\sigma}_\gamma$ suggests that the variability of fatigue life can be explained by either of the corresponding model parameters.

The contour plots for the model parameters reflect the correlations between parameter estimators. Narrower contours indicate higher correlations. Figures 3 and 4 give contour plots for two pairs of parameters. The contour plots presented here indicate single peaks and, thus, the absence

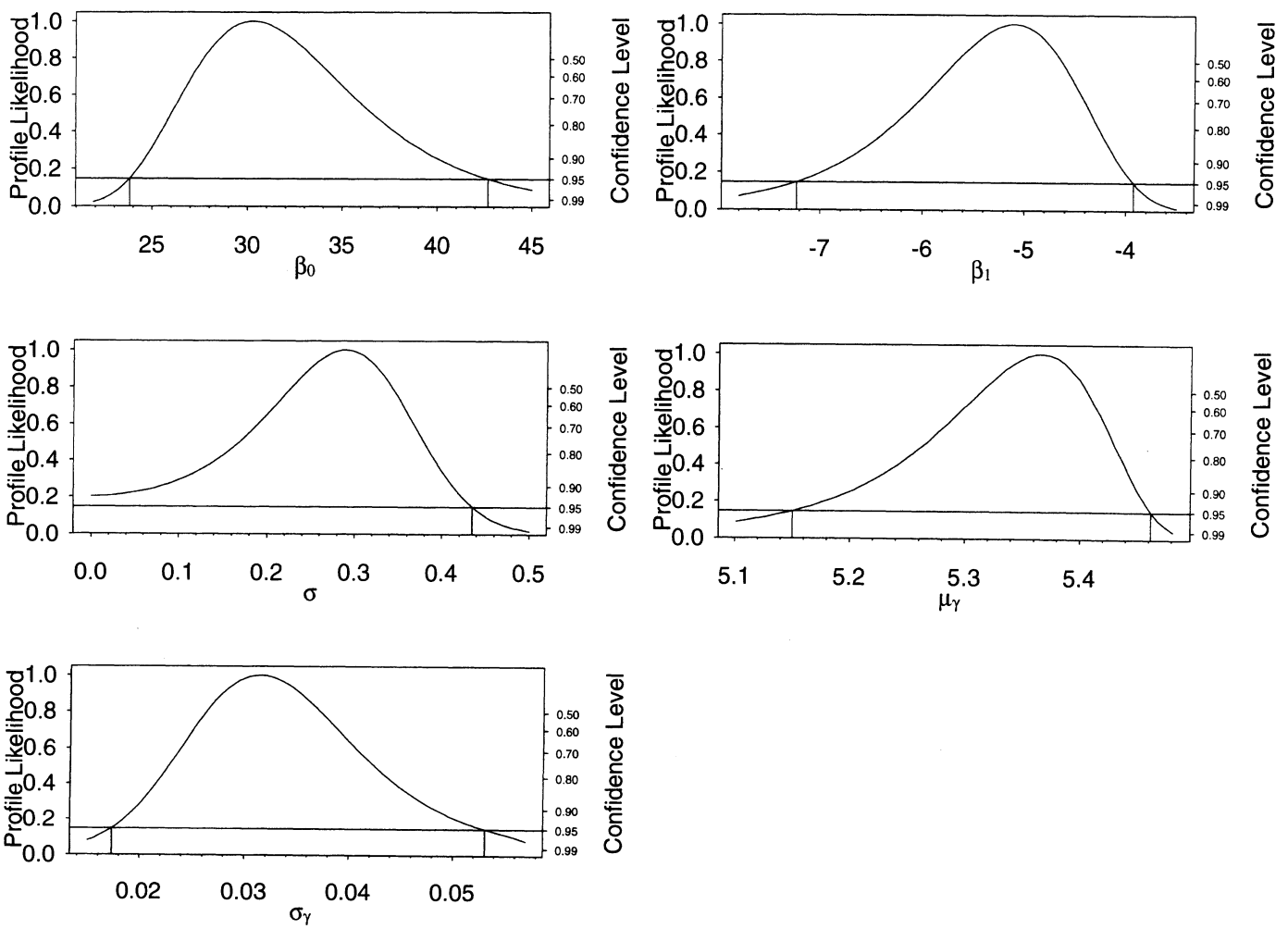

Figure 2. Profile Plots for Model Parameters for the Laminate Panel Data.

Table 3. Estimated Correlations Between Model Parameter Estimates for the Laminate Panel Data

	β_0	β_1	σ	μ_γ	σ_γ
β_0	1.0000	−.9997	−.35	−.99	.63
β_1		1.0000	.35	.99	−.63
σ			1.0000	.35	−.82
μ_γ				1.0000	−.63
σ_γ					1.0000

of multiple likelihood extrema.

4.2 Approximate 95% Lower Confidence Bounds for the .05 and .01 Quantiles of a Fatigue Life Distribution

Shen et al. (1996b) reviewed different methods of computing "design curves" that include quantile curves and tolerance limits for quantiles. In many applications, low quantiles of the life distribution are of primary interest. Using likelihood ratio methods we compute pointwise approximate 95% lower confidence bounds for the .05 and .01 quantiles of fatigue life based on the four distribution combinations discussed previously. Lower bounds for these quantities can be used to characterize fatigue strength. Figures 5 and 6 give plots of the lower confidence bounds for the .05 and .01 quantiles, respectively. In both plots, the sev-sev model gives the lowest confidence bounds. For the .05 quantile, the normal-normal, sev-normal, and normal-sev models give similar bounds. There are clearer differences among these three models for the .01 quantile. Here, the normal-normal model gives the highest bounds.

Figure 1 shows that, for the range of stress in the data, there is a one-to-one correspondence between the .05 quantile and the applied stress. In particular, the .05 quantile increases as stress is reduced. Thus, Figure 5 also provides approximate 95% upper confidence bounds for the stress level yielding a particular .05 quantile. To obtain these bounds graphically, we locate the point on the curve corresponding to the desired quantile value and read off the stress level on the vertical axis. A similar comment can be made about the .01 quantiles.

4.3 Probability (P-P) Plots

It is important to have a method for assessing distributional fit. A commonly used tool for comparing fits of competing distributions is the quantile-quantile (Q-Q) probability plot in which the sample quantiles are plotted against the corresponding quantiles of a hypothesized distribution. We can use Q-Q plots to assess the fit of the distribution based on the ML estimates of model parameters. If a model is appropriate, the corresponding plot should be roughly linear. There are, however, several difficulties in using Q-Q plots. For instance, points in the distribution tails in Q-Q plots are those with the greatest variability. In general, these plots work well with location-scale distributions such as the normal, smallest extreme value and logistic distributions. For other distributions, like the random fatigue-limit model, separate Q-Q plots with different plotting axes are necessary for different parameter values, making it more difficult to display and compare the plots across different distribution combinations. Probability (P-P) plots provide a convenient alternative for this application. See Crowder, Kimber, Smith, and Sweeting (1991, chaps. 2 and 3) for more information on P-P plots.

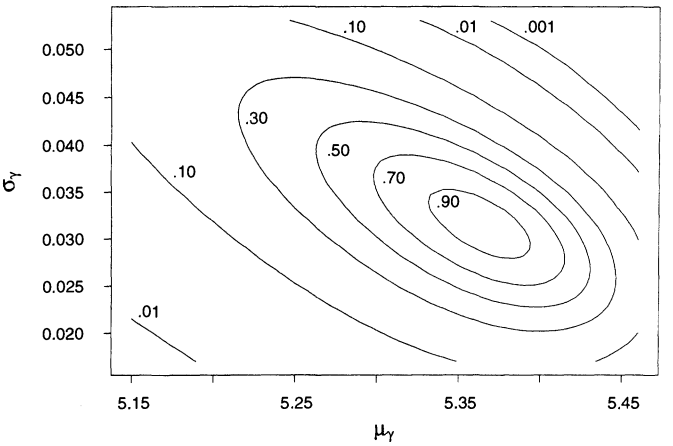

Figure 4. Contour Plot for μ_γ and σ_γ for the Laminate Panel Data.

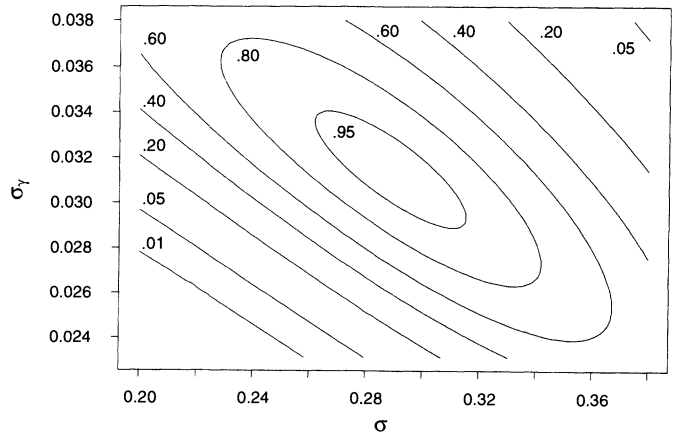

Figure 3. Contour Plot for σ and σ_γ for the Laminate Panel Data.

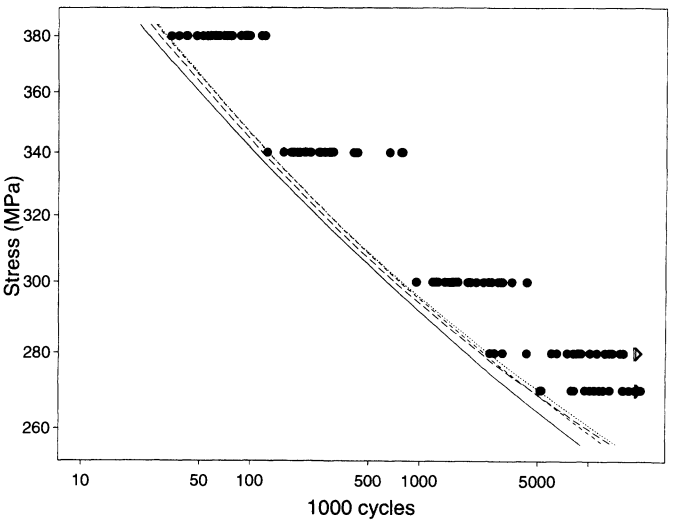

Figure 5. Approximate 95% Lower Confidence Lower Bounds for the .05 Quantile of Fatigue Life for the Laminate Panel Data: ——, Sev-Sev; - - - -, Normal-Normal; – – –, Sev-Normal; — —, Normal-Sev.

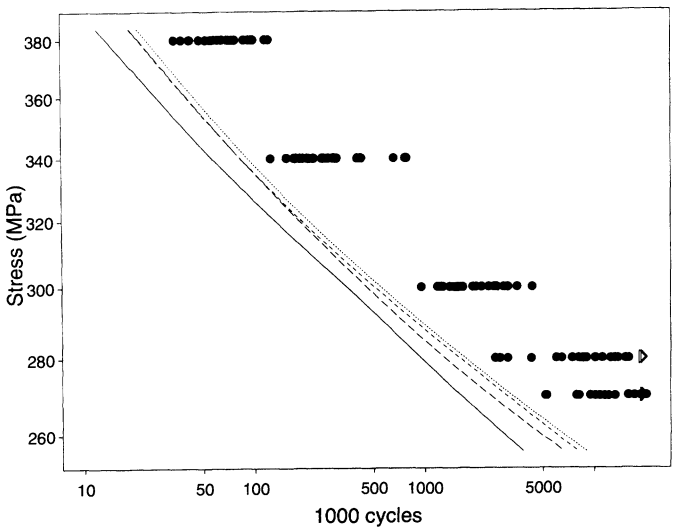

Figure 6. Approximate 95% Lower Confidence Lower Bounds for the .01 Quantile of Fatigue Life for the Laminate Panel Data: ———, Sev-Sev; — —, Normal-Normal; – – –, Sev-Normal; — —, Normal-Sev.

For the random fatigue-limit model, the plotting points of a Q-Q or P-P plot are based on the Kaplan–Meier estimate of the survival probability and the ML estimate of the fatigue life distribution. Let $y_1 \leq \cdots \leq y_n$ be the ordered observations at log stress x. Let d_i be the number of failures at time y_i and n_i be the total number of unfailed and uncensored observations just prior to y_i. The Kaplan–Meier estimate of the survival probability $S(y) = \Pr(Y \geq y)$ is given by

$$\hat{S}(y) = \prod_{\{i:y_i<y\}} \left(1 - \frac{d_i}{n_i}\right).$$

Suppose that failures occur at distinct times $y'_1 < \cdots < y'_r$, where $r \leq n$. Let

$$p_i = 1 - \frac{1}{2}\{\hat{S}(y'_i) + \hat{S}(y'_{i+1})\}$$

for $i = 1, \ldots, r$.

Let $\hat{\boldsymbol{\theta}}$ be the ML estimate of $\boldsymbol{\theta}$. The plotting points of a Q-Q plot are given by

$$(\log(y'_i), F_W^{-1}(p_i; x, \hat{\boldsymbol{\theta}}))$$

for $i = 1, \ldots, r$. In practice, if the parent distribution F_W is location-scale, the location parameter is set to 0 and the scale parameter to 1. The plotting points of a P-P plot are given by

$$(p_i, F_W(\log(y'_i); x, \hat{\boldsymbol{\theta}}))$$

for $i = 1, \ldots, r$. Wilk and Gnanadesikan (1968) remarked that P-P plots are sensitive to discrepancies in the center of the distribution. In contrast to Q-Q plots, the extreme tail points in P-P plots have lower variability than those in the center. Michael (1983) suggested constructing a stabilized probability (S-P) plot with points,

$$\left(\frac{2}{\pi}\sin^{-1}(p_i), \frac{2}{\pi}\sin^{-1}[F_W(\log(y'_i); x, \hat{\boldsymbol{\theta}})]\right),$$

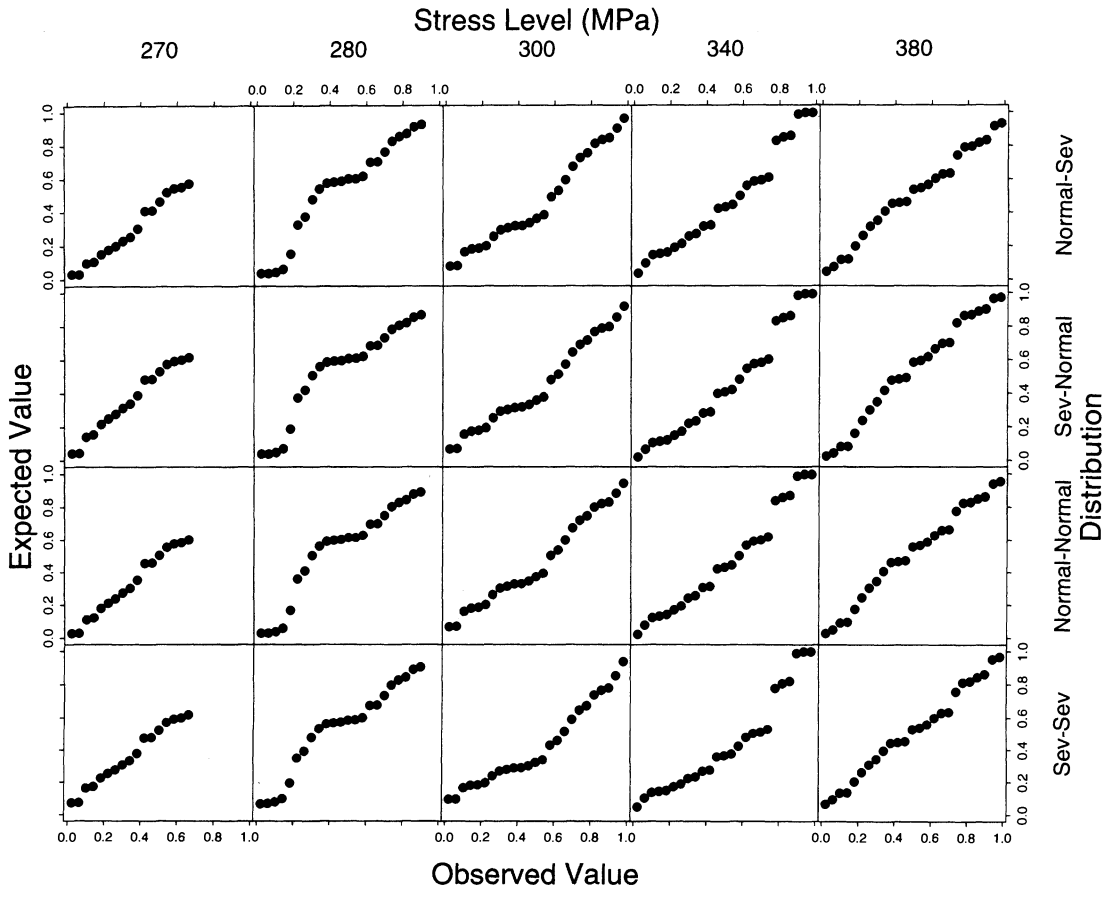

Figure 7. P-P Plots for the Laminate Panel Data.

Table 4. Kolmogorov–Smirnov D-Statistic Values for the Fits of the Random Fatigue-Limit Models to the Laminate Panel Data

Stress (MPa)	Sev-sev	Normal-normal	Sev-normal	Normal-sev
270	.4	.4	.5	.6
280	1.1	1.2	1.2	1.1
300	1.1	.9	.9	.9
340	1.2	.7	.8	.8
380	.4	.3	.4	.4

to stabilize the variation in plotting points.

We feel that no particular method (P-P, Q-Q, or S-P plot) has any big advantage over the others for any particular part of the distribution. To know what can be expected from statistical noise, one only needs a method of calibration. Crowder et al. (1991, p. 66) suggested using Monte Carlo methods to aid in the interpretation of P-P and Q-Q plots. They also provided pertinent references.

The axes for the P-P plots are the same for any set of parameter values. This facilitates comparison of P-P plots under competing models. Figure 7 gives P-P plots by stress level for the panel data. Linearity in the plots indicates a good fit. A careful inspection of the plots reveals that the sev-sev combination does not fit as well as the rest, particularly for stress levels 300 and 340 MPa. It appears that the normal-normal model has the best fit. The plots indicate a possible anomaly in stress levels 280 and 300 MPa. There are concentrations of points around 8.5 million cycles for 280 MPa and around 1.5 million cycles for 300 MPa. The Q-Q and S-P plots (not shown here) for the data yield similar information. The Q-Q, P-P, and S-P plots agree with the AIC statistics in Table 1 that the sev-sev and normal-normal models provide, respectively, the worst and best fits to the data.

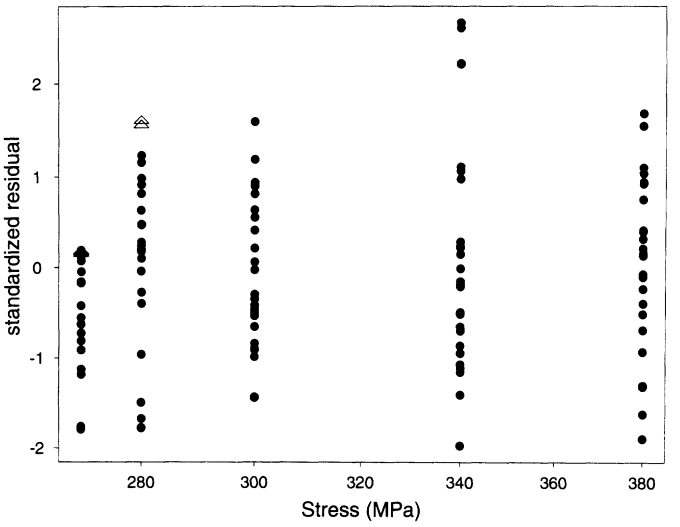

Figure 9. Plot of Standardized Residuals Versus Stress Levels for the Normal-Normal Model Fit to the Laminate Panel Data.

4.4 Goodness-of-Fit Tests

We assess the statistical significance of departures from the random fatigue-limit model by performing empirical distribution function goodness-of-fit tests. We use Kolmogorov–Smirnov D statistics to perform these tests. For each stress level and distribution combination, we test the null hypothesis that the data obtained come from the corresponding random fatigue-limit distribution. To do this, we adapt the methods discussed by D'Agostino and Stephens (1986, chap. 4). Let $w_1 = \log(y_1) \leq \cdots \leq w_n = \log(y_n)$ be the ordered observations at log stress x. Let $z_i = F_W(w_i; x, \boldsymbol{\theta})$ for $i = 1, \ldots, n$. Under the true value $\boldsymbol{\theta}, Z_i = F_W(W_i; x, \boldsymbol{\theta})$ are ordered uniform random variables. If one or more components of $\boldsymbol{\theta}$ are unknown, these components are replaced by estimates—for example, the ML estimate $\hat{\boldsymbol{\theta}}$. However, $z_i = F_W(w_i; x, \hat{\boldsymbol{\theta}})$ will not be an

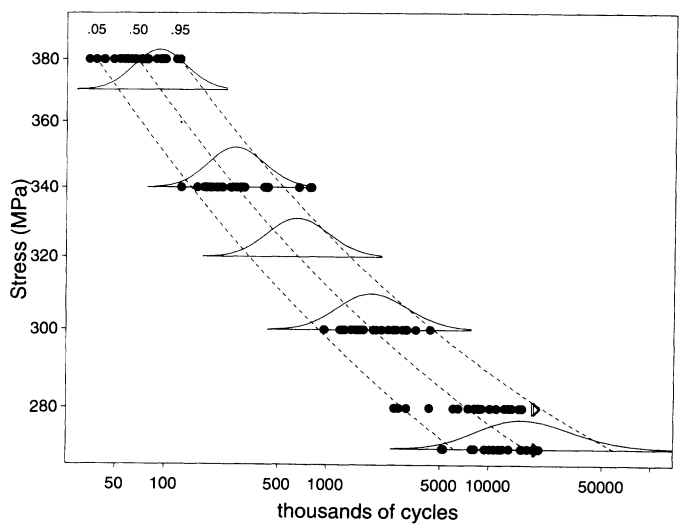

Figure 8. Log-Log S-N Plot for the Laminate Panel Data With ML Estimates of Density Curves and the .05, .50, and .95 Quantile Estimates Under the Fitted Normal-Normal Model.

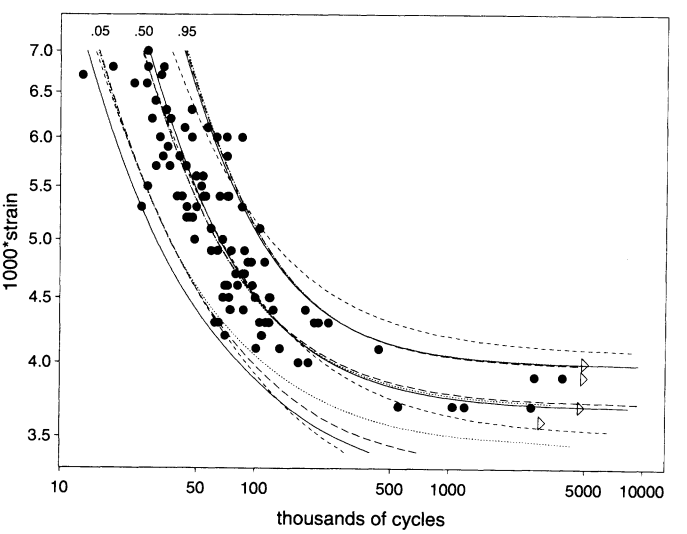

Figure 10. Log-Log S-N Plot for the Nickel-Base Superalloy Data With ML Estimates of the .05, .50, and .95 Quantiles (•, failure; ▷, censored observation); ———, Sev-Sev; - - - -, Normal-Normal; – – –, Sev-Normal; — —, Normal-Sev.

Table 5. ML Results for the Nickel-Base Superalloy Data

		Model			
		Sev-sev	Normal-normal	Sev-normal	Normal-sev
Log-likelihood	log[$L(\theta)$]	−67.895	−62.082	−66.352	−63.161
AIC statistic		145.790	134.164	142.704	136.322
Parameters	β_0	4.552	4.370	4.705	4.387
	β_1	−.923	−.928	−1.105	−.938
	σ	.227	.315	.203	.305
	μ_γ	1.324	1.309	1.257	1.329
	σ_γ	.052	.044	.089	.046
Quantiles	$\hat{y}_{.05}(3.5)$	226.0	1131.8	219.9	354.3
	$\hat{y}_{.05}(4)$	82.8	110.2	90.0	96.8
	$\hat{y}_{.05}(5)$	30.7	36.0	36.2	36.0
	$\hat{y}_{.05}(6)$	18.8	21.6	21.5	21.8
	$\hat{y}_{.05}(7)$	13.6	15.5	15.0	15.7

ordered uniform sample even when the null hypothesis is true. D'Agostino and Stephens (1986) suggested modifications of the test statistic to account for the use of $\hat{\boldsymbol{\theta}}$ in place of $\boldsymbol{\theta}$. The modifications are functions of the test statistic and sample size. They give tables of percentage points for the modified test statistic. We shall replace unknown parameter values with ML estimates and compute modified K–S test statistic values.

For complete (no censoring) datasets, the K-S D statistic is given by

$$D = \max_{1\le i\le n}\left\{\frac{i}{n} - z_i, z_i - \frac{i-1}{n}\right\}.$$

The statistic is modified using the formula

$$D^* = D\left(\sqrt{n} + .12 + \frac{.11}{\sqrt{n}}\right).$$

From D'Agostino and Stephens (1986, p. 105, table 4.2), the reject region for the .05 level of significance is $\{D^* > 1.358\}$.

Suppose that the dataset contains r failures and $n - r$ Type I censored observations that have log failure times above w_t. Let $z_t = F_W(w_t; x, \hat{\boldsymbol{\theta}})$. The K–S statistic adapted for Type I censoring is given by

$$D = \max_{1\le i\le r}\left\{\frac{i}{n} - z_i, z_i - \frac{i-1}{n}, z_t - \frac{r}{n}\right\}.$$

When unknown parameters are replaced by estimators, D'Agostino and Stephens (1986) suggested the modification

$$D^* = \sqrt{n}D + \frac{.19}{\sqrt{n}}.$$

The percentage points for D^* were given by D'Agostino and Stephens (1986, p. 112, table 4.4).

In the laminate panel data, there are $n = 25$ observations at each stress level. There are $r = 8$ and 2 censored observations at stress levels 270 and 280 MPa, respectively. In both cases, we choose $w_t = \log(20 \text{ million cycles})$ and compute modified K–S statistic for Type I censoring. At stress levels 300, 340, and 380 MPa, all test units failed and, for these, we compute the modified K–S statistic for complete data.

Table 4 gives the test statistic values for each stress level. None of the tests are significant at the .05 level of significance. Thus, there is not enough evidence to rule out any of these distributions. The observed departures from linearity in the P-P plots may be explained by variability under the hypothesized models.

4.5 Residual Analysis

To assess the validity of models and, in particular, detect problems with the life/stress relationship part of the models, we can study plots of residuals versus the stress levels. We follow methods suggested by Nelson (1973) to perform residual analysis on the fit of the normal-normal model. Figure 8 gives a plot of the fit of this model to the data. The plot includes density curves at selected stress levels.

For the log stress level x_i, define the raw residual e_i by

$$e_i \equiv \log(y_i) - \hat{\mu}(x_i),$$

where $\hat{\mu}(x_i)$ is the ML estimate of the mean log fatigue life at log stress level x_i conditioned on the specimen failing—that is, conditioned on the fatigue limit falling below the stress level. Because the standard deviation $\sigma(x_i)$ of fatigue life varies with the log stress x_i, we define standardized residuals

$$e_i^* \equiv \frac{e_i}{\hat{\sigma}(x_i)},$$

where $\hat{\sigma}(x_i)$ is ML estimate of the standard deviation of log fatigue life at log stress level x_i given that the specimen is going to fail. We use standardized residuals in the plots and refer to them as residuals henceforth.

The plots of residuals versus the stress levels should appear patternless. Figure 9 gives the plot of the residuals versus the stress levels for the normal-normal model. The plot does not show any clear patterns in the residuals.

5. NICKEL–BASE SUPERALLOY DATA

In this section, we fit the random fatigue-limit model to a nickel-base superalloy fatigue dataset given by Shen (1994).

Table 6. Confidence Intervals for the Parameters of the Normal-Normal Model for the Nickel-Base Superalloy Data

	Normal approximation			
Parameter	Confidence interval	Standard error	Coefficient of variation (%)	Likelihood-ratio confidence interval
β_0	(4.2110, 4.5283)	.0809	1.85	(4.2334, 4.5809)
β_1	(−1.0938, −.7624)	.0845	−9.11	(−1.1266, −.7810)
σ	(.2620, .3688)	.0273	8.64	(.2651, .3740)
μ_γ	(1.2751, 1.3424)	.0172	1.31	(1.2626, 1.3374)
σ_γ	(.0247, .0637)	.0099	22.46	(.0292, .0708)

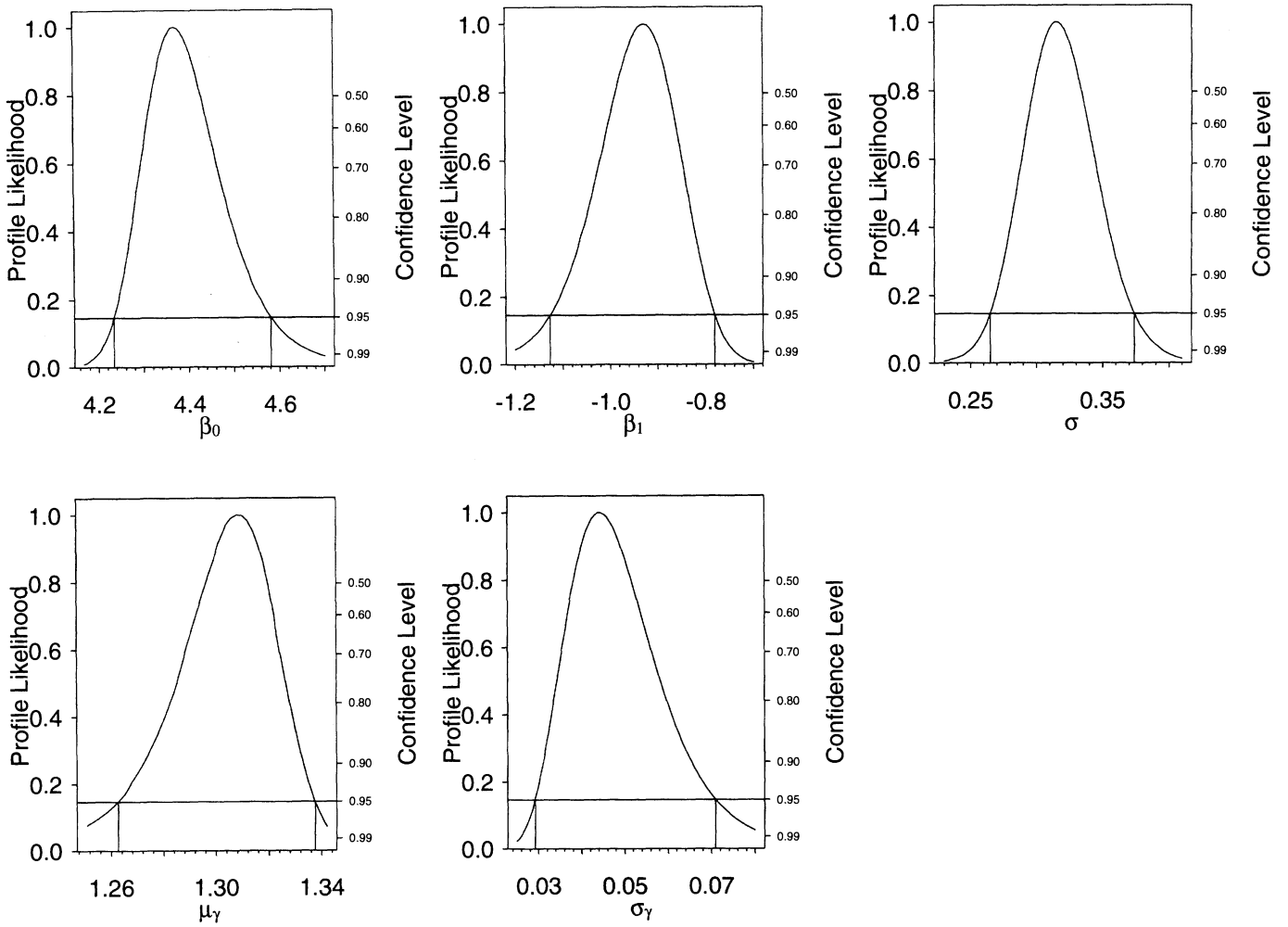

Figure 11. Profile Plots for Model Parameters for the Nickel-Base Superalloy Data.

The original dataset consists of 246 observations with four runouts. We fitted the model to the whole dataset and observed that the model did not fit the data well in the low-strain region. One possible explanation for this is that the character of the failure mechanisms could be different at high strain levels. We refitted the model to the 115 observations for which strain is below .007 units. This improved the fit for the lower strain levels, implying that the omitted high-strain observations were influential, biasing estimates of fatigue life at low levels of strain.

Figure 10 gives a plot of the reduced dataset on a log-log scale. In this plot, "•" and "▷" represent failures and censored observations, respectively. The applied force here is in strain units for which there are 32 unique values in contrast to 5 levels of stress in the laminate panel data. To assess the fits of the models to the data, we again construct P-P plots to compare combinations of distributions for $W|V$ and V, and we perform residual analysis.

Table 7. Estimated Correlations Between Model Parameter Estimates for the Nickel-Base Superalloy Data

	β_0	β_1	σ	μ_γ	σ_γ
β_0	1.0000	.81	.12	−.87	.49
β_1		1.0000	.11	.71	−.42
σ			1.0000	.15	−.33
μ_γ				1.0000	−.54
σ_γ					1.0000

5.1 Maximum Likelihood Estimates of the Random Fatigue-Limit Model Parameters

Table 5 gives the ML estimates of the model parameters and the values of the log-likelihood under the ML estimates. The table includes AIC values and ML estimates $\hat{y}_{.05}(s)$ of the .05 quantiles at strain levels $x = 3.5, 4, 5, 6$, and 7 for each model. The quantile estimates are in units of thousands of cycles. As in the previous example, the AIC values indicate that the normal-normal and the sev-sev combinations provide, respectively, the best and worst fit to the data.

Figure 10 shows curves for the ML estimates of the .05, .50, and .95 quantiles of fatigue life under the different distribution combinations. The models give similar estimates of the .50 and .95 quantiles. At the lower strain levels, the sev-sev model gives lower estimates of the .05 quantile, whereas the normal-normal model gives higher estimates.

The plot of the data in Figure 10 indicates that the standard deviation of fatigue life decreases with increasing strain. Even though the standard deviation is not written explicitly as a function of the strain, the random fatigue-limit model describes the increased variability at lower levels of

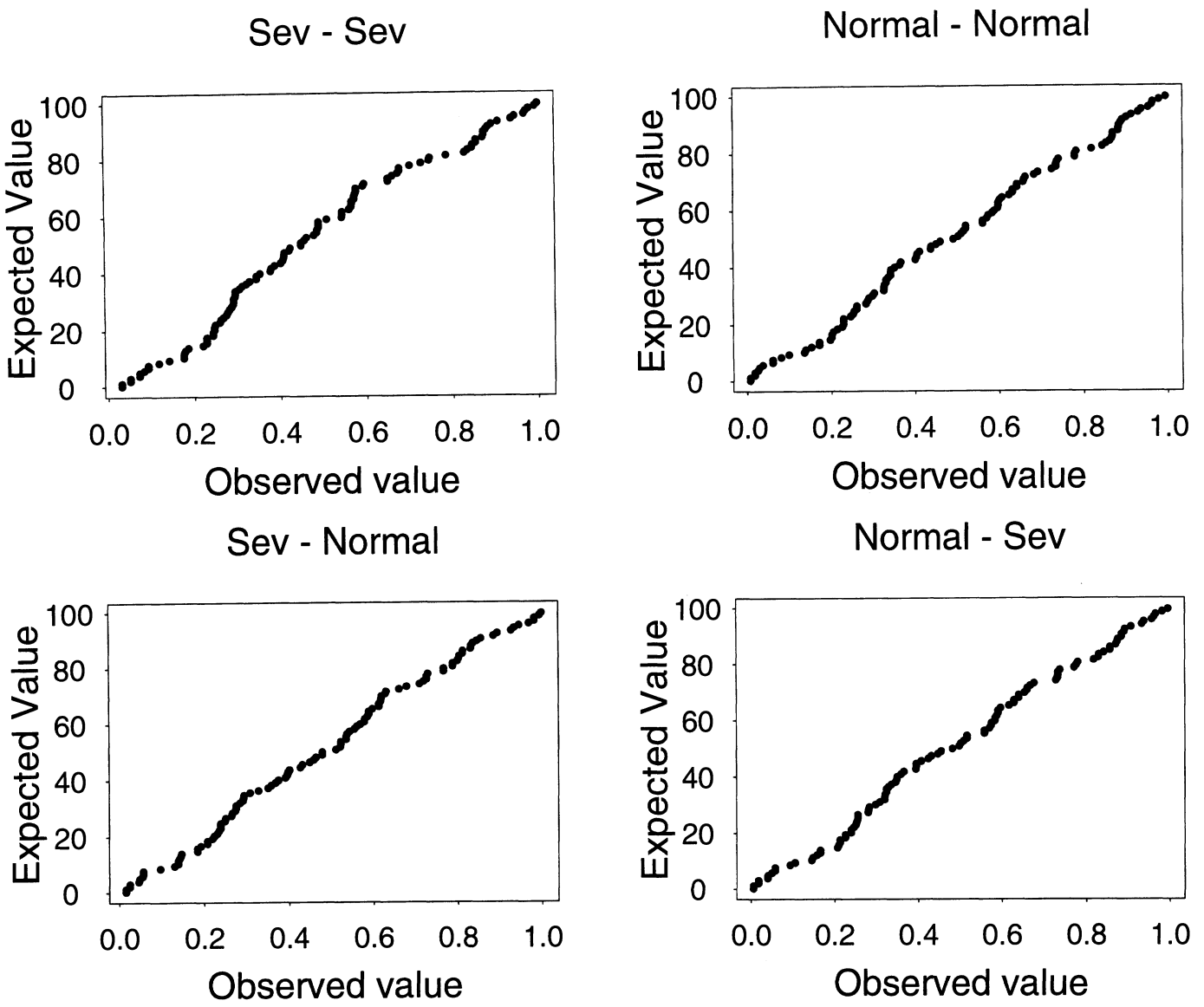

Figure 12. P-P Plots for the Nickel-Base Superalloy Data.

strain. Although not as obvious, a similar observation can be made for the laminate panel example in Section 4.

Table 6 gives approximate confidence intervals for normal-normal model parameters based on large-sample asymptotics and likelihood ratio methods. It also gives the asymptotic standard errors and coefficient of variation of the estimators. In contrast to the laminate panel results, the coefficients of variation are lower and the confidence intervals for σ do not contain 0. If the normal-normal model is appropriate, the estimated (log) fatigue-limit distribution is normal with mean 1.309 and standard deviation .044 log(1,000 $\times$ strain units). Figure 11 gives the profile likelihood plots for the model parameters for the superalloy data. The profile likelihood plots and contour plots (not shown here) suggest that the likelihood has a unique maximum.

Table 7 gives the estimates of the correlations between parameter estimators for the nickel-base superalloy data. In contrast to the laminate panel results, the estimators $\hat{\beta}_0, \hat{\beta}_1$, and μ_γ are not as highly correlated. Observe that there is substantial curvature in the superalloy S-N plot at the lower stress/strain levels unlike in the laminate panel S-N plot. Having a significant curvature in the empirical S-N plot improves the ability to estimate certain functions of the parameters. Thus, test runs should be conducted in a range of stress or strain that will produce a clear curvature in the S-N plot of the resulting data.

We constructed contour plots (not included here) for different pairs of parameter estimators. The plots reflect the correlations as given in Table 7. As in the previous example, the contour plots do not suggest possible multiple likelihood extrema.

Because the superalloy data are not grouped by levels of stress or strain, we cannot construct P-P plots by strain levels. Instead, we take the following approach. Let W_i be the log fatigue life at log strain level x_i [i.e., $W_i = \log(Y_i)$, for $i = 1, \ldots, n$, where n is the sample size]. If $F_W(W_i; x_i, \boldsymbol{\theta})$ is the true distribution of W_i, then $Z_1 = F_W(W_1; x_1, \boldsymbol{\theta}), \ldots, Z_n = F_W(W_n; x_n, \boldsymbol{\theta})$ are independently and identically distributed UNIF(0, 1). Using this result, we construct P-P plots to assess whether if

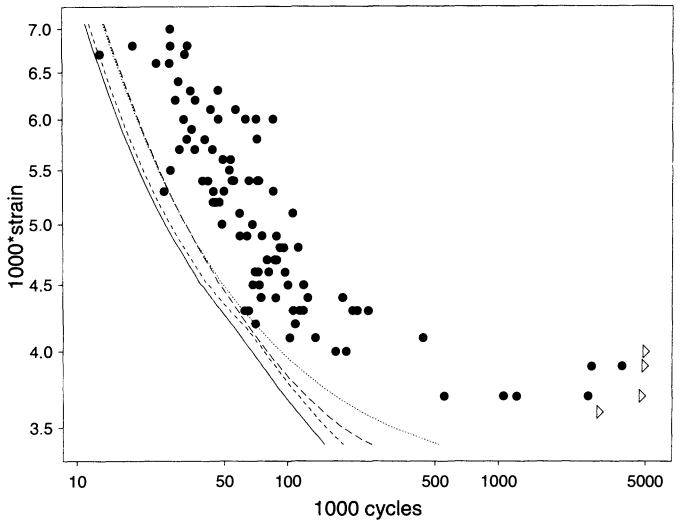

Figure 13. Lower 95% Confidence Lower Bounds for the .05 Quantile of Fatigue Life for the Nickel-Base Superalloy Data: ———, Sev-Sev; - - - -, Normal-Normal; – – –, Sev-Normal; — —, Normal-Sev.

$z_1 = F_W(w_1; x_1, \boldsymbol{\theta}), \ldots, z_n = F_W(w_n; x_n, \boldsymbol{\theta})$ agrees with what we would expect to see from a random sample from a uniform distribution. See Figure 12. The plots show that the normal-normal and normal-sev models provide better fits than the other two distribution combinations. The S-P plots for the data yield similar information. The P-P and S-P plots agree with the AIC statistics in Table 5 that the normal-normal and normal-sev models provide the best fits.

Figures 13 and 14 give plots of the lower confidence bounds for the .05 and .01 quantiles, respectively. For either quantile, the normal-normal model gives wider confidence intervals than the other models, and the sev-sev model gives narrower confidence intervals than the others. At the lower strain levels, there are similarities between the sev-sev and sev-normal models and between the normal-normal and normal-sev models. At the higher strain levels, however, there are similarities between the sev-sev and normal-sev and between normal-normal and sev-normal.

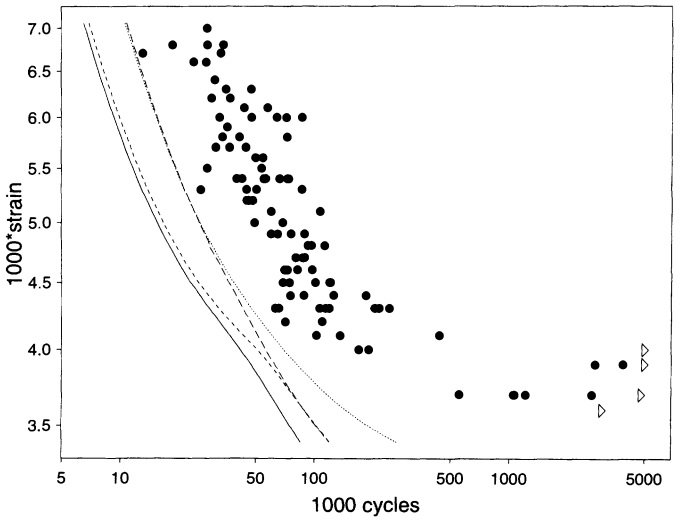

Figure 14. Lower 95% Confidence Lower Bounds for the .01 Quantile of Fatigue Life for the Nickel-Base Superalloy Data: ———, Sev-Sev; - - - -, Normal-Normal; – – –, Sev-Normal; — —, Normal-Sev.

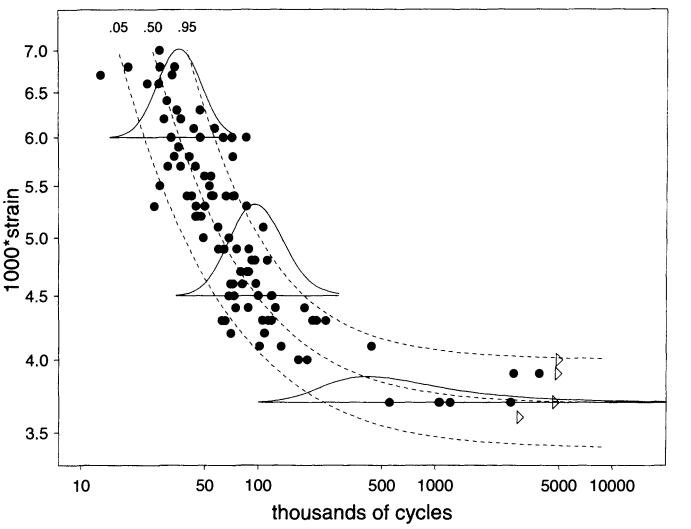

Figure 15. Log-Log S-N Plot for the Nickel-Base Superalloy Data With ML Estimates of Density Curves and the .05, .50, and .95 Quantile Estimates Under the Fitted Normal-Normal Model.

5.2 Residual Analysis

Figure 15 gives a plot of the data and the fitted normal-normal model. The plot includes density curves at selected strain levels. Figure 16 gives the corresponding plot of the residuals versus the strain levels. The residual plot does not show any clear patterns in the residuals. The normal-normal model appears to provide an adequate description of the life-strain relationship.

6. CONCLUSIONS AND AREAS FOR FURTHER RESEARCH

When fatigue limits exist, plots of fatigue life versus stress/strain often exhibit curvature at lower stress/strain levels. Moreover, in most fatigue experiments, the variance of fatigue life decreases as stress/strain increases and the standard deviation is often modeled as a monotonic function of stress/strain. Curvature in the fatigue life versus

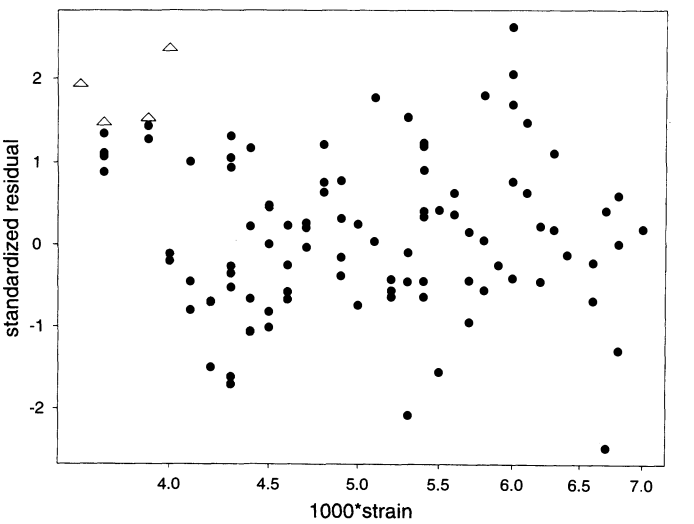

Figure 16. Plot of Standardized Residuals Versus Strain Levels for the Normal-Normal Model Fit to the Nickel-Base Superalloy Data.

stress/strain relationship can be modeled by including a constant fatigue-limit parameter in statistical models for fatigue life. Fixed fatigue-limit models, however, do not address the possible variability of the fatigue limit. This variability could be expected to be due to dependence of the fatigue limit on material structural properties that may vary from specimen to specimen. The random fatigue-limit model provides a description of the commonly observed increase in variability in log fatigue life at low levels of stress/strain and suggests a possible physical explanation for this behavior. These issues are the main motivations for the random fatigue-limit model. The examples considered here show that the random fatigue-limit model is able to address these issues adequately. The model has performed just as well for other datasets relating fatigue life to stress or strain.

Our examples compared the fits of the random fatigue-limit model under different distributional assumptions for both stress and strain fatigue data. In both cases, the normal-normal distribution combination provided the best fit. We compared the ML estimates of the .05, .50, and .95 quantiles and the lower confidence bounds for the .01 and .05 quantiles for different combinations of distributional assumptions. Similar computations and comparisons are possible if the experimenter is interested in other quantities.

There are several possible extensions that deserve to be explored further:

1. Our examples have used combinations of normal and sev distributions for $W|V$ and V. It would be useful to explore the use of other distributions, perhaps motivated from physical theory.

2. It would be useful to combine the analytical approach to testing goodness of fit in Section 4.3 with the P-P plots, giving confidence bands to help one assess the lack of fit of the model distribution. The approach of Nair (1984) could be adapted for this purpose, or corresponding simulation-based methods could be developed.

3. In the preceding examples, we use the AIC statistic to judge which distributional combination of the random fatigue-limit model best approximates the true model. In both examples, the AIC statistics agree with the information given by Q-Q, P-P, and S-P plots. The AIC may perhaps provide a viable method of comparing different random fatigue-limit models. Linhart (1988) presented a procedure to test whether or not two AIC's differ significantly. The test is based on the asymptotic distribution of the difference between the discrepancies of two AIC's from the true AIC. Using this method instead of simply comparing AIC values provides an objective criterion for comparing competing models.

4. There are important questions about how to design fatigue experiments under the random fatigue-limit model. Traditional methods will have to be extended to account for the nonlinear relationship between life and stress. Large-sample approximations would provide easy-to-compare evaluations of test plan properties with respect to the efficiency of estimating quantities of interest. Simulation studies require much more computer time but can be conducted to study the small-sample properties of the test plans. This is currently under investigation.

ACKNOWLEDGMENTS

We thank Wayne Nelson for suggesting the problem to us and making helpful suggestions on an earlier version of this article and Paul Wirsching of the University of Arizona for providing datasets and important references. Thanks to the editor, associate editor, and referees for valuable comments and suggestions. Computing for the research reported in this article was run using the equipment in the Department of Statistics, purchased with funds provided by an National Science Foundation SCREMS grant award DMS 9707740 to Iowa State University.

[Received August 1997. Revised February 1999.]

REFERENCES

Akaike, H. (1973), "Information Theory and an Extension of the Maximum Likelihood Principle," in *2nd International Symposium on Information Theory*, Budapest: Akademiai Kiado, pp. 267–281.

Colangelo, V. J., and Heiser, F. A. (1974), *Analysis of Metallurgical Failures*, New York: Wiley.

Collins, J. A. (1993), *Failure of Materials in Mechanical Design*, New York: Wiley.

Crowder, M., Kimber, A., Smith, R., and Sweeting, T. (1991), *Statistical Analysis of Reliability Data*, London: Chapman & Hall.

D'Agostino, R., and Stephens, M. (1986), *Goodness-of-Fit Techniques*, New York: Marcel Dekker.

Dieter, G. E. (1976), *Mechanical Metallurgy*, New York: McGraw-Hill.

Hirose, H. (1993), "Estimation of Threshold Stress in Accelerated Life-Testing," *IEEE Transactions on Reliability*, 42, 650–657.

Klesnil, M., and Lukáš, P. (1992), *Fatigue of Metallic Materials*, New York: Elsevier.

Little, R. E. (1974), "The Up-and-Down Method for Small Sample With Extreme Value Response Distributions," *Journal of the American Statistical Association*, 69, 803–806.

——— (1990), "Optimal Stress Amplitude Selection in Estimating Median Fatigue Limits Using Small Samples," *Journal of Testing and Evaluation*, 18, 115–122.

Michael, J. R. (1983), "The Stabilized Probability Plot," *Biometrika*, 70, 11–17.

Nair, V. N. (1984), "Confidence Bands for Survival Functions With Censored Data: A Comparative Study," *Technometrics*, 26, 265–275.

Nelson, W. (1973), "Analysis of Residuals from Censored Data," *Technometrics*, 15, 697–715.

——— (1984), "Fitting of Fatigue Curves With Nonconstant Standard Deviation to Data With Runouts," *Journal of Testing and Evaluation*, 12, 69–77.

——— (1990), *Accelerated Testing: Statistical Models, Test Plans, and Data Analyses*, New York: Wiley.

Ostrouchov, G., and Meeker, W. Q. (1988), "Accuracy of Approximate Confidence Bounds Computed From Interval Censored Weibull and Lognormal Data," *Journal of Statistical Computing and Simulations*, 29, 43–76.

Pascual, F. G., and Meeker, W. Q. (1997), "Analysis of Fatigue Data With Runouts Based on a Model With Nonconstant Standard Deviation and a Fatigue Limit Parameter," *Journal of Testing and Evaluation*, 25, 292–301.

Ross, J. S. (1990), *Nonlinear Estimation*, New York: Springer-Verlag.

Shen, C. L. (1994), "Statistical Analysis of Fatigue Data," unpublished Ph.D. dissertation, University of Arizona, Dept. of Aerospace and Mechanical Engineering.

Shen, C. L., Wirsching, P. H., and Cashman, G. T. (1996a), "An Advanced Comprehensive Model for the Statistical Analysis of S-N Fatigue Data,"

unpublished manuscript, University of Arizona, Dept. of Aerospace and Mechanical Engineering.

——— (1996b), "Design Curve to Characterize Fatigue Strength," *Journal of Engineering Materials and Technology*, 118, 535–541.

Shimokawa, T., and Hamaguchi, Y. (1987), "Statistical Evaluation of Fatigue Life and Fatigue Strength in Circular-Holed Notched Specimens of a Carbon Eight-Harness-Satin/Epoxy Laminate," in *Statistical Research on Fatigue and Fracture* (Current Japanese Materials Research, Vol. 2), eds. T. Tanaka, S. Nishijima, and M. Ichikawa, London: Elsevier, pp. 159–176.

Symonds, J. (1996), "Mechanical Properties of Materials," in *Marks' Standard Handbook for Mechanical Engineers*, eds. E. A. Avallone and T. Baumeister III, New York: McGraw-Hill, Chap. 5, pp. 2–15.

Vander Wiel, S. A., and Meeker, W. Q. (1990), "Accuracy of Approximate Confidence Bounds Using Censored Weibull Regression Data From Accelerated Life Tests," *IEEE Transactions on Reliability*, R-39, 346–351.

Wilk, M. B., and Gnanadesikan, R. (1968), "Probability Plotting Methods for the Analysis of Data," *Biometrika*, 55, 1–19.

Discussion

Charles ANNIS

United Technologies Pratt & Whitney
West Palm Beach, FL 33410-9600
(*annis@pwfl.com*)

Pascual and Meeker present an innovative approach to a longstanding difficulty in modeling fatigue curves—to wit, the ballooning variance as fatigue lifetime exceeds, say, 10^7 cycles. And they may have given us even more—an approach for parametric modeling of competing damage mechanisms that constitute fatigue.

First, some background. For over a century, before much of the physics of fatigue was understood, fatigue has been described using an S-N diagram, relating demand (stress or strain, S) and capability (cycles-to-failure, N). All engineering S-N curves use a logarithmic axis for cycles, and the *de*pendent variable, cycles, is plotted on the x axis. S-N curves are sigmoidal over the entire life range from monotonic tensile strength (plotted at 1/4 cycle) and stress at fatigue runout, if such a fatigue limit exists. A linear relationship between log s and log N is at best appropriate in the middle life region, but this is often the range of interest.

Because, ceteris paribus, it costs 10 times as much in time and money to run a test to 10^8 cycles as to 10^7 cycles, an endurance limit may well exist outside the range of the data that is prohibitively costly to pursue. As part of their model, Pascual and Meeker posit a distribution (with random location and scale parameters) for runout strength (or fatigue limit). But their formulation also suggests to me an expanded use in which the location is parametrically related to another of the fatigue-controlling parameters—stress ratio, for example. Before a more detailed discussion of their model, perhaps an overview of metal fatigue would be worthwhile.

Most engineering alloys are polycrystalline. (Some are not: Modern turbine blades are grown from a single crystal that can be six inches or more in length.) The grainsizes can be large, a quarter inch in cast parts, or small, .001 inch in fine-grain wrought superalloys. (For comparison, the diameter of a human hair is .003 inch.) The micromechanics of fatigue proceeds as an involved sequence of events that begins with microslip in shear between adjacent planes of atoms in the crystal structure of a single grain, and usually oriented 45° to the normal (perpendicular) load, and culminates with the propagation of a macrocrack many times larger than the grainsize and oriented approximately normal to the load. The behavior of the material in aggregate can be rather different from the behavior of an individual crystal. Still it is useful to describe the aggregate behavior in gross terms by relating some measure of durability like cycles-to-failure, with one or more of the durability-controlling parameters, such as applied stress or strain range. (Stress is the applied force normalized to a unit area; strain is elongation normalized to undeformed length.)

The material response is a function of more than just stress (or strain) range. Stress ratio, R (R = min stress/max stress), hold-times-at-load, and temperature (if isothermal) or thermal cycle also influence fatigue capability. [Frequency does not seem to have a great influence on S-N behavior (initiation), although it does influence crack propagation, at least for moderate values of ΔK in the so-called Region II of the crack growth rate $(da/dN, \Delta K)$ curve, but not so much for the growth at low ΔK.] The material's chemistry and processing history also play a role in initiating a crack, as does the local three-dimensional geometry of the fatigued part. (It can be argued that the three-dimensional loading and resulting geometry-influenced state of stress are outside the purview of materials modeling, however.) Finally, the relative influence of all of these factors changes throughout the fatigue process.

S-N models cannot deal with cracks. After a fatigue crack is formed, the S-N curve is no longer useful. [This is somewhat ironic because some fraction of the life of a fatigue specimen is composed of a propagating macrocrack, and the mechanics of crack propagation in low-cycle fatigue (LCF) is well understood.] Fracture mechanics (F/M) considers the stress *field* (not just an average stress) *and* its synergism with a material discontinuity (crack). [The stress intensity parameter, $K, \Delta K = \Delta\sigma\sqrt{\pi a} f$ (geometry), where $\Delta\sigma$ is stress range and a is cracklength, reflects this synergism.]

TECHNOMETRICS, NOVEMBER 1999, VOL. 41, NO. 4

For two decades F/M has been used with great success to describe the behavior of a potential crack and thus mitigate its threat. Because of the very large variability in time-to-crack, many component lifetimes are determined using the anticipated behavior of a propagating crack that is assumed to be present from cycle 1 (cf. Larsen, Schwartz, and Annis, 1980.)

If F/M is so wonderful, who cares about S-N curves? Some of the emphasis on F/M may change with the recognition that high-cycle fatigue (HCF; i.e., $N \gg 10^7$ cycles) is *not* LCF at a higher frequency (cf Annis 1999). The ability to describe the behavior of a crack may not be useful if the propagation time is measured in minutes, even if the cycle count is measured in millions. (At 20 KHz, you can accumulate more than a million cycles in less than a minute.) Furthermore, as Pascual and Meeker show, the variance of fatigue lifetime increases (or appears to increase) for longer and longer lives as a consequence of a random fatigue limit. In any event, the precision for predicting fatigue lives under HCF is quite poor and inadequate for component design. But we still may be able to predict the probability of *having* the HCF excitation. Thus, there is a potential shift in emphasis from estimating a runout stress under LCF to understanding the conditions under which HCF loading will produce failure. (Some HCF excitation cannot be avoided, but it can be mitigated—with damping, for example.)

The purpose of this digression is to call to the reader's attention that fatigue is significantly more complex than the conventional two-parameter S-N curve and to suggest that the authors' formulation may help describe the complexity of these physical processes and provide a tool useful where F/M is wanting.

My initial impression of Pascual and Meeker's model was that its publication would further a healthy debate over the appropriateness of the statistical techniques being discussed, but *I* would not have used it. (What an embarrassing, erroneous rush to judgment!) My preference then was to estimate a runout stress and its confidence interval. For example, the authors' Figure 10 suggests an endurance limit near 3.5 microstrain. Were I to estimate an endurance limit, I would consider a binary response such as "Failed in fewer than 500,000 cycles, yes/no?" and use a generalized linear model (GLM) with microstrain as the controlling covariate. Experience suggests an asymmetric link, such as complementary log-log, because the yes/no behavior can change more precipitously with respect to strain (or stress) on one side of the median than the other. A direct result of such a GLM is an immediate estimate of probability of failure (before some cycle count) as a function of some design parameter—microstrain here. There is another advantage to this approach: It obviates the discussion of heteroscedasticity and with it the problem of placing "lower bounds" on the S-N curve. Remember that it is the stress below which it is safe to design, rather than some cycle count, that is of primary interest.

Alas, I was overly hasty in dismissing the potential utility of Pascual and Meeker's model. In the 18 months that their article was being considered, the flight propulsion industry has generated some interesting S-N data at high cycle counts, some exceeding 10^9 cycles. For these high-cycle fatigue data, the behavior for $N > 10^7$ cycles strongly suggests the existence of a fatigue limit. It is amazing how your *own* data can refocus your perspective.

The authors graciously shared their software for estimating the model parameters, and preliminary analysis of the data produces "lower bound" S-N behavior significantly more plausible than that produced by more conventional regression techniques.

The software works with S-PLUS, and can be downloaded from Meeker's Web site, *http://www.public.iastate.edu/~stat533/slida.html*. After installation, the estimates are invoked from the command line—for example,

>hcf.fit<-get.rflm.mle(data.ld=hcf.ld, cond.dist="sev", fl. dist="normal")

>censored.data.plot(data.ld=hcf.ld, x.axis="log", y.axis ="log", response.on.yaxis=F)

>rflm.mle.plot(hcf.fit, response.on.yaxis=F)

There are other useful options in the *get.rflm.mle*( ) routine, and the resulting fit object contains almost everything useful, including the variance–covariance matrix of the parameter estimates.

So what? Pascual and Meeker are not just offering another curve-fit. Their formulation is based on physics, not just arithmetic, and may provide new insight into the interplay of parameters that primarily influence LCF ($N < 10^7$ cycles) and those influencing HCF ($N \gg 10^7$ cycles), or how the influence of a given parameter changes throughout cyclic lifetime. Stress ratio R, for example, may be seen to have a more pronounced influence on the fatigue limit, the authors' Equation (2), than on the LCF portion of the curve [their Eq. (1)].

The authors also conduct a clinic on thoroughness in assessing the utility of *any* fatigue model. Their findings on the likely form of the location, scale distribution of the log of the fatigue limit, and the log of the conditional cycles to failure distribution (ϕ_V and $\phi_{W|V}$) provide a foundation for future researchers. (Actually, their software is so easy to use that various combinations of distributional forms are readily investigated.) I am grateful to Pascual and Meeker for providing a significant step forward in modeling fatigue and am looking forward to using their model to help elucidate the complex behavior of HCF.

[Received August 1997. Revised February 1999.]

ADDITIONAL REFERENCES

Annis, C. (1999), "HCF, g-Functions, Probabilistic Engineering Analysis: A Checklist for Probabilists," unpublished paper presented at the 4th National Turbine Engine High Cycle Fatigue Conference, Monterey, CA, February 9–11.

Larsen, J. M., Schwartz, B. J., and Annis, C. G. (1980), "Cumulative Damage Fracture Mechanics Under Engine Spectra," Technical Report AFML-TR-79-4159, Air Force Materials Laboratory.

Discussion

Paul H. KVAM
School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, GA 30332-0205
(*pkvam@isye.gatech.edu*)

Eric P. KVAM
School of Materials Engineering
Purdue University
West Lafayette, IN 47907-1289
(*kvam@purdue.edu*)

1. EXISTENCE OF RANDOM FATIGUE LIMITS

The S-N curve (or Whöler curve) relates in a simple way the effect of a constant stress (S) on the test item to the number of cycles to failure (N). Although the S-N curve has been around for over 100 years, it carries important material-design information and is still used in engineering applications involving the lifetime estimation of machine parts. In tests for which the S-N curve has a random and unobservable fatigue limit, estimation of the fatigue limit curves provides the greatest challenge in the statistical analysis. Pascual and Meeker have provided a template for the practical analysis of fatigue life in material tests in which the fatigue limit is known to vary.

As the authors note, the relationship between stress and cycles until failure is highly nonlinear. Analytical representations might have the form $N = cS^{-d}$, where (c, d) are positive material parameters. With a random fatigue limit, an unknown stress threshold (S_0) exists, below which no material damage occurs, and the material assumes an infinite life; that is, $N(S) = c(S - S_0)^{-d}$, where $S > S_0$, and $N(S) = \infty$ when $S \leq S_0$. The authors have correctly emphasized the importance of this threshold because it represents an asymptote in the curve estimate. Such singularities are known to greatly influence the estimation of curve parameters, and most often curve estimation is relegated to nonlinear regression methods.

The existence of such a threshold is a point of concern in all materials testing and worthy of examination. It is likely that every material has a fatigue limit (although in some nonengineering materials it may be ridiculously small). Crack propagation requires breakage of atomic bonds, which cannot happen unless their relative displacement is sufficiently large. In fatigue, atomic displacements become magnified due to local *plastic deformation*—that is, unrecoverable rearrangements of the atomic structure at or near the crack tip. Below some stress level, this plastic deformation will no longer occur, and the crack will reach a steady state. Thus, the assumption of a nonzero fatigue limit is justified. With iron and titanium alloys, for example, dissolved gas atoms become bound to the *dislocations*, which are defects whose motion causes plastic deformation. The stress required to move the dislocations is markedly increased because this binding energy must be overcome before any dislocation motion can occur. This effectively interrupts the progression of the crack. High fatigue limits can thus be understood for several particularly important metal alloy systems.

The dependence of fatigue life on many of the parameters cited by Pascual and Meeker are also related to the phenomenon of dislocation motion. Crystal size, composition, orientation, and heat treatment will all have an impact on dislocation mobility, as will other parameters. (One not mentioned is the R value, the ratio of minimum to maximum applied stress level.)

The theory of plastic deformation makes known the fatigue limit to certain metal alloys under prespecified conditions, geometries, and stresses. This suggests that the asymptote in the S-N curve can be identified in some cases, or at least partially bounded, and the random fatigue-limit model is not required to model the metal's strength. In other cases, such as testing fatigue strength of composites (e.g., laminate panel), random fatigue limits make sense. In such materials, the combination of viscoelastic and brittle behaviors makes them a requirement.

If the fatigue limit is unknown, models to guide the relationship between S and N are unrefined. In contrast to Castillo and Hadi (1995), whose mathematical models were derived based on physical principles of test material, Pascual and Meeker consider an empirical approach to modeling fatigue life and emphasize tests for model fit rather than leaning on assumptions based on material properties. The empirical approach, it has been shown, is a logical course for fatigue tests of modern composite materials for which physical properties fail to suggest more refined models.

2. ALTERNATIVES TO THE S-N MODEL

The S-N curve is part of the *safe-life design* approach to materials testing. Although still widely used in some practices, it is rarely applied by the aircraft industry, which made the S-N curve application popular initially. The safe-life design approach fails to distinguish the crack-initiation phase of the test from the crack-propagation phase; only the total time until a failure event is considered. This *infinite-life design* is not economically feasible in aircraft design, due to the very low fatigue limits of aluminum alloys. Current practice is to use *damage-tolerant* design for air frames. In general, the crack-initiation phase takes longer and has much greater relative variability. Each S-N curve test begins with an uncracked item, and the time to a crack initiation is often a strong function of laboratory conditions

TECHNOMETRICS, NOVEMBER 1999, VOL. 41, NO. 4

and other uncontrolled covariates, making this phase highly unpredictable.

Alternatively, damage-tolerant design involves two major aspects, one of which is the inclusion of crack-arrest structures to halt the catastrophic growth of a large crack. The other aspect is the use of fracture mechanics, based on the assumption that cracks already exist in a structure. Knowing the maximum possible size for these existing cracks, a conservative prediction of their growth rates may be made. Crack growth rates are deterministic more than stochastic and are tabulated in da/dN versus ΔK plots, where a is the crack length, N is the number of cycles, and ΔK is a measure of stress-intensity variation dependent on the crack length, the applied stress-level differences, and known geometrical factors. A typical curve is sketched in Figure 1.

At very low ΔK, the stress reversals that cause the microscopic material changes are not strong enough, and the crack is halted, as it would be in static loading. At the very highest ΔK, the material is put through severe plastic distortions, and the crack reaches critical size (static-load propagation) after only a few cycles. Most loading of interest occurs in the near-linear central region (often referred to as Paris-law crack growth). For a known material and loading stress/frequency, the crack growth can be predicted, and the maintenance, repair, and inspection schedules can be established reliably. In S-N curve measurement, the material preparation can be crucial. It is possible to sharply limit existing surface flaws, except for a few microstructural effects. In the da/dN versus ΔK, on the other hand, a crack of known size is fabricated prior to beginning the actual experiment. The worst S-N results (fastest failures) represent an asymptote to the da/dN-ΔK behavior.

Another limiting property of the S-N curve is its inability to model strength when stresses change. Stress amplitude varies during the lifetime of most materials, especially in applications outside the laboratory such as turbulence and take-off/landing events for airframes. The S-N curve assumes a constant stress amplitude on the test item, which narrows its area of application significantly.

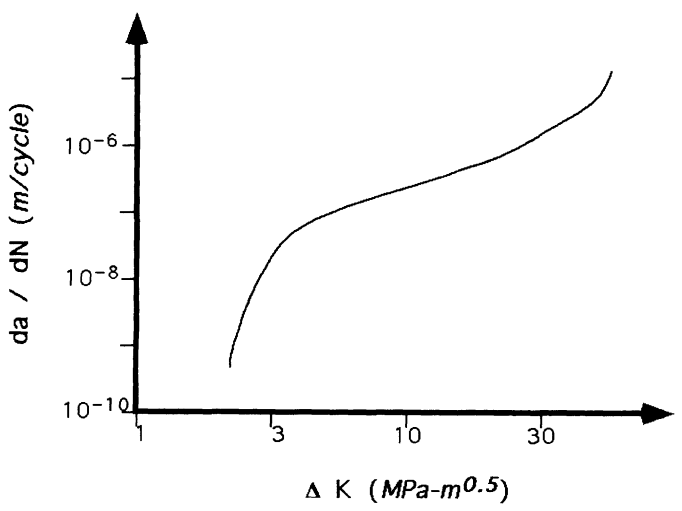

Figure 1. Damage-Tolerant Design Graph of da/dN Versus ΔK.

There are applications for which the S-N curve makes practical sense, of course. Ceramic materials generally have ionic or covalent bonds (i.e., no free electrons), which makes them intrinsically stronger than metals. Ceramics tend to fracture rather than bend, however, when stresses are applied to them. In this case, the number of cycles until crack nucleation is reached means more to the researcher than the phase of crack growth, which might be nearly instantaneous.

Finally, damage-tolerant design models for fatigue crack growth are more easily derived using properties of the material and basic rules of cumulative damage. Models are numerous in the literature; the works of Goto (1998) and Zhao, Gao, and Sun (1998) contain many recent references. In summary, many fatigue-critical applications are moving to fracture mechanics rather than S-N testing. At the same time, new and more appropriate applications for S-N and S-N-P testing may be appearing.

3. EXAMINING S-N MODEL FIT

Unlike similar fatigue models found in the literature, Pascual and Meeker's analysis emphasizes the importance of assessing the selected model's fit with the data. Probability plots, Nelson's (1973) graphical residual analysis, and the more analytic goodness-of-fit statistics are not applied enough in engineering practice involving fatigue testing. The probability plots and goodness-of-fit tests are intelligibly and succinctly explained by the authors, which should encourage model verification in future applications. These methods also serve as tools for model selection. None of the tests follow perfectly the assumptions needed to guarantee their validity, but violations are centered on the correlation of residuals, which should be mitigated somewhat with large samples.

In addition to problems with correlation, one can argue that probability plots can potentially mislead the practitioner due to their subjective nature. Statisticians often regard the practice of interpreting such plots with suspicion because it is well understood that one can sometimes take repeated logarithms of the plotted variables until we observe a quasi-linear relationship between them. The original S-N curve is already based on logarithmic transformations, which makes the P-P plot more suspect in such applications. Still, the plots are an excellent device to detect misfit models.

One could argue at length about the choices of distributions to characterize the variation in the fatigue limit and the dependence on the stress level. Different failure models have appeared in the engineering and materials-science literature using basic assumptions of the test material and the environment. For example, Nakazawa and Kodama (1987) stated that the Japan Society of Mechanical Engineers adopted a simple log-linear model to describe the relationship between S and N and dealt with the fatigue limit separately. In sharp contrast, Castillo and Hadi (1995) used various physical principles to derive a generalized reversed Pareto distribution to characterize the S-N curve.

The authors showed that, empirically, the model fit is not overly sensitive to the choice of distributions among sensible alternatives. It is not clear, however, that this is always true at the extreme percentiles of the distributions. Important to the practitioner, in any case, is that the parameters keep some operational meaning. The models chosen by the authors accomplish this. All of the competing models have numerous parameters, leading to difficulties in the maximum likelihood procedure. Castillo and Hadi (1995) considered this a serious flaw in the estimation of the S-N curve and derived simpler regression-type estimation alternatives.

From Table 3, we might infer that the extreme correlation of the maximum likelihood estimates (MLE's) might be due to overparameterization in the fatigue-limit model. Although the authors carefully address the problems associated with such correlations, it made the task of computing MLE's a difficult one for us and makes awkward any discussion of the joint distribution of the model-parameter estimates. One consequence of the model's lack of fit can be seen in the S-N plots (Figs. 1 and 10) for both examples, along with the plots of the residuals (Figs. 9 and 16).

For the laminate panel data, the model seems to fit the data better at the higher stress levels (280 MPa and above) compared to the lowest stress level (270 MPa). This cannot be reflected in the overall tests of model fit. Although it is not stated in the article (or the corresponding references), we might assume that the fit of the model is important at low (regular-use) stresses. This assumption is analogous to those made in accelerated life tests, for example.

Overall, the guidelines laid out by Pascual and Meeker are thorough, succinct, and carefully explained. The article entices statisticians with an interesting problem—estimating a random fatigue limit in an S-N curve. It also invites material scientists and engineers to apply a sensible algorithm to characterize and verify the S-N curve. We enjoyed reading the article and found the examples especially motivating.

ADDITIONAL REFERENCES

Castillo, E., and Hadi, A. S. (1995), "Modeling Lifetime Data With Application to Fatigue Models," *Journal of the American Statistical Association*, 90, 1041–1054.

Goto, M. (1998), "Statistical Investigation of the Effect of Laboratory Air on the Fatigue Behavior of a Carbon Steel," *Fatigue & Fracture of Engineering Materials & Structures*, 21, 705–715.

Nakazawa, H., and Kodama, S. (1987), "Statistical S-N Testing Method With 14 Specimens: JSME Standard Method for Determination of S-N Curves," in *Statistical Research on Fatigue and Fracture*, eds. T. Tanaka, S. Nishijima, and M. Ichikawa, London: Elsevier Applied Science, pp. 59–69.

Zhao, Y.-X., Gao, Q., and Sun, X.-F. (1998), "A Statistical Investigation of the Fatigue Lives of Q235 Steel-welded Joints," *Fatigue & Fracture of Engineering Materials & Structures*, 21, 781–790.

Discussion

Valen E. JOHNSON
Institute of Statistics and
Decision Sciences
Duke University
Durham, NC 27708-0251
(*valen@isds.duke.edu*)

Mark FITZGERALD
Group TSA-1 Statistical Sciences
Los Alamos National Laboratory
Los Alamos, NM 87545
(*markfitz@lanl.gov*)

Harry F. MARTZ
Group TSA-1 Statistical Sciences
Los Alamos National Laboratory
Los Alamos, NM 87545
(*hfm@lanl.gov*)

In an extremely readable article, Pascual and Meeker (PM) have developed a new random fatigue-limit model for use in analyzing fatigue data. The authors use a random fatigue limit to model the curvature that often exists in fatigue curves as well as to capture the inverse relationship between the standard deviation of fatigue life and stress. Using maximum likelihood (ML) methods, they fit the model to two different datasets and thoroughly examine the fits using a variety of diagnostic probability and residual plots as well as goodness-of-fit techniques. These two examples clearly illustrate the potential widespread application of the model, and we compliment the authors for an interesting and useful article.

Because the fatigue limit associated with each specimen is considered to be a random effect, it occurred to us that hierarchical Bayes (HB) might represent a useful alternative way to model and analyze such data. We now describe this approach and present the results for the same laminate-panel fatigue data considered by PM. We also present some important advantages and disadvantages of this alternative approach in Section 2.

TECHNOMETRICS, NOVEMBER 1999, VOL. 41, NO. 4

1. AN ALTERNATIVE HIERARCHICAL BAYES APPROACH

Using the PM notation, a directed graph for the random fatigue-limit model is shown in Figure 1. A solid arrow indicates a stochastic dependency, and a dashed arrow indicates a logical function. The two undirected dashed lines represent logical range constraints resulting from the censored data and the restriction that $\gamma < s$ for uncensored observations.

For illustration, we consider the laminate-panel fatigue data in conjunction with the normal-normal model. Our HB approach proceeds as follows. We first specify prior distributions for the five parameters outside the "specimen (observation) i" box in Figure 1. Inference on these five unknown parameters requires that we obtain samples from the joint posterior $\pi(\beta_0, \beta_1, \sigma^2, \mu_\gamma, \sigma_\gamma^2 | \mathbf{Y})$, which of course also depends on the observed stresses s as well as knowledge of which observations have been censored (denoted by **Y.cen**). We then use empirical summary statistics from these samples to draw the desired inferences. To calculate samples from this joint posterior, we must successively sample from the full conditional distributions; that is, we must successively sample from the conditional distribution of each stochastic node given all the others in the graph.

Consider the joint prior distribution $\pi(\beta_0, \beta_1, \sigma^2, \mu_\gamma, \sigma_\gamma^2)$. Our initial calculations suggest that, for the "standard" noninformative prior $\pi(\beta_0, \beta_1, \sigma^2, \mu_\gamma, \sigma_\gamma^2) \propto 1/(\sigma^2 \sigma_\gamma^2)$, the posterior is, in fact, unbounded if $v_i = \mu_\gamma$ for all i and $\sigma_\gamma \to 0$. In this case, the posterior distributions on the model parameters are highly irregular as well. A similar situation also occurs when $\sigma = 0$ and all observations are predicted exactly by inclusion of the random effects γ_i. In both of these cases we suspect that this posterior is improper and places infinite mass at the modes mentioned previously. Similarly, certain other relationships between the parameters may also lead to an improper posterior distribution. A similar problem exists for the PM likelihood approach—the likelihood function has unbounded modes at $\sigma^2 = 0$ (when the random fatigue limits are permitted to provide an exact fit to the data) and $\sigma_\gamma^2 = 0$.

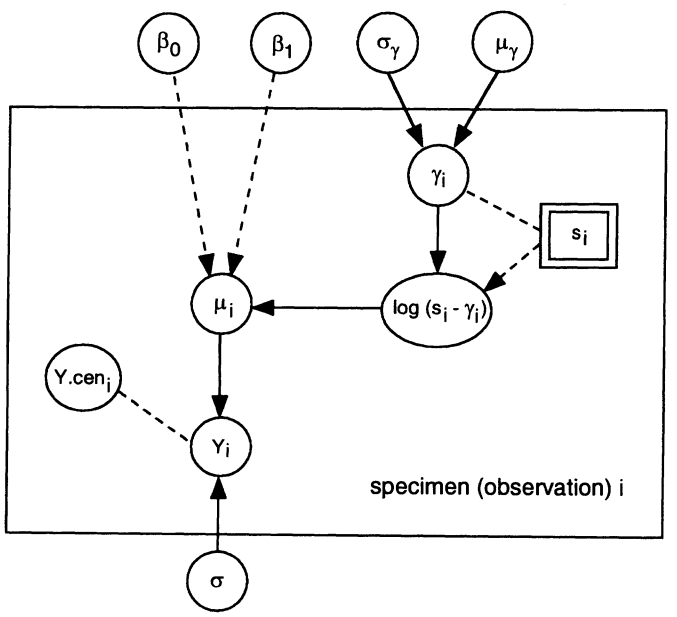

Figure 1. Directed Graph for the Hierarchical Bayes Analysis of the Random Fatigue-Limit Model.

Table 1. Hierarchical Bayes Laminate Model Parameter Estimates

Parameter	Mean	Median	Standard error	95% interval
β_0	28.201	27.979	2.610	(23.701, 34.438)
β_1	−4.723	−4.686	.472	(−5.833, −3.898)
σ	.467	.473	.052	(.341, .556)
μ_γ	5.396	5.400	.039	(5.301, 5.460)
σ_γ	.011	.009	.006	(.004, .026)

To ensure a proper posterior distribution, we consider the following informative prior distribution. The marginal prior on μ_γ was taken to be $\mathrm{N}(\mu_0, \sigma_0)$. Based on the stress levels chosen in the experiment, we empirically choose $\mu_0 = \log(250) = 5.52$. We have also chosen the prior standard deviation to be approximately 50 on the original scale, which implies that $\sigma_0 = .5[\log(300) - \log(200)] = .20$.

The informative marginal prior on σ_γ^2 is taken to be an inverse gamma $\mathrm{IG}(\alpha_0, \lambda_0)$ distribution. Examination of the functional equation producing the mean fatigue life μ indicates that random errors in the fatigue limit affect the predictions of mean life by a factor of

$$\exp\{\beta_1 \log[s - \exp(v)]\} = [s - \exp(\mu_\gamma + \text{error})]^{\beta_1}.$$

For β_1 near $-5, s = 300$ and $\mu_\gamma = \log(250)$, which is its prior mean, the change in mean fatigue life associated with a random error of .01 is

$$[300 - 250\exp(.01)]^{-5}/(300 - 250)^{-5},$$

or approximately 23%. Assuming that such changes in mean represent a two-standard-deviation change beyond that from a zero random effect leads to our assumption that the prior standard deviation on the random fatigue limit should be about .005, thus implying a prior variance of the order of magnitude of .000025.

Taking $\alpha_0 = 3$ ensures that the IG prior has a finite variance. Furthermore, setting $\lambda_0 = .0001$ implies a prior mode for σ_γ^2 of $\lambda_0/(\alpha_0 + 1) = .000025$, a prior mean of $\lambda_0/(\alpha_0 - 1) = .00005$, and a prior standard deviation of $\lambda_0/[(\alpha_0 - 1)(\alpha_0 - 2)^{.5}] = .00005$. Because this distribution is relatively heavy-tailed, it places substantial mass around values of σ_γ^2 beyond the mean. At the same time, the small positive value of λ_0 eliminates a singularity in the posterior distribution that occurs in the likelihood when all random effects are identically 0.

The marginal prior on σ^2 was likewise assumed to be an informative $\mathrm{IG}(\xi_0, \eta_0)$ distribution. An analysis of vari-

Table 2. Correlations Between the Hierarchical Bayes Model Parameters for the Laminate Data

	β_0	β_1	σ	μ_γ	σ_γ
β_0	1.000	−.999	.107	−.983	−.119
β_1		1.000	−.103	.974	.114
σ			1.000	−.113	−.724
μ_γ				1.000	.131
σ_γ					1.000

ance of the logarithm of fatigue life was performed treating stress as a factor variable and excluding censored observations. This resulted in a preliminary estimate of σ^2, unadjusted for random effects, of $(.45)^2$ with 110 df. Based on this analysis, we took $\xi_0 = 1$ and $\eta_0 = .405$. Although this IG prior has neither a mean nor a variance, it has a mode at the desired value of $(.45)^2$.

Finally, marginal uniform priors were assumed for β_0 and β_1, the regression parameters. Although the preceding choice of priors is rather arbitrary, this choice bounds the marginal posteriors on σ_γ^2 and σ^2 away from 0.

To simulate samples from the corresponding posterior distribution, we used a modified Gibbs/Metropolis–Hastings Markov-chain Monte Carlo sampling technique (Gilks, Richardson, and Spiegelhalter 1996). After 5,000 burn-in iterations, 20,000 simulated parameter values for these five parameters were recorded every 50 iterations in a chain run for 10^6 iterations. Total computational time for this simulation required less than 30 minutes on a midpriced Unix workstation.

Table 1 contains the posterior means, medians, standard deviations, and 95% credible intervals for the five parameters. By comparing these estimates with those in Tables 1 and 2 of PM we see that the HB posterior mean estimate of σ is 50% larger than PM's ML estimate, but the HB mean estimate of σ_γ is 50% smaller than the ML estimate. The HB 95% credible intervals for both σ and σ_γ are significantly shorter than the corresponding likelihood ratio confidence intervals, and the standard errors of the HB estimates of β_0, β_1, and μ_γ are also smaller than those of PM. This may be due to the use of large-sample asymptotic results by PM and/or the possible effect of the informative prior.

Table 2 gives the correlations between the HB model parameters for the laminate data. The extremely high correlations between β_0, β_1, and μ_γ compare favorably with those in Table 3 of PM, but the remaining correlations are all smaller (in absolute value) than those reported by PM.

Figure 2 displays fitted values under the HB model. The centermost line in this figure is an estimate of the posterior median fatigue life as a function of the stress level. The two lines immediately surrounding the median estimate represent a 90% posterior credible interval for the mean fatigue life as a function of stress. The most extreme lines provide a 90% posterior predictive region for future fatigue-life observations as a function of stress level. These lines were all easily obtained by recording appropriate sample data on fatigue life and mean fatigue life while executing the Markov-chain simulation.

2. CONCLUSIONS

There are several advantages in using an HB approach for modeling and analyzing random fatigue-limit data. The primary advantage is the ability to use prior information in the analysis beyond that contained in the sample data. New materials, whose fatigue-life properties are of interest, are often evolutionary developments of existing materials whose mechanical properties are well known. These existing properties can provide the basis for informative prior distributions.

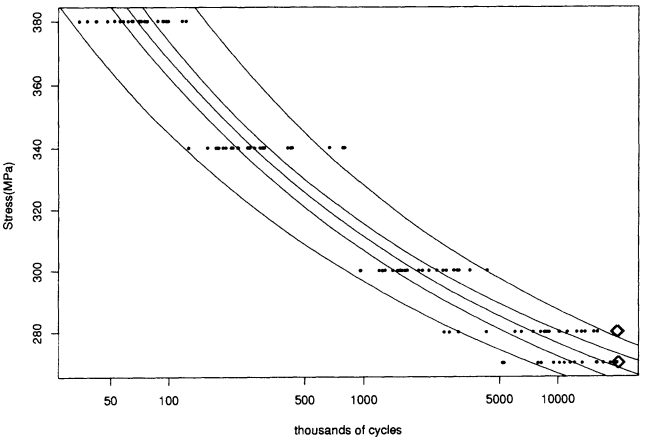

Figure 2. Log-Linear S-N Plot for the Laminate Panel Data with Hierarchical Bayes Estimates of the Median Fatigue Life, 90% Posterior Credible Interval on the Mean Fatigue Life, and 90% Prediction Limits for Future Fatigue Life Observations.

Furthermore, the posterior distribution provides great flexibility with which posterior inferences can be summarized. For example, the 90% posterior predictive limits in Figure 2 are easily obtained as a simple function of the simulated data. This simple, straightforward approach is in contrast to the more complicated methods that must often be used when computing lower confidence bounds on fatigue-life distribution quantiles. In addition, posterior credible intervals are direct probability statements about a parameter, but confidence intervals do not have this desirable interpretation. We also note here that the posterior predictive limits in Figure 2 were nearly invariant to several different choices of informative and noninformative marginal prior distributions on σ^2 that we considered.

Because all the information regarding the parameters is contained in the posterior distribution, there is no need to resort to asymptotic approximations for calculating such quantities as standard errors and interval estimates. In the HB approach these are readily empirically estimated from the simulated posterior data to any desired degree of precision. There is also no need for numerical integration of the likelihood components as required by PM.

Although not illustrated here, there is another important advantage of an HB approach. Model validation and checking are easily accomplished within the HB paradigm. Gelman, Carlin, Stern, and Rubin (1995, chap. 6) illustrated various ways to accomplish this validation.

The primary disadvantage of the HB analysis of the random fatigue-limit model is that a suitable prior distribution that guarantees a proper posterior must be identified. As we have observed, this is sometimes not a trivial task.

ADDITIONAL REFERENCES

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995), *Bayesian Data Analysis*, London: Chapman & Hall.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996), *Markov Chain Monte Carlo in Practice*, London: Chapman & Hall.

Response

We thank the discussants for their interesting and useful comments and their compliments to our article. We appreciate the opportunity to extend and amplify some of these and to offer our own perspective on some of the new issues that have been raised.

1. RESPONSE TO ANNIS

1.1 Fracture Mechanics

Annis has provided a clear and concise description of the science of fracture mechanics. Fracture mechanics has been one of the important success stories of physical modeling. Fracture mechanics methods have been used successfully in a wide variety of important, safety-critical design/reliability applications, from jet engines to buildings and bridges. It is often said that reliability predictions would be much more accurate if more failure mechanisms were understood as well as fatigue/fracture. We would like to mention that, despite the fact that the crack-growth phenomena are often (to a very good approximation) viewed as deterministic, there are plenty of interesting stochastic and statistical problems in this area, especially for predicting life in variable environments. In this regard, Bogdanoff and Kozin (1985), Sobczyk and Spencer (1992), and Sobczyk (1993) are useful references.

We agree with Annis's view about the relative importance of fracture mechanics as a model for components in situations in which a substantial part of life is in the crack-growth phase (i.e., after a detectable initiation). Jet-engine components made of ductile materials like titanium alloys are good examples. In other applications, particularly when inspection is not economically feasible, larger strength/weight ratios can be obtained with materials that are more brittle (e.g., hardened steel components used in certain automobile components). S-N curves are still widely used in such applications, as well as in applications in which components like springs and mounts are being tested directly (as opposed to the testing of material specimens to obtain estimates of materials properties).

1.2 Binary Responses

Annis suggests the use of a general linear model (GLM) and binary response to estimate a probability of failure versus stress curve, for a specified number of cycles. It is always possible to take a continuous-response model and replace it with a binary-response model, but in more complicated models it may not be possible to estimate all of the parameters in the original model without imposing some restrictions. Although some such models do not fall within the GLM family, maximum likelihood (ML) can always be used if the parameters are estimable. With one of the standard link functions and data having a mix of failures and nonfailures at the specified number of cycles at two or more levels of stress, estimability should not be a problem within the GLM family of binary-response models.

We encourage the use of alternative, sensible models, especially if they require fewer assumptions. Comparing the results of different models is an important part of the sensitivity analysis that is so important in practical applications. The binary response model certainly falls into this category. The GLM approach requires the specification of the link function, and it may not be possible, without much larger samples, to distinguish effectively among the alternatives. Sensitivity analysis among different plausible (i.e., seems to fit the data well) link functions should be used, as we have done in our modeling.

One disadvantage of this approach is that it reduces the observed response (cycles to failure) to a binary response. Thus, we expect precision to be sacrificed. It would be interesting to investigate the amount of precision lost as a function of the particular inference being made and the experimental design procedure being used. Moreover, if more than one number-of-cycles level is of interest, the more comprehensive random fatigue-limit (RFL) model might have more appeal (provided that it provides a useful description of the data).

Finally, if one contemplates the use of a binary response, the design of the experiment would likely be different. In particular, assuming that the model is adequate over some range of stresses, taking data at the extremes of the explanatory variable generally provides maximum information in a continuous-response experiment, whereas in a binary response model, testing at extremes where either all fail or all survive provides relatively little information for estimation. See Meeker and Hahn (1977) for information of design of such experiments.

1.3 The SLIDA Software

We thank Annis for mentioning the SLIDA software. We should add that SLIDA is a large collection of S-PLUS functions for reliability data analysis, designed as a complement for and to provide access to the methods used by Meeker and Escobar (1998). At present it is available only for S-PLUS 4.5 running in the Windows environment. Probability plotting, ML estimation, and graphical presentation of model estimates (often along with data) are key features. Much of the functionality is currently available through the extendable S-PLUS graphical user interface (GUI) so that users do not have to remember commands. By the time this article is in print, the RFL model can also be fitted from

TECHNOMETRICS, NOVEMBER 1999, VOL. 41, NO. 4

within the GUI without having to use commands like those shown in Annis's discussion.

2. RESPONSE TO KVAM AND KVAM

2.1 Background

Kvam and Kvam (K&K) have nicely filled in some gaps in our brief introduction to S-N modeling and the philosophy of fatigue limits, complementing the material given by Annis. Their lucid description of the underlying mechanisms of crack growth is most helpful.

2.2 The Castillo and Hadi Model

We thank K&K for bringing to our attention the important article by Castillo and Hadi (C&H) (1995). The C&H model is based on a more extensive set of physical assumptions and less of the empirical flexibility that we allow. If the C&H model fits one's data well, then it might be better able to predict outside the range of the data (which is often necessary, but which was not a purpose of our analysis). C&H illustrated and compared their model with other models using data on concrete failure.

We thought that it would be of interest to fit our model to the concrete-failure data and compare with the C&H results. The data consisted of five stress levels with 15 measurements each. Table 1 gives the RFL model-parameter ML estimates, values of the log-likelihood maxima, and the corresponding Akaike information criterion (AIC) statistics for the concrete data. Based on the log-likelihood (or, equivalently, the AIC), the sev-normal and normal-normal RFL model best describes the data.

Quantile curves under each RFL model fit are given in Figure 1. The sev-sev RFL model gives conservative (lower) estimates of the .05 quantiles, and all the models give similar estimates of upper quantiles. The quantile estimates for the best fits, sev-normal and normal-normal RFL models, almost coincide.

Castillo and Hadi (1995) listed five models proposed in the literature and compared them using the concrete-fatigue data. For this dataset, let x_j be the jth stress level, $j = 1, \ldots, 5$, and $y_{ij}, i = 1, \ldots, 15$, be the ith smallest observation (order statistic) at stress x_j. (Note that C&H used x to denote fatigue life and y to denote applied stress.) Table 2 gives equations for quantile estimates and the number of estimated parameters for each model. Here, $\hat{y}_{ij}$ is the estimate of the $p_i = (i - .5)/15$ quantile and z_{p_i} is the p_i quantile of the standard normal distribution. The table also includes the quantile estimate under the RFL model (obtained by inverting the marginal cdf).

Table 1. ML Results for the Concrete-Fatigue Data

		RFL Model			
		Sev-sev	*Normal-normal*	*Sev-normal*	*Normal-sev*
Log-likelihood	log[$L(\theta)$]	−75.583	−72.883	−72.878	−74.746
AIC statistic		161.166	155.766	155.757	159.491
Parameters	β_0	−8.893	−9.370	−9.328	−8.971
	β_1	−7.495	−8.346	−8.562	−7.359
	σ	.249	.295	.184	.359
	μ_γ	−.573	−.634	−.647	−.560
	σ_γ	.029	.033	.036	.026

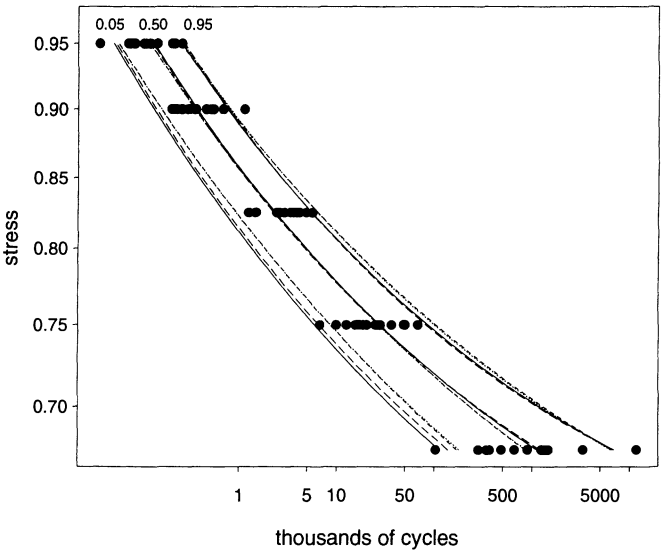

Figure 1. Log-Log S-N Plot for the Concrete-Fatigue Data With ML Estimates of the .05, .50, and .95 Quantiles (• Failure, ▷ Censored Observation): ——, Sev-Sev; - - - -, Normal-Normal; – – –, Sev-Normal, — — —, Normal-Sev.

To compare fits of different models to the data, C&H used the absolute prediction error criterion

$$E = \sum_{j=1}^{5} \sum_{i=1}^{15} |\log(y_{ij}) - \log(\hat{y}_{ij})|.$$

We computed the prediction error under each RFL model fit to the data. Table 3 gives the values of E for the five models in Table 2, C&H's model, and the RFL model. The first five prediction error values in Table 3 are different from those in table 7 of C&H because, in the latter article, $\hat{y}_{ij}$ had been taken to be the estimated mean. In recent correspondence with Castillo and Hadi, we concurred that it would be more appropriate to compare sample and predicted-order statistics for all of the models being compared. Castillo and Hadi generously provided us the new prediction error values based on this criterion for the first five models in Table 3.

With respect to the prediction-error criterion, the RFL model under any distribution combination performs better than the first six models. In particular, the normal-normal and sev-normal combinations give the best E values. These two combinations also yield similar likelihood maxima with sev-normal being slightly higher.

2.3 Calibrating Probability Plots

Perhaps some statisticians regard probability plots with suspicion, as suggested by K&K. We do not! Tools are needed, however, to deal with the subjectivity of graphical analysis so that analysts can effectively separate signal from noise. Although we did not use them in our article, the tools that we favor for helping to interpret probability plots are simultaneous nonparametric confidence bands (generally obtained by inverting an appropriate goodness-of-fit

Table 2. Models for the Relationship Between Applied Stress and Lifetime

Model	*Quantile estimate*	*Number of parameters*
Little and Ekvall (1981)	$\log(\hat{y}_{ij}) = \hat{A} + \hat{B}x_j + z_{p_i}\hat{\sigma}$	3
Little and Ekvall (1981)	$\log(\hat{y}_{ij}) = \hat{A} + \hat{B}\log(x_j) + z_{p_i}\hat{\sigma}$	3
Spindel and Haibach (1981)	$\log(\hat{y}_{ij}) = \log(\hat{Y}_0) + \hat{A}\log\left(\frac{x_j}{\hat{X}_0}\right) + \hat{B}\left\{\log\left(\frac{x_j}{\hat{X}_0}\right) + \frac{1}{\hat{\alpha}}\log\left[1+\left(\frac{x_j}{\hat{X}_0}\right)^{-2\hat{\alpha}}\right]\right\} + z_{p_i}\hat{\sigma}$	6
Bastenaire (1972)	$\log(\hat{y}_{ij}) = \hat{Y}_0 + \hat{A}\frac{\exp[-\hat{C}(x_j-\hat{X}_0)]}{x_j-\hat{X}_0} + z_{p_i}\hat{\sigma}$	5
Castillo et al. (1985)	$\log(\hat{y}_{ij}) = \log(\hat{Y}_0) + \frac{\hat{A}}{\log(x_j)-\log(\hat{X}_0)} + z_{p_i}\hat{\sigma}$	4
Castillo and Hadi (1995)	$\log(\hat{y}_{ij}) = \frac{p_i^{\hat{\alpha}}-\hat{D}-\hat{C}x_j}{\hat{A}x_j+\hat{B}}$	5
RFL model (1999)	$\log(\hat{y}_{ij}) = F_W^{-1}(p_i; \log(x_j), \hat{\theta})$	5

test) and simulation. These tools were described and illustrated in chapters 3 and 6 of Meeker and Escobar (1998).

2.4 Parameterization and Parameter Interpretation

In our opinion, it is convenient, but not important, for parameters of a model to have operational meaning. Sometimes, either tradition or numerical stability will be the determining factors in parameterization choice. What we must be able to do easily is to compute estimates and confidence intervals for the functions of the parameters that are of primary or other operational interest. For a simple example, when fitting a Weibull distribution to data with 1% or 2% failing where interest centers on the lower tail of the distribution, most software report the traditional Weibull shape and characteristic life (approximately the .63 quantile) parameters, even though the .63 quantile would be of little operational value. Well-designed software will, however, allow computation of lower tail-quantile or failure-probability estimates and confidence intervals.

The parameters, as used in our model, follow from tradition but do have useful interpretations. Other functions of these parameters, like quantiles at particular levels of stress, are, however, also of interest and estimates of and confidence intervals for these quantities can be computed with our software.

2.5 Correlations Between Parameters

K&K mention the (estimated) correlations among the parameters estimates. When these correlations are large, the resulting problems are similar to those encountered with multicollinearity in regression analysis. Suppose that, as in our model, the parameters have some operational meaning or interpretation. As with multicollinearity, high correlations suggest that certain parameters cannot be estimated independently of others, limiting interpretability. It is important to recognize such limitations. This behavior can be visualized, for pairwise correlations, by using the profile contours like those in Figures 3 and 4 of our article. As with multicollinearity, high correlations are a problem resulting from limitations in the data and not the model.

Table 3. Comparison of Models Using Absolute Prediction Error

Reference	*E*	*Number of parameters*
Little and Ekvall (1981)	41.13	3
Little and Ekvall (1981)	31.17	3
Spindel and Haibach (1981)	17.35	6
Bastenaire (1972)	20.52	5
Castillo et al. (1985)	20.27	4
Castillo and Hadi (1995)	18.12	5
Sev-sev RFL	14.77	5
Normal-normal RFL	12.84	5
Sev-normal RFL	12.96	5
Normal-sev RFL	14.28	5

There is not much that can be done about the interpretational difficulties other than obtaining more or different data or using outside information (e.g., fixing or putting a prior distribution on one or more parameters).

High correlations between parameters are symptomatic of ridge-like behavior on the likelihood surface. These ridges probably caused the numerical difficulties encountered by K&K (although computation of the integrals as $\sigma \to 0$ can also be tricky). In Section 4.1 we mentioned the use of "stable parameters" and referenced Ross (1990) but did not elaborate. Internally, in our software, we replace $[\beta_0, \beta_1, \sigma, \mu_\gamma, \sigma_\gamma]$ with $[\beta_0^*, \beta_1, \log(\sigma), \mu_\gamma^*, \log(\sigma_\gamma)]$, where $\beta_0^* = \beta_0 + \beta_1\bar{x}, \mu_\gamma^* = (\mu_\gamma - x_0)/\sigma_\gamma, \bar{x}$ is the average log stress, and x_0 is a number between the minimum log stress at which there was an observed failure and $\bar{x}$. We chose x_0 to be the mean of the two smallest log stress levels for the concrete data. We use $\log(\sigma)$ and $\log(\sigma_\gamma)$ in place of σ and σ_γ, respectively, to avoid parameter-range constraints. This alternative parameterization is used in maximizing the likelihood and to find estimates of the variance–covariance matrix (using numerical perturbation). This reparameterization has the effect of making the computations much more stable. Because of the invariance of ML estimates, the ML estimates of the original parameters are easy to compute. A simple delta-method computation is used to compute the estimates of the variance–covariance matrix (and correlations) for the original parameter estimators. By default the user sees only the parameters with "operational meaning."

Table 4 gives the estimated correlations between model parameters for the concrete-fatigue data normal-normal RFL model fit. Table 5 gives the estimates for the transformed parameters. The reparameterization reduced 7 of the 10 correlations (in particular, those associated with β_0, β_1, and μ_γ). The other three correlations remained the same

Table 4. Estimated Correlations Between Original Parameter Estimators for the Concrete-Fatigue Data

	β_0	β_1	σ	μ_γ	σ_γ
β_0	1.00	.92	.58	.84	−.66
β_1		1.00	.61	.99	−.77
σ			1.00	.60	−.83
μ_γ				1.00	−.77
σ_γ					1.00

up to the number of digits shown. Similar improvements occurred in fitting the RFL model to other datasets.

2.6 Starting Values

The use of good starting values is usually important in nonlinear fitting. Typically one can find useful starting values based on simple moment, bounding, or graphical methods. In determining starting values $\boldsymbol{\theta}^\diamond$ for the ML estimation process of $\boldsymbol{\theta}$, we choose $\mu_\gamma^\diamond$ to be somewhat smaller than the minimum log stress at which there was an observed failure. More curvature in the S-N plot suggests a value closer to the minimum log stress at which there was an observed failure. We take the $\beta_0^\diamond$ to be the intercept, $\beta_1^\diamond$ to be the slope, and $\sigma^\diamond$ to be the square root of the mean squared error in the least squares (linear) regression of $\log y$ on $\log[s - \exp(\mu_\gamma^\diamond)]$, where y is fatigue life and s is applied stress. We choose $\sigma_\gamma^\diamond$ to be a fraction, say one third, of the distance between the minimum log stress at which there was an observed failure and $\mu_\gamma^\diamond$. When there is not a lot of curvature in the S-N plot, we use different values of $\mu_\gamma^\diamond$, compare the corresponding likelihoods, and take the set of starting values that gives the largest likelihood.

3. RESPONSE TO JOHNSON, FITZGERALD, AND MARTZ

3.1 The Hierarchical Bayes Approach

We thank Johnson, Fitzgerald, and Martz (JFM) for their illustration of the hierarchical Bayes (HB) approach to the RFL model. We heartily agree that, when credible prior information is available, this is a most appropriate approach to use. We hasten to add, however, that in the absence of prior information, and without further investigation, we would feel more comfortable with the ML approach presented in our article. Given the large-sample approximations and the large number of failures in our datasets, we have a good understanding of the operational properties of the procedures used in our article. We know much less about the properties of the Bayesian methods under a prior distribution like that suggested by JFM.

Table 5. Estimated Correlations Between Transformed Parameter Estimators for the Concrete-Fatigue Data

	β_0^*	β_1	$\log(\sigma)$	μ_γ^*	$\text{Log}(\sigma_\gamma)$
β_0^*	1.00	.65	.44	.21	−.42
β_1		1.00	.61	.45	−.77
$\text{Log}(\sigma)$			1.00	−.25	−.83
μ_γ^*				1.00	.21
$\text{Log}(\sigma_\gamma)$					1.00

The repeated-sampling properties of inference procedures are important, even to some Bayesians (e.g., Carlin and Louis 1996, chap. 4). It would be interesting to compare the properties of the Bayes and the likelihood-based intervals to see which have coverage probabilities more closely approximating the nominal confidence level. Previous simulations of likelihood methods (e.g., Vander Weil and Meeker 1990; Jeng and Meeker in press) suggest that, in situations with sample sizes as large as those used here, the likelihood methods would be hard to beat. Bootstrap calibration of the likelihood methods would be expected to improve the approximation further.

It is interesting that the vast improvements that we have seen in computing and statistical technology have sharpened the issues involved in choosing among competing inference methods. Today we have the computational power to do a full Bayesian analysis for a wide range of important statistical models, as demonstrated by JFM. At the same time the modern asymptotic approximations that are often available to the frequentist (especially likelihood ratio methods and bootstrap methods with second-order accuracy) provide very good (likelihood ratio) to excellent (bootstrap-t or likelihood ratio methods calibrated with simulation) asymptotic approximations.

3.2 Global Maxima and the Density Approximation

JFM comment that "the maxima cited in the article are not truly global maxima" (p. 000). Numerically, this may be true, but practically and statistically, our maxima are the correct ML estimates. As either σ or σ_γ approach 0, the RFL pdf and cdf have well-defined limits. When both σ and σ_γ approach 0, it may be possible for the likelihood, as written in our article, to become unbounded. The probability of the data in this case is, however, approaching 0 and thus such aberrations in the likelihood function can be ignored or otherwise avoided.

The root of the potential problem is the use of a continuous-model likelihood for discrete data. It is unfortunately not so well appreciated that *all* data are discrete. In the case of fatigue data, in which the response is the number of cycles to failure, even the underlying model for the response would usually be considered to be discrete. In other situations, such as when hours to failure of light bulbs are recorded, the underlying response may be continuous, but the data are rounded and are thus discrete. The "correct likelihood" accounts for the discreteness in data. Giesbrecht and Kempthorne (1976) eloquently explained this (and gave references to others who had previously recognized this important issue) in the context of the three-parameter lognormal distribution. Over the years we have encountered several continuous-response statistical models in which the density approximation to the likelihood contribution for observation i,

$$[F(t_i + \Delta_i/2; \boldsymbol{\theta}) - F(t_i - \Delta_i/2; \boldsymbol{\theta})] = \int_{t_i - \Delta_i/2}^{t_i + \Delta_i/2} f(t; \boldsymbol{\theta})\, dt \approx f(t_i; \boldsymbol{\theta})\Delta_i, \quad (1)$$

breaks down, typically, as one approaches some boundary of the parameter space. We have been drawn to these examples by claims of failure of the ML method. For other examples, see Friedman and Gertsbakh (1980) and Meeker and Escobar (1994, sec. 8.2). We note that Castillo and Hadi (1995), in section 3 of their article, mentioned the problem of an "unbounded likelihood" for their model, dismissed ML, and proceeded to use other methods of estimation.

The solution to this problem is, in general, to use the "correct likelihood" contributions on the left side of (1) for failures reported as "exact." Indeed, we have such an option built into our software. Likelihoods constructed in this manner are proportional to a probability (the constant of proportionality not depending on any of the unknown parameters) and thus are bounded. In the vast majority of models used in practice (including the RFL model) the density approximation is, however, perfectly adequate, at least if one does not venture too close to the boundary of the parameter space in trying to evaluate the likelihood. When the approximation is adequate, true maxima of the likelihood and likelihood ratio statistics based on the density approximation are approximately equal to those computed with the correct likelihood. The values of Δ_i chosen can be based on the amount of rounding in the data. As long as the Δ_i values are small relative to the spread in the data, likelihood-based inferences will be insensitive to the choice of Δ_i.

Note the profile likelihood for σ in our Figure 2. As $\sigma \to 0$, the profile likelihood is constant at approximately .2. For the concrete data, the profile for σ levels off at around .6, and for another (proprietary) dataset it levels off at a value very close to 1. Even if $\sigma = 0$, the probability of the data is positive and, indeed, can be large, relative to the value at the maximum. This implies that the variability in the data can be explained by the variability in the fatigue limit alone. These results suggest that, even with the discrete-data likelihood, using the "standard" noninformative prior initially suggested by JFM will also result in an improper posterior, due to problems in identifying the sources of variability from the data. This is similar to what happens in the simple random-effects model (e.g., see Gelman, Carlin, Stern, and Rubin 1995, sec. 13.1).

3.3 Differences in Results

JFM comment on differences in results. Modulo the effect of the informative (but diffuse and thus approximately noninformative) prior, the posterior is proportional to the likelihood. The likelihood, however, is not symmetric. It is therefore not surprising that the Bayes point estimates (expectations of the marginal posterior) differ from the ML estimates (maximum of the likelihood). For similar reasons (maximization is not the same as marginalization), the Bayes credible intervals differ from our likelihood-based confidence intervals.

The estimates of standard errors and correlations that we give are based on the local estimate of the Fisher information matrix (which seems to be most commonly used in practice). Because these estimates are based on a quadratic approximation for the log-likelihood, we view these standard errors as only rough indicators of parameter uncertainty. If they were important, bootstrap sampling could provide more accurate estimates of these correlations.

3.4 Prediction

JFM illustrate the use of the Bayes predictive distribution for fatigue life. For some applications, such a distribution would be useful (particularly if one were interested in predicting the outcome of a single future unit). To get results more parallel with what has been used in product design applications and what we presented in Figures 1, 5, and 6 of our article, it would be useful to obtain the marginal posterior distribution and the credible intervals for particular quantiles of the fatigue-life distribution as a function of stress. The same could be done to estimate failure probabilities for a given number of cycles and stress. Correspondingly, if the focus were prediction for the failure time of a single future unit (or some other future realization of a random variable), the bootstrap prediction methods described by Escobar and Meeker (1999) could be extended so that they work for the RFL model. Practical reasons for choosing among confidence, tolerance, and prediction intervals and methods for computing these different kinds of intervals were described by Hahn and Meeker (1991).

3.5 Parameterization

In our response to K&K, we suggested an alternative parameterization that improved the numerical stability of the ML iterations. This same parameterization would also improve the rate of convergence of the Markov-chain Monte Carlo computations used by JFM.

Again we thank all of the discussants for their valuable contributions.

ADDITIONAL REFERENCES

Bastenaire, F. A. (1972), "New Method for the Statistical Evaluation of Constant Stress Amplitude Fatigue-Test Results," in *Probabilistic Aspects of Fatigue (ASTM STP 511)*, ed. R. A. Heller, Philadelphia: American Society for Testing and Materials, pp. 3–28.

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

Carlin, B. P., and Louis, T. A. (1996), *Bayes and Empirical Bayes Methods for Data Analysis*, New York: Chapman & Hall.

Castillo, E., Fernández-Canteli, A., Esslinger, V., and Thürlimann, B. (1985), "Statistical Model for Fatigue Analysis of Wires, Strands and Cables," in *International Association for Bridge and Structural Engineering Proceedings P-82/85*, Zürich, International Association for Bridge and Structural Engineering, pp. 1–40.

Castillo, E., and Hadi, A. S. (1995), "Modeling Lifetime Data With Application to Fatigue Models," *Journal of the American Statistical Association*, 90, 1041–1054.

Escobar, L. A., and Meeker, W. Q. (1999), "Statistical Prediction Based on Censored Life Data," *Technometrics*, 41, 113–124.

Friedman, L. B., and Gertsbakh, I. (1980), "Maximum Likelihood Estimation in a Minimum-type Model With Exponential and Weibull Failure Modes," *Journal of the American Statistical Association*, 75, 460–465.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995), *Bayesian Data Analysis*, London: Chapman & Hall.

Giesbrecht, F., and Kempthorne, O. (1976), "Maximum Likelihood Estimation in the Three-Parameter Lognormal Distribution, *Journal of the Royal Statistical Society*, Ser. B, 38, 257–264.

Hahn, G. J., and Meeker, W. Q. (1991), *Statistical Intervals: A Guide for*

Practitioners, New York: Wiley.

Jeng, S. L., and Meeker, W. Q. (in press), "Comparisons of Approximate Confidence Interval Procedures for Type I Censored Data," *Technometrics*, 42.

Little, R. E., and Ekvall, J. C. (eds.) (1981), *Statistical Analysis of Fatigue Data (ASTM STP 744)*, Philadelphia: The American Society for Testing and Materials.

Meeker, W. Q., and Escobar, L. A. (1994), "Maximum Likelihood Methods for Fitting Parametric Statistical Models to Censored and Truncated Data," in *Probabilistic and Statistical Methods in the Physical Sciences*, eds. J. Stanford and S. Vardeman, New York: Academic Press, pp. 211–244.

——— (1998), *Statistical Methods for Reliability Data*, New York: Wiley.

Meeker, W. Q., and Hahn, G. J. (1977), "Asymptotically Optimum Over-Stress Tests to Estimate the Survival Probability at a Condition With a Low Expected Failure Probability" (with discussion), *Technometrics*, 19, 381–399.

Sobczyk, K. (ed.) (1993), *Stochastic Approach to Fatigue: Experiments, Modelling, and Reliability Estimation*, New York: Springer-Verlag.

Sobczyk, K., and Spencer, B. F. (1992), *Random Fatigue: From Data to Theory*, Boston: Academic Press.

Spindel, J. E., and Haibach, E. (1981), "Some Considerations in the Statistical Determination of the Shape of the S-N Curves," in *Statistical Analysis of Fatigue Data* (ASTM STP 744), eds. R. E. Little and J. C. Ekvall, Philadelphia: The American Society for Testing and Materials, pp. 89–113.