Decelerated Testing: A Hierarchical Bayes Approach

Nozer D. Singpurwalla

The Institute for Reliability and Risk Analysis
Department of Statistics
The George Washington University
Washington, DC 20052
(*nozer@gwu.edu*)

The problem discussed here has arisen from an industrial scenario involving the potential failure of an element of building structures. The element carries with it a warranty of several years. The scenario considered is not specific to buildings and occurs under other circumstances; it goes under the label "product stewardship." Its statistical content, however, is inference from accelerated tests in reverse, that is, testing at low stress levels and making predictions at high use conditions ones. Inference under stress testing for lifetimes is a well-discussed topic in the statistical literature, but what is different here are three features: not using time-transformation functions relating parameters of failure models with stress, using material strength instead of lifetimes as a metric for reliability, and fusing information consisting of test data and expert testimonies about the degradation of strength with cumulative stress. Because stress testing is also germane to survival analysis, with immunity replacing strength, the scope of this work could extend to biometry. All of the data used here are real, including that pertaining to expert testimonies. However, in the interest of avoiding potential lawsuits, the owners of the data have mandated that neither their source nor their nature be made public.

KEY WORDS: Accelerated testing; Degradation factors; Expert testimony; Information integration; Reliability; Survival analysis; Strength of materials; Warranties.

1. MOTIVATION AND BACKGROUND: THE FAILURE OF SOLIDS

1.1 Motivation

Manufacturers of consumer goods often warranty their products for long periods of time, sometimes 25–30 years. The advantage to a manufacturer who provides long warranties is increased sales; the disadvantage is increased risk of financial loss. The choice of an optimal warranty is therefore of central concern.

"Product stewardship" is the task of nurturing a product from inception to warranty termination in a manner that maximizes a manufacturer's expected utility. Thus, inherent to product stewardship are the dual tasks of designing an optimum warranty and of assessing the probability of failure during warranty. These two tasks are interrelated; however, the focus of this article is the latter issue. Our aim here is to contribute to the science of product stewardship by providing some statistical technology. The specific scenario that has motivated our work entails some nuances, and for this it is useful to start with a discussion of how solid materials fail, and why is it that informed testimonies on the strength of materials could be germane to the prediction problem discussed here.

1.2 Stress, Strength, and Thresholds

We start our discussion by introducing the notions of stress and strength, and discussing their role in determining failure. By *stress*, we mean the applied force per unit area. The applied force could be a deterministic function of time, cyclical or otherwise, or it could be random. The *strength* of an object is the level of stress at which it experiences an *instantaneous* failure. Because the notion of an instantaneous failure is an abstraction, strength per se can only be a conceptual entity. An item fails whenever the applied level of stress equals or exceeds its strength. This forms the basis of what is known as a "stress-strength" based approach to reliability modeling and failure data analysis. However, there is another way to look at the failure-causing process, based on the notions of material degradation and threshold. This alternative perspective goes under the label "degradation modeling," which has now become an actively researched topic. Lu, Meeker, and Escobar (1996) have provided an excellent perspective. In what follows we attempt to trace the commonalities between the two approaches to make the case that strengths and thresholds are, de facto, parallel concepts. A consequence is that the notion of strength can be given an interpretive flavor with the advantage that expert testimonies on strength can be elicited with a concrete picture in mind.

By *degradation* (or *damage*), we mean the tearing apart of the chemical bonds that hold together the atoms of an object. This tearing apart is caused by the cumulative effect of the stresses that act on the object. The cause of failure of a solid object is a degradation of its material properties beyond their endowed "thresholds." An endowed threshold is a conceptual entity (Singpurwalla 2004) that is viewed as being fixed but unknown. Thus it cannot be observed and measured. In principle, degradation cannot be directly observed either. However, degradation manifests itself via what are known as "markers"; examples of such markers are cracks, corrosion, and wear (the instructive paper of Whitmore, Crowder, and Lawless 1998).

TECHNOMETRICS, NOVEMBER 2005, VOL. 47, NO. 4
DOI 10.1198/004017005000000175

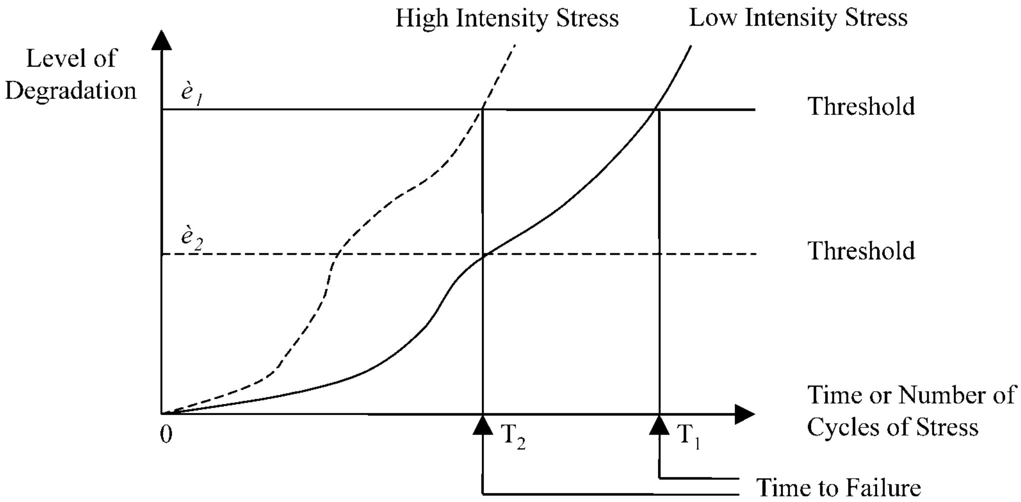

Figure 1. Degradation Curves for Two Types of Stress-Loading Protocols.

The amount of degradation experienced by an object subjected to any type of stress is nondecreasing in time, and an object's life-length is the time at which the level of degradation exceeds its threshold, say θ_1; see Figure 1. Because thresholds are random, and degradation levels, being dependent on an interaction between an object's material properties and the applied stress, are also random, life-lengths are random quantities. As an illustration, the bold curve of Figure 1 represents the damage caused by a stress regimen of low intensity, and the dotted curve represents the degradation caused by a stress regimen of high intensity. The corresponding life-lengths are T_1 and T_2. Because strength is a level of stress that results in an instantaneous failure, the degradation curve corresponding to an instantaneous failure would be a point mass at time 0. The essence of Figure 1 can be better appreciated when degradation is viewed as a measurable quantity, like crack length. With the foregoing as a background, we next use the likes of Figure 1 to make the case that the notions of threshold and stress are isomorphic to one another, so that strength can serve as a substitute for life-length.

1.3 Substituting Life-Lengths by Stress

We start by noting (see Fig. 1) that for a fixed threshold, say θ_1, the effect of a high-intensity stress is to shorten the life-length from T_1 to T_2, where T_1 corresponds to a stress of low intensity. Inversely, for a fixed life-length, say T_2, the effect of a high-intensity stress is to lower the threshold from θ_1 to θ_2. But this feature parallels the view that the strength of an object decreases with increased levels of cumulative applied stress. This, plus the fact that both strength and thresholds are endowed properties, enables us to conceptualize a one-to-one relationship between thresholds and strengths. Because thresholds go to determine life-lengths, we may substitute strengths for life-lengths, and by doing so elicit expert testimonies on strengths instead of life-lengths. Indeed, for the application scenario that motivated this work, the engineers and material scientists were willing to provide testimonies on strength. The notion of providing testimony on parameters of probability distributions of life-lengths appeared to be alien to their thinking. The purpose of this section has been to provide a justification as to why expert testimonies on strengths are as meaningful as testimonies on life-lengths.

In what follows the discussion focuses on strength, and assumes that the applied stress is cyclical with the cumulative number of cycles as a metric for the level of stress experienced by an item.

1.4 Why Decelerated Testing and What Does It Mean?

Stress cycles caused by the effects of the environment, like temperature, wind speeds, or other attributes of the weather, tend to have large amplitudes and large periods. Such stress cycles cause a degradation of an item experiencing stress, the level of degradation being a function of the cumulative number of stress cycles experienced. Furthermore, the total number of cycles needed to degrade an item to its threshold could run to the thousands. Thus a laboratory replication of the failure-causing phenomenon covering the duration of a several-year warranty is not practical. Time accelerating the degradation process (Meeker and Escobar 1998; Nelson 1990) by increasing the amplitude and/or decreasing the period is an option, but for the scenario of interest to us here, this option turned out to be problematic. The main obstacle is that it was not physically possible to develop an instrument that increases the amplitude, nor was it possible to develop an instrument that decreases the period. In other words, it was not possible to physically construct a mechanism that is able to accelerate the degradation phenomenon.

In view of the foregoing, the likelihood of failure within the duration of the warranty is assessed via a combination of informed testimonies and data from laboratory tests involving a small number of stress cycles whose amplitudes and periods mimic those actually encountered in practice. Testing at low levels of the applied stress (in our case, a small number of cumulative stress cycles) and using the test results as a basis for extrapolation to conditions stress, is termed *decelerated testing*. To the best of our knowledge, this terminology has not heretofore appeared in the literature, at least in the manner described here. In our particular case, inference under decelerated testing entails studying the degradation of strength in going from one level of stress to a higher level, using the cumulative number of stress cycles as a metric for the level of the applied stress. It is important to bear in mind that extrapolating from low levels to high levels of stress entails risks that parallel those encountered in traditional accelerated life testing, wherein one extrapolates from high stress levels to low stress levels. An answer to the question as to which of these two risks is the greater depends on the nature of the item being tested.

It is easy to argue that the decelerated testing scenario described earlier can arise in several other contexts as well. Indeed, all newly designed products that carry long warranties and whose degradation is a consequence of prolonged use (e.g., dishwashers and refrigerators) would be subjected to some form of inference under both an accelerated test and a decelerated test before their mass manufacture. The accelerated tests not only help shorten test times, but also allow assessment of robustness of the product against misuse. The decelerated tests play a role in setting warranties that are equitable to both the consumer and the producer. Thus the accelerated and the decelerated tests serve useful, but different, purposes. In the biomedical context, decelerated testing may come into play when physicians prescribe drugs with side effects. For example, cholesterol-lowering drugs often have the disadvantage of having an adverse effect on liver function. Consequently, physicians often prescribe a 10-mg dose for a trial period of, say, 3 months (to explore the drug's effectiveness as well as its side effects) before prescribing a desired higher-level dose of, say, 20 mg for subsequent normal use.

2. A MODEL FOR DECELERATED TESTING BASED ON STRENGTHS

Let $C_1 \prec C_2 \prec \cdots \prec C_m$ denote the m stress levels at which seemingly identical (the technical term is "exchangeable") specimens of the item are tested. "$C_i \prec C_j$" denotes that C_i is less severe than C_j. Here each C_i denotes the cumulative number of (identical) stress cycles of an environmental stress experienced by each of several specimens. Also, let $0 \equiv C_0 \prec C_1$, where C_0 is labeled as a "background" stress, that is, a stress that is always present and that is outside the control of the tester. In our particular application (see Sec. 4.1), $m = 8$ and the C's are equally spaced. Also, 30 specimens were tested at each value of C, giving a total number of specimens tested of 240.

Let Y_i denote the strength of an item that has experienced a stress of C_i cycles, $i = 0, 1, \dots, m$. That is, Y_i encapsulates the additional damage that the item can withstand before it fails. We suppose that $Y_0 \ge Y_1 \ge Y_2 \ge \cdots$. The Y_i's, being conceptual entities, are unknown, and each $Y_i \in [0, \infty)$. Let $\widehat{Y}_i, i = 1, 2, \dots, m$, be the *observed* strength of an item that has experienced a stress of C_i cycles. In practice, each $\widehat{Y}_i$ is obtained by selecting an item that has experienced stress C_i, and then determining, via laboratory means, the additional stress needed to cause its instantaneous failure. We assume that for $i = 1, \dots, m$, $\widehat{Y}_i < \infty$. It may be useful to remark that $\widehat{Y}_i$ is not a realization of Y_i, because it is not an observed value of Y_i. Rather, $\widehat{Y}_i$ can be viewed as an estimate of Y_i. In our particular application, we deal with $\widehat{Y}_{ij}$, $i = 1, \dots, 8$, and $j = 1, \dots, 30$, because 30 specimens were tested at each C_i. Then $\widehat{Y}_i$ will be the sample average of $\widehat{Y}_{ij}$, $j = 1, \dots, 30$. The eight values of $\widehat{Y}_i$ turned out to be 392, 354, 310, 285, 289, 287, 286, and 285. Their respective standard deviations were 45.1, 32.0, 43.7, 40.3, 38.2, 31.9, 39.4, and 28.1.

Our aim is to make a predictive inference about Y_{m+n}, $n = 1, 2, \dots,$ the strength of an item experiencing condition C_{m+n}, the cumulative number of stress cycles that the item will experience during its *service life*. There are two points to note. The first point is that in addition to the $\widehat{Y}_i$'s, there exist informed judgments about Y_0, the *background strength*, that is, the strength of an item that experiences only the background stress C_0. The second point is that it is not necessary that $\widehat{Y}_1 \ge \widehat{Y}_2 \ge \cdots \ge \widehat{Y}_m$; rank reversals are possible.

Of the informed views that engineers may have, the one of interest here pertains to the degradation of strength in going from one stress level to a higher one. The form of this judgment accords well with some theories of crack growth—namely, the bigger the crack, the larger the growth (Gordon 1976). Specifically, the subject matter specialists assumed that for some unknown constant $p_i \in (0, 1]$, $i = 1, 2, \dots,$

$$Y_i = p_i Y_{i-1}, \tag{1}$$

where $(1 - p_i)$ is the *proportion loss in strength* in going from C_{i-1} to C_i. Furthermore, the initial degradation of strength is supposedly more drastic than the latter, so that the strength as a function of large stresses remains de facto a constant. A strategy for capturing this, especially because the C_i's are equally spaced, is a version of the power law relationship

$$p_i = p_{i-1}^{\alpha}, \qquad i = 2, 3, \dots, \tag{2}$$

where $\alpha \in [0, 1]$ is a kind of an omnibus (and unknown) constant that encapsulates the amount by which the strength degrades in going from one stress level to the next. Motivated by a terminology used by engineers and material scientists who collaborated with us, we label the constant α, a *degradation factor*.

A question may be raised about the justification of (2), especially because we offer no physical basis for it. Our response is that this equation is meaningful because it captures the feature (claimed by the physical scientists) that the initial degradation of strength is more drastic than the latter degradation. Furthermore, the relationship of (2) brings into play, via α, the notion of a degradation factor to which the scientists repeatedly allude, but are unable to make precise. We feel that (2) provides a suitable vehicle for articulating the notion of a "degradation factor." We have more to say about this later, but for now we need to address the philosophical issue of whether it makes sense to invoke relationships—like (2)—that do not have a physical basis. The nonempiricist point of view is that all relationships are reasonable judgments whose value depends on how useful they are with respect to either their predictive capability or their ability to describe phenomena judged to be meaningful. For example, it is often the case that parameters in probability models cannot be given a physical interpretation. Yet they are used because their purpose is to impart independence between the unknown random variable(s) and the known background information that could be of infinite dimension (Lindley 2000). In other words, a parameter is a convenient think piece, which physicists often call a "hidden variable."

Under (2), p_i increases in i, and for any value of $\alpha \in [0, 1]$, p_i increases to 1 as i gets large. The increase of p_i to 1 is fast when α is in the vicinity of .5; it is slow when α is in the vicinity of 1. When $\alpha = 0$, there is no degradation of strength with stress. Thus the value that α takes provides scientists with a clue about the pattern of degradation of strength with stress; therefore, its assessment could serve a useful purpose. Equations (1) and (2) constitute our proposed two-stage (or *hierarchical*) model for decelerated testing for strength. The model is written in terms of strength and degradation, notions that engineers can relate to; and thus can provide informed testimony about. Such testimony are based on empirical evidence and/or a knowledge of material science.

If the C_i's were not equally spaced, then the assumption of a constant α would be flawed. In such cases we would have to propose a model that describes how the α's change (by, e.g., a dynamic linear model), which this would result in a three-stage hierarchy. The disadvantage now would be twofold: (1) added complications to an elaborate statistical development, and (2) the search for an alternate parameter that one can meaningfully interpret as a degradation factor. Because in a decelerated (or an accelerated) test the C_i's are under the control of an experimenter, the advantage of making them equally spaced outweighs the disadvantages mentioned here, albeit at the cost of generality.

3. PREDICTING SERVICE LIFE: OVERVIEW OF THE BAYESIAN APPROACH

When invoked recursively, (1) and (2) imply that for $i = 2, 3, \ldots,$

$$p_i = p_1^{\alpha^{(i-1)}} \tag{3}$$

and

$$Y_i = Y_0 \cdot p_1^{(1-\alpha^i)/(1-\alpha)}, \qquad i = 1, 2, \ldots, \tag{4}$$

where $(1 - p_1)$ is the proportion loss of background strength in going from C_0 to C_1. If we let

$$\widehat{p}_1 = \frac{\widehat{Y}_1}{Y_0} \qquad \text{and} \qquad \widehat{p}_i = \frac{\widehat{Y}_i}{\widehat{Y}_{i-1}}, \qquad i = 2, 3, \ldots, m,$$

then knowing Y_0 and $\widehat{\underline{\mathbf{Y}}} = (\widehat{Y}_1, \ldots, \widehat{Y}_m)$ is tantamount to knowing $\widehat{\underline{\mathbf{p}}} = (\widehat{p}_1, \ldots, \widehat{p}_m)$.

The quantity $\widehat{p}_1$ is a substitute for p_1, but to obtain $\widehat{p}_1$, we must know Y_0. The assessment of Y_0, based on expert testimony, is described in Section 3.1.1. Also, the $\widehat{p}_i$'s need not fall within the interval (0, 1], and (4) implies that $\lim_{i\to\infty} Y_i = Y_0 p_1^{1/(1-\alpha)} > 0$. This latter feature suggests that the product has some residual strength even after an infinite number of cycles. Because such a feature is counter to intuition, our position is that the proposed model is meaningful when the number of cycles is not very large.

Section 3.1 describes the assessment of α based on the data $\widehat{\underline{\mathbf{Y}}}$ and informed testimony on α, Y_0, and p_1. The subsections of Section 3.1 entail some delicate maneuvers and some judgments of conditional independence; the latter are characteristic of much statistical practice. The manner in which the testimonies and the data are fused together is illustrated in Figure 2; hopefully, it provides a convenient road map to Sections 3.1.1–3.1.5. Section 3.2 describes the prediction of service life using the assessments of Section 3.1; Section 3.3 addresses the computational issues. Readers not interested in the detailed developments of Sections 3.1 and 3.2 may skip these and proceed directly to Section 4, which pertains to the application.

3.1 Assessing the Degradation Factor α

Information about α comes from the data $\widehat{p}_1, \ldots, \widehat{p}_m$ and informed testimonies about α. We start by noting that

$$P(\alpha|\widehat{p}_1, \ldots, \widehat{p}_m) = \int_{p_1} P(\alpha|p_1, \widehat{p}_2, \ldots, \widehat{p}_m) P(p_1|\widehat{p}_1)\, dp_1, \tag{5}$$

assuming that p_1 is independent of $\widehat{p}_2, \ldots, \widehat{p}_m$, given $\widehat{p}_1$, denoted by $(p_1 \perp (\widehat{p}_2, \ldots, \widehat{p}_m)|\widehat{p}_1)$, and $(\alpha \perp \widehat{p}_1 | p_1, \widehat{p}_2, \ldots, \widehat{p}_m)$. But, by Bayes' law,

$$P(p_1|\widehat{p}_1) \propto L(p_1; \widehat{p}_1) P(p_1),$$

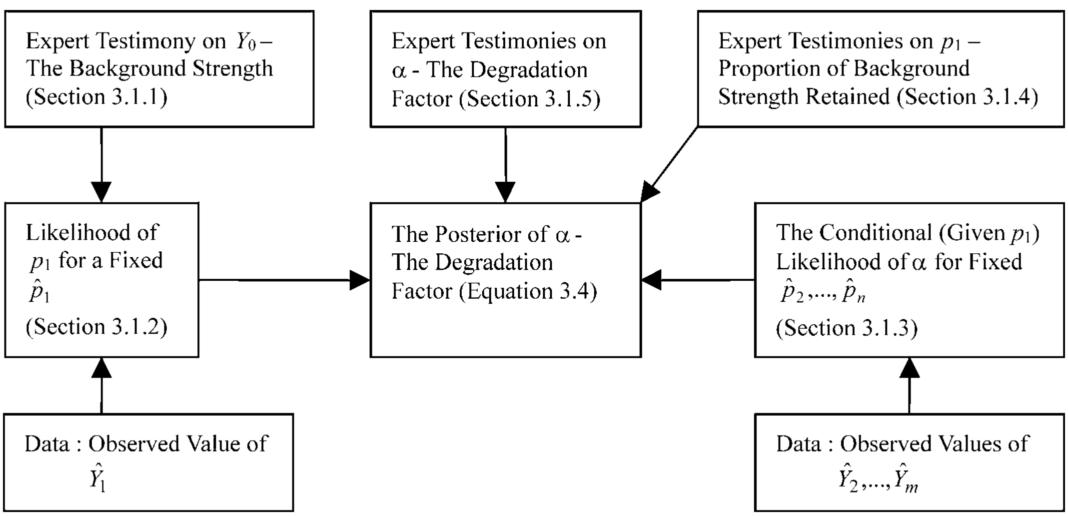

Figure 2. Assessing the Degradation Factor by Integrating Data and Expert Testimony.

where $P(p_1)$ is the prior on p_1 and $L(p_1;\widehat{p}_1)$ is the likelihood of p_1. Similarly (by Bayes' law),

$$P(\alpha|p_1,\widehat{p}_2,\dots,\widehat{p}_m) \propto L(\alpha|p_1;\widehat{p}_2,\dots,\widehat{p}_m)P(\alpha),$$

where $L(\alpha|p_1;\widehat{p}_2,\dots,\widehat{p}_m)$ is the conditional likelihood for α given p_1 and $P(\alpha)$ is the prior on α. Thus we may write (5) as

$$P(\alpha|\widehat{p}_1,\dots,\widehat{p}_m) \propto \int_{p_1} L(\alpha|p_1;\widehat{p}_2,\dots,\widehat{p}_m)L(p_1;\widehat{p}_1)P(\alpha)P(p_1)\,dp_1. \quad (6)$$

Equation (6) entails two likelihoods and two priors. The priors are based on expert testimony, which we discuss in Sections 3.1.4 and 3.1.5. For the likelihoods part, we first need to assess Y_0, because $\widehat{p}_1$ is obtained via Y_0; this is done in Section 3.1.1.

3.1.1 Expert Testimony on Background Strength Y_0. There is much empirical evidence which indicates that the frequency distribution of observed life-lengths is skewed to the right. Because strength is a proxy for life-length, we suppose that an expert $\mathcal{E}$'s judgment about Y_0 can be conceptualized by a lognormal distribution whose mean and median are a^* and b^*. Thus $\mathcal{E}$'s lognormal distribution for Y_0 has parameters $a = \ln(b^*)$ and $b^2 = 2\ln(a^*/b^*)$. To account for the possibility that $\mathcal{E}$ may have biases and tendencies to exaggerate the declared a^* and b^*, we may introduce the *modulating parameters*, say $\overline{a}$, $\overline{b}$, and $\overline{\gamma}$, and obtain the quantities (see Lindley 1983)

$$\widehat{a} = \frac{\ln(b^*) - \overline{b}}{\overline{a}} \qquad \text{and} \qquad \widehat{b} = \frac{\overline{\gamma}(2\ln(a^*/b^*))^{1/2}}{\overline{a}}. \quad (7)$$

The role of the modulating parameters is to facilitate an adjustment for $\mathcal{E}$'s biases and tendencies to exaggerate. When we have full faith in $\mathcal{E}$'s expertise about Y_0, we set $\overline{b} = 0$ and $\overline{a} = \overline{\gamma} = 1$. If our previous opinion about Y_0 is flat over the region wherein the likelihood of Y_0 for fixed values of $\widehat{a}$ and $\widehat{b}$ is appreciable, then a straightforward application of Bayes' law yields the result that Y_0 has a lognormal distribution with parameters $\widehat{a}$ and $\widehat{b}$. Specifically,

$$P(Y_0|a^*,b^*,\overline{a},\overline{b},\overline{\gamma}) = \frac{1}{\sqrt{2\pi}\,\widehat{b}Y_0}\exp\left[-\frac{1}{2}\left(\frac{\ln Y_0 - \widehat{a}}{\widehat{b}}\right)^2\right], \qquad Y_0 > 0; \quad (8)$$

the details, which are relatively straightforward, have been given by Lindley (1983). Note that by a flat prior where the likelihood is appreciable, we do not mean an improper prior; rather, it is a proper prior that happens to be flat on a certain support of the likelihood function. A choice of such priors allows the data to play a dominant role in the inferential process. The down side is that the prior will be chosen after the data are observed, an action objectionable to staunch Bayesians.

The result of (8) can be compactly written as $(Y_0|\widehat{a},\widehat{b}) \sim \Lambda(\widehat{a},\widehat{b})$ or, equivalently, $(\ln Y_0|\widehat{a},\widehat{b}) \sim \mathrm{N}(\widehat{a},\widehat{b})$, where $\Lambda(\mathrm{N})$ denotes a lognormal (normal) distribution. The assumption of lognormality of Y_0 is not crucial; being $\mathcal{E}$'s conceptualized distribution, it captures the essentials and provides a closed-form expression.

3.1.2 A Model for the Likelihood $L(p_1;\widehat{p}_1)$. Because $\widehat{Y}_i$ is an estimated value of Y_i, we say that $\widehat{Y}_i$ is a proxy for Y_i. Suppose that, like Y_0, the $\widehat{Y}_i$'s have a lognormal distribution with parameters μ_i and σ, $i = 1,\dots,m$. The assumption of a common value σ is not essential, but it simplifies the analyses. Assuming that $E(\widehat{Y}_i) = Y_i$, it is easy to see that for all i,

$$(\ln\widehat{Y}_i|Y_i,\sigma^2) \sim \mathrm{N}\left(\ln Y_i - \frac{\sigma^2}{2},\sigma^2\right), \quad (9)$$

and, in particular,

$$(\ln\widehat{Y}_1|Y_1,\sigma^2) \sim \mathrm{N}\left(\ln Y_1 - \frac{\sigma^2}{2},\sigma^2\right). \quad (10)$$

Using the assertion that $(\ln Y_0|\widehat{a},\widehat{b}) \sim \mathrm{N}(\widehat{a},\widehat{b})$ [see Sec. 3.1.1 and (10)] and, the fact that $(\widehat{Y}_1 \perp Y_0|p_1)$ and $Y_1 = p_1Y_0$, we have

$$(\ln\widehat{p}_1|p_1,\widehat{a},\widehat{b},\sigma^2) \sim \mathrm{N}\left(\ln p_1 - \frac{\sigma^2}{2},\sigma^2 + 2\widehat{b}^2\right).$$

If we let $s^2 = \sigma^2 + 2\widehat{b}^2$, then, suppressing all constants except p_1, the likelihood of p_1 is of the form

$$L(p_1;\widehat{p}_1,\cdot) = \frac{1}{\sqrt{2\pi}s\widehat{p}_1}\exp\left[-\frac{1}{2s^2}\left(\ln\left(\frac{\widehat{p}_1}{p_1}\right) + \frac{\sigma^2}{2}\right)^2\right]. \quad (11)$$

3.1.3 A Model for the Conditional Likelihood of α. Using arguments that parallel those of Section 3.1.2, plus the fact that $(\widehat{Y}_i \perp \widehat{Y}_{i-1}|p_i)$ and $p_i = p_1^{\alpha^{(i-1)}}$, we have, for $i = 2,3,\dots,$

$$(\widehat{p}_i|p_i,\alpha,\sigma^2) \sim \Lambda\left(\ln p_1^{\alpha^{(i-1)}}, 2\sigma^2\right). \quad (12)$$

Because $(\widehat{p}_2 \perp \widehat{p}_3 \perp \dots \perp \widehat{p}_m|p_1,\alpha)$, the conditional likelihood for α, given p_1, is of the form

$$L(\alpha|p_1;\widehat{p}_2,\dots,\widehat{p}_m) = \prod_{i=2}^{m}\frac{1}{2\sigma\widehat{p}_i\sqrt{\pi}}\left(\frac{p_1^{\alpha^{(i-1)}}}{\widehat{p}_i}\right)^{1/(2\alpha^2)}. \quad (13)$$

3.1.4 Eliciting and Coding Experts Testimonies on p_1 *via Log-Odds.* Because of the Markovian structure of the p_i's, it suffices to consider only p_1. Accordingly, suppose that we consult k experts, $\mathcal{E}_1,\dots,\mathcal{E}_k$, and focus on expert $\mathcal{E}_i$. Suppose that $\mathcal{E}_i$ declares m_i and v_i as measures of the location and scale of $\mathcal{E}_i$'s assuming distribution of p_1; then our task is to assess $P(p_1;m_i,v_i)$.

Because $p_1 \in (0,1]$, it is convenient to consider the quantity $q_1 = \log(p_1/(1-p_1))$, the "log-odds." The normal distribution for the log odds is standard. By Bayes' law,

$$P(q_1;m_i,v_i) \propto L(q_1;m_i,v_i)P(q_1), \quad (14)$$

where $P(q_1)$ is the prior on q_1 and $L(q_1;m_i,v_i)$ its likelihood. To propose a model for this likelihood, we need to deduce the location and scale for q_1 in terms of m_i and v_i. Unlike p_1, q_1 does not have an interpretive content.

Deducing Measure of Location and Scale for the Log-Odds. Because $p_1 \in (0, 1]$, we suppose that $\mathcal{E}_i$ conceptualizes the uncertainty about p_1 via a beta distribution with parameters α_i and β_i. Thus in $\mathcal{E}_i$'s view,

$$E(p_1|\alpha_i, \beta_i) = \frac{\alpha_i}{\alpha_i + \beta_i} \qquad \text{and}$$
$$V(p_1|\alpha_i, \beta_i) = \frac{\alpha_i \beta_i}{(\alpha_i + \beta_i)^2(\alpha_i + \beta_i + 1)}. \tag{15}$$

We obtain α_i and β_i by setting $m_i = E(p_1|\alpha_i, \beta_i)$ and $v_i^2 = V(p_1|\alpha_i, \beta_i)$. A knowledge of α_i and β_i enables us to deduce $\mathcal{E}_i$'s measure of location and scale for q_i, m_i^*, and s_i^*. This is because

$$E(q_1|\alpha_i, \beta_i) = E(\ln p_i|\alpha_i, \beta_i) - E\big(\ln(1 - p_i)|\alpha_i, \beta_i\big) = m_i^*, \tag{16}$$

and there exist numerical routines (such as Maple) for evaluating

$$\int_0^1 f(p) \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha - 1}(1 - p)^{\beta - 1}\, dp$$

for any function $f(p)$ of p. In our case, $f(p)$ is $\ln p_1$ or $\ln(1 - p_1)$. Similarly, we may deduce $(s_i^*)^2$ using the fact that

$$V(q_1|\alpha_i, \beta_i) = V[\ln p_1 - \ln(1 - p_1)] = (s_i^*)^2. \tag{17}$$

To summarize, evaluating $P(q_1; m_i, v_i)$ is equivalent to evaluating $P(q_1; m_i^*, s_i^*)$, and because $q_1 \in (-\infty, +\infty)$, we follow a strategy analogous to that of Section 3.1.1 for prescribing a likelihood of q_1 in light of m_i^* and s_i^*. In particular, we may find it useful to introduce the a_i, b_i, and γ_i to incorporate $\mathcal{E}_i$'s biases and tendencies to overestimate (or underestimate).

The Case of Multiple Experts. With k experts, we would be eliciting $2k$ values (m_i^*, s_i^*), $i = 1, \ldots, k$, and associated with each pair would be the triplet of modulating parameters (a_i, b_i, γ_i), $i = 1, \ldots, k$. However, there is likely to be correlation between testimonies, so we introduce parameters ρ_{ij}, where $\rho_{ij} \in [-1, 1]$ encapsulates relationships between the declared values of $\mathcal{E}_i$ and $\mathcal{E}_j$, as perceived by us.

Once the foregoing has been done, we may use standard methodology (see Lindley 1983) to develop a model for the likelihood of q_1 in the light of (m_i^*, s_i^*), (a_i, b_i, γ_i), $i = 1, \ldots, k$, and the ρ_{ij}'s. Here again, if our prior on q_1 is flat where the likelihood is appreciable, then

$$P\big(q_1|(m_i^*, s_i^*), (a_i, b_i, \gamma_i), \rho_{ij}, i, j = 1, \ldots, k\big) = \frac{1}{\sqrt{2\pi}\,\overline{\sigma}} \exp\left(-\frac{1}{2\overline{\sigma}^2}(q_1 - \overline{\mu})^2\right), \qquad -\infty < q_1 < +\infty,$$

where

$$\overline{\mu} = \sum_i \sum_j b_i \sigma^{ij}(m_j^* - a_j) \Big/ \sum_i \sum_j b_i \sigma^{ij} b_j \qquad \text{and}$$
$$\overline{\sigma}^2 = \left(\sum_i \sum_j b_i \sigma^{ij} b_j\right)^{-1}. \tag{18}$$

The σ^{ij}'s are the elements of a matrix that is the inverse of a matrix with elements $\sigma_{ii} = (\gamma_i s_i^*)^2$ and $\sigma_{ij} = \rho_{ij}(\sigma_{ii}\sigma_{jj})^{1/2}$, $i, j = 1, \ldots, k$ (see Lindley 1983 for details).

Once $P(q_1|(m_i^*, s_i^*), i = 1, \ldots, k, \cdot)$ is obtained, we use a transformation of variables to describe our uncertainty about p_1, for $0 < p_1 \le 1$, as

$$P\big(p_1|(m_i^*, s_i^*), (a_i, b_i, \gamma_i), \rho_{ij}, i, j = 1, \ldots, k\big) = \frac{1}{\sqrt{2\pi}\,\overline{\sigma} p_1(1 - p_1)} \times \exp\left[-\frac{1}{2\overline{\sigma}^2}\left(\ln\left(\frac{p_1}{1 - p_1}\right) - \overline{\mu}\right)^2\right]. \tag{19}$$

3.1.5 Expert Testimony on the Degradation Factor α. In many cases of practical interest, the convergence of p_i to 1 is supposedly slow; thus it is reasonable to assume that α should take values between .5 and 1. A suitable prior for α is therefore a translated beta distribution. That is, for any θ having the distribution

$$\pi(\theta|\lambda, \delta) = \frac{\Gamma(\lambda + \delta)}{\Gamma(\lambda)\Gamma(\delta)} \theta^{\lambda - 1}(1 - \theta)^{\delta - 1}, \qquad 0 < \theta < 1, \tag{20}$$

α should equal $a + (b - a)\theta$, with $a = .5$ and $b = 1$.

To pin down λ and δ, we may elicit from each of the k experts their assessed values of p_{j-1} and p_j for any suitable value of j, $j \ge 2$. Because p_1 has already been assessed (see Sec. 3.1.4), it is convenient to consider $j = 2$. Let us focus on $\mathcal{E}_i$ and suppose that $\mathcal{E}_i$'s testimony for the measures of location for p_{j-1} and p_j are $m_{(j-1)i}$ and m_{ji}.

The proper Bayesian approach would be to assess p_{j-1} and p_j via a scheme like the one of Section 3.1.4, and then to deduce a distribution for α using the fact that $\alpha = \ln(p_j)/\ln(p_{j-1})$. However, this could be cumbersome given the nonstandard nature of (19). A more direct strategy would be to consider the sample mean and the sample variance of the α_i's, where $\alpha_i = \ln(m_{ji})/\ln(m_{(j-1)i})$, $i = 1, \ldots, k$; there is no assurance that $\alpha_i \in (.5, 1]$. Let $\widehat{\alpha} = \sum_{i=1}^k \alpha_i/k$ and $V(\widehat{\alpha})$ be the sample variance of the α_i's. Then suitable values for λ and δ can be found by setting $\widehat{\alpha} = E(\alpha) = a + \lambda/(\lambda + \delta)$ and $V(\widehat{\alpha}) = V(\alpha) = \lambda\delta/[(\lambda + \delta)^2(\lambda + \delta + 1)]$, and solving for λ and δ. In taking the sample mean of the α_i's, we are gaining some robustness by using shrinkage to $\widehat{\alpha}$. Alternatively, we also could have chosen the trimmed mean of the α_i's, but doing so would be tantamount to a denial of judgments that are out of kilter with the rest.

Once the foregoing has been done, our elicited prior on α can be obtained by a standard transformation of variables such as

$$\pi(\alpha|\lambda, \delta) = \frac{\Gamma(\lambda + \delta)}{\Gamma(\lambda)\Gamma(\delta)} \left(\frac{\alpha - a}{b - a}\right)^{\lambda - 1} \frac{1}{b - a} \left(\frac{b - \alpha}{b - a}\right)^{\delta - 1}, \qquad a < \alpha < b. \tag{21}$$

3.1.6 Summary of Section 3.1. We have outlined an approach for specifying the four components of (6). This equation enables us to assess the degradation factor α based on actual data as well as expert testimonies on the background strength Y_0; the proportion of strength lost in going from C_0 to C_1, namely $(1 - p_1)$; and the degradation factor itself. As mentioned earlier, knowledge of the degradation factor is useful, because this is an indicator of the service life of the item and plays a role in predicting strength at service life. Figure 2 is the heart of our development; it illustrates the manner in which the data and the testimonies come together to constitute (6).

3.2 An Architecture for Predicting Strength at Service Life

Recall that our main aim is to make statements of uncertainty about $Y_{m+n}, n = 1, 2, \dots,$ in the light of $\underline{\widehat{\mathbf{Y}}} = (\widehat{Y}_1, \dots, \widehat{Y}_m)$ and $\boldsymbol{\mathcal{E}}^*$, where $\boldsymbol{\mathcal{E}}^*$ denotes the testimony of all experts. For convenience, let $\boldsymbol{\mathcal{E}}^* = (\mathcal{E}_1^*, \mathcal{E}_2^*)$, where $\mathcal{E}_1^*$ is the expert testimony of the k experts about p_1 and α, and $\mathcal{E}_2^*$ is expert testimony about Y_0. To assess $P(Y_i; \underline{\widehat{\mathbf{Y}}}, \boldsymbol{\mathcal{E}}^*)$ for $i = m + n$, we write

$$P(Y_i; \underline{\widehat{\mathbf{Y}}}, \boldsymbol{\mathcal{E}}^*) = \int_{Y_0} \int_\alpha \int_{p_1} P(Y_i | Y_0, \alpha, p_1; \underline{\widehat{\mathbf{Y}}}, \boldsymbol{\mathcal{E}}^*) \times P(Y_0, \alpha, p_1; \underline{\widehat{\mathbf{Y}}}, \boldsymbol{\mathcal{E}}^*)\, dp_1\, d\alpha\, dY_0. \quad (22)$$

Because $(Y_i \perp (\underline{\widehat{\mathbf{Y}}}, \boldsymbol{\mathcal{E}}^*) | Y_0, \alpha, p_1)$, the first term on the right side of (22) is simply

$$P(Y_i \approx y_i | Y_0, \alpha, p_1) = \begin{cases} 1 & \text{whenever } Y_0 \cdot p_1^{(1-\alpha^i)/(1-\alpha)} \approx y_i, \\ 0 & \text{otherwise,} \end{cases}$$

for all $y_i \in [0, \infty)$ [see (4)]. Thus

$$P(Y_i \approx y_i; \underline{\widehat{\mathbf{Y}}}, \boldsymbol{\mathcal{E}}^*) = \iiint_\Omega P(Y_0, \alpha, p_1; \underline{\widehat{\mathbf{Y}}}, \boldsymbol{\mathcal{E}}^*)\, dp_1\, d\alpha\, dY_0, \quad (23)$$

where Ω is the set of all values of Y_0, α, and p_1 such that

$$Y_0 \cdot p_1^{(1-\alpha^i)/(1-\alpha)} \approx y_i.$$

Now, by the multiplication rule,

$$P(Y_0, \alpha, p_1; \underline{\widehat{\mathbf{Y}}}, \boldsymbol{\mathcal{E}}^*) = P(\alpha | p_1, Y_0; \underline{\widehat{\mathbf{Y}}}, \boldsymbol{\mathcal{E}}^*) P(p_1 | Y_0; \underline{\widehat{\mathbf{Y}}}, \boldsymbol{\mathcal{E}}^*) P(Y_0; \underline{\widehat{\mathbf{Y}}}, \boldsymbol{\mathcal{E}}^*), \quad (24)$$

and because $(\alpha \perp (Y_0, \mathcal{E}_2^*) | p_1, \underline{\widehat{\mathbf{Y}}})$, $(p_1 \perp (\widehat{p}_2, \dots, \widehat{p}_m, \mathcal{E}_2^*) | \widehat{p}_1)$, and $(Y_0 \perp (\underline{\widehat{\mathbf{Y}}}, \mathcal{E}_1^*) | \mathcal{E}_2^*)$,

$$P(Y_0, \alpha, p_1; \underline{\widehat{\mathbf{Y}}}, \boldsymbol{\mathcal{E}}^*) = P(\alpha | p_1; \underline{\widehat{\mathbf{Y}}}, \mathcal{E}_1^*) P(p_1 | Y_0; \widehat{p}_1, \mathcal{E}_1^*) P(Y_0; \mathcal{E}_2^*). \quad (25)$$

Because Y_0 has a lognormal distribution with parameters $\widehat{a}$ and $\widehat{b}$,

$$P(Y_0; \mathcal{E}_2^*) = \frac{1}{\sqrt{2\pi}\, \widehat{b} Y_0} \exp\left[-\frac{1}{2} \left(\frac{\ln Y_0 - \widehat{a}}{\widehat{b}} \right)^2 \right], \qquad Y_0 \ge 0. \quad (\text{A})$$

and, by Bayes' law,

$$P(\alpha | p_1; \underline{\widehat{\mathbf{Y}}}, \mathcal{E}_1^*) \propto L(\alpha | p_1; \underline{\widehat{\mathbf{Y}}}) P(\alpha; \mathcal{E}_1^*)$$

and

$$P(p_1 | Y_0; \widehat{p}_1, \mathcal{E}_1^*) \propto L(p_1 | Y_0; \widehat{p}_1) P(p_1 | Y_0; \mathcal{E}_1^*).$$

From (13) and (21),

$$L(\alpha | p_1; \underline{\widehat{\mathbf{Y}}}) \equiv L(\alpha | p_1; \widehat{p}_2, \dots, \widehat{p}_m) = \prod_{i=2}^m \frac{1}{2\sigma \widehat{p}_i \sqrt{\pi}} \left(\frac{p_1^{\alpha^{(i-1)}}}{\widehat{p}_i} \right)^{1/(2\sigma^2)} \quad (\text{B})$$

and

$$P(\alpha; \mathcal{E}_i^*) = \frac{\Gamma(\lambda + \delta)}{\Gamma(\lambda)\Gamma(\delta)} \left(\frac{\alpha - a}{b - a} \right)^{\lambda - 1} \frac{1}{b - a} \left(\frac{b - \alpha}{b - a} \right)^{\delta - 1}, \quad a < \alpha < b. \quad (\text{B}')$$

Similarly, from (11) and (19),

$$L(p_1 | Y_0; \widehat{p}_1) = \frac{1}{\sqrt{2\pi}\, s \widehat{p}_1} \exp\left[-\frac{1}{2s^2} \left(\ln\left(\frac{\widehat{p}_1}{p_1} \right) + \frac{\sigma^2}{2} \right)^2 \right] \quad (\text{C})$$

and

$$P(p_1 | Y_0; \mathcal{E}_1^*) = \frac{1}{\sqrt{2\pi}\, \overline{\sigma} p_1 (1 - p_1)} \times \exp\left[-\frac{1}{2\overline{\sigma}^2} \left(\ln\left(\frac{p_1}{1 - p_1} \right) - \overline{\mu} \right)^2 \right]. \quad (\text{C}')$$

Equation (23) can now be written as

$$P(Y_i \approx y_i; \underline{\widehat{\mathbf{Y}}}, \boldsymbol{\mathcal{E}}^*) \propto \iiint_\Omega ((\text{C}) \cdot (\text{C}'))((\text{B}) \cdot (\text{B}'))(\text{A})\, dp_1\, d\alpha\, dY_0, \quad (26)$$

where (A), (B), (B′), (C), and (C′) are as specified here and Ω is the set of all values $0 \le Y_0 \le \infty, a \le \alpha \le b$, and $0 < p_1 \le 1$, such that

$$Y_0 \cdot p_1^{(1-\alpha^i)/(1-\alpha)} \approx y_i. \quad (27)$$

3.3 Computational Matters

Although (6) may be tackled using Markov chain Monté Carlo (MCMC), we have chosen to use numerical integration after discretization. There are two reasons for doing so. The first reason pertains to convergence of the MCMC when the underlying terms involve products of Gaussian, lognormal, and transformed beta terms; the second is that the numerical integration can be easily done on a desktop computer. Software to implement it has been written by Kong (1999) and is available (as freeware) to interested users from the author on request.

For the posterior distribution of α, we discretize α over the range [.5, 1], p_1 over the range (0, 1], and use (6), which, in the light of the expert testimony $\boldsymbol{\mathcal{E}}^*$ and the material of Section 3.2, can be compactly written as

$$P(\alpha | \underline{\mathbf{Y}}, \boldsymbol{\mathcal{E}}^*) \propto \int_{p_1} ((\text{B}) \cdot (\text{B}'))((\text{C}) \cdot (\text{C}'))\, dp_1.$$

Thus for each $\alpha_j \in [.5, 1]$ and $p_{1j} \in (0, 1]$,

$$P(\alpha \approx \alpha_j | \underline{\mathbf{Y}}, \boldsymbol{\mathcal{E}}^*) \propto \sum_{p_{1j}} ((\text{B}) \cdot (\text{B}'))((\text{C}) \cdot (\text{C}')), \quad (28)$$

with α_j replacing α and p_{1j} replacing p_1 in (B), (B′), (C), and (C′).

4. APPLICATION: DECELERATED TESTING OF AN INDUSTRIAL PRODUCT

As stated in the abstract, the proposed methodology was motivated by an example involving the decelerated testing of a building materials product. The product's identity cannot be disclosed; the data given here are actual (with scaling to provide some camouflage).

4.1 The Test Data

The testing scenario and the data were introduced in Section 2. For convenience, here we highlight some of the essentials. Let $\widehat{Y}_{ij}, j = 1, \ldots, 30$, denote the observed strengths of all of the specimens experiencing stress, C_i, $i = 1, \ldots, 8$, and let $\widehat{Y}_i$ denote the sample average of the $\widehat{Y}_{ij}$'s. The $\widehat{Y}_i$'s are 392, 354, 310, 285, 289, 287, 286, and 285, and the corresponding standard deviations are 45.1, 32.0, 43.7, 40.3, 38.2, 31.9, 39.4, and 28.1. We assume (see Sec. 3.1.2) that the $\widehat{Y}_{ij}$'s, $j = 1, \ldots, 30$, have a lognormal distribution with a scale parameter σ_i that is common for all values of i. The σ_i's, $i = 1, \ldots, 8$, turn out to be .11, .08, .13, .12, .13, .11, .14, and .10. Thus their common value σ was taken as the average of the foregoing, namely $\sigma = .12$.

4.2 Elicitation of Expert Testimony

The next step in our analysis was eliciting expert testimony on Y_0 (see Sec. 3.1.1). This entailed eliciting, from a single expert $\mathcal{E}$, the mean a^* and the median b^* of the distribution of Y_0, which we conceptualize as being lognormal; these turn out to be 430 and 419. For this expert, the modulating parameters $\overline{a}$, $\overline{b}$, and $\overline{\gamma}$ were chosen as 1, 0, and 1, so that the values of $\widehat{a}$ and $\widehat{b}$ [see (7)] turn out to be 6.06 and .227. The numbers 430 and 419 declared by $\mathcal{E}$ are based on $\mathcal{E}$'s exposure to previous data on several specimens subjected to only a background stress.

For the elicitation of expert testimony on p_1 (see Sec. 3.1.4) seven experts, $\mathcal{E}_1, \ldots, \mathcal{E}_7$, were consulted, with each asked to declare m_i and v_i, $i = 1, \ldots, 7$, their measures of location and scale for their assumed distributions of p_1. Whereas the experts had no difficulty declaring their m_i's, the v_i's were deduced by us from the expert's declared uncertainties about their m_i's as reflected in statements of the form $m_i \pm \epsilon$, where ϵ typically ranged between .03 and .10. The modulating constants a_i, b_i, and γ_i associated with an m_i and v_i were provided by a group of individuals knowledgeable about the expertise of the experts and their perceived relationships between the experts, which are encapsulated by correlations. The correlations, subjectively specified, were given as $\rho_{13} = \rho_{31} = .7$, $\rho_{14} = \rho_{41} = .7$, $\rho_{34} = \rho_{43} = .8$, and $\rho_{56} = \rho_{65} = .5$; all of the other remaining correlations were assessed as 0. In principle, one could endow prior distributions to the correlations and use the values given earlier to obtain their posterior distributions, but this again would make the analysis more elaborate. Table 1 gives the values of the m_i, the v_i, and the modulating constants a_i, b_i, and γ_i, $i = 1, \ldots, 7$.

Table 1. Expert Declarations and Their Modulating Constants for the Parameter p_1

	Declared values of E_i		Modulating constants		
Expert	m_i	v_i	a_i	b_i	γ_i
E_1	.90	.051	0	1.0	1.0
E_2	.95	.051	0	1.15	1.5
E_3	.75	.099	0	1	.9
E_4	.75	.150	0	1	.9
E_5	.70	.150	0	1	.9
E_6	.75	.099	0	.9	1.1
E_7	.70	.051	0	1	.8

Finally, for expert testimony on the degradation factor α (see Sec. 3.1.5), six of the seven experts were consulted ($\mathcal{E}_7$ was unable to participate), with each asked to declare the proportion of the background strength Y_0 retained by a specimen if it were to experience a stress of C_{10} and C_{20} cycles. Recall that in actuality, testing was done only up to C_8 cycles. The six pairs of declared values turn out to be (.5, .4), (.9, .87), (.65, .36), (.7, .65), (.4, .35), and (.4, .1). Using the fact that $\ln(.4)/\ln(.5) = (\alpha^{10} + 1)$ [a result that can be verified from (4), it is seen that for expert $\mathcal{E}_1$, $\alpha = .892$. Similarly, for the other experts, α turns out to be .892, 1, .855, .825, and 1. The average value of α is therefore .911, and its variance is .005. Setting $.911 = .5 + \lambda/(\lambda + \delta)$ and $.005 = \lambda\delta/[(\lambda + \delta)^2(\lambda + \delta + 1)]$, we find that $\lambda \approx 18$ and $\delta \approx 26$; the .5 given earlier corresponds to our choice of $a = .5$ (and $b = 1$) in (21).

4.3 The Posterior and the Predictive Analysis

In our particular application, m is 8 and n is chosen to be 2, so that, given the expert testimony and the data on strength tests conducted at eight equally spaced stress cycles, we need to predict the strength of a specimen were it to experience a stress of C_{10} cycles. This predictive distribution, based on the constants discussed in the previous two sections, is shown in Figure 3. Its mean and standard deviation turn out to be 218 and 40, with and the latter according reasonably well with the observed data. The mean is smaller than what we would expect to see, given that $\widehat{Y}_6$, $\widehat{Y}_7$, and $\widehat{Y}_8$ are 287, 286, and 285, but the shift to 218 is in the right direction and is a consequence of expert testimony, which in actuality could have erred on the conservative side. However, conservative estimates of strength correlate well with a designer's need for safety and are preferable to exaggerated estimates of strength. Also shown in Figure 3 (by the dotted curve) is a normal distribution with mean 218 and standard deviation 40; this is for the purpose of comparison. Because the C_i's are equally spaced and because C_{10} is not too different from C_8, we surmise that had we used C_1–C_7 as providing the data, and predicted strength at C_8, our predicted strength would be much lower than 285, the observed average.

For the purpose of comparison, we also obtain the predictive distributions of strength for C_{20} and C_1 cycles of stress

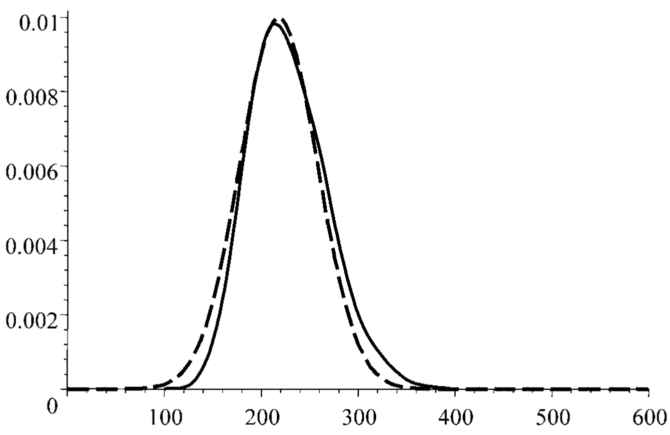

Figure 3. The Predictive Distribution of Strength Supposing C_{10} Cycles of Stress (—) and Its Comparison With a Normal Distribution With Mean 218 and Standard Deviation 40 (- - -).

to see the directions in which their means and standard deviations move, and for comparison with the observed strengths under C_1, that is, for cross-validatory purposes. For the case C_{20}, the mean and standard deviation turn out to be 216 and 39; the predicted mean affirms the thesis that the degradation of strength with stress reaches a plateau. For the case C_1, the mean and the standard deviation turn out to be 366 and 60. The predicted mean of 366 is again conservative, because $\widehat{Y}_1$ is 392. The predicted standard deviation of 60 is also conservative, given that the observed one is 45.1.

Whereas the predictive distribution of strength is the feature of interest to designers and users, a knowledge of the degradation factor α is of concern to material scientists. Based on the elicited constants discussed earlier, Figure 4 shows the posterior distribution of α, using (28). Also shown, for the purpose of comparison, is its prior distribution [see (21)]. The effect of the data on the expert testimony about α can be appreciated by comparing the solid and dotted curves of Figure 4. Specifically, we observe that the data tend to suggest a value of α smaller than that implied by the expert testimonies. Recall that small values of α suggest that the degradation of strength with stress diminishes at a faster rate than that implied by the larger values of α. This feature is borne out by the observed values $\widehat{Y}_i, i = 1, \dots, 8$, and also the predicted strengths under C_{10} and C_{20}.

4.4 Sensitivity to Inputs on Background Strength

Recall that the Bayesian engine behind the model is started off by Y_0, the background strength, for which a substantial amount of subjective information has been elicited. Consequently, a question arises abut the precipitating effects of the elicited prior of Y_0 on the subsequent analysis. This question is germane, because there are no data on Y_0. To address it, we start by noting that the direct impact of the prior of Y_0 [see (8)] is on the likelihood of p_1, as encapsulated via (11). The parameters describing this likelihood are s^2 and σ^2, where $s^2 = \sigma^2 + 2\widehat{b}^2$. Thus what matters are the values of σ^2 and $\widehat{b}$, where, with $\overline{\alpha} = \overline{\gamma} = 1$ and $\overline{b} = 0, \widehat{b} = \ln(a^*/b^*)$ [see (7)]. Now σ^2 has been empirically estimated using the observed data on strengths, and so the only subjectively elicited inputs are a^* and b^*, the mean and the median of the perceived (by us) lognormal distribution of Y_0. Because $\widehat{b} = \ln(a^*/b^*)$, the value that $\widehat{b}$ takes will be negligible compared with the value of σ^2, particularly if a^* and b^* are close to each other; in our particular case, $\widehat{b} = .034$. Thus, as is to be expected, sensitivity will become a matter of concern only if the disparity between a^* and b^* is large and the life-test data are such that the estimated σ^2 is small. Otherwise, the procedure is robust to prior inputs. What seems to contribute greatly toward making the procedure robust is a consideration of the ratios of the Y_i's via (1) and the recursive feature in (2) of the proposed model.

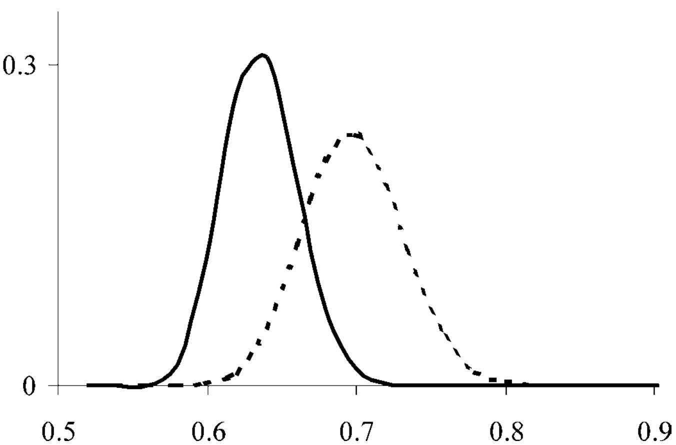

Figure 4. Prior (- - -) and Posterior (—) Distributions of the Degradation Factor α.

In principle, the question of sensitivity and robustness should also be addressed against a range of plausible models of the type given by (2). This we are unable to do, because unlike models used in accelerated life testing that have a physical basis, the model of (4) is conceptual, with simplicity and reasonableness as its main virtue. Alternate models relating the p_i's will end up being elaborate, with the resulting analysis more cumbersome than the one described here. It is questionable whether using more elaborate models will produce results that are distinctly superior to those obtained here.

5. SUMMARY, CONCLUSIONS, AND MERITS OF THE APPROACH

This article has introduced an approach for inference from decelerated tests using the degradation of strength with stress as a guiding physical phenomenon. This phenomenon has a foundation in the theory of crack growth, and the notion of strength is one that engineers can relate to. The existing approaches to stress testing (Meeker and Escobar 1998; Nelson 1990) postulate a relationship between the parameters of a failure model and stress, a strategy that would require the elicitation of expert testimony on parameters; this could be difficult. Of course, the approaches of Meeker and Escobar and of Nelson are centered around the use of data alone and thus do not call for expert testimonies. Our model for degradation is a natural one involving a single parameter α; it is encapsulated by two equations, giving it a hierarchical flavor. Our approach is based on the fusion of expert testimony and test data via Bayes' law. Even though the two-stage hierarchical model is motivated by a specific scenario, it could be viewed as being general enough for a broad-based use.

This article began with an explanation of the notions of stress, strength, damage, and threshold, notions that facilitate an appreciation of the proposed model. It ends with an application illustrating the proposed methodology. Central to this are the elicitation, modulation, and codification of expert testimonies.

The idea of using a hierarchical model to describe the degradation of strength with stress may have ramifications beyond the physical sciences that have motivated it. For example, in the biological sciences involving dose-response experiments, if an argument can be made relating, say, immunity to strength, and if the immunity can be conceptualized as a threshold, then the approach of this article could have applications in biomedical contexts as well.

Questions can be raised as to how the methods described here can be applied and why such methods would be superior to the existing methods. These questions have merit, because the proposed methodology entails heavy (but not difficult) technical machinery within which are embedded assumptions of conditional independence. Our response is that the independence assumptions are meaningful, being conditional on (unknown) constants; that the existing archetypal methods do not consider the fusion of hard data with scientific knowledge (in the form of informed testimonies); and that using the method boils down to collecting data, eliciting testimonials, and feeding this information into software packages. The complicated technical machinery underlying the foregoing is the price one must pay to formally integrate data with other types of information. Informal approaches may do just as well, but defending them could be onerous. Perhaps, when all is said and done, and as a minimum, this article could be seen as an attempt to unravel how we need to think about approaches for systematically pulling in diverse sources of information to address a common problem in reliability.

ACKNOWLEDGMENTS

The author thanks the associate editor for comments and for eliciting the testimonies of two excellent referees, one who focused on the practicalities and raised significant awareness, and the other who focused on the technical aspects of Bayesian inference and helped smooth out many details. The author also thanks Karen Kafadar for her detailed inputs to a preliminary version of this manuscript. In going through the review process of this article, the author has come to realize that there is an art to writing papers in applied statistics—an art that a referee appears to have mastered and has been generous enough to transmit to the author. This Research was supported in part by U.S. Army Research Office grants DAAD 19-01-1-0502 and DAAD 19-02-1-0195, and Office of Naval Research contract N00014-00-1-0046.

[Received October 2000. Revised October 2004.]

REFERENCES

Gordon, J. E. (1976), *The New Science of Strong Materials*, London: Penguin Books.

Kong, C. W. (1999), "DeLTA: A Software Package for Decelerated Life Testing," technical report, The Institute for Reliability and Risk Analysis, The George Washington University.

Lindley, D. V. (1983), "Reconciliation of Probability Distributions," *Operations Research*, 31, 866–880.

——— (2000), "The Philosophy of Statistics," *The Statistician*, 49, 293–337.

Lu, C. J., Meeker, W. Q., and Escobar, L. A. (1996), "Using Degradation Measures to Estimate a Time-to-Failure Distribution," *Statistica Sinica*, 6, 531–546.

Meeker, W. Q., and Escobar, L. A. (1998), *Statistical Methods for Reliability Data*, New York: Wiley.

Nelson, W. B. (1990), *Accelerated Testing: Statistical Models, Test Plans and Data Analysis*, New York: Wiley.

Singpurwalla, N. D. (2004), "The Hazard Potential," technical report, The Institute for Reliability and Risk Analysis, The George Washington University.

Whitmore, G. A., Crowder, M. J., and Lawless, J. F. (1998), "Failure Inference From a Marker Process Based on a Bivariate Wiener Process Model," *Lifetime Data Analysis*, 4, 229–251.