# MATHEMATICAL RELIABILITY: AN EXPOSITORY PERSPECTIVE

Edited by
REFIK SOYER
Department of Management Science
The George Washington University
Washington, DC 20052

THOMAS A. MAZZUCHI
Department of Engineering Management and Systems Engineering
The George Washington University
Washington, DC 20052

NOZER D. SINGPURWALLA
Department of Statistics
The George Washington University
Washington, DC 20052

**Kluwer Academic Publishers**
Boston/Dordrecht/London

)5). *Applied Reliability,* Second
ld.

. (1995). Using Degradation Data
Fluorescent Lamp Reliability, *J.*

· *for Accelerated Life Testing,*

gradation by a Wiener Diffusion
*etime Data Analysis,* 1: 307–319.

: (1997). Modeling Accelerated
on with a Time Scale Transfor-

.

*Experiments: Planning, Analysis*
v York: Wiley.

ine Procedure for Terminating an
*Sinica,* 8: 207–220.

gning a Degradation Experiment,

Chapter 9

# INTEGRATED ANALYSIS OF COMPUTER AND PHYSICAL EXPERIMENTAL LIFETIME DATA

Alyson G. Wilson
*Group D-1, Statistical Sciences*
*Los Alamos National Laboratory*
agw@lanl.gov

C. Shane Reese
*Group D-1, Statistical Sciences*
*Los Alamos National Laboratory*
reese@lanl.gov

Michael S. Hamada
*Group D-1, Statistical Sciences*
*Los Alamos National Laboratory*
hamada@lanl.gov

Harry F. Martz
*Group D-1, Statistical Sciences*
*Los Alamos National Laboratory*
hfm@lanl.gov

**Abstract** Recent advances in computational capabilities often make engineering simulations of lifetime tractable. We consider the case in which there exist lifetime data from a computational model as well as data from a physical reliability experiment. In addition, there may also exist one or more expert opinions about the expected lifetime for selected factor settings. We simultaneously analyze the combined data using a hierarchical Bayes model. In this integrated approach we recognize important differences, such as possible biases, in these experimental data and expert opinions.

We illustrate the methodology by means of an example. Hellstrand [6] designed and conducted an experiment to study the effect of three categorical design parameters on ball bearing lifetime. In addition to the lifetime data from a $2^3$ full factorial experiment, we assume the existence of computationally produced lifetimes for four of the eight factor settings for the same three factors. We also assume there are expert opinion data for seven of the eight factor settings. The integrated data are used to estimate the reliability functions for the eight factor settings. The results indicate that reliability is more precisely estimated by using this integrated data approach.

**Keywords:** Information integration, recursive Bayesian hierarchical model, reliability function.

## 1. Introduction

The purpose of a lifetime (or time-to-failure) reliability experiment is to quantify the effect of one or more factors on the lifetime of some device of interest. Recent advances in computational capabilities often make engineering simulations of lifetime tractable. For example, degradation models are commonly used in which the corresponding lifetime is the random time before the degradation reaches some critical threshold (or limiting) value. The primary reason for using computational models is that they often offer lower cost/time means to explore parameter effects than physical experiments. However, physical experimentation is often used as a means of validating computational results while at the same time adding noise effects.

We assume here that there exists related computer experimental lifetime data on the same, or a subset of the same, physical experimental factors. We can often maximize return on development costs through the integrated analysis of such hybrid experimental design data. It is also statistically efficient to fit a single, integrated model that statistically expresses the effect that the factors of interest have on lifetime.

The situation we consider is quite general in that we do not require that we have computed (or measured) lifetimes at the same factor values in both experiments. We only require that there exist some common set of factors (either all or at least some) for both experiments. For example, it is sometimes the case that a broad (or screening) computer experiment is first performed, that is followed later by a physical reliability experiment in a smaller region of particular interest of the overall computer experiment design space.

In addition, there may also exist one or more expert opinions regarding reliability. Traditional statistical approaches consider each of these sets of data separately with corresponding separate analyses and results. A compelling argument can be made that better, more powerful statistical results can be obtained if we simultaneously analyze the combined data using a recursive Bayesian hierarchical method (RBHM). As we will illustrate, the simultaneous analysis of the combined data permits the unknown factor effects to be more precisely estimated. We propose an RBHM for the integrated statistical analysis of expert opinion, physical, and computational experimental data that recognizes important differences (such as biases) in these data.



Consider the following example. In an experiment to improve the reliability of a standard ball bearing design in a certain application, Hellstrand [6] designed and conducted an experiment to study the effect of three categorical design parameters on ball bearing lifetime. Based on past experience, the inner ring heat treatment (Factor A), the outer ring osculation (the ratio of the ball diameter and the radius of the outer ring raceway) (Factor B), and the cage design (Factor C) were thought to have an effect on the performance and life of the bearing in this application. The experiment employed a $2^3$ design, given in Table 9.1, where the two levels are standard ($-$) and modified ($+$) values, respectively. The response is the natural logarithm of the observed lifetime.

*Table 9.1.* Ball Bearing Physical Experimental Design Matrix and Lifetimes

| *Run* | Factor *A* | Factor *B* | Factor *C* | log(*lifetime*) |
|---|---|---|---|---|
| 1 | $-$ | $-$ | $-$ | 2.83 |
| 2 | $+$ | $-$ | $-$ | 3.26 |
| 3 | $-$ | $+$ | $-$ | 3.22 |
| 4 | $+$ | $+$ | $-$ | 4.44 |
| 5 | $-$ | $-$ | $+$ | 2.94 |
| 6 | $+$ | $-$ | $+$ | 2.77 |
| 7 | $-$ | $+$ | $+$ | 3.04 |
| 8 | $+$ | $+$ | $+$ | 4.85 |

In addition, suppose there is a computational model for predicting bearing lifetime that also contains the same three factors. However, we assume that the effect of the modified inner ring heat treatment (Factor A, $+$ level) has not yet been implemented in the computational model. Table 9.2 contains the corresponding $2^2$ design and associated log(lifetimes) for the computer experiment.

An expert in such bearing designs was also available and her opinions regarding the effect of these three factors on lifetime were elicited. However, she was either unwilling or unable to express her opinion about the expected lifetime for the fully modified bearing design; thus, only seven factor combinations were elicited. Table 9.3 gives her corresponding log(expected lifetime), the subjective 0.90 quantile, log(q), on the lifetime, and the equivalent worth, $m$, of each opinion relative to an equivalent physical experimental result for the

*Table 9.2.* Ball Bearing Computer Experimental Design Matrix and Lifetimes

| | | Factor | | |
|---|---|---|---|---|
| *Run* | *A* | *B* | *C* | log(*lifetime*) |
| 1 | – | – | – | 2.12 |
| 2 | – | + | – | 3.07 |
| 3 | – | – | + | 2.13 |
| 4 | – | + | + | 3.07 |

seven combinations that were elicited. Further discussion regarding these three values is given in Reese et al. [10].

*Table 9.3.* Ball Bearing Expert Judgment Experimental Design Matrix and Lifetimes

| | | Factor | | | | |
|---|---|---|---|---|---|---|
| *Run* | *A* | *B* | *C* | $m$ | log(*expected lifetime*) | log($q$) |
| 1 | – | – | – | 0.5 | 2.51 | 2.58 |
| 2 | + | – | – | 0.75 | 2.83 | 3.05 |
| 3 | – | + | – | 0.5 | 2.76 | 3.29 |
| 4 | + | + | – | 0.5 | 4.17 | 4.38 |
| 5 | – | – | + | 0.5 | 2.37 | 2.43 |
| 6 | + | – | + | 1.0 | 2.59 | 2.66 |
| 7 | – | + | + | 0.75 | 2.67 | 3.02 |

We present the basics of the RBHM in Section 2, and we apply it to the above example in Section 3. Finally, we present our conclusions in Section 4.

## 2. The Basics of Data Integration Using RBHM

The design and analysis of computer experiments has evolved as the power of computers has grown (although it has certainly not kept pace!). Sacks et al. [11] provide a review of techniques used in the analysis of output from complex computer codes as well as issues for design. Latin hypercube sampling had its genesis in the design of computer experiments (McKay, Beckman, and Conover [7]). Bayesian treatment of design and analysis of computer experiments is presented in Currin et al. [1]. These papers are primarily concerned with issues when the only source of information is the output from a complex computer model.

Data integration had its genesis in the meta analytic literature. Zeckhauser [13] provides an early treatment of meta analysis. Hedges and Olkin [5] provide a nice review of meta analytic techniques. However, meta analysis has not been viewed without strong criticism (Shapiro [12] and discussion). Muller et al. [9]



present a Bayesian hierarchical modeling approach for combining case-control and prospective studies, where effects due to different studies as well as different centers are allowed.

The statistical notion of pooling data (sometimes also known as "borrowing strength") underlies the RBHM and analysis to be discussed. Modern methods used to borrow strength have their basis in hierarchical Bayes modeling. A nice introduction to both hierarchical Bayes modeling and borrowing strength is given by Draper et al. [2]. The basic idea involves the notion that, when information concerning some response of interest arises from several independent, but not identical, data sources, a hierarchical model is often useful to describe relationships involving the observed data and unobserved parameters of interest. For example, unobserved parameters might be the coefficients and error variance in an assumed regression model. Each source of data provides perhaps biased information about these parameters, in which case methods that borrow strength will be useful. The practical advantages of borrowing strength for estimating the unknown parameters will be illustrated in Section 3.

We propose fitting lifetime models using information from three distinct sources: expert opinion, computational, and physical experiments. The problem is difficult because the information sources are not necessarily all available for the same set of design points. For example, physical experiments may be performed according to a statistically designed experiment, while computer runs may be made using a different design. In addition, expert opinions may only be available at a very limited set of design points, such as the center of the statistical design region. Our goal is to combine these sources of information using an appropriately flexible integration methodology that considers (and automatically adjusts) for the uncertainties and possible biases in each of these three data sources.

Reese et al. [10] describe the RBHM that we will use here to combine the three sets of lifetime data given in Section 1. The physical experimental lifetimes, $y_p$ are assumed to follow a standard lognormal linear regression model; namely,

$$\log(y_p) = N(X_p\beta, \sigma^2 I), \tag{9.1}$$

in which $X_p$ is a known model (or design) matrix, $\beta$ is a vector of unknown coefficients that must be estimated, and the subscript $p$ denotes "physical experiment." We see that each physical lifetime is independent of the others and has common variance $\sigma^2$, which must also be estimated.

If physical experimental lifetimes were the only information source considered, this model would typically be fit using either standard least-squares regression methods (Draper and Smith [3]) or standard Bayesian linear model methods (Gelman et al. [4]. However, we want to incorporate information both from experts and computer experimental data to "improve" our estimates of $\beta$ and $\sigma^2$.

Suppose also that there are $e$ expert opinions, which do not have to be from distinct experts. The $i^{th}$ expert opinion ($i = 1, \ldots, e$) is assumed to be elicited at design point $x_i$. Each expert opinion consists of the following information:

- the expected lifetime, $y_{oi}$
- a subjective coverage probability on the physical lifetime $y_{oi}$, $\psi_i$, and the quantile associated with that probability, $q_{\psi i}$ (i.e., $P(y_{oi} \leq q_{\psi i}) = \psi_i$).

In addition, we consider the elicited "worth" of the opinion in units of equivalent physical experimental data observations, $m_{oi}^{(e)}$. In order to use these data, we need to transform these individual pieces of information into probability distributions that provide information about $\beta$ and $\sigma^2$. Assume for the moment that the three quantities above can be used to create expert opinion "data" in accordance with the following model:

$$\log(y_o) = N(X_o\beta + \delta_o, \sigma^2\Sigma_o). \tag{9.2}$$

As with the physical experimental lifetimes, the expert opinion data $y_o$ are assumed to follow a lognormal distribution. However, the mean is now $X_o\beta + \delta_o$, where $\delta_o$ is a vector of possible expert-specific location biases. The variances are also biased, and the matrix $\Sigma_o$ contains the scale biases for each expert.

Besides location biases, in which an expert's average value is high or low relative to the true mean, scale biases often occur due to information over-valuation and are well-documented in the elicitation literature. For example, an expert may be asked to provide what they think is a 0.90 quantile but which in reality is actually only a 0.60 quantile (Meyer and Booker [8]. Such over-valuation of information may be expressed in the model as the scale bias parameter $\Sigma_o$. Although responses from experts can be correlated by having non-diagonal elements in $\Sigma_o$, we consider only the case of uncorrelated responses. Thus,

$$\Sigma_0 = \begin{bmatrix} 1/k_{o_1} & 0 & \cdots & 0 \\ 0 & 1/k_{o_2} & 0 & \cdots \\ \vdots & 0 & \ddots & \cdots \\ 0 & \cdots & \cdots & 1/k_{o_e} \end{bmatrix}.$$

Reese et al. [10] describe how the three expert-elicited quantities above (for each expert opinion) can be described by the model given in (9.2).

Now consider an analogous model for the computationally-produced lifetime data; namely,

$$\log(y_c) = N(X_c\beta + \delta_c, \sigma^2\Sigma_c). \tag{9.3}$$

In addition to possible location biases, computer experimental lifetime data are likely to have scale biases, as these data usually tend to be less variable



than physical lifetime data; in fact, there is often no stochastic variability for given values of the factors, as a computer code is often deterministic. The variability occurs relative to the assumed model. Another reason for the reduced variability relative to physical experimental lifetimes is that we know that not all factors generating the physical lifetimes are incorporated into the computer code—perhaps all of the factors causing variability are unknown. Although we consider biases only in the intercept term of this model, more general bias structures for the parameters can also be considered.

The RBHM provides a convenient way to sequentially integrate data. We begin by assigning informative but diffuse priors on all the unknown model parameters including the biases. These priors are then updated with the expert opinion data using Bayes theorem to form Stage 1 posterior distributions. These Stage 1 posteriors are then likewise updated using the computationally-produced lifetimes to form Stage 2 posteriors. At Stage 2, these posteriors represent the combined use of only the expert opinion and computational data. Finally, these posteriors become the priors for Stage 3 and these are again updated using the physical experimental lifetimes to produce the final posterior distributions of interest. In this way, all the available data are recursively used within the model context to successively (and more precisely) estimate the desired effects. Although these calculations cannot be done in closed form, they can be accomplished using Markov Chain Monte Carlo (MCMC) simulation. Reese et al. [10] describes the details of these three steps and also provides more general information on MCMC and the particular Metropolis-Hastings algorithm used.

## 3. Ball Bearing Example

Hellstrand [6] describes a $2^3$ experiment to improve the reliability of a standard ball bearing design. As stated in Section 1, the inner ring heat treatment (Factor A), the outer ring osculation (Factor B), and the cage design (Factor C) were thought to have an effect on the performance and life of the bearing in this application. We analyzed the data from Tables 9.1–9.3 to illustrate the RBHM framework described in Section 2.

In particular, the $X\beta$ terms in Equations 9.1–9.3 are parameterized with a "grand mean" and linear treatment effects. The results are presented in Table 9.4. This table contains the maximum likelihood estimates for the $\beta$s and $\sigma^2$, fit with only the physical experimental data (Table 9.1), in the ML column. The confidence intervals for the ML estimates are not presented as they are not directly comparable to the RBHM estimates. In particular, expert judgment is used to develop an informative prior for $\sigma^2$, as detailed in Reese et al. [10]. Table 9.4 also contains posterior means and 95% highest posterior density (HPD) regions for the physical experimental data and the combined experimental data

(Tables 9.1–9.3). Notice that the HPD regions for the multiple data sources combined using the RBHM are smaller than for the physical experimental data alone. Figures 9.1 *(a)* and *(b)* illustrate the same change in precision by displaying the prior distribution and posterior distributions using physical experimental and combined data for $\sigma^2$ and $\beta_4$.

*Table 9.4.* Posterior Distributions for Physical and Combined Data

| *Parameter* | *Physical Post. Mean* | *Physical 95% HPD* | *Combined Post. Mean* | *Combined 95% HPD* | *ML* |
|---|---|---|---|---|---|
| $\beta_0$ | 3.40 | (3.03,3.81) | 3.36 | (3.05,3.65) | 3.42 |
| $\beta_1$ | 0.40 | (0.035,0.78) | 0.43 | (0.15,0.74) | 0.41 |
| $\beta_2$ | 0.46 | (0.091,0.82) | 0.50 | (0.21, 0.77) | 0.47 |
| $\beta_3$ | -0.010 | (-0.37,0.37) | -0.027 | (-0.31, 0.27) | -0.017 |
| $\beta_4$ | 0.34 | (-0.070,0.74) | 0.31 | (0.0042, 0.59) | 0.35 |
| $\beta_5$ | 0.077 | (-0.32,0.48) | 0.062 | (-0.20, 0.33) | -0.0016 |
| $\beta_6$ | -0.00078 | (-0.40,0.35) | 0.0087 | (-0.27, 0.31) | 0.076 |
| $\sigma^2$ | 0.30 | (0.13,0.65) | 0.22 | (0.12, 0.41) | 0.17 |

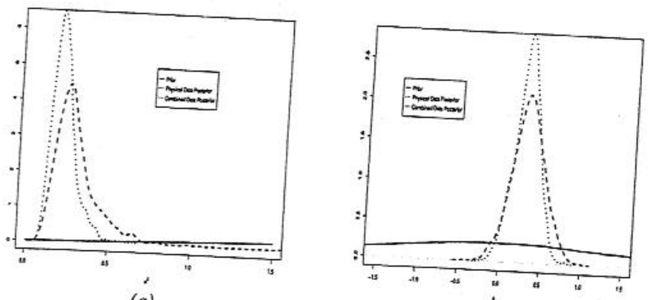
Figure 9.2 illustrates the prior and posterior distributions for $\delta_{c4}$, the location bias for the fourth computer observation. Although there is very little data to estimate this parameter, by "borrowing strength" through the model, the posterior mean (-0.14) has shifted left.

Figures 9.3 *(a)* and *(b)* are median reliability functions for the standard settings and modified settings, respectively, with corresponding 95% HPD regions. Notice for each plot that the probability bands are smaller when more data are incorporated. Figure 9.4 plots the reliability functions using the combined data for the standard and modified settings on the same scale. For example, the probability of the lifetime exceeding 80 hours with the standard settings is 0.0065, with a 95% HPD of (0,0.062); with the modified settings is 0.68, with a 95% HPD of (0.12,0.99).

## 4. Conclusions

We have presented an RBHM that can be used to combine expert opinions, computationally-produced lifetimes, and physically observed lifetimes in an experimental design setting. Available expert opinion data are used to "sharpen" the initial informative, but diffuse, prior distributions on the unknown coefficients, biases, and prior parameters. The example results clearly show that significantly more precise estimates of the factor effects and error variance can be obtained using this method. In addition, the marginal posterior distributions of the computer model biases can be used as diagnostic indicators for assessing the validity of the computational model. That is, the more the location and scale





(a) (b)

*Figure 9.1.* Prior and Posterior Distributions for *(a)* $\sigma^2$ and *(b)* $\beta_4$

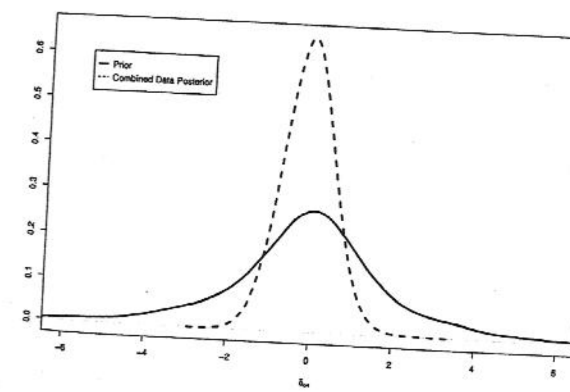
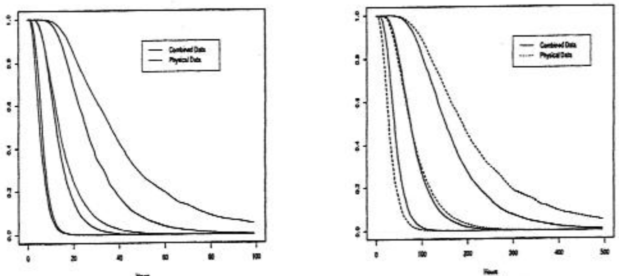



*Figure 9.2.* Prior and Posterior Distributions for a Computer Experiment Location Bias



(a) (b)

*Figure 9.3.* Reliability Function, *(a)* All Standard Settings and *(b)* All Modified Settings

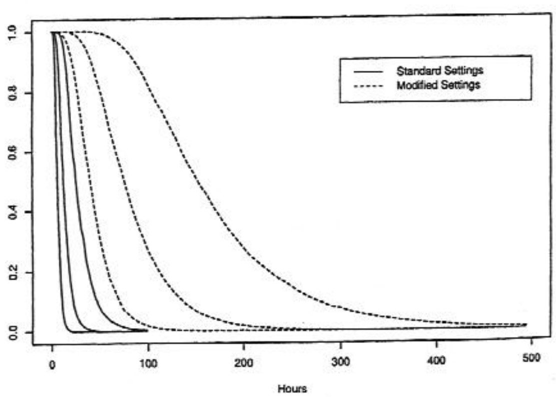



*Figure 9.4.* Reliability Functions for the Combined Data



bias posteriors overlap 0 and 1, respectively, the more valid the computational model.

Information from more than three sets of such data can likewise easily be combined by continued use of the RBHM, once for each data set. Finally, biases that are not of particular interest can simply be marginalized; that is, averaged out of the analysis using their respective prior distributions. Although we have considered categorical factors here, the use of a linear regression model permits more complicated mixed integrated models to be analyzed (see Reese et al. [10]).

## References

[1] Currin, C., Mitchell, T. J., Morris, M. D., and Ylvisaker, D. (1991). Bayesian Prediction of Deterministic Functions, With Applications to the Design and Analysis of Computer Experiments, *Journal of the American Statistical Association*, 86: 953–963.

[2] Draper, D., Gaver, D. P., Goel, P. K., Greenhouse, J. B., Hedges, L. V., Morris, C. N., Tucker, J. R., and Waternaux, C. M. (1992). Chapter 4: Selected Statistical Methodology for Combining Information (CI), in *Combining Information: Statistical Issues and Opportunities for Research*, Washington, DC: National Research Council.

[3] Draper, N. R. and Smith, H. (1981). *Applied Regression Analysis*, New York: John Wiley.

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

[5] Hedges, L. V. and Olkin, I. (1987). *Statistical Methods for Meta Analysis*, New York: John Wiley.

[6] Hellstrand, C. (1989). The Necessity of Modern Quality Improvement and Some Experience with its Implementation in the Manufacture of Rolling Bearings, *Philosophical Transactions of the Royal Society of London, Series A, Mathematical and Physical Sciences*, 347: 1596, 529–535.

[7] McKay, M. D., Beckman, R. J., and Conover, W. J. (1979). A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code, *Technometrics*, 21: 239–245.

[8] Meyer, M. A. and Booker, J. M. (1991). *Eliciting and Analyzing Expert Judgement: A Practical Guide*, New York: Academic Press.

[9] Muller, P., Parmigiani, G., Schildkraut, J., and Tardella, L. (1999). A Bayesian Hierarchical Approach for Combining Case-Control and Prospective Studies, *Biometrics*, 55: 858–866.

[10] Reese, C. S., Wilson, A. G., Hamada, M. S., Martz, H. F., and Ryan, K. J. (2001). Integrated Analysis of Computer and Physical Experiments, to appear in *Technometrics*.

[11] Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989). Design and Analysis of Computer Experiments, *Statistical Science*, 4: 409–423.

[12] Shapiro, S. (1994). Meta-analysis/Shmeta-analysis," *American Journal of Epidemiology*, 140: 771–791.

[13] Zeckhauser, R. (1971). Combining Overlapping Information, *Journal of the American Statistical Association*, 66: 91–92.

IV

# THE FAILURE RATE FUNCTION AND BURN-IN