How I Do Research

WESEP 594
Real-Time Research Seminar
Fall 2013

William Q. Meeker
Department of Statistics
Center for Nondestructive Evaluation
Iowa State University

Overview

- My background and areas of research
- Where do my research problems come from?
- A PhD dissertation in Statistics
- Examples of my research (actually my former students' research)
- Research as a process
- Concluding remarks

My Background

- BS in Industrial Management, Clarkson College of Technology, 1972
- MS in Operations Research, Union College, 1973
- PhD in Administrative and Engineering Systems (dissertation in Statistics), 1975
- Intern GE Corporate Research and Development summers 1973-1975
- At ISU since August 1975. Working on CNDE projects since 1989.
- Visitor, GE Global Research Center 1976-1977, 1992-2004
- Visiting professor, AT&T Bell Laboratories, summers 1978-1992 working in telecommunications reliability
- Faculty affiliate, Los Alamos National Laboratory, 1999-present
- Consultant for various companies and organizations on problems in applied statistics (especially the analysis and interpretation of reliability and related data)

Some Notable Accomplishments

- Six co-authored papers emanated from my GE CRD internship experiences
- Developed FORTRAN software for reliability data analysis 1975-1988
- Co-authored book on *Statistical Intervals*, published in 1991 (revision expected in 2015)
- Co-authored book *Statistical Methods for Reliability Data*, 1998 (revision and new advanced books due 2016 and 2017, respectively)
- Developed Splus/R based software for reliability data analysis 1998-present
- Approximately 200 publications (approximately 120 in refereed archival journals)

My Current Research Areas

- Statistical Methods in Reliability
  - Accelerated test planning and analysis
  - Analysis of field failure data
  - Failure modeling and prediction
- Nondestructive Evaluation (NDE)
  - Planning experiments to evaluate NDE systems
  - Estimation of probability of detection (POD)
- Combining Physical and Statistical Modeling

Where Do My Research Problems Come From?

- Problems that arise when consulting with "Industry"
- Problems that arise through other contacts with "Industry"
- On campus collaboration
- Resolving technical issues that arise in practical applications and previous research
- Important extensions of previous work (almost all of my research papers end with a section "Areas for Further Research")

A PhD Dissertation in Statistics

- Depending on a student’s advisor, the dissertation may be highly abstract and theoretical or very applied. My students’ dissertations are somewhere in between.
- Most Statistics students use the “Alternative Dissertation Format” where most chapters correspond to research papers that have been or are about to be submitted for publication (between 2 and 7 papers)
- Personal: I like to have at least one of the papers (preferably more) driven by a motivating application

The Annals of Applied Statistics
2009, Vol. 3, No. 2, 857–879
DOI: 10.1214/00-AOAS231

PREDICTION OF REMAINING LIFE OF POWER TRANSFORMERS BASED ON LEFT TRUNCATED AND RIGHT CENSORED LIFETIME DATA

BY YILI HONG, WILLIAM Q. MEEKER AND JAMES D. MCCALLEY

Iowa State University

Prediction of the remaining life of high-voltage power transformers is an important issue for energy companies because of the need for planning maintenance and capital expenditures. Lifetime data for such transformers are complicated because transformer lifetimes can extend over many decades and transformer designs and manufacturing practices have evolved. We were asked to develop statistically-based predictions for the lifetimes of an energy company's fleet of high-voltage transmission and distribution transformers. The company's data records begin in 1980, providing information on installation and failure dates of transformers. Although the dataset contains many units that were installed before 1980, there is no information about units that were installed and failed before 1980. Thus, the data are left truncated and right censored. We use a parametric lifetime model to describe the lifetime distribution of individual transformers. We develop a statistical procedure, based on age-adjusted life distributions, for computing a prediction interval for remaining life for individual transformers now in service. We then extend these ideas to provide predictions and prediction intervals for the cumulative number of failures, over a range of time, for the overall fleet of transformers.

1. Introduction.

Supplementary materials for this article are available online. Please click the Technometrics link at http://pubs.amstat.org.

Field-Failure and Warranty Prediction Based on Auxiliary Use-Rate Information

Yili Hong

Department of Statistics
Virginia Polytechnic Institute
and State University
Blacksburg, VA 24060
(*yilihong@vt.edu*)

William Q. Meeker

Department of Statistics
Iowa State University
Ames, IA 50011

Assessment of risk due to product failure is important both for purposes of finance (e.g., warranty costs) and safety (e.g., potential loss of human life). In many applications a prediction of the number of future failures is an important input to such an assessment.

Usually the field-data response used to make predictions of future failures is the number of weeks (or another unit of real time) in service. Use-rate information usually is not available (automobile warranty data are an exception, where both weeks in service and number of miles driven are available for units returned for warranty repair). With new technology, however, sensors and smart chips are being installed in many modern products ranging from computers and printers to automobiles and aircraft engines. Thus the coming generations of field data for many products will provide information on how the product was used and the environment in which it was used. This article was motivated by the need to predict warranty returns for a product with multiple failure modes. For this product, cycles-to-failure/use-rate information was available for those units that were connected to the network. We show how to use a cycles-to-failure model to compute predictions and prediction intervals for the number of warranty returns. We also present prediction methods for units not connected to the network. To provide insight into the reasons that use-rate models provide better predictions, we also present a comparison of asymptotic variances comparing the cycles-to-failure and time-to-failure models. This article has supplementary material online.

KEY WORDS: Calibration; Cycles to failure; Multiple failure modes; Prediction intervals; Product reliability; Risk analysis.

Research Article

Applied Stochastic Models in Business and Industry

Received 25 September 2008, Revised 3 January 2010, Accepted 6 January 2010 Published online in Wiley Interscience

(www.interscience.wiley.com) DOI: 10.1002/asmb.830

The importance of identifying different components of a mixture distribution in the prediction of field returns

Yili Hong[a,*,†] and William Q. Meeker[b]

Data from a mixture of distributions with two different increasing hazard functions can behave, over some period of time, like a distribution with decreasing hazard functions. As a result, reliability predictions based on data from a mixture of units with two or more different physical designs could be seriously wrong if the pooled data are used to extrapolate in time. Thus, it is important to identify components of the mixture and do statistical inference based on the stratified data. In this paper, the importance of this principle is investigated analytically and illustrated with lifetime data on high-voltage power transformers. From engineering knowledge, the lifetime distribution of a transformer has an increasing hazard due, largely, to insulation aging. However, data from a population of units could indicate a decreasing hazard due to a mixture of units with different designs or environmental conditions. Comparisons are made between the predictions based on the pooled-data and stratified-data models and the importance of correct stratification in practice is shown. Some suggestions for practitioners are also given.

Keywords: hazard function; maximum likelihood; stratification; transformer reliability; Weibull

Accelerated Destructive Degradation Test Planning

Ying Shi

Department of Statistics
Iowa State University
Ames, IA 50011
(yshi@iastate.edu)

Luis A. Escobar

Department of Experimental Statistics
Louisiana State University
Baton Rouge, LA 70803
(luis@lsu.edu)

William Q. Meeker

Department of Statistics
Iowa State University
Ames, IA 50011
(wqmeeker@iastate.edu)

Accelerated destructive degradation tests (ADDTs) provide reliability information quickly. An ADDT plan specifies factor-level combinations of an accelerating variable (e.g., temperature) and evaluation time and the allocations of test units to these combinations. This article describes methods for finding good ADDT plans for an important class of destructive degradation models. First, a collection of optimum plans is derived. These plans minimize the large sample approximate variance of the maximum likelihood (ML) estimator of a specified quantile of the failure-time distribution. The general equivalence theorem is used to verify the optimality of these plans. Because an optimum plan is not robust to the model specification and the planning information used in deriving the plan, a more robust and useful compromise plan is proposed. Sensitivity analyses show the effects that changes in sample size, time duration of the experiment, levels of the accelerating variable, and misspecification of the planning information have on the precision of the ML estimator of a failure-time quantile. Monte Carlo simulations are used to evaluate the statistical characteristics of the ADDT plans. The methods are illustrated with an application for an adhesive bond.

KEY WORDS: Compromise accelerated destructive degradation test plan; General equivalence theorem; Large-sample approximate variance; Monte Carlo simulation; Optimum accelerated destructive degradation test plan; Reliability.

1. INTRODUCTION

1.1 Motivation

Manufacturers often conduct up-front reliability tests on materials and components as their products are being designed. Because degradation data provide more information on reliability than traditional failure-time data (where time to failure is the response), especially in applications where few or no failures are expected, degradation tests are used in manufacturing industries to obtain the reliability information of product components and materials. For most applications, however, degradation rates at normal use conditions are so low that appreciable degradation will not be observed in a test of practical time length. For this reason, degradation tests often are accelerated to get reliability information more quickly. Generally, information from tests conducted with high levels of accelerating variables is extrapolated to obtain estimates of lifetime or degradation rates at lower, normal use conditions based on a physically reasonable statistical model.

1.2 Accelerated Destructive Degradation Test

For some applications, the degradation measurement process destroys or changes the physical/mechanical characteristics of test units, so that only one meaningful measurement can be taken on each unit. An accelerated degradation test with such degradation data is called an *accelerated destructive degradation test* (ADDT).

Escobar et al. (2003) described an application of an ADDT to evaluate an adhesive bond (designated adhesive bond B). The response was the strength (in Newtons) of the adhesive bond over time. The measurement process was destructive because the strength of a test unit could be measured only once. In addition, there was special interest in estimating the time at which 1% of the devices would have a strength less than 40 Newtons when operating at room temperature of 25 °C (i.e., the 0.01 quantile of the failure-time distribution). To obtain information about the 0.01 quantile of the failure-time distribution, these authors used an ADDT. As a baseline, 8 units with no aging were measured at the start of the experiment. A total of 80 additional units were aged and measured according to the temperature and time schedule presented in Table 1.

1.3 Related Literature

Nelson (1981; 1990, chap. 11) introduced basic models and methods for analyzing ADDT data. Escobar et al. (2003) provided an application for accelerated destructive degradation data and introduced a more general class of models. There is

TECHNOMETRICS, FEBRUARY 2009, VOL. 51, NO. 1
DOI 10.1198/TECH.2009.0001

Methods for Planning Repeated Measures Degradation Studies

Brian P. Weaver

Statistical Sciences Group
Los Alamos National Laboratory
Los Alamos, NM 87545
(theguz@lanl.gov)

William Q. Meeker

Department of Statistics
Iowa State University
Ames, IA 50010
(wqmeeker@iastate.edu)

Luis A. Escobar

Department of Experimental Statistics
Louisiana State University
Baton Rouge, LA 70803
(luis@lsu.edu)

Joanne Wendelberger

Statistical Sciences Group
Los Alamos National Laboratory
Los Alamos, NM 87545
(joanne@lanl.gov)

Repeated measures degradation studies are used to assess product or component reliability when there are few or even no failures expected during a study. Such studies are often used to assess the shelf life of materials, components, and products. We show how to evaluate the properties of proposed test plans. Such evaluations are needed to identify statistically efficient tests. We consider test plans for applications where parameters related to the degradation distribution or the related lifetime distribution are to be estimated. We use the approximate large-sample variance–covariance matrix of the parameters of a mixed effects linear regression model for repeated measures degradation data to assess the effect of sample size (number of units and number of measurements within the units) on estimation precision of both degradation and failure-time distribution quantiles. We also illustrate the complementary use of simulation-based methods for evaluating and comparing test plans. These test-planning methods are illustrated with two examples. We provide the R code and examples as supplementary materials (available online on the journal web site) for this article.

KEY WORDS: Aging and degradation; Degradation distributions; Lifetime distributions; Repeated measures planning.

1. INTRODUCTION

1.1 Motivating Examples

Engineers often need to quantify the failure-time distribution of highly reliable items. Traditional life tests, where the response is time to failure, typically yield few or no failures. Instead, engineers can sometimes use methods that measure the degradation of an item, providing more information than the traditional life tests. One such method is to use nondestructive repeated measurements over time on the degradation of each item. Given a degradation model and a relationship between degradation and failure, a failure-time distribution can be established. Before the test is performed, however, the engineers need to decide how many items should be measured and how often should these measurements be made to achieve a certain level of precision.

This work is motivated by two different applications that we have encountered. The first application involved a long-term shelf-life study on the chemical degradation of a certain compound in a particular environment. A sample of 12 items were randomly selected from a much larger population of items in storage. The engineers would then make annual measurements of the concentration of the chemical compound in units of parts per million volume (ppmv). Because of the importance of the application, the available data would be analyzed and a summary report would be prepared annually. Since the data were sensitive and not available for release, Figure 1 shows data that were simulated on a modified scale to mimic the original study. The question asked by the engineers was, "Given the pattern of the observations in Figure 1 (from a previous similar study), how should the next shelf-life study be performed?"

The second application involves a study involving inkjet printer heads. The engineers involved in this example were interested in performing a system reliability study for which the printheads were a component. The engineers wanted an estimate of the failure-time distribution where failure time depends on the degradation level of the printhead. Degradation was defined to be the amount of diffusion of an ink-related substance in the printheads. As time progresses, if this substance reaches a certain location in the printhead, a failure will soon follow. In the experiment, measurements were taken periodically on a sample of 12 units. At each inspection time, the units were measured to determine how far this substance had moved (in millimeters) after a certain amount of time. Figure 2 shows a scatterplot of the printhead degradation data. Again, the data were scaled to

TECHNOMETRICS, MAY 2013, VOL. 55, NO. 2
DOI: 10.1080/00401706.2012.715838

Supplementary materials for this article are available online. Please go to http://www.tandfonline.com/r/TECH

Physical Model-Assisted Probability of Detection of Flaws in Titanium Forgings Using Ultrasonic Nondestructive Evaluation

Ming Li

Applied Statistics Lab
GE Global Research
Niskayuna, NY 12309
(*limi@ge.com*)

William Q. Meeker

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

R. Bruce Thompson

Department of Aerospace Engineering and Center for Nondestructive Evaluation
Iowa State University
Ames, IA 50011
Q1 (*xxx*)

Nondestructive evaluation is used widely in many engineering and industrial areas to detect defects or flaws such as cracks inside parts or structures during manufacturing or for products that need to be inspected while in service. The commonly used standard statistical model for such data is a simple empirical linear regression between the (possibly transformed) signal response variables and the (possibly transformed) explanatory variable(s) such as defect size. For some applications, such a simple empirical approach is inadequate. An important alternative approach is to use knowledge of the physics of the inspection process to provide information about the underlying relationship between the response and the explanatory variable or variables. Use of such knowledge can greatly increase the power and accuracy of the statistical analysis and enable, when needed, limited extrapolation outside the range of the observed explanatory variables. This article describes a set of physical model-assisted analyses to study the capability of two different ultrasonic testing inspection methods to detect synthetic hard alpha inclusion defects in titanium forging disks. Supplementary materials for this article are available online.

KEY WORDS: Bayesian analysis; Censored data; Extrapolation; Hard alpha inclusion; Kirchhoff approximation; Mixed effects; Ultrasonic testing.

1. INTRODUCTION

1.1 Background

Nondestructive evaluation (NDE) is used to characterize the status or properties of components or structures without causing any permanent physical damage. The aerospace industry is one important NDE application area where failing to detect defects inside engine or airframe components can lead to disasters (see e.g., NTSB/AAR-89/03 1989; NTSB/AAR-90/06 1990). In virtually all NDE applications, there are random effects and errors involved in the measurements and statistical models are needed to analyze the NDE datasets. Sources of variability in NDE studies include operator effects, differences in samples used for calibration, and flaw morphology. MIL-HDBK-1823A (2009) described the standard statistical approaches used in NDE studies. Given a sufficient amount of data over an appropriate region of interest for the explanatory variables (e.g., flaw size and depth), simple empirical statistical models are often adequate to describe the relationship between the response and the explanatory variables. In many applications, however, including the one that motivated this research, the available data are not sufficient to address the questions that need to be answered. Under such circumstances, a physics-based statistical model can sometimes be used to extract the needed information from the limited data. In addition, the physics-based model, when compared to commonly used empirical models, provides a more solid basis for limited extrapolation outside the range of the available data. Of course even physics-based models are only approximations and may be incorrect if faulty assumptions are made. Thus, care is always needed when extrapolating.

As exemplified in the literature by NTSB/AAR-90/06 (1990), hard alpha inclusions in titanium alloy aircraft engine disks can lead to serious accidents. A hard alpha inclusion is a brittle nitrogen-based contamination that could cause fatigue cracks to form and grow more rapidly than what would be otherwise expected in the usually ductile titanium alloy. To develop better NDE tools for detection of hard alpha inclusions, a synthetic inclusion forging disk (known as the SID) was fabricated (details are given in Section 6.2 of the literature by Margetan et al. 2007). The SID contains numerous types of synthetic hard alpha (SHA) inclusions and flat bottom holes (FBHs) of different known

TECHNOMETRICS, xxxx 2013, VOL. 00, NO. 00
DOI: 10.1080/00401706.2013.818580

Statistical Methods for Estimating the Minimum Thickness Along a Pipeline

Shiyao Liu
Department of Statistics
Iowa State University
lsy641@iastate.edu

William Q. Meeker
Department of Statistics
Iowa State University
wqmeeker@iastate.edu

July 23, 2013

Abstract

Pipeline integrity is important because leaks can result in serious economic or environmental losses. Inspection information from a sample of locations along the pipeline can be used to estimate corrosion levels. The traditional parametric model method for this problem is to estimate parameters of a specified corrosion distribution and then to use these parameters to estimate the minimum thickness in a pipeline. Inferences using this method are, however, highly sensitive to the distributional assumption. Extreme value modeling provides a more robust method of estimation if a sufficient amount of data is available. For example, the block-minima method produces a more robust method to estimate the minimum thickness in a pipeline. To use the block-minima method, however, one must carefully choose the size of the blocks to be used in the analysis. In this paper we use simulation to compare the properties of different models for estimating minimum pipeline thickness, investigate the effect of using different size blocks, and illustrate the methods using pipeline inspection data.

Key Words: Block minima; Estimation; Extreme value; Maximum likelihood; Simulation.

Using Degradation Models to Assess Pipeline Life

Shiyao Liu, William Q. Meeker

Department of Statistics

Iowa State University

September 11, 2013

Abstract

Longitudinal inspections of pipeline thickness at particular locations along the pipeline provide useful information to assess the lifetime of the pipeline. In applications with different mechanisms of corrosion processes, we have observed various types of general degradation paths. In one application, we used a degradation model to describe the corrosion initiation and growth behavior in the pipeline, and employed a Bayesian approach for parameter estimation for the degradation model. We also built a hierarchical model to quantify the pipeline corrosion rate for similar circuits within a single facility, under the assumption that the corrosion rates at particular locations are constant over time within a circuit in the facility. The failure-time and remaining lifetime distributions are derived from the degradation model, and we compute Bayesian estimates and credible intervals of the failure-time and remaining lifetime distribution.

Key Words: Bayesian; longitudinal data; pipeline reliability.

Understanding and Addressing the Unbounded "Likelihood" Problem

Shiyao Liu, Huaiqing Wu, William Q. Meeker

Department of Statistics

Iowa State University

August 2, 2013

Abstract

The joint probability density function, evaluated at the observed data, is commonly used as the likelihood function to compute maximum likelihood estimates. For some models, however, there exist paths in the parameter space along which this density-approximation likelihood goes to infinity and maximum likelihood estimation breaks down. In applications, all observed data are discrete due to the round-off or grouping error of measurements. The "correct likelihood" based on interval censoring can eliminate the problem of an unbounded likelihood. This paper categorizes the models leading to unbounded likelihoods into three groups and illustrates the density breakdown with specific examples. We also study the effect of the round-off error on estimation, and give a sufficient condition for the joint density to provide the same maximum likelihood estimate as the correct likelihood, as the round-off error goes to zero.

Key words: Density approximation; Interval censoring; Maximum likelihood; Round-off error; Unbounded likelihood.

Research as a Process

- Problem definition: What is the question that needs to be answered?
- What related things have been done previously (literature search)?
- What data are available (getting appropriate data can be a challenge)?
- What assumptions are needed? What is an appropriate model?
  - Describe the data
  - Answer the problem question
- What is an appropriate level of abstraction?
- What algorithms will be needed? How will they be implemented (R, C, FORTRAN)?
- Begin formalization and writing up rough notes early (also helps communication)
- Validation of statistical methods
  - Large-sample theory
  - Monte Carlo simulation
- Write-up, presentations, and publication(s)
- Areas for future related research
- Feedback and revision

Research Interpersonal Relationships with Students and Collaborators

- Initial ideas for my students’ research projects often come from my experiences
- Start with something simple (get experience, insight, build confidence), but be planning for appropriate extensions
- Students are sometimes asked to abstract, generalize, extend, and improve incomplete things I started on my own or with other collaborators
- When I enter a new area, I like to find an experienced collaborator to work with
- I generally schedule hour-long face-to-face meetings with each my students each week, but invite them to contact me anytime if they have a question that cannot wait
- I encourage my students to consult with each other on technical matters, when useful and appropriate
- Email communication between meetings is strongly encouraged

Other Ideas and Suggestions That Have Worked for Me

- I prefer to work on real problems, because of the motivation and knowing that the probability for impact is high
- For many of us, some hours of the day are better for creative thinking that others. Try to understand what are your best hours and use them accordingly.
- Reserve big blocks of time to do research, but take frequent breaks
- When at conferences seek out individuals who might have interesting problems to discuss and engage them
- Some things that work for me:
  – Think about a challenging problem before going to bed. Continue thinking about it while going to sleep.
  – If you wake up and have a good idea in mind, get up and write it down.
  – Thinking about a challenging problem before certain kinds of physical exercise can also be productive.

Concluding Remarks

- The most difficult problems arising in research require broad knowledge, creativity, and take time
- Certain parts of the research process can be highly frustrating (getting stuck on a problem for a long time), but with success comes great satisfaction. Sometimes it is good to take a break for a while and work on things where progress will be more rapid.
- Research should be fun. Having a project that is of high interest to you helps make the research fun.