Field Failure Prediction Using Dynamic Environmental Data

Yili Hong[1] and William Q. Meeker[2]

[1] Department of Statistics, Virginia Tech, Blacksburg, VA, 24060 yilihong@vt.edu
[2] Department of Statistics, Iowa State University, Ames, IA, 50011 wqmeeker@iastate.edu

Summary. Due to the dynamics of the environment and the variability on the product usage, product units in the field are usually exposed to varying failure-causing stresses. Some products are equipped with sensors and smart chips that measure and record usage/environmental information over the life of the product. For some products, it is possible to track environmental variables dynamically, even in real time, providing useful information for field-failure prediction. In many applications, predictions are needed for individual units, giving the remaining life of individuals, and for the population, giving the cumulative number of failures at a future time. It is always desirable to obtain more accurate predictions for both the population and the individuals. This paper outlines a model and methods that can be used for field-failure prediction using dynamic environmental data. Multivariate time series models are also used to describe the dynamic covariate information. The cumulative exposure model is used to link the explanatory variables which are recorded as a multivariate time series, and the failure-time model.

Key Words: Cumulative exposure model; covariate process; failure time data; multivariate time series; reliability; usage history.

1 Introduction

1.1 Background

Laboratory tests that are conducted to obtain product reliability information are often done under a constant stress. Product units in the field, however, are usually exposed to varying failure-causing stresses due to the dynamics of the environment and the variability in product usage. These variations include environmental variables such as temperature, humidity, vibration, UV intensity and spectrum, and usage variables such as loading and use rate, which vary from unit to unit and over time within each unit.

For some products, it is possible to track environmental variables dynamically, even in real time. This usage/environmental information can be obtained from sensors and smart chips that are installed in a product to measure and record such information over the life of the product. For products that are connected to a network or installed with a wireless transmission device, such information is available dynamically or periodically. For products that are not connected to a network, this information is available at the time of product inspection, return, or repair.

Several examples of products/systems that provide dynamic information are as follows.

- OnStar™ [OnS09] is an in-vehicle safety and security system created to help protect automobile occupants. The system consists of various sensors and has the ability to communicate vehicle information to the driver as well as to a central location, via a satellite wireless connection. The system also collects usage and environmental information and, with the vehicle owner's permission, transmits this information periodically to the central location.
- Large medical systems, such as CT scanners, have sensors and devices that can provide real-time system information to those who do system maintenance.
- Hahn and Doganaksoy [HD08, Section 9.9] describe an application involving modern locomotive engines installed with sensors that indicate operating status variables such as oil pressure, oil temperature, and water temperature. Such information is automatically recorded and transmitted to a central location and can be used to shutdown an engine, should a dangerous condition arise. Aircraft subsystems also have similar sensors.
- High-voltage power transformers can be monitored by an automatic dissolved gas analyzer (DGA) system (e.g., [STW+05]). DGA automatically performs periodic analyses (typically every hour) to indicate the presence of different kinds of dissolved gases in the transformer insulating oil and moisture content. Certain combinations of gas mixtures are known to be a precursor of a failure event. In addition, the DGA system reports real-time dynamic loading and thermal information. This information is automatically transmitted to a control center for monitoring and analyses.
- Computers and high-end printers with smart chips can record the usage history and the environmental condition such as operating temperature. This information is available dynamically through the network or other communications channels and can, in cooperation with the owner, be downloaded periodically.

1.2 Applications in Prediction

One important reason for outfitting products with sensors, smart chips, and communications channels is to assist in the delivery of timely maintenance actions and to increase system availability. It can be expected, however, that using dynamic usage/environmental information in modeling and data analysis will also provide stronger statistical methods and more accurate inferences or predictions of field failures. These improvements can be realized when one or more of the important sources of variability the field data can be explained by the additional information. In applications, predictions for the number of field failures or warranty returns for the population is important for financial planning decisions, such as setting warranty reserves for a manufactured product or capital budgeting for a company's fleet of assets. For example, after a product has been introduced into the field for a certain period of time (e.g., one year), the finance department is often interested to know what will be the total number of returns for some future period of time (i.e., the next three years), based on early warranty returns of the product. By taking advantage of the dynamic information available from the product, one can expect to get more accurate predictions than what would be obtained by using only the traditional failure-time data.

The prediction of the remaining life of individual units sometimes is also of interest, especially for fleets of assets for a company (e.g., locomotives in Section 9.9 of [HD08], and high-voltage power transformers in [HMM09]). [HMM09] give the prediction intervals for the remaining life of high-voltage power transformers based only on currently available failure-time data. The prediction intervals given there are wide for individual units. The dynamic usage/environmental information can be expected to improve the accuracy of prediction intervals for individuals.

1.3 Related Literature

[Nel90, Chapter 10] describes the cumulative exposure model in the context of life tests. The cumulative exposure model is equivalent to the time scale accelerated failure time model with time-dependent covariates used in [RT92]. [RT92], however, used a nonparametric estimation method that does not require specification of the baseline failure-time distribution.

[Nel01] describes prediction for field reliability of units under dynamic stresses using the cumulative exposure model. The problem considered in our paper, however, is different. We consider predictions and prediction intervals (PIs) for both the population and individual units based on the distribution of remaining life. The uncertainty in the covariate process is also considered.

In the area of warranty prediction involving dynamic stress, [GMMO09] consider warranty prediction with stress that is random from unit to unit but constant within a unit. Stress information, however, is not available for individual units. [HM10] consider a warranty prediction problem where the the average use-rate is available for both failed and censored units. However, warranty prediction procedures using the dynamic information need to be developed.

2 Data and Model

2.1 Notation

Let T be the time to failure random variable. The usage/environmental information at time t is denoted by a random vector $X(t) = [X_1(t), \cdots, X_p(t)]'$ where p is the number of covariates. The history of the covariate process is denoted by $\boldsymbol{X}(t) = \{X(s) : 0 \leq s \leq t\}$ which records the dynamic information from time 0 to time t. Because of the dynamic information on usage and environmental conditions, observations of $X(t)$ are available for each small time interval with length Δ. $X(t)$ is assumed to be constant over these intervals. Thus, the covariate history is recorded as a multivariate time series.

The data are denoted by $\{\, t_i, \delta_i, \boldsymbol{x}_i(t_i) \,\}$ for $i = 1, 2 \cdots, n$ where n is the number of observations in the dataset. Here t_i is the failure time (time in service) for unit i if it failed (did not fail). The censoring indicator $\delta_i = 1$ if unit i failed and $\delta_i = 0$ otherwise. Let $x(t)$ be the observed covariate information at time t. Then $\boldsymbol{x}_i(t_i) = \{x(s) : 0 \leq s \leq t_i\}$ is the observed covariate history from the time origin to t_i for unit i.

2.2 Cumulative Exposure Model

We use the cumulative exposure model, as described in [Nel90], to model the failure-time data with covariates which were recorded as a multivariate time series. In particular, the cumulative exposure U for a unit with failure time T is defined as

$$U = \int_0^T g[X(s); \boldsymbol{\beta}] ds \sim F_0(u, \boldsymbol{\theta}_0) \tag{1}$$

given the covariate entire history $\boldsymbol{X}(\infty) = \boldsymbol{x}(\infty)$. Here $g[X(t); \boldsymbol{\beta}]$ is the time scale acceleration rate which is a function of the covariate process history with parameter $\boldsymbol{\beta}$, and $F_0(u, \boldsymbol{\theta}_0)$ is the baseline cumulative distribution function (cdf). $g[X(t); \boldsymbol{\beta}]$ gives the instantaneous effect of the stress/exposure on the product life from both the usage and the environment at time t. If a unit is operated under harsh environmental conditions and/or has a large use rate, then $g[X(t); \boldsymbol{\beta}] > 1$. That is, the calender time scale is accelerated. The unit would be expected to fail sooner than those used under mild conditions.

There needs to be a restriction on $g[X(t); \boldsymbol{\beta}]$ in order for the parameters to be estimable. In particular, the function needs to have $g[X(t); \boldsymbol{\beta}] = 1$ when $\boldsymbol{\beta} = \mathbf{0}$. Given the covariate history $\boldsymbol{X}(\infty) = \boldsymbol{x}(\infty)$ and $\boldsymbol{\beta} = \mathbf{0}$, the cumulative exposure is $U = \int_0^T g[x(s);\boldsymbol{\beta}]ds = \int_0^T 1ds = T$. That is, the cumulative exposure has the same scale as the calender time scale.

We assume that lifetimes of product units that are all used at the same constant use rate and environmental conditions can be adequately described by the same distribution. This is because the failure mechanisms of those units are similar. The cumulative exposure model converts units under different usage and environmental conditions into a comparable scale which is called the cumulative exposure. We assume that failure time of the population under the cumulative exposure time scale can be adequately described by a single distribution.

2.3 Modeling the Time Scale Acceleration Rate

The following log-linear relationship is widely used as an acceleration factor

$$g[X(t);\boldsymbol{\beta}] = \exp[\boldsymbol{\beta}' X(t)]. \tag{2}$$

This model assumes the effect on the time scale acceleration is proportional for different values of $X(t)$. This model is sometimes called the proportional quantiles (PQ) model or the scale accelerated failure-time (SAFT) model (e.g., [ME98, Chapter 17]). Here $X(t)$ might be transformed values of the original explanatory variable.

- If use-rate information is available, one possible relationship is $g[X(t);\boldsymbol{\beta}] = \exp[\beta_0 + \beta_1 X(t)]$ where $X(t) = \log(\text{use rate})$, which is the inverse power law (e.g., [ME98, Page 480]).
- If information on temperature is available, the Arrhenius relationship (e.g., [ME98, Page 472]) can be used, in which case $g[X(t);\boldsymbol{\beta}] = \exp[\beta_0 + \beta_1 X(t)]$ where $X(t) = -11605/(\texttt{temp} + 273.6)$ is a transformation of Celsius temperature `temp`. β_1 can be interpreted as the effective activation energy in electron volts. Both β_0 and β_1 are product or material characteristics.
- If information on both temperature and use rate is available, the following relationship can be used,

  $$g[X(t);\boldsymbol{\beta}] = \exp[\beta_0 + \beta_1 X_1(t) + \beta_2 X_2(t)]$$

  where $X_1(t)$ is the transformed temperature and $X_2(t)$ is the transformed use rate.

2.4 Modeling the Baseline Distribution

The baseline cdf is defined as the cdf of the failure-time distribution for a unit which is used at typical fixed conditions. These baseline conditions are similar to the use conditions in a standard life test or accelerated life test. We will model the baseline cdf $F_0(u;\boldsymbol{\theta}_0)$ as a log-location-scale distribution. The general log-location-scale cdf is

$$F_0(u;\boldsymbol{\theta}_0) = \Phi\left[\frac{\log(u)-\mu}{\sigma}\right], \quad u>0. \tag{3}$$

Here $\boldsymbol{\theta}_0 = (\mu,\sigma)'$, μ is the location parameter, σ is the scale parameter, and $\Phi(z)$ is the standard cdf for the location-scale family of distributions (location 0 and scale 1). The corresponding probability density function (pdf) is

$$f_0(u;\boldsymbol{\theta}_0) = \frac{dF_0(u)}{du} = \frac{1}{\sigma u}\phi\left[\frac{\log(u)-\mu}{\sigma}\right]$$

where $\phi(z) = d\Phi(z)/dz$. The Weibull and lognormal distributions are the most commonly used distributions for modeling of failure-time data from this family of distributions. The cdf and pdf of T, given the entire history $\boldsymbol{X}(\infty) = \boldsymbol{x}(\infty)$, is

$$F(t;\boldsymbol{\beta},\boldsymbol{\theta}_0) = F_0\left(\int_0^t g[x(s);\boldsymbol{\beta}]ds;\boldsymbol{\theta}_0\right) \text{ and } f(t;\boldsymbol{\beta},\boldsymbol{\theta}_0) = g[x(t);\boldsymbol{\beta}]f_0\left(\int_0^t g[x(s);\boldsymbol{\beta}]ds;\boldsymbol{\theta}_0\right)$$

respectively.

2.5 Modeling the Covariate Process

For the purpose of prediction of failure times, a parametric model for $X(t)$ is needed. This allows prediction of the future covariate vector for an individual unit. $X(t)$ is modeled as

$$X(t) = m(t;\boldsymbol{\eta}) + a(t)\,. \tag{4}$$

Here $m(t;\boldsymbol{\eta})$ is the mean function with parameter $\boldsymbol{\eta}$ and $a(t)$ is the error term which is assumed to be a stationary process. The parametric form for the mean function $m(t;\boldsymbol{\eta})$ needs to be specified according to the particular application. Some components of $\boldsymbol{\eta}$ can be random to allow for population nonhomogeneity of the covariate process. Also, depending on the application, the following gives possible models for the distribution of $a(t)$.

- In a simple case, $a(t)$ for different values of t can be modeled as independently and identically distributed (iid) with $N(0,\Sigma)$ where Σ is the covariance matrix.
- To allow for more complicated models for $a(t)$, the vector autoregressive (VAR) moving average time series models in [Rei03] can be used. For example, the VAR(1) model is represented as

$$a(t) = \Psi\, a(t-1) + \varepsilon(t) \tag{5}$$

where Ψ is an unknown coefficient matrix. For the purpose of prediction, a parametric distribution assumption is needed for the noise term $\varepsilon(t)$. One common choice is that $\varepsilon(t)$ are iid with $N(0, \nu^2 I)$ where I is the identity matrix and ν^2 is the variance factor.

3 Parameter Estimation

In this section, we use the method of maximum likelihood (ML) to obtain estimates for unknown model parameters. The ML estimates for the failure-time distribution parameters are obtained by conditioning on the observed covariate history. The ML estimates for the parameters of the covariate history can also be obtained by assuming that the data were generated from a specific class of multivariate time series models.

3.1 ML Estimate for Parameters of the Failure-time Distribution

The likelihood of the failure-time data, conditional on the observed covariate history, is

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}_0 | DATA) = \prod_{i=1}^{n} \left\{ g[x_i(t_i); \boldsymbol{\beta}] f_0 \left(\int_0^{t_i} g[x_i(s); \boldsymbol{\beta}] ds; \boldsymbol{\theta}_0 \right) \right\}^{\delta_i} \left\{ 1 - F_0 \left(\int_0^{t_i} g[x_i(s); \boldsymbol{\beta}] ds; \boldsymbol{\theta}_0 \right) \right\}^{1-\delta_i}. \quad (6)$$

The maximum likelihood (ML) estimate $\left(\widehat{\boldsymbol{\beta}}', \widehat{\boldsymbol{\theta}}'\right)'$ is obtained by finding those values of $(\boldsymbol{\beta}', \boldsymbol{\theta}')'$ that maximize (6).

3.2 ML Estimate for Parameters of the Covariate Process

The likelihood for the covariate history, assuming the parameter $\boldsymbol{\eta}$ in the mean function $m(t; \boldsymbol{\eta})$ is non-random and the error term $a(t)$ is distributed with $N(0, \Sigma)$, is

$$L(\boldsymbol{\eta}, \Sigma | DATA) = \prod_{i=1}^{n} \prod_{s \le t_i} \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} [x_i(s) - m(s; \boldsymbol{\eta})]' \Sigma^{-1} [x_i(s) - m(s; \boldsymbol{\eta})] \right\}. \quad (7)$$

The ML estimates are denoted by $\widehat{\boldsymbol{\eta}}$ and $\widehat{\Sigma}$. For more complicated models describing $a(t)$, the parameters can also be estimated using the method of ML. The ML estimation techniques described, for example, in [Rei03, Chapter 5] can be used.

4 Predictions

As described in Section 1.2, predictions are needed for both the cumulative number of field failures and the remaining life of individual units. There is need for accurate predictions on both the population and individuals in business and industry. The availability of dynamic environmental information can be expected to improve the accuracy of predictions, especially for individuals. In applications, predictions need to correspond to the real time scale, after the data-freeze date (DFD). These predictions will be based on the distribution of remaining life of units that have survived until the DFD.

4.1 Distribution of Remaining Life

The distribution of remaining life provides the basis for calculating the prediction for the population and the individuals. The distribution of T_i, given $T_i > t_i$ and the covariate history $\boldsymbol{X}_i(t_i)$, is

$$\rho_i(t_w;\boldsymbol{\theta}) = \Pr[t_i < T_i \leq t_w | T_i > t_i, \boldsymbol{X}_i(t_i)], \quad t_w > t_i. \tag{8}$$

Here $\boldsymbol{\theta}$ is the collection of parameters including $\boldsymbol{\beta}, \boldsymbol{\theta}_0$, and the parameters for the covariate process. In particular,

$$\begin{aligned}\rho_i(t_w;\boldsymbol{\theta}) &= \mathbf{E}_{\boldsymbol{X}_i(t_i,t_w)|\boldsymbol{X}_i(t_i)}\left\{\Pr[t_i < T_i \leq t_w | T_i > t_i, \boldsymbol{X}_i(t_i), \boldsymbol{X}_i(t_i,t_w)]\right\} \\ &= \frac{\mathbf{E}_{\boldsymbol{X}_i(t_i,t_w)|\boldsymbol{X}_i(t_i)}\left\{F_0\left(\int_0^{t_w} g[X_i(u);\boldsymbol{\beta}]du;\boldsymbol{\theta}_0\right)\right\} - F_0\left(\int_0^{t_i} g[x_i(u);\boldsymbol{\beta}]du;\boldsymbol{\theta}_0\right)}{1 - F_0\left(\int_0^{t_i} g[x_i(u);\boldsymbol{\beta}]du;\boldsymbol{\theta}_0\right)}\end{aligned} \tag{9}$$

where $\boldsymbol{X}_i(t_1,t_2) = \{X_i(u) : t_1 < u \leq t_2\}$. When the model for $X(t)$ is complicated, the distribution of $\boldsymbol{X}_i(t_i,t_w)|\boldsymbol{X}_i(t_i)$ may be mathematically intractable. Numerical methods can, however, be applied to evaluate $\rho_i(t_w;\boldsymbol{\theta})$.

4.2 Prediction for the Population

When focusing on the overall population, we need to generate predictions for the cumulative number of failures for the units in the field. The prediction for the warranty returns can also be obtained in a similar way but there is a need to adjust the risk set for the length of the warranty period. Prediction intervals are also needed for quantifying the statistical uncertainties.

Let $N(s)$ be the number of field failures at s time units after the DFD. $N(s) = \sum_{i\in RS} I_i(s)$ where RS is the risk set and $I_i(s) \sim \text{Bernoulli}[\rho_i(t_i+s;\boldsymbol{\theta})]$. The point prediction for $N(s)$ is $\widehat{N}(s) = \sum_{i\in RS}\rho_i(t_i+s;\widehat{\boldsymbol{\theta}})$. A prediction interval (PI) for $N(s)$ is denoted by $\left[\underset{\sim}{N}, \widetilde{N}\right]$. The naive (plug-in) PI is obtained by solving

$$F_N(\underset{\sim}{N};\widehat{\boldsymbol{\theta}}) = \frac{\alpha}{2}, \quad \text{and} \quad F_N(\widetilde{N};\widehat{\boldsymbol{\theta}}) = 1 - \frac{\alpha}{2}. \tag{10}$$

Here $F_N(n_k;\boldsymbol{\theta}), n_k = 0, 1, \cdots, n^*$ is the cdf of $N(s)$ where n^* is the number of units in the RS at the DFD. $1-\alpha$ is the desired coverage probability. Note that $N(s)$ is a sum of non-identically distributed Bernoulli random variables. The cdf of N_k does not have a simple closed-form expression. An approximation is usually needed in applications. The Volkova approximation ([Vol96]), which is based on a refined normal approximation with correction for the skewness of $N(s)$, is used by [HMM09] for a prediction problem. The Poisson approximation is also used in the literature (e.g., [EM99, Section A.3]) when the expected number of failure (after the DFD in setting of this paper) is small (e.g., less than 10).

4.3 Prediction for Individuals

When focusing on individuals, we will compute prediction intervals for each individual. The naive prediction interval for the individual remaining life is denoted by $\left[\underset{\sim}{T}_i, \widetilde{T}_i\right]$ and can be obtained by solving

$$\rho_i(\underset{\sim}{T}_i;\widehat{\boldsymbol{\theta}}) = \frac{\alpha}{2}, \text{ and } \rho_i(\widetilde{T}_i;\widehat{\boldsymbol{\theta}}) = 1 - \frac{\alpha}{2} \tag{11}$$

where $\rho_i(\,\cdot\,,\boldsymbol{\theta})$ is given in (9). Note here the PI of remaining life is conditional on the individual's current time in service t_i and its observed covariate process $\boldsymbol{x}_i(t_i)$. Thus each individual will have a distinct PI.

5 Calibration of Prediction Intervals

The PIs in (10) and (11) ignore the uncertainty in $\widehat{\boldsymbol{\theta}}$. Thus the coverage probability is generally smaller than the nominal $1-\alpha$ level. These PIs can be calibrated to improve the coverage probability property. We will use simulations to do the calibration.

5.1 Bootstrapping the Distribution of $\widehat{\boldsymbol{\theta}}$

To account for the uncertainty in $\widehat{\boldsymbol{\theta}}$, we use a parametric bootstrap simulation to approximate the distribution of $\widehat{\boldsymbol{\theta}}$. The calibration has two parts: first, we use bootstrap to generate the bootstrap version of the covariate process $\boldsymbol{x}_i^*(t_i), i = 1, 2, \cdots, n$. Because we assume a parametric model for the covariate process as in (4), parametric simulation methods can be used here to generate $\boldsymbol{x}_i^*(t_i)$. Repeating the ML estimation procedure in Section 3.2, one obtains the bootstrap version of estimates of the parameters for the covariate process.

The second part of the calibration process is to obtain the bootstrap version estimates of parameters for the failure-time distribution. The traditional bootstrap method that uses simple random sampling with replacement can be problematic with heavy censoring, as it can result in bootstrap samples without enough failures for the estimation of the parameters. Here we use the random weighted bootstrap method (e.g., [NR94], [JYW01]) to obtain the bootstrap version estimates of the parameters. See [HMM09] for another application of random weighted bootstrap in calibration PIs. In particular, with a set of random weights Z_i generated from any positive continuous distribution with $\mathbf{E}(Z_i) = \sqrt{\mathrm{Var}(Z_i)}$, the random weighted likelihood is

$$L^*(\boldsymbol{\beta}, \boldsymbol{\theta}_0 | DATA) =$$
$$\prod_{i=1}^{n} \left\{ g[x_i^*(t_i); \boldsymbol{\beta}] f_0 \left(\int_0^{t_i} g[x_i^*(s); \boldsymbol{\beta}] ds; \boldsymbol{\theta}_0 \right) \right\}^{\delta_i Z_i} \left\{ 1 - F_0 \left(\int_0^{t_i} g[x_i^*(s); \boldsymbol{\beta}] ds; \boldsymbol{\theta}_0 \right) \right\}^{(1-\delta_i) Z_i} .$$

Here $\boldsymbol{x}_i^*(t_i)$ is the bootstrap sample generated in the first part. The bootstrap versions of the parameter estimates for the failure-time distribution can be obtained by maximizing the random weighted likelihood. Combining with the bootstrap version estimates of the parameter for the covariates, we obtain the bootstrap version of $\widehat{\boldsymbol{\theta}}$, which is denoted by $\widehat{\boldsymbol{\theta}}^*$.

5.2 Calibration for Prediction Intervals

With B bootstrap samples of $\widehat{\boldsymbol{\theta}}^*$, the calibration of PIs for the population can be done by using a procedure similar to the procedure described in Section 6.2 of [HMM09]. Here B is usually chosen to be a large number (e.g., $B = 10,000$). The calibration of PIs for individuals can be done by using a procedure similar to the procedure described in Section 5.4 of [HMM09].

6 Conclusions and Areas for Future Research

In this paper, we outline a model and methods that can be used for field-failure prediction using dynamic environmental data. We also describe predictions of the cumulative number of field failures and the remaining life for individuals. Prediction intervals are also given and the associated calibration procedures are also described.

In future work, we will consider more general modeling of effects of covariate processes on the failure-time distribution. For example, one can model the time scale acceleration rate function g in (1) as $g[\boldsymbol{X}(t), \boldsymbol{\beta}]$. This means the time scale acceleration rate function depends on the history from the time origin to time t. For example, some environmental variables may have delayed effect on the failure-time distribution. Alternative models, such as the proportional hazards model with time dependent covariates could also be considered. Parametric models for the baseline hazard function and the covariate process will be needed if prediction is the main goal of the application.

Modern sensor technology also allow us to obtain dynamic degradation measurements (or indirect measurements) for products or components of products on the field. Prediction and intervals using dynamic degradation can be expected to have some advantages and provide more useful information. Some research has been done in this direction. [GP08] used dynamic environmental data to update the distribution of remaining life under a Bayesian frame work. [VTM09] developed a statistical model for linking field and laboratory exposure data that measure the chemical degradation processes of a coating system, where environmental variables such as UV spectrum and intensity, temperature, and relative humidity were also measured repeatedly. More general models and methods for prediction and prediction intervals, however, need to be developed for these situations.

Acknowledgements

We would like to thank Katherine Meeker for helpful comments on an earlier version of this paper.

References

[EM99] L. A. Escobar and W. Q. Meeker. Statistical prediction based on censored life data. *Technometrics*, 41:113–124, 1999.

[GMMO09] Huairui Guo, Arai Monteforte, Adamantios Mettas, and Doug Ogden. Warranty prediction for products with random stresses and usages. In *Proceedings Annual Reliability and Mantainability Symposium*, pages 72–77, Fort Worth, TX, 2009.

[GP08] Nagi Gebraeel and Jing Pan. Prognostic degradation models for computing and updating residual life distributions in a time-varing environment. *IEEE Transactions on Reliability*, 57:539–550, 2008.

[HD08] Gerald J. Hahn and Necip Doganaksoy. *The Role of Statistics in Business and Industry.* John Wiley & Sons, Inc., New Jersey, Hoboken, 2008.

[HM10] Y. Hong and W. Q. Meeker. Field-failure and warranty prediction based on auxiliary use-rate information. *Technometrics*, 52:xxx–xxx, 2010.

[HMM09] Y. Hong, W. Q. Meeker, and J. D. McCalley. Prediction of remaining life of power transformers based on left truncated and right censored lifetime data. *The Annals of Applied Statistics*, 3:857–879, 2009.

[JYW01] Z. Jin, Z. Ying, and L. J. Wei. A simple resampling method by perturbing the minimand. *Biometrika*, 88(2):381–390, 2001.

[ME98] W. Q. Meeker and L. A. Escobar. *Statistical Methods for Reliability Data.* John Wiley & Sons, Inc., New York, 1998.

[Nel90] W. Nelson. *Accelerated Testing: Statistical Models, Test Plans, and Data Analyses, (Republished in a paperback in Wiley Series in Probability and Statistics, 2004)*. John Wiley & Sons, New York, 1990.

[Nel01] W. Nelson. Prediction of field reliability of units, each under differing dynamic stresses, from accelerated test data. *Handbook of Statistics 20: Advances in Reliability, Chapter IX*, pages edited by N. Balakrishnan and C. R. Rao, Amsterdam: North–Holland, 2001.

[NR94] Michael A. Newton and Adrian E. Raftery. Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society: Series B*, 56:3–48, 1994.

[OnS09] OnStar. *The OnStar Official site*. http://www.onstar.com/us_english/jsp/index.jsp, 2009.

[Rei03] G. C. Reinsel. *Elements of Multivariate Time Series Analysis*. Springer-Verlag, New York, 2nd edition, 2003.

[RT92] J. Robins and A. A. Tsiatis. Semiparametric estimation of an accelerated failure time model with time-dependent covariates. *Biometrika*, 79:311–319, 1992.

[STW+05] K. Spurgeon, W. H. Tang, Q. H. Wu, Z. J. Richardson, and G. Moss. Dissolved gas analysis using evidential reasoning. *IEE Proceedings - Science, Measurement & Technology*, 152:110–117, 2005.

[Vol96] A. Y. Volkova. A refinement of the central limit theorem for sums of independent random indicators. *Theory of Probability and its Applications*, 40(4):791–794, 1996.

[VTM09] I. Vaca-Trigo and W. Q. Meeker. A statistical model for linking field and laboratory exposure results for a model coating. In Jonathan Martin, Rose A. Ryntz, Joannie Chin, and Ray A. Dickie, editors, *Service Life Prediction of Polymeric Materials*. Springer, 2009.