Statistical Modeling of Multivariate Destructive Degradation Tests with Blocking

Qiuzhuang Sun[†], Zhi-Sheng Ye[†], and Yili Hong[‡]

[†]Department of Industrial Systems Engineering and Management, National University of Singapore

[‡]Department of Statistics, Virginia Tech

Abstract

In degradation tests, the test units are usually divided into several groups, with each group tested simultaneously in a test rig. Each rig constitutes a rig-layer block from the perspective of design of experiments. Within each rig, the test units measured at the same time further form a gauge-layer block. Due to the uncontrollable factors among test rigs and the common errors incurred for each measurement, the degradation measurements of the test units may differ among various blocks. On the other hand, the degradation should be more homogeneous within a block. Motivated by an application of emerging contaminants (ECs), this study proposes a multivariate statistical model to account for the two-layer block effects in destructive degradation tests. A multivariate Wiener process is first used to model the correlation among different dimensions of degradation. The rig-layer block effect is modeled by a one-dimensional frailty motivated by the degradation physics, while the gauge-layer block effect at each measurement epoch is captured by a common additive measurement error. We develop an Expectation-Maximization (EM) algorithm to obtain the point estimates of the model parameters and construct confidence intervals for the parameters. A procedure is proposed to test significance of the block effects in the degradation data. Through a case study on an EC degradation dataset, we show the existence of the two-layer block effects from the test. By making use of the proposed model, decision makers can readily make risk assessment of each contaminant and determine the minimal water treatment time for removal of the contaminants.

Key words: Blocking, destructive degradation test, EM algorithm, multivariate Wiener process, hypothesis testing, emerging contaminants.

1 Introduction

1.1 Background and Motivating Example

Degradation is a cumulative and irreversible change of a subject's performance characteristic over time. Degradation studies are often used to assess reliability of products subject to degradation-induced failures (Whitmore and Schenkelberg, 1997; Mercier et al., 2012; Rafiee et al., 2014; Si, 2015; Sun et al., 2019). They are also important tools for risk assessment in chemical, environmental, and biomedical engineering. For example, Duan et al. (2017) investigated photodegradation of polymeric materials such as organic paints and coatings caused by exposure to ultraviolet radiation. Zeng et al. (2016) studied the biodegradation of scaffold in tissue engineering to aid the replacement decision of the tissues or organs in human bodies. Xu et al. (2011) used degradation for risk assessment of emerging contaminants (ECs), a new environmental threat due to the increasing consumption of newly synthesized compounds. In all these studies, degradation data are essential for the modeling and analysis.

Degradation testing is probably the most important source of degradation data. When the degradation can be measured in a non-intrusive way, a subject is usually repeatedly measured over time, leading to repeated measures degradation test data (Bae and Kvam, 2004; Liao and Elsayed, 2006; Weaver et al., 2013; Peng et al., 2018). In many degradation tests, however, a test subject has to be destroyed in order to measure its degradation (Li and Doganaksoy, 2014). Implementation of such destructive degradation tests (DDTs) can be found in measuring the strength of an adhesive bond (Shi et al., 2009), the breakdown dielectric strength of electrical insulation (Nelson, 1981), a tissue-engineered scaffold (Zeng et al., 2016), and the concentration of emerging contaminants (Xu et al., 2011). Motivated by EC applications, we confine our interest to DDT data in this study.

Since it is common that a test rig is able to accommodate more than one unit, the test units in a DDT are usually divided into groups with each group assigned to a test rig. At each of a set of predetermined test times, several units are taken out from a rig and their degradation is destructively measured. Such an arrangement of the test units and the measurement scheme introduce blocking in the DDT. According to Box et al. (2005, Section 3.3), the units in a block are "expected to be more homogeneous than the aggregate..... Runs made close together in time or space are likely to be more similar than runs made further apart and hence can often provide a basis for blocking." As an analogy, units tested in the same rig are the "runs made close together in *space*", and they share a common experiment environment. Within a test rig, units measured at the same time are the "runs made close together in *time*", and they may share a common measurement error during the destructive measurement process. As a result, there are two layers of block effects in a DDT, which are called the rig-layer and the gauge-layer block effects in this study. Blocking results in incomplete randomization (Blsgaard and Steinberg, 1997; Ye and Sun, 2019) in the degradation tests, which complicates data analysis.

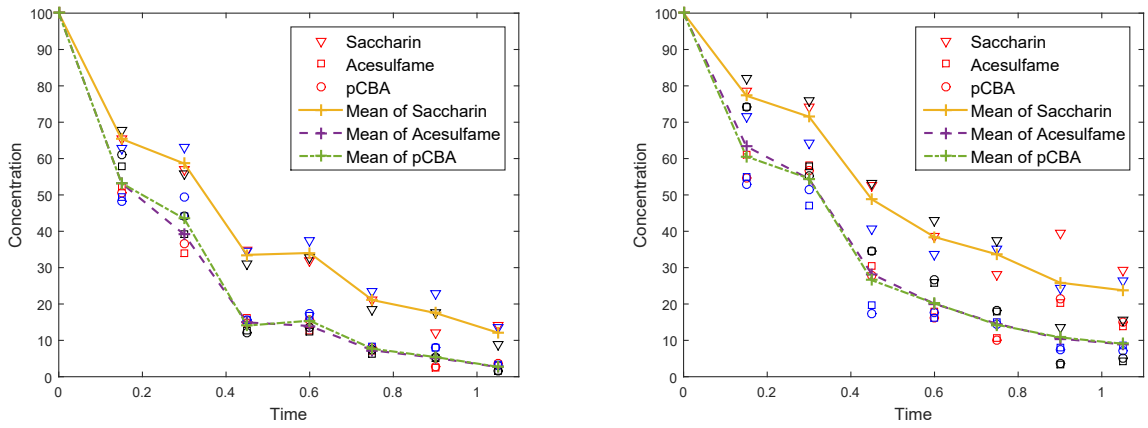
Another challenge is the multivariate nature of degradation in many applications. There are usually multiple criteria to evaluate the performance or condition of a product. For example, degradation of a railway track is defined in terms of wear in both the longitudinal and the transversal directions (Mercier et al., 2012). For a multi-component system, each component has its own degradation characteristic, while degradation of different components is normally correlated because of the interaction and coordination within the system. For instance, wear processes of the four tires in a car are positively correlated since the four tires share a common operating environment. Both scenarios above lead to multivariate degradation that requires joint modeling to capture the correlation among different dimensions. However, there is scant literature in multivariate degradation analysis (Hong et al., 2018b).

A motivating example of multivariate degradation with two-layer block effects comes from a DDT for ECs. The purpose of studying EC degradation is twofold. The first is to

quantify the persistence in the environment for risk assessment of each individual EC, and the second is to determine a suitable treatment time for removing a bundle of high-risk contaminants in water treatment (Hong et al., 2018a). A DDT is used to collect EC degradation data for the two goals. In the experiment, a test solution of the contaminants of interest is prepared and distributed to 21 tubes. These tubes are then put in a thermostat, and the ECs degrade over time. At each of seven predetermined epochs, three tubes are removed from the test rig, and concentrations of all contaminants in the three tubes are simultaneously measured by a measurement system called liquid chromatography tandem mass spectrometry (LC-MS/MS). The above experiment is replicated six times to ensure reproductivity of the results. Each replication can be regarded as a rig-layer block, and tubes measured at the same time constitute a gauge-layer block. The experiment condition is approximately constant in each replication, but different replications can be slightly different due to the prepared solution and the experimenter. Similarly, the measurement of contaminant concentration can be affected by reagents used for quenching and gauging (Petyuk et al., 2008), leading to a common measurement error at each measurement time. For illustration, Figure 1 shows the multivariate DDT data for a bundle of three contaminants from two replications, where the data are masked by a change in time scale. The degradation in the left panel looks faster than the right, indicating the rig-layer block effects. The pseudo first-order kinetics (Steinfeld et al., 1989) for the chemical contaminant degradation implies that the mean degradation path is log-linear, which tallies with the curves in the figures. However, simultaneous humps and dips are observed in the three curves in each plot, which may be a result of the common measurement error.

1.2 Related Literature

Most of the literature on degradation analysis focuses on one-dimensional degradation. See Bae and Kvam (2004); Giorgio et al. (2010); Wang (2010); Rafiee et al. (2014), to name a few. Because of a lack of data, there are relatively few studies on multivariate degradation



(a) Rig 1 (b) Rig 2

Figure 1: EC degradation data from two test rigs. Each test unit is a tube of solution with three contaminants. There are 21 units on each test rig, and seven predetermined measurement times are employed. Contaminant concentrations of three tubes are measured simultaneously at each measurement time. Colors of the points at each measurement time indicate the degradation characteristics from the same test unit. The lines represent the empirical mean degradation level of each contaminant.

analysis. Mercier et al. (2012) used the trivariate reduction method to construct correlation between different dimensions of degradation in a railway track. They showed that negligence of the correlation between different dimensions may result in an optimistic estimation of the system residual useful life. An alternative way for multivariate degradation modeling is the copula-based method. See Wang et al. (2015), Hong et al. (2018b), and Peng et al. (2018) for some examples. The copula models generally do not retain the infinite divisibility property, which can lead to model inconsistency in a multivariate degradation setting. The data-driven nature further makes it difficult to incorporate the block effects from physical mechanism into the copula models. In this study, we employ the multivariate Wiener degradation process (Hong et al., 2018a) as a building block to account for block effects. The multivariate Wiener process preserves the infinite divisibility property, and it is straightforward to include the two-layer block effects based on the physical information from the EC experiment. As will be seen in Section 3, the maximum likelihood (ML) estimation can be readily done by an Expectation-Maximization (EM) algorithm.

Blocking is commonly observed in reliability experiments. Numerous studies in lifetime tests have revealed serious biases and efficiency loss in estimating important reliability characteristics when overlooking the block effects. León et al. (2009) found that overlooking the block effect can lead to a large estimation bias of model parameters. They proposed a Weibull regression model where the rig-layer block effect is modeled as a lognormal random variable. The model was further extended by Freeman and Vining (2010); Kensler et al. (2015); Seo and Pan (2017). Blocking is also common in degradation tests. In a coating degradation test reported in Duan et al. (2017), for example, four specimens were tested in the same chamber, and the chamber served as a rig-layer block. However, blocking in a degradation test is more complicated than that in a life test because of the gauge-layer block. In one-dimensional degradation modeling, Li and Doganaksoy (2014) used a common additive term for the gauge-layer block effect, while Zhai and Ye (2018) used a common stochastic time scale to model the rig-layer block effect. Both studies have found significant improvements in data fitting when blocking is considered, but they only considered one layer of block effects. Since both layers of the block effects might exist in a degradation test, it is more reasonable to take both into account during data analysis. This task becomes challenging when the degradation is multivariate.

1.3 Overview

The main objective of the study is to develop a multivariate degradation model with two-layer blocking for DDT data. Motivated by the EC application, the rig-layer block effect is modeled as a frailty. Within a test rig, the gauge-layer block effect is modeled by a common measurement error. We develop an EM algorithm for point and interval estimation of the model parameters. Some hypothesis testing procedures are proposed to examine the significance of the block effects.

The rest of the paper is organized as follows. Section 2 introduces the settings of our DDT and proposes a degradation model with the block effects. Section 3 develops an EM algorithm

to obtain the ML estimates of model parameters and discusses methods for the interval estimation. Section 4 proposes a procedure to examine the significance of the block effects. Section 5 evaluates the performance of the proposed methods and illustrates the importance of considering the block effects through simulations. Section 6 applies the degradation model to the EC degradation data. Concluding remarks are available in Section 7.

2 DDT and Degradation Model

2.1 DDT Setting and Data

Consider a subject with d-dimensional performance characteristics degrading over time. A DDT is used to collect its degradation data. Suppose the DDT data consist of n rig-layer blocks. This includes the scenarios where the DDT is conducted using n rigs simultaneously, and where the experiment is repeated n times using a single rig. Without loss of generality, suppose all rigs share a common set of predetermined measurement times $\{t_1, \cdots, t_m\}$. Let $[I] \equiv \{1, \cdots, I\}$ be the set of integer indexes up to I. Consider rig i, $i \in [n]$. At t_j, $j \in [m]$, totally K test units are removed from rig i, and their degradation levels are destructively measured. Let $\boldsymbol{Y}_{ijk} \in \mathcal{R}^d$ be the corresponding measured degradation level of the kth unit, $k \in [K]$. The K units from the same rig and measured at the same time form a gauge-layer block. Therefore, each rig accommodates mK test units, and the total number of test units is $N = nmK$. The observed DDT data are denoted as $\mathbb{D} \equiv \{\boldsymbol{Y}_{ijk}; i \in [n], j \in [m], k \in [K]\}$. A schematic of the above DDT setting is given in Figure 2.

2.2 Degradation Model with Two-Layer Block Effects

Consider the kth unit measured at t_j from rig i, $i \in [n]$, $j \in [m]$, and $k \in [K]$. Let $\{\boldsymbol{X}_{ijk}(t) \in \mathcal{R}^d,\ t \geq 0\}$ be the underlying degradation process, where the subscript "ijk" is the label for the unit. The value of $\boldsymbol{X}_{ijk}(t_j)$ is different from $\boldsymbol{Y}_{ijk}$ due to a measurement error. We use a multivariate Wiener process to capture the dynamics of the degradation of

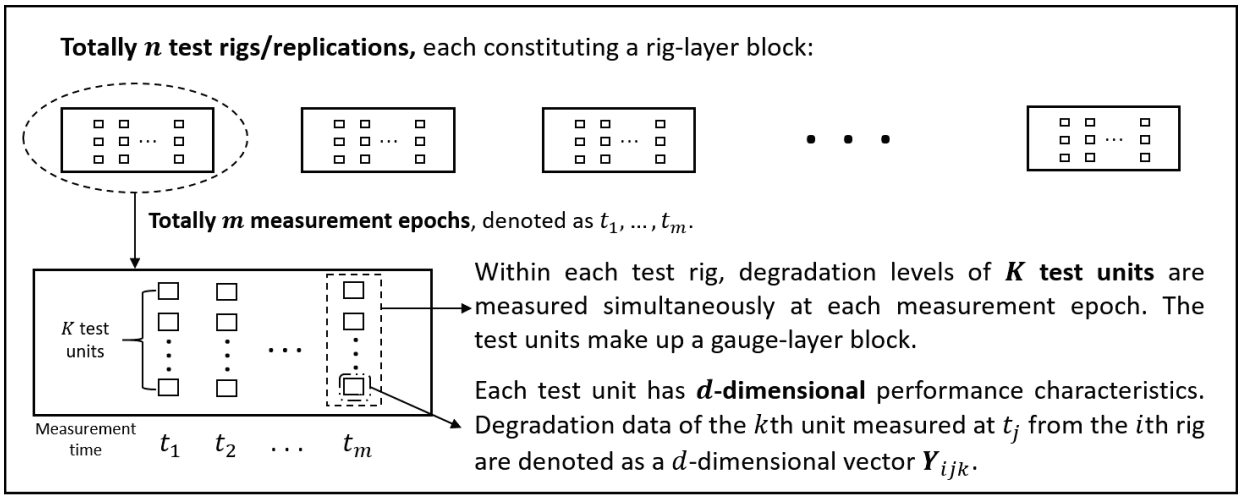


Figure 2: The experiment setting and data scheme of the DDT.

the unit, which is given by

$$\boldsymbol{X}_{ijk}(t) = \boldsymbol{v}_{ijk}t + \boldsymbol{\Sigma}^{1/2}\boldsymbol{\mathcal{B}}_{ijk}(t), \quad i \in [n],\ j \in [m],\ k \in [K], \tag{1}$$

where $\boldsymbol{v}_{ijk} \in \mathcal{R}^d$ is the vector of degradation rates, $\boldsymbol{\Sigma} \in \mathcal{R}^{d\times d}$ is the covariance matrix introducing possible correlation between different dimensions, and $\boldsymbol{\mathcal{B}}_{ijk}(t) \in \mathcal{R}^d$ is the standard d-dimensional Brownian motions. The ith diagonal element of $\boldsymbol{\Sigma}$ is σ_i^2 and the off-diagonal element is $\rho_{ij}\sigma_i\sigma_j$ for the (i,j)th entry, where $|\rho_{ij}| \leq 1$ is the correlation coefficient of degradation between the ith and the jth dimension. Here, t should be understood as a time-scale transformation (Whitmore and Schenkelberg, 1997); see details in Park and Padgett (2005); Wang (2010); Ye et al. (2013); Yan et al. (2016), to name a few. In our EC application, the time scale does not contain unknown parameters since we can linearize the data based on the physical mechanism as shown in Section 6. Such physical laws exist in many cases, e.g., the first-order chemical reaction law for the degradation of printed-circuit boards (Meeker et al., 1998). In other applications, we can further allow the possibility that each dimension of degradation has its own time scale transformation function $h_l(t)$, $l \in [d]$. When there are unknown parameters in $h_l(t)$, the estimation can still be readily done.

Next, the two layers of block effects are introduced into (1). Within the same rig, the

degradation is homogeneous and the degradation rate parameters of the units in the same rig are usually assumed identical (Li and Doganaksoy, 2014). Between rigs, a one-dimensional frailty is used to model the heterogeneity due to the rig-layer block effects. Specifically, we model the degradation rate $\boldsymbol{\upsilon}_{ijk}$ in the ith rig as $\boldsymbol{\upsilon}_{ijk} = \zeta_i \boldsymbol{\mu}$, where $\boldsymbol{\mu} = [\mu_1, \cdots, \mu_d]'$ is the mean degradation rate, and ζ_i, $i \in [n]$, are i.i.d. mean-one random variables representing the block effect. Since ζ_i is a rig-layer random effect, we follow the common practice of random-effects modeling for degradation processes (Wang, 2010) and let $\zeta_i \sim \mathcal{N}(1, \omega^2)$. The normality assumption to introduce such randomness has been widely used (Bae and Kvam, 2004; Weaver et al., 2013). To model the gauge-layer block effect, we introduce ϵ_{ij} for units from rig i and measured at t_j. Assume that ϵ_{ij}, $i \in [n]$, $j \in [m]$, are i.i.d. normal, i.e., $\epsilon_{ij} \sim \mathcal{N}(0, \kappa^2)$, and ϵ_{ij} are independent of the degradation process and ζ_i. Therefore, the observed degradation levels $\boldsymbol{Y}_{ijk}$ are given by

$$\boldsymbol{Y}_{ijk} = \boldsymbol{X}_{ijk}(t_j) + \epsilon_{ij}\mathbf{1}_d = \zeta_i \boldsymbol{\mu} t_j + \boldsymbol{\Sigma}^{1/2}\boldsymbol{\mathcal{B}}_{ijk}(t_j) + \epsilon_{ij}\mathbf{1}_d, \quad i \in [n],\ j \in [m],\ k \in [K], \tag{2}$$

where $\mathbf{1}_d$ is a d-column vector with all ones. The above model formulation is reasonable when all the d performance characteristics have a similar physical interpretation and are of the same order of magnitudes. This is the case for our EC motivating example because each dimension of degradation in Figure 1 corresponds to the concentration of a contaminant. Such multivariate degradation data are commonly seen in reality, e.g., the degradation of organic paints measured at different wavenumbers of infrared spectroscopy (Hong et al., 2018b), the degradation of high-strength steel measured at different selected points on the material surface (Si et al., 2018), and the degradation of railway track in longitudinal and transversal directions (Mercier et al., 2012). This formulation saves the number of model parameters to estimate, and the resulting model is tractable, parsimonious, and flexible.

In the proposed model, the variation in a degradation observation is the synergistic result of the rig-layer block effect, the gauge-layer block effect, as well as the inherent volatility of

the degradation process. By letting $\omega = 0$ or $\kappa = 0$, the rig-layer or the gauge-layer block effect is excluded from the model. Hence, the proposed model includes several models as its special cases, which facilitates the detection of possible block effects in DDTs. For example, when $\omega = \kappa = 0$, the proposed model degenerates to the basic multivariate Wiener process (Liu et al., 2014) commonly used for fitting multivariate degradation data. The rationale of the above setting can be justified using the degradation of d kinds of contaminants in a solution. Let $\boldsymbol{\mathcal{C}}_{ijk}(t) = [\mathcal{C}_{ijk}^{(1)}(t), \cdots, \mathcal{C}_{ijk}^{(d)}(t)]'$ be the concentrations of the d contaminants of the test unit indexed with "ijk" at time t. The pseudo first-order kinetics (Steinfeld et al., 1989, Chapter 1) suggests log-linear paths for all contaminants under ideal conditions,

$$\mathrm{d}\mathcal{C}_{ijk}^{(l)}(t) = -\tilde{\zeta}_i r_l \mathcal{C}_{ijk}^{(l)}(t)\mathrm{d}t, \quad l \in [d], \tag{3}$$

where $\tilde{\zeta}_i$ is the concentration of hydroxyl radicals in the ith repetition of the DDT, and r_l is the bimolecular rate constants for the lth contaminant. The above differential equation gives $\log\left(C_{ijk}^{(l)}(0)/C_{ijk}^{(l)}(t)\right) = \tilde{\zeta}_i r_l t$. Let $X_{ijk}^{(l)}(t) = \log[\mathcal{C}_{ijk}^{(l)}(0)/\mathcal{C}_{ijk}^{(l)}(t)]$ and $\boldsymbol{X}_{ijk}(t) = [X_{ijk}^{(1)}(t), \cdots, X_{ijk}^{(d)}(t)]'$. First, we model $\{\boldsymbol{X}_{ijk}(t), t \geq 0\}$ as a multivariate Wiener process $\boldsymbol{X}_{ijk}(t) = \tilde{\zeta}_i \boldsymbol{r} t + \boldsymbol{\Sigma}^{1/2}\boldsymbol{\mathcal{B}}_{ijk}(t)$ to account for the increasing uncertainty in $\boldsymbol{X}_{ijk}(t)$ over t, where $\boldsymbol{r} = [r_1, \cdots, r_d]'$. Second, the hydroxyl radicals cause all contaminants to degrade, while preparation of the solution and the ambient temperature may lead to slight difference of $\tilde{\zeta}_i$ in each repetition. Hence, we treat $\tilde{\zeta}_i$, $i \in [n]$, as realizations of a normal distribution, i.e., $\tilde{\zeta}_i \sim \mathcal{N}(\varphi, \tilde{\omega}^2)$. To make the model identifiable, the model is rewritten as $\boldsymbol{X}_{ijk}(t) = \zeta_i \boldsymbol{\mu} t + \boldsymbol{\Sigma}^{1/2}\boldsymbol{\mathcal{B}}_{ijk}(t)$, where $\boldsymbol{\mu} = \varphi \boldsymbol{r}$ and $\zeta_i \sim \mathcal{N}(1, \omega^2)$ with $\omega = \tilde{\omega}/\varphi$. Third, we add the same chemicals to quench the degradation and then use the same measurement system (LC-MS/MS) to measure the EC concentration, leading to measurement errors for the observed degradation levels. This block effect is factored in by adding $\epsilon_{ij}\mathbf{1}_d$ for the measurement taken at time t_j. This culminates in the proposed model given by (2).

3 Maximum Likelihood Estimation

This section uses the observed data $\mathbb{D}$ to estimate the unknown model parameters in (2), denoted as $\boldsymbol{\theta} \equiv (\boldsymbol{\mu}, \boldsymbol{\Sigma}, \omega, \kappa)$. For rig i, $i \in [n]$, we vectorize the observed degradation from the mK units as $\boldsymbol{\mathcal{Y}}_i \equiv [\boldsymbol{Y}'_{i11}, \cdots, \boldsymbol{Y}'_{i1K}, \cdots, \boldsymbol{Y}'_{im1}, \cdots, \boldsymbol{Y}'_{imK}]'$. Conditional on ζ_i, $\boldsymbol{X}_{ijk}(t) \sim \mathcal{N}_d(\zeta_i \boldsymbol{\mu} t, \boldsymbol{\Sigma} t)$ for all j and k, and thus $\boldsymbol{\mathcal{Y}}_i$ follows an M-dimensional multivariate normal distribution $\mathcal{N}_M(\zeta_i \boldsymbol{\mu}_{\boldsymbol{\mathcal{Y}}}, \boldsymbol{\Sigma}_{\boldsymbol{\mathcal{Y}}})$, where $M = dmK$,

$$\boldsymbol{\mu}_{\boldsymbol{\mathcal{Y}}} = [\underbrace{\boldsymbol{\mu}' t_1, \cdots, \boldsymbol{\mu}' t_1}_{K \text{ repetitions}}, \cdots, \underbrace{\boldsymbol{\mu}' t_m, \cdots, \boldsymbol{\mu}' t_m}_{K \text{ repetitions}}]' = \boldsymbol{t} \otimes \mathbf{1}_K \otimes \boldsymbol{\mu},$$

and

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{Y}}} &= \operatorname{diag}(\operatorname{diag}(\underbrace{\boldsymbol{\Sigma} t_1, \cdots, \boldsymbol{\Sigma} t_1}_{K \text{ repetitions}}) + \kappa^2 \mathbf{1}_{dK} \mathbf{1}'_{dK}, \cdots, \operatorname{diag}(\underbrace{\boldsymbol{\Sigma} t_m, \cdots, \boldsymbol{\Sigma} t_m}_{K \text{ repetitions}}) + \kappa^2 \mathbf{1}_{dK} \mathbf{1}'_{dK}) \\ &= \operatorname{diag}(\boldsymbol{t} \otimes \mathbf{1}_K) \otimes \boldsymbol{\Sigma} + \mathbf{I}_m \otimes (\kappa^2 \mathbf{1}_{dK} \mathbf{1}'_{dK})\end{aligned}$$

is a block diagonal matrix. Here, $\boldsymbol{t} = [t_1, \cdots, t_m]'$, $\mathbf{I}_m$ is the identity matrix of size m, and $\otimes$ denotes the Kronecker product. Conditional on ζ_i, the corresponding moment generating function (MGF) of $\boldsymbol{\mathcal{Y}}_i$ is given by $\mathcal{M}_{\boldsymbol{\mathcal{Y}}_i|\zeta_i}(\boldsymbol{z}) = \exp\left(\zeta_i \boldsymbol{\mu}'_{\boldsymbol{\mathcal{Y}}} \boldsymbol{z} + \frac{1}{2} \boldsymbol{z}' \boldsymbol{\Sigma}_{\boldsymbol{\mathcal{Y}}} \boldsymbol{z}\right)$. Integrating ζ_i out, we obtain the unconditional MGF of $\boldsymbol{\mathcal{Y}}_i$ as

$$\begin{aligned}\mathcal{M}_{\boldsymbol{\mathcal{Y}}_i}(\boldsymbol{z}) &= \int_{-\infty}^{\infty} \exp\left(\zeta_i \boldsymbol{\mu}'_{\boldsymbol{\mathcal{Y}}} \boldsymbol{z} + \frac{1}{2} \boldsymbol{z}' \boldsymbol{\Sigma}_{\boldsymbol{\mathcal{Y}}} \boldsymbol{z}\right) \frac{1}{\sqrt{2\pi}\omega} \exp\left[-\frac{(\zeta_i - 1)^2}{2\omega^2}\right] \mathrm{d}\zeta_i \\ &= \exp\left[\boldsymbol{\mu}'_{\boldsymbol{\mathcal{Y}}} \boldsymbol{z} + \frac{1}{2} \boldsymbol{z}' (\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{Y}}} + \omega^2 \boldsymbol{\mu}_{\boldsymbol{\mathcal{Y}}} \boldsymbol{\mu}'_{\boldsymbol{\mathcal{Y}}}) \boldsymbol{z}\right].\end{aligned}$$

The MGF of $\boldsymbol{\mathcal{Y}}_i$ suggests that $\boldsymbol{\mathcal{Y}}_i \sim \mathcal{N}_M(\boldsymbol{\mu}_{\boldsymbol{\mathcal{Y}}}, \boldsymbol{\Sigma}_{\boldsymbol{\mathcal{Y}}} + \omega^2 \boldsymbol{\mu}_{\boldsymbol{\mathcal{Y}}} \boldsymbol{\mu}'_{\boldsymbol{\mathcal{Y}}})$, $i \in [n]$. Hence, the log-likelihood function (up to a constant) of $\boldsymbol{\theta}$ is

$$\ell(\boldsymbol{\theta}) = -\frac{1}{2} \sum_{i=1}^{n} \left\{\ln[\det(\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{Y}}} + \omega^2 \boldsymbol{\mu}_{\boldsymbol{\mathcal{Y}}} \boldsymbol{\mu}'_{\boldsymbol{\mathcal{Y}}})] + (\boldsymbol{\mathcal{Y}}_i - \boldsymbol{\mu}_{\boldsymbol{\mathcal{Y}}})' (\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{Y}}} + \omega^2 \boldsymbol{\mu}_{\boldsymbol{\mathcal{Y}}} \boldsymbol{\mu}'_{\boldsymbol{\mathcal{Y}}})^{-1} (\boldsymbol{\mathcal{Y}}_i - \boldsymbol{\mu}_{\boldsymbol{\mathcal{Y}}})\right\}. \quad (4)$$

Direct maximization of (4) is difficult because the dimension of $\boldsymbol{\theta}$ is relatively high, and the log-likelihood may not be concave in $\boldsymbol{\theta}$. In this section, we use an EM algorithm by treating the unobserved block effects as missing data.

3.1 The EM Algorithm

There are two layers of block effects. Both the rig-layer and the gauge-layer block effects, $\{\zeta_i; i \in [n]\}$ and $\{\epsilon_{ij}; i \in [n], j \in [m]\}$, are unobserved and treated as missing data. With the complete data $\mathbb{D}_{\mathrm{c}} = \mathbb{D} \cup \{\zeta_i, \epsilon_{ij}; i \in [n], j \in [m]\}$, the Q-function resembles the multivariate normal log-likelihood, making the maximization straightforward. The log-likelihood function (up to a constant) based on $\mathbb{D}_{\mathrm{c}}$ is $\ell_{\mathrm{c}}(\boldsymbol{\theta}; \mathbb{D}) = \ell_{\mathrm{c}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) + \ell_{\mathrm{c}}(\kappa) + \ell_{\mathrm{c}}(\omega)$, where

$$\ell_{\mathrm{c}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = -\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{K} \{\ln[\det(\boldsymbol{\Sigma})] + (\boldsymbol{Y}_{ijk} - \epsilon_{ij}\mathbf{1}_d - \zeta_i \boldsymbol{\mu} t_j)' (\boldsymbol{\Sigma} t_j)^{-1} (\boldsymbol{Y}_{ijk} - \epsilon_{ij}\mathbf{1}_d - \zeta_i \boldsymbol{\mu} t_j)\},$$

$$\ell_{\mathrm{c}}(\kappa) = -nm \ln(\kappa) - \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{\epsilon_{ij}^2}{2\kappa^2}, \quad \text{and} \quad \ell_{\mathrm{c}}(\omega) = -n \ln(\omega) - \sum_{i=1}^{n} \frac{(\zeta_i - 1^2)}{2\omega^2}.$$

Let $\boldsymbol{\theta}^{(\tau)}$ be the EM estimates of $\boldsymbol{\theta}$ at the τth iteration. Then at the $(\tau+1)$st iteration, we compute the Q-function $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(\tau)}) \equiv \mathbb{E}[\ell_{\mathrm{c}}(\boldsymbol{\theta}; \mathbb{D})|\boldsymbol{\theta}^{(\tau)}, \mathbb{D}]$ in the E-step and maximize it with respect to $\boldsymbol{\theta}$ to obtain $\boldsymbol{\theta}^{(\tau+1)}$ in the M-step. By taking the first-order derivatives of the Q-function with respect to $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, κ, and ω, and letting them equal zero, we can update the estimates of model parameters in the M-step at the $(\tau+1)$st iteration. The M-step needs values of some expectations relevant to the missing data. Therefore, we derive the distribution of missing data in the E-step. Fortunately, our proposed block-effects model admits *closed-form* expectations and maximizers in the E- and the M-step, respectively. It significantly facilitates the application of our model in reality. The details of the E- and the M-step are provided in Section S.1.1 of the supplement.

We also make two remarks on the numerical computation of the developed EM algorithm. First, Section S.1.1 of the supplement shows that the E-step needs to invert two matrices of

size $M \times M$. One of them is a diagonal matrix. The inverse is computed by replacing each diagonal element with its reciprocal. The other matrix is block diagonal with each block of size $d \times d$. The computational complexity for its inverse is $\mathcal{O}(Md^2)$ with the singular value decomposition. It is significantly smaller than the general case where the computational complexity is of order $\mathcal{O}(M^3)$. Therefore, the computational burden at each iteration is low, ensuring a fast implementation of the EM algorithm in reality. Second, when the data size is large, the likelihood function may have numerous local optima. Therefore, it is important to start the EM algorithm with an educated guess of the parameters. In Section S.1.2 of the supplement, we propose a workable procedure to find such an educated guess.

When each dimension of degradation has its own transformed time scale $h_l(t)$, $l \in [d]$, and there are unknown parameters in these functions, the EM algorithm is still applicable for statistical inference. Details of the EM algorithm with a time scale transformation in each dimension of degradation are provided in Section S.1.3 of the supplement. For the proposed EM algorithm, we can show that its limit points are stationary points of the incomplete data log-likelihood ℓ in (4), because the corresponding Q-function $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(\tau)})$ is continuous in both $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^{(\tau)}$. To check the performance of the EM algorithm, a commonly adopted strategy is to compute the biases and mean squared errors numerically through simulations (Pena et al., 2001; van Ryzin and Vulcano, 2017), which is adopted in Section 5.2.

3.2 Interval Estimation

The Fisher information matrix is needed in interval estimation, when the large-sample normal approximation or the bootstrap-t is used. Since each observation is multivariate, it is not a simple task to compute the Fisher information matrix. For notation convenience, we rewrite the model parameters as $\boldsymbol{\theta} = [\theta_1, \cdots, \theta_{d(d+3)/2+2}]'$, where $\theta_1, \cdots, \theta_d$ denote the parameters in $\boldsymbol{\mu}$, $\theta_{d+1}, \cdots, \theta_{d(d+3)/2}$ denote the parameters in $\boldsymbol{\Sigma}$, $\theta_{d(d+3)/2+1} \equiv \omega$, and $\theta_{d(d+3)/2+2} \equiv \kappa$. Based on a similar idea in Besson and Abramovich (2013), we compute the Fisher information

matrix $\mathcal{I}(\boldsymbol{\theta})$ entry-by-entry. The (k,l)th entry is given by

$$[\mathcal{I}(\boldsymbol{\theta})]_{k,l} = \frac{n}{2}\mathrm{tr}(\mathbf{A}_k\mathbf{A}_l) + n\left(\frac{\partial \boldsymbol{\mu}_{\boldsymbol{\mathcal{Y}}}}{\partial \theta_k}\right)'(\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{Y}}} + \omega^2\boldsymbol{\mu}_{\boldsymbol{\mathcal{Y}}}\boldsymbol{\mu}_{\boldsymbol{\mathcal{Y}}}')^{-1}\left(\frac{\partial \boldsymbol{\mu}_{\boldsymbol{\mathcal{Y}}}}{\partial \theta_l}\right), \tag{5}$$

for $k,l \in [d(d+3)/2+2]$. The derivation of (5), the expressions of $\mathbf{A}_k$, and the partial derivatives in (5) are given in Section S.2.1 of the supplement.

In addition to $\boldsymbol{\theta}$, we are usually interested in some population characteristic $g(\boldsymbol{\theta})$ that is a function of $\boldsymbol{\theta}$. Typical examples include the expected remaining useful life of a machinery system and the bimolecular rate constant of an EC with respect to a specific reactive intermediate. Its ML estimator is given by $g(\widehat{\boldsymbol{\theta}})$. The asymptotic variance of $g(\widehat{\boldsymbol{\theta}})$ can be obtained using the delta method with the Fisher information matrix (5), which is given by

$$\mathrm{Avar}(g(\widehat{\boldsymbol{\theta}})) = \left[\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]'\mathcal{I}(\boldsymbol{\theta})^{-1}\left[\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]\Bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}. \tag{6}$$

The large-sample normal approximation can be used to construct a confidence interval for $g(\boldsymbol{\theta})$. When the sample size is moderate, a confidence interval based on the bootstrap-t is usually helpful in improving the coverage probability. The procedure is recapitulated in Section S.2.2 of the supplement.

4 Testing Significance of the Block Effects

This section concerns validation of the block effects using hypothesis testing. To quantitatively detect the block effects, we extend the method of Wilk's Lambda for multivariate analysis of variance (MANOVA) (Rencher, 2003, Chapter 6) to accommodate the scenario of multiple comparisons due to multiple measurement times. Consider the nK degradation measurements $\boldsymbol{Y}_{ijk}$, $i \in [n]$ and $k \in [K]$, at time t_j. Let

$$\boldsymbol{\mathcal{H}}_j = K\sum_{i=1}^{n}(\bar{\boldsymbol{Y}}_{ij\cdot} - \bar{\boldsymbol{Y}}_{\cdot j\cdot})(\bar{\boldsymbol{Y}}_{ij\cdot} - \bar{\boldsymbol{Y}}_{\cdot j\cdot})' \quad \text{and} \quad \boldsymbol{\mathcal{E}}_j = \sum_{i=1}^{n}\sum_{k=1}^{K}(\boldsymbol{Y}_{ijk} - \bar{\boldsymbol{Y}}_{ij\cdot})(\boldsymbol{Y}_{ijk} - \bar{\boldsymbol{Y}}_{ij\cdot})',$$

where $\bar{\boldsymbol{Y}}_{ij\cdot} = \sum_{k=1}^{K} \boldsymbol{Y}_{ijk}/K$ and $\bar{\boldsymbol{Y}}_{\cdot j\cdot} = \sum_{i=1}^{n}\sum_{k=1}^{K} \boldsymbol{Y}_{ijk}/nK$. The ratio

$$\Lambda_j = \frac{\det(\boldsymbol{\mathcal{E}}_j)}{\det(\boldsymbol{\mathcal{E}}_j + \boldsymbol{\mathcal{H}}_j)} \tag{7}$$

is called Wilk's Lambda. Here, the matrices $\boldsymbol{\mathcal{H}}_j$ and $\boldsymbol{\mathcal{E}}_j$ measure the variance between and within the rig-layer blocks, respectively. Under the null hypothesis that both the rig- and the gauge-layer block effects are negligible at time t_j, all the nK multivariate measurements at time t_j should be i.i.d. In this case, Λ_j follows a Wilk's Lambda distribution with $(d, n-1, n(K-1))$ degrees of freedom (Rencher, 2003, Chapter 6) if the measurements are normally distributed. If either the rig- or the gauge-layer block effect is significant, $\det(\boldsymbol{\mathcal{E}}_j + \boldsymbol{\mathcal{H}}_j)$ should be much larger than $\det(\boldsymbol{\mathcal{E}}_j)$. Hence, we reject the null hypothesis if the test statistic computed from the DDT data is too small.

Since there are m measurements in each test rig, one possible method to examine the block effects is to extend the test with Wilk's Lambda to the scenario of multiple comparisons problem (Bretz et al., 2016). In this case, the m measurement times give rise to m hypothesis tests where the jth test is for the block effects at the jth measurement time, $j \in [m]$. The associated Λ_j, $j \in [m]$, are i.i.d. under the null. To deal with the multiple hypothesis tests, we may fix the family-wise error rate α and allocate a common significance level $1-(1-\alpha)^{1/m}$ to every hypothesis. When any of the m tests rejects its corresponding null, we reject the null hypothesis that no block effects exist. In our problem, note that all the m tests share the same null hypothesis. When the block effect is significant, the null for all the tests is false and all the Λ_j's will tend to be small. In this scenario, a more powerful approach to this multiple comparison problem is to pool the m test statistics as $\lambda = \sum_{j=1}^{m} \log(\Lambda_j)$, and base the hypothesis test on λ. Under the null hypothesis, the distribution of λ only depends on (n, m, K), and is free of the model parameters. The reason to use such a test statistic is that Λ_j's follow independent Wilk's Lambda distributions under the null. According to Rencher (2003, Chapter 6), $-[n(K-1)-0.5(d-n+2)]\log(\Lambda_j)$ is approximately chi-squared

distributed with $d(n-1)$ degrees of freedom, i.e., $\chi^2_{d(n-1)}$. This motivates us to pool the Λ_j's as $\sum_{j=1}^{m} \log(\Lambda_j)$, so that $-[n(K-1) - 0.5(d-n+2)]\lambda$ is approximately a chi-squared random variable with $dm(n-1)$ degrees of freedom. The distribution of λ can be obtained through the chi-squared approximation or the simulation. For the latter, we can generate nK realizations from a d-dimensional standard multivariate normal distribution and compute a realization of Λ_j using (7). Use the above routine to generate Λ_j, $j \in [m]$, and obtain a realization $\lambda = \sum_{j=1}^{m} \log(\Lambda_j)$. We then repeat the procedure for a large number of times and use the empirical quantiles to estimate the quantile of λ. If either the rig- or the gauge-layer block effect is significant, λ should be small. Thus, we reject the null hypothesis if the test statistic from data is smaller than the αth sample quantile of λ obtained above.

Since the model (2) includes the scenario of no block effect as special cases, it can be used to double validate the block effects through the likelihood ratio (LR) test. Details of the LR test are provided in Section S.3 of the supplement. A graph that plots the degradation data as in Figure 1 can also reveal the block effects. In addition, we can make use of a quantile-quantile (Q-Q) plot to assess the goodness-of-fit of the proposed model (2). In Section S.3 of the supplement, we also illustrate the details to construct the Q-Q plot. To compare the proposed model with other models overlooking the block effects, we adopt the Akaike information criterion (AIC), AIC $= 2|\boldsymbol{\theta}| - 2\ell$, where $|\boldsymbol{\theta}|$ is the number of model parameters and ℓ is the corresponding maximum log-likelihood.

5 Simulation Study

5.1 Power of the Hypothesis Test

Section 4 advocated testing the block effect using λ. An alternative is to conduct m hypothesis tests simultaneously, with the size of each hypothesis equal to $1-(1-\alpha)^{1/m}$. A comprehensive simulation using the model in (2) is conducted to compare the power of these two methods. As with the EC example, we consider three-dimensional degradation, i.e.,

$d = 3$, and set $(n, m, K) = (6, 7, 3)$. The measurement times are $t_j = j$, $j \in [m]$. In the simulation, we set $\boldsymbol{\mu} = [5, 8, 10]'$ and $(\sigma_1^2, \sigma_2^2, \sigma_3^2, \rho_{12}, \rho_{13}, \rho_{23}) = (1, 1.5, 2, 0.5, 0.6, 0.7)$. For the parameters ω and κ, we consider four scenarios for the block effects. In the first scenario, we let $\kappa = 0$, i.e., no gauge-layer block effect, and change the value of ω. The second scenario fixes $\omega = 0$ and changes the value of κ. In the other two scenarios, we examine the power when the two-layer block effects are both significant. We fix $\kappa = 1$ and $\omega = 0.05$, respectively, in the third and the fourth scenario, and plot the corresponding power curve as a function of ω or κ. In the simulation, the Type-I error is fixed at $\alpha = 0.05$. The powers of both methods are obtained based on $10,000$ Monte Carlo replications for each parameter setting, as shown in Figure 3. The plots reveal that the proposed test procedure based on λ has a higher power in nearly all parameter settings. This is expected because when the block effects are significant, all the values of Λ_j drift towards 0 simultaneously. Aggregation retains the mean of the drift while decreasing the variance.

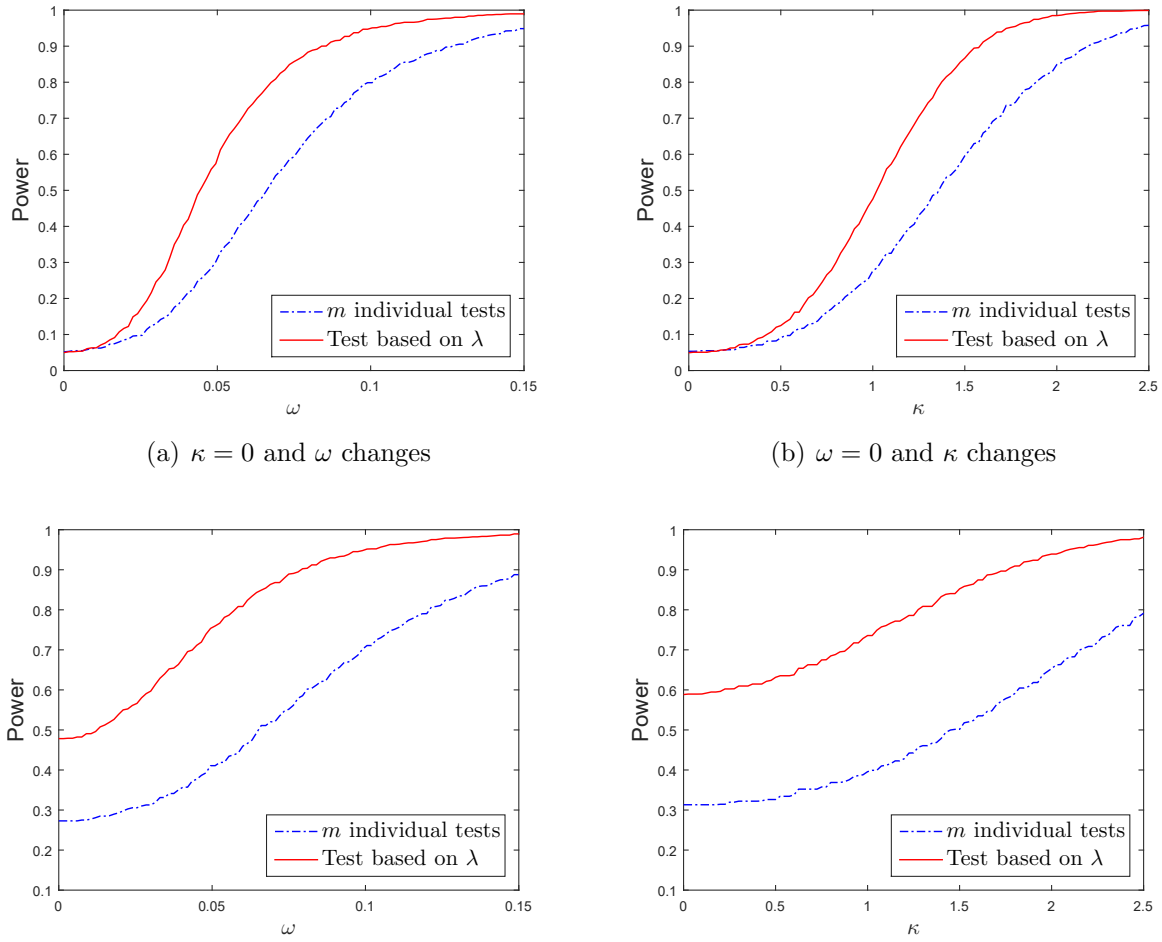
5.2 Performance of the EM Algorithm

Next, we evaluate the performance of the EM algorithm. The values of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the same as in Section 5.1, and we let $\omega = 0.2$, and $\kappa = 0.7$. We consider $n = 5$ and 10, $m = 5$ and 10, and $K = 1$ and 5. This leads to eight combinations of the sample size (n, m, K). The measurement times are $t_j \in [m]$. For each simulated dataset, the initial value of the EM algorithm is obtained from the procedure in Section S.1.2 of the supplement. The computational time for the EM algorithm to converge ranges from 2 to 13 seconds under different (n, m, K) on a personal computer with Intel Core i5-6600 CPU 3.30 GHz processor, which is satisfactory from a practical view point. The biases and root mean square errors (RMSEs) for $\boldsymbol{\theta}$ are computed based on $1,000$ Monte Carlo replications, as displayed in Tables S.1 and S.2 of the supplement. Generally, the biases and RMSEs decrease with the increase of sample size. The improvement in estimating $\boldsymbol{\mu}$ and ω is significant when increasing the number of rigs n. It is intuitive since ω characterizes the rig-layer block effect.

In comparison, the increase of m and K only slightly improves the estimation of $\boldsymbol{\mu}$ and ω. For the parameters in $\boldsymbol{\Sigma}$, the increase of K has the most significant effect on improving the estimation accuracy. It may be attributed to a more accurate estimation of the gauge-layer block effect parameter κ when K increases. The increase of the measurement times m can lead to a more accurate estimation of $\boldsymbol{\Sigma}$ and κ, but the improvement is not as significant as increasing K. Given a fixed number of the test units $N = nmK$ in a DDT, the simulation implies that larger values of n and K and a relatively small m can improve the estimation accuracy. In Section S.4.1 of the supplement, we further provide the simulation result to show that the proposed EM algorithm also performs well when d increases to 10.

5.3 Consequence of Overlooking the Block Effects

In the proposed model (2), all units in the same rig share an identical ζ to capture the rig-layer block effect, while all units destructively measured at the same time share an identical ϵ to capture the gauge-layer block effect. To understand the consequence of overlooking the blocking, we consider a pure random-effects model without blocking, where each test unit has an independent realization of (ζ, ϵ). We first use this model to fit the simulated data generated in Section 5.2. The corresponding biases and RMSEs are shown in Section S.4.1 of the supplement. The results show a significant increment in the biases and RMSEs of the estimated model parameters $\boldsymbol{\Sigma}$ and κ. We then compare the coverage probabilities of the two-sided 95% confidence intervals for $\boldsymbol{\theta}$ by using the two models when the data are generated from (2). The confidence interval is obtained from the bootstrap-t with $B = 2,000$ bootstrap replications. Based on $1,000$ Monte Carlo replications, Table S.6 in the supplement shows the coverage probability of the confidence intervals when $(n, m, K) = (5, 10, 5)$ and $(10, 10, 5)$, respectively. The bootstrap-t generally works well for the proposed model. When neglecting the block effects, on the other hand, the coverage probabilities of the confidence intervals are systematically inaccurate. The coverage probabilities for $\boldsymbol{\mu}$ and ω are significantly lower than the standard level. This implies that overlooking the blocking effects leads to a biased



(a) $\kappa = 0$ and ω changes

(b) $\omega = 0$ and κ changes

(c) $\kappa = 1$ and ω changes

(d) $\omega = 0.05$ and κ changes

Figure 3: Power curves for the two methods of hypothesis testing under the four scenarios. The powers are estimated using 10,000 Monte Carlo replicates.

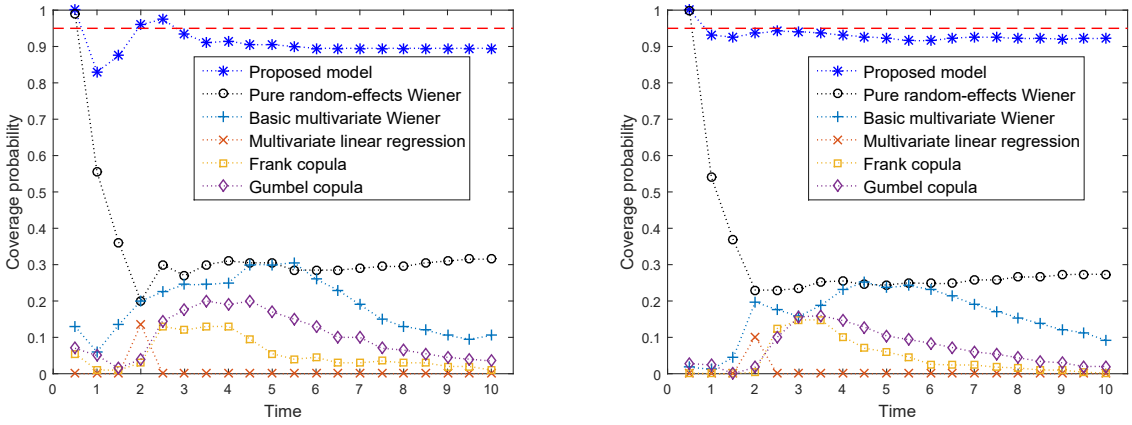
estimation of the Fisher information matrix and the asymptotic variance.

Motivated by the application of ECs, we also conduct a simulation to examine the coverage probability of the confidence intervals for two population characteristics, i.e., the ratio of degradation rates μ_{l_1}/μ_{l_2}, $l_1 \neq l_2$, and the probability that each dimension of degradation exceeds a corresponding threshold at time t given by $R(t) \equiv P[X^{(1)}(t) > D_1, \cdots, X^{(d)}(t) > D_d]$. Here, $X^{(l)}(t)$ is the true degradation level of the lth dimension at time t, and D_l is a predetermined threshold, $l \in [d]$. The detailed relationship between the two population characteristics and the EC application is elaborated in Section 6.1. Table 1 shows the corresponding empirical coverage probability of the 95% confidence intervals for μ_1/μ_3 and μ_2/μ_3

Table 1: Coverage probability (in %) of the 95% interval estimators for μ_1/μ_3 and μ_2/μ_3 based on $1,000$ simulation replications, where $(m, K) = (10, 5)$.

n	Candidate model	μ_1/μ_3	μ_2/μ_3	Candidate model	μ_1/μ_3	μ_2/μ_3
5	Proposed model	94.3	94.9	Random-effects model	92.5	95.2
	Basic Wiener process	92.0	95.5	Linear regression	79.6	81.5
	Frank copula	49.2	81.7	Gumbel copula	55.3	78.2
10	Proposed model	95.0	94.7	Random-effects model	92.9	94.6
	Basic Wiener process	93.5	94.5	Linear regression	79.6	80.6
	Frank copula	21.3	70.6	Gumbel copula	24.6	66.5

obtained from the proposed two-layer block-effects model, the pure random-effects model, the basic Wiener process without block effect (Liu et al., 2014), the multivariate linear regression (Rencher, 2003, Chapter 10), and two copula-based models with the Frank and the Gumbel copula functions (Wang et al., 2015), respectively. The copula models here are without block effects because there is no copula-based model with block effects to the best of our knowledge, and it remains unknown how to incorporate the block effects into the copula models. The coverage probabilities are based on $1,000$ simulation replications with $B = 1,000$ bootstrap resamples, and (n, m, K) is set to $(5, 10, 5)$ and $(10, 10, 5)$, respectively. Meanwhile, Figure 4 shows the corresponding coverage probabilities of the 95% bootstrap percentile confidence intervals for $R(t)$ at selected time points from the six candidate models. The thresholds are $(D_1, D_2, D_3) = (10, 15, 20)$ and the results are based on $B = 1,000$ resamples. For the degradation rate ratio μ_{l_1}/μ_{l_2}, $l_1 \neq l_2$, the coverage probability is generally satisfactory if the model is based on the multivariate Wiener process. On the other hand, the interval estimator of $R(t)$ is poor for all the models overlooking the block effects in terms of the significantly lower coverage probabilities. This could lead to inferior decisions in practice as shown in the water treatment example in Section 6.3. The bad performance of the benchmarks is because of the biases in estimation of the asymptotic variance of model parameters when the block effects are incorrectly overlooked.



(a) $(n, m, K) = (10, 10, 5)$ (b) $(n, m, K) = (20, 10, 5)$

Figure 4: The coverage probabilities of the point-wise bootstrap percentile confidence intervals for $R(t)$ in $1,000$ simulation replications.

6 Case Study

6.1 Overview of the EC Degradation Test

The proposed model is applied to degradation of $d = 3$ contaminants, i.e., Saccharin, Acesulfame, and p-Chlorobenzoic acid (pCBA). There is one test stand available in laboratory. To conduct the DDT on the test stand (as shown in Figure 5(a)), we first dissolve the three contaminants and hydrogen peroxide (H_2O_2) in deionized water, where initial concentrations of the three contaminants are all 100 μg/L. The solution of the contaminant mixture is distributed into 21 test tubes, and these tubes are placed into the test stand for the DDT. Within the test stand, there is a solar simulator that provides simulated natural sunlight. The photochemically produced reactive intermediate from H_2O_2, i.e., hydroxyl radicals, then induces degradation of the three contaminants during the test. At each of the $m = 7$ predetermined measurement times, $t_j = 0.15, 0.3, \cdots, 1.05$, we remove $K = 3$ tubes from the test stand. Degradation is quenched with the addition of *tertiary*-butanol into these tubes, and their degradation levels are then simultaneously measured by LC-MS/MS (as shown in Figure 5(b)). The above experiment is replicated six times, leading to degradation data from $n = 6$ rigs. Hence, there are $N = nmK = 126$ test units for the degra-

(a) Thermostat (b) LC-MS/MS

Figure 5: The test rig and the measurement system used in the EC degradation test.

dation test. The observed concentrations of Saccharin, Acesulfame, and pCBA are denoted as $\boldsymbol{C}_{ijk} = [C_{ijk}^{(1)}, C_{ijk}^{(2)}, C_{ijk}^{(3)}]'$, $i = 1, \cdots, 6$, $j = 1, \cdots, 7$, and $k = 1, \cdots, 3$. We have briefly discussed the possible reasons for the two-layer block effects for our DDT in Section 1, and Figure 1 has shown the degradation data from two rigs. A detailed physical interpretation for the block effects and the rest data are included in Section S.4.2 of the supplement. Based on the pseudo first-order kinetics (3), a logarithmic transformation is applied to $\boldsymbol{C}_{ijk}$. Since initial concentrations of all contaminants are 100 μg/L, we let

$$Y_{ijk}^{(l)} = \log\left(100/C_{ijk}^{(l)}\right), \quad l = 1, 2, 3, \tag{8}$$

and fit the data $\boldsymbol{Y}_{ijk} = [Y_{ijk}^{(1)}, Y_{ijk}^{(2)}, Y_{ijk}^{(3)}]'$ with the two-layer block-effects model (2).

The (transformed) DDT data $\boldsymbol{Y}_{ijk}$ can be used for risk assessment of the contaminants in terms of their environmental persistence. A standard measure of persistence is the bimolecular rate constant (Xu et al., 2011). Let r_l and μ_l be the respective bimolecular rate constant and the (transformed) mean degradation rate of the lth contaminant. The bimolecular rate constant r_l is defined to be μ_l divided by the concentration of hydroxyl radicals. Since the concentration of hydroxyl radicals is difficult to measure, computing r_l is not straightforward. It is customary to use a reference contaminant with known bimolecular rate constant, which

is pCBA in the above DDT, to facilitate the estimation of r_l of the other contaminants (Buxton et al., 1988). For convenience, let the third contaminant be pCBA in the DDT above. Then the bimolecular rate constants of the other two contaminants can be calculated as

$$r_l = \frac{\mu_l}{\mu_3} r_3, \quad l = 1 \text{ and } 2. \tag{9}$$

Another purpose of the DDT for ECs is to determine a suitable treatment time in a waterworks, which is the shortest time for achieving satisfactory removal of harmful contaminants. The DDT above is for estimating r_l. Nevertheless, a DDT for determining the treatment time is almost the same as the DDT above. The only difference is the types of ECs and their initial concentrations. Thus, we also use the data to demonstrate the estimation of treatment times by assuming that the purpose is to remove the three contaminants. The minimal treatment time $t_{\min}$ depends on the emission standards of the three contaminants and the required probability $\underline{p}$ with which the quality of the treated water meets the emission standard. Let $X^{(l)}(t)$ be the transformed concentration of the lth contaminant after t time units of degradation treatment, and D_l be the corresponding transformed emission threshold, $l = 1, 2, 3$. With a treatment time t, the probability of meeting the emission standard is $R(t) \equiv P[X^{(1)}(t) > D_1, X^{(2)}(t) > D_2, X^{(3)}(t) > D_3]$. In reality, the concentration of hydroxyl radicals for each replication of the water treatment can be different. Since the DDT is replicated n times, this variation can be captured by the rig-layer block effect. Hence, $R(t)$ can be expressed as a function of the model parameters $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, and ω as

$$R(t) = \int_0^\infty \frac{1}{\sqrt{2\pi}\omega} \exp\left[-\frac{(\zeta-1)^2}{2\omega^2}\right] \left(\int_{L_1}^\infty \int_{L_2}^\infty \int_{L_3}^\infty \phi_{\boldsymbol{\Sigma}}(z_1, z_2, z_3) \mathrm{d}z_3 \mathrm{d}z_2 \mathrm{d}z_1 \right) \mathrm{d}\zeta, \tag{10}$$

where $L_l = (D_l - \zeta\mu_l t)/\sqrt{t}$, $l = 1, 2, 3$, and $\phi_{\boldsymbol{\Sigma}}$ is the probability density function of the three-dimensional normal distribution $\mathcal{N}_3(\mathbf{0}, \boldsymbol{\Sigma})$. Given $\underline{p}$, the water treatment time can be obtained as $t_{\min} = R^{-1}(\underline{p})$. Here, $R^{-1}(\cdot)$ is the inverse function of $R(\cdot)$, which exists since $R(t)$ is strictly monotone increasing in t.

Table 2: AIC values for the model selection.

Candidate models	Log-likelihood	# of parameters	AIC
Multivariate linear regression	234.003	9	−450.007
Multivariate Wiener process without block effects	236.691	9	−455.382
Pure random-effects model	259.555	11	−497.110
Block-effects model	325.360	11	−628.720
Block-effects model (time scale transformation)	326.282	14	−624.565
Frank copula	200.295	7	−386.59
Gumbel copula	213.877	7	−413.754

6.2 Data Analysis and Risk Assessment

The raw data after transformation using (8) are shown in Figure S.1 of the supplement. The aggregate Wilk's Lambda λ, as developed in Section 4, is employed to examine the significance of the block effects. Based on $100,000$ Monte Carlo replications, the critical value given $\alpha = 0.05$ is -10.58. The test statistic computed from the DDT data is -21.96 with a p-value smaller than 1×10^{-5}. As discussed in Section 4, Rencher (2003, Chapter 6) also suggested that the statistic $-[n(K-1) - 0.5(d-n+2)]\lambda = -12.5\lambda$ approximately follows a chi-squared distribution with $dm(n-1) = 105$ degrees of freedom. The corresponding critical value is -10.39 for $\alpha = 0.05$, and the p-value from our data is 0. The test results indicate statistically significant block effects in the DDT data. The likelihood ratio tests based on the proposed model (2) are further used to examine the two layers of block effects. The tests for both $H_0 : \kappa = 0$ and $H_0 : \omega = 0$ lead to p-values near 0. The results reveal that both layers are significant in our DDT data, which validates our discussion in Section 1.

The proposed model (2) is then applied to fit the EC degradation data. The ML estimates of the model parameters are $\hat{\boldsymbol{\mu}} = [1.658, 2.892, 2.874]'$,

$$\hat{\boldsymbol{\Sigma}} = \begin{bmatrix} 0.0425 & 0.0784 & 0.0718 \\ & 0.152 & 0.142 \\ \text{Symmetric} & & 0.145 \end{bmatrix},$$

$\hat{\omega} = 0.139$, and $\hat{\kappa} = 0.123$. The computational time for the EM algorithm to converge is

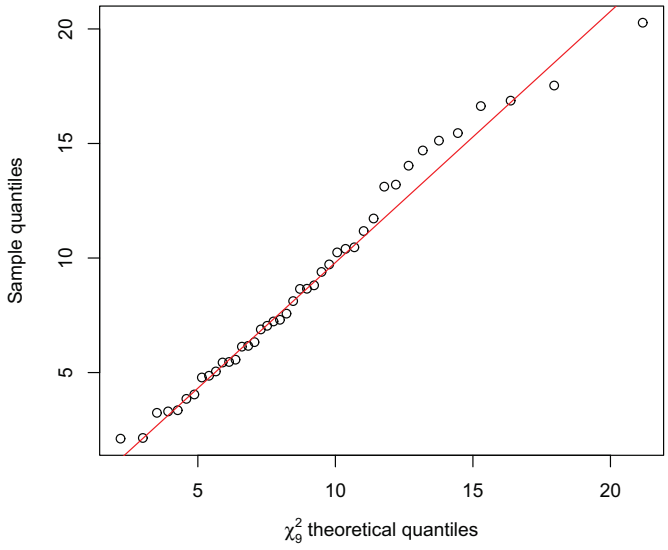


Figure 6: The χ^2_9 Q-Q plot for the proposed model using the EC degradation data.

around two seconds with starting points from Section S.1.2 of the supplement. For comparison, the DDT data are also fitted by the candidate models used in Section 5.3. Meanwhile, we also fit the data with a time scale transformation $h_l(t) = t^{q_l}$, $l = 1, 2, 3$. The resulting AIC values are given in Table 2. The AIC favors the proposed model, again indicating the significance of the block effects in the data. It is worth mentioning that when we use the proposed model with a time scale transformation in each dimension to fit the data, the ML estimates for the parameters in $h_l(t)$ are $(\hat{q}_1, \hat{q}_2, \hat{q}_3) = (1.06, 1.04, 1.04)$. It implies that the linearization based on the pseudo first-order kinetics (3) is satisfactory. To examine the goodness-of-fit, we construct the χ^2_{dK} Q-Q plot as introduced in Section 4 with $dK = 3 \times 3 = 9$. The result is shown in Figure 6. There is no obvious departure of the proposed model to the data, as the empirical quantiles scatter around the straight line.

Risk assessment of the contaminants is made based on $\boldsymbol{\mu}$ and $\boldsymbol{r} = [r_1, r_2, r_3]'$. As introduced in Section 6.1, the bimolecular rate constant of the reference contaminant pCBA with respect to the hydroxyl radical is known to be $r_3 = 5 \times 10^9$ $\mathrm{M}^{-1} \cdot \mathrm{s}^{-1}$ (Buxton et al., 1988). From (9), we can readily obtain the ML estimates of r_1 and r_2. The results are shown in

Table 3: The ML estimates and the 95% confidence intervals/bounds of the degradation rate, the bimolecular rate constant, and the correlation coefficient between the contaminants' degradation in the DDT. Index of the contaminants: 1–Saccharin; 2–Acesulfame; 3–pCBA.

	ML estimate	Standard error	Upper-tailed	Lower-tailed	Two-sided	Length
μ_1	1.658	0.100	(1.441, ∞)	($-\infty$, 1.859)	(1.388, 1.903)	0.516
μ_2	2.892	0.171	(2.504, ∞)	($-\infty$, 3.237)	(2.416, 3.321)	0.905
μ_3	2.874	0.170	(2.491, ∞)	($-\infty$, 3.219)	(2.405, 3.302)	0.897
$r_1(\times 10^{-9})$	2.884	0.0271	(2.842, ∞)	($-\infty$, 2.929)	(2.834, 2.937)	0.103
$r_2(\times 10^{-9})$	5.031	0.0220	(4.997, ∞)	($-\infty$, 5.066)	(4.990, 5.072)	0.082
ρ_{12}	0.975	5.05×10^{-3}	(0.967, ∞)	($-\infty$, 0.983)	(0.965, 0.984)	0.019
ρ_{13}	0.915	0.0250	(0.889, ∞)	($-\infty$, 0.940)	(0.883, 0.946)	0.063
ρ_{23}	0.959	7.31×10^{-3}	(0.946, ∞)	($-\infty$, 0.969)	(0.944, 0.971)	0.027

Table 3. For interval estimations, we run a bootstrap-t with $2,000$ bootstrap replications. The two-sided 95% confidence intervals, the upper-, and the lower-tailed confidence bounds of μ_l and r_l are also shown in Table 3. Based on these estimates, we can rank the potential environmental persistence of the three contaminants in an ascending order as Acesulfame, pCBA, and Saccharin, where a longer persistence implies a higher risk. As a byproduct, the correlation coefficients between different contaminants of degradation can also be estimated. We can see from Table 3 that degradation of the three contaminants has a strong positive correlation, which justifies the needs for a multivariate degradation model. One possible reason for such positive correlation is that all the three contaminants react with the same reactive intermediate, i.e., hydroxyl radicals, in the test solution (cf. Equation (3)). As a result, their degradation processes during the experiment are all affected by the concentration of hydroxyl radicals in the solution, leading to our estimation result.

6.3 Determination of Water Treatment Time

For illustration, we set the transformed degradation emission thresholds of Saccharin, Acesulfame, and pCBA as $(D_1, D_2, D_3) = (2, 3, 3)$. The minimal water treatment time $t_{\min}$ in (10) is a function of the required probability $\underline{p}$ that the treated water meets the emission standard, where a large value of $\underline{p}$ is usually set to ensure a complete removal of the potentially harmful contaminants. We examine $\underline{p} \in (0.9, 1)$ and compute the corresponding

$t_{\min}$ based on the ML estimates in Section 6.2. The result is shown in Figure 7. For comparison, we also show the minimal water treatment time obtained separately from the basic multivariate Wiener process model without block effects and the pure random-effects model in Section 5.3. The corresponding 95% point-wise upper confidence bounds using the parametric percentile bootstrap with $2,000$ bootstrap replications are also shown in Figure 7. We can see that the basic multivariate Wiener process model without block effects tends to give a short treatment time, which might lead to an incomplete removal of contaminants in water. By contrast, the water treatment time obtained from the pure random-effects model may not be cost-effective because it could be unnecessarily long. The above results could be attributed to the fact that $t_{\min}$ can be interpreted as the $\underline{p}$th quantile of the time at which degradation level of each contaminant has exceeded a respective (transformed) emission standard. The basic multivariate Wiener process incorrectly overlooks the block effect, so that it underestimates the variation of degradation in real applications. As a result, we could underestimate the corresponding $\underline{p}$th quantile of the treatment time. On the other side, simulation result in Section S.4.1 of the supplement suggests that the pure random-effects model tends to overestimate the inherent degradation volatility when the data are from a DDT with two-layer blocking. It may explain the possibly unnecessarily long treatment time from this model. The proposed model correctly captures the variation of degradation by considering both the heterogeneity among blocks and the homogeneity within a block. Consequently, the resulting treatment time is in between of the two extreme cases.

7 Conclusions and Discussions

Blocking in a DDT introduces undesired incomplete randomization in the experiment. This study has successfully proposed a multivariate statistical model for block effects in DDTs. Under the proposed model, the variation of each degradation observation is a synergistic result of the rig-layer block effect, the gauge-layer block effect, and the inherent degra-

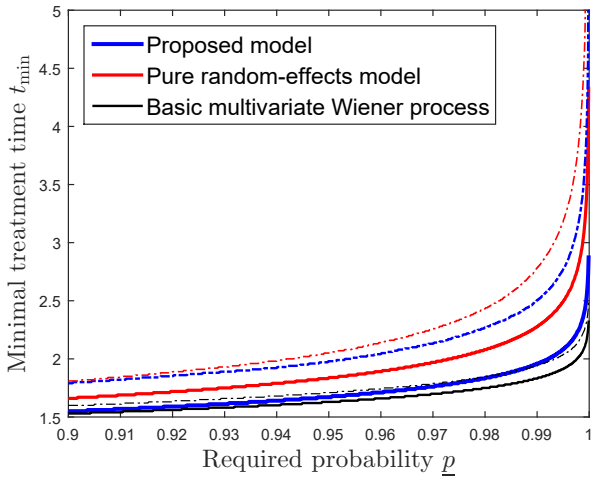


Figure 7: The minimal water treatment time $t_{\min}$ as a function of the required probability $\underline{p}$ that the treated water meets the emission standard. The dot-dashed curves are the corresponding 95% point-wise upper confidence bounds of the three models.

dation volatility of the test unit. The block effects are examined through the aggregate MANOVA that aggregates information from all measurement times. Comprehensive simulations showed that this test procedure is more powerful than traditional methods in multiple comparison problems. If blocking is incorrectly overlooked, the resulting estimates may have large estimation errors for point estimation and inaccurate coverage probabilities for interval estimation. We demonstrated the proposed model with two-layer block effects using an EC degradation dataset. The inference results were used for risk assessment of the contaminants in terms of their potential environmental persistence and for determining the minimal water treatment time for a waterworks to remove the contaminants. The statistical model in this study was developed based on the experiment setting that the measurement times for all rigs/replications are the same. Such an experiment setting is common in practice since it is easy to check the reproducibility of the experiment result. Nevertheless, our proposed estimation method is still workable when the measurement time for each rig layer is different. Despite that the test based on Wilk's Lambda may not be applicable in this case, we can still use the LR test to examine the significance of the two-layer block effects.

A common ϵ_{ij} was added in (2) for all the d performance characteristics of degradation at

t_j for the gauge-layer block effect. We have verified the assumption when all the performance characteristics have a similar physical interpretation and are of the same order of magnitude. In case that the order of magnitude for each dimension of degradation varies significantly from each other, we may suggest introducing a common multiplicative measurement error at t_j by multiplying ϵ_{ij} to $\boldsymbol{X}_{ijk}(t_j)$. This model could be more difficult for parameter estimation, and we may investigate its inference procedure as a possible future research topic. The simulation study in Section 5.2 showed that, with a fixed number of test units N, we may allocate a larger value on n and K, and a relative small value on m to improve the statistical estimation accuracy. Further consideration may be possible to investigate the optimal allocation for the DDT with blocking. In addition, the experimental conditions, such as the illumination strength and the concentration of H_2O_2, were fixed in our test. We may change the experimental conditions in each replication of the test to accelerate the degradation, which is known as the accelerated degradation tests (Feiveson and Kulkarni, 2000; Tseng and Wen, 2000). The statistical inference and the development of the optimal testing plan in accelerated destructive degradation tests with blocking are also worth investigation.

Supplementary Materials

The PDF file includes details of the EM algorithm, the interval estimation, the goodness-of-fit test, and the LR test; additional simulation results in Section 5.3; and graphical illustrations of the linearized data used in Section 6. The zipped package includes data and codes to implement the EM algorithm and to compute the Fisher information matrix.

Acknowledgment

The authors would like to thank the Editor, the Associate Editor, and three anonymous reviewers for their constructive comments which have led to a substantial improvement to an earlier version of the paper. Sun and Ye were supported by Singapore MOE AcRF Tier

2 under Grant R-266-000-125-112 and the National Science Foundation of Jiangsu Province under Grant BK20180232. Hong was partially supported by the National Science Foundation under sub-award CMMI-1904165 to Virginia Tech.

References

Bae, S. J. and Kvam, P. H. (2004), "A nonlinear random-coefficients model for degradation testing," *Technometrics*, 46(4), 460–469.

Besson, O. and Abramovich, Y. I. (2013), "On the Fisher information matrix for multivariate elliptically contoured distributions," *IEEE Signal Processing Letters*, 20(11), 1130–1133.

Blsgaard, S. and Steinberg, D. M. (1997), "The design and analysis of $2^{k-p} \times s$ prototype experiments," *Technometrics*, 39(1), 52–62.

Box, G. E., Hunter, W. G., and Hunter, J. S. (2005), *Statistics for Experimenters: an Introduction to Design, Data Analysis, and Model Building*, Wiley, 2nd ed.

Bretz, F., Westfall, P., and Hothorn, T. (2016), *Multiple Comparisons Using R*, Chapman and Hall/CRC.

Buxton, G. V., Greenstock, C. L., Helman, W. P., and Ross, A. B. (1988), "Critical review of rate constants for reactions of hydrated electrons, hydrogen atoms and hydroxyl radicals ($\cdot$OH/$\cdot$O$^-$) in aqueous solution," *Journal of Physical and Chemical Reference Data*, 17(2), 513–886.

Duan, Y., Hong, Y., Meeker, W. Q., Stanley, D. L., and Gu, X. (2017), "Photodegradation modeling based on laboratory accelerated test data and predictions under outdoor weathering for polymeric materials," *The Annals of Applied Statistics*, 11(4), 2052–2079.

Feiveson, A. H. and Kulkarni, P. M. (2000), "Reliability of space-shuttle pressure vessels with random batch effects," *Technometrics*, 42(4), 332–344.

Freeman, L. J. and Vining, G. G. (2010), "Reliability data analysis for life test experiments with subsampling," *Journal of Quality Technology*, 42(3), 233–241.

Giorgio, M., Guida, M., and Pulcini, G. (2010), "A state-dependent wear model with an application to marine engine cylinder liners," *Technometrics*, 52(2), 172–187.

Hong, L., Ye, Z.-S., and Ling, R. (2018a), "Environmental risk assessment of emerging contaminants using degradation data," *Journal of Agricultural, Biological, and Environmental Statistics*, 23(3), 390–409.

Hong, Y., Zhang, M., and Meeker, W. Q. (2018b), "Big data and reliability applications: The complexity dimension," *Journal of Quality Technology*, 50(2), 135–149.

Kensler, J. L., Freeman, L. J., and Vining, G. G. (2015), "Analysis of reliability experiments with random blocks and subsampling," *Journal of Quality Technology*, 47(3), 235–251.

León, R. V., Li, Y., Guess, F. M., and Sawhney, R. S. (2009), "Effect of not having homogeneous test units in accelerated life tests," *Journal of Quality Technology*, 41(3), 241–246.

Li, M. and Doganaksoy, N. (2014), "Batch variability in accelerated-degradation testing," *Journal of Quality Technology*, 46(2), 171–180.

Liao, H. and Elsayed, E. A. (2006), "Reliability inference for field conditions from accelerated degradation testing," *Naval Research Logistics*, 53(6), 576–587.

Liu, X., Al-Khalifa, K. N., Elsayed, E. A., Coit, D. W., and Hamouda, A. S. (2014), "Criticality measures for components with multi-dimensional degradation," *IIE Transactions*, 46(10), 987–998.

Meeker, W. Q., Escobar, L. A., and Lu, C. J. (1998), "Accelerated degradation tests: modeling and analysis," *Technometrics*, 40(2), 89–99.

Mercier, S., Meier-Hirmer, C., and Roussignol, M. (2012), "Bivariate Gamma wear processes for track geometry modelling, with application to intervention scheduling," *Structure and Infrastructure Engineering*, 8(4), 357–366.

Nelson, W. (1981), "Analysis of performance-degradation data from accelerated tests," *IEEE Transactions on Reliability*, 30(2), 149–155.

Park, C. and Padgett, W. J. (2005), "New cumulative damage models for failure using stochastic processes as initial damage," *IEEE Transactions on Reliability*, 54(3), 530–540.

Pena, E. A., Strawderman, R. L., and Hollander, M. (2001), "Nonparametric estimation with recurrent event data," *Journal of the American Statistical Association*, 96(456), 1299–1315.

Peng, W., Ye, Z.-S., and Chen, N. (2018), "Joint online RUL prediction for multivariate deteriorating systems," *IEEE Transactions on Industrial Informatics*, 15(5), 2870–2878.

Petyuk, V. A., Jaitly, N., Moore, R. J., Ding, J., Metz, T. O., Tang, K., Monroe, M. E., Tolmachev, A. V., Adkins, J. N., Belov, M. E., et al. (2008), "Elimination of systematic mass measurement errors in liquid chromatography-mass spectrometry based proteomics using regression models and a priori partial knowledge of the sample content," *Analytical Chemistry*, 80(3), 693–706.

Rafiee, K., Feng, Q., and Coit, D. W. (2014), "Reliability modeling for dependent competing failure processes with changing degradation rate," *IIE Transactions*, 46(5), 483–496.

Rencher, A. C. (2003), *Methods of Multivariate Analysis*, John Wiley & Sons.

Seo, K. and Pan, R. (2017), "Data analysis of step-stress accelerated life tests with heterogeneous group effects," *IISE Transactions*, 49(9), 885–898.

Shi, Y., Escobar, L. A., and Meeker, W. Q. (2009), "Accelerated destructive degradation test planning," *Technometrics*, 51(1), 1–13.

Si, W., Yang, Q., Wu, X., and Chen, Y. (2018), "Reliability analysis considering dynamic material local deformation," *Journal of Quality Technology*, 50(2), 183–197.

Si, X.-S. (2015), "An adaptive prognostic approach via nonlinear degradation modeling: Application to battery data," *IEEE Transactions on Industrial Electronics*, 62(8), 5082–5096.

Steinfeld, J. I., Francisco, J. S., and Hase, W. L. (1989), *Chemical Kinetics and Dynamics*, vol. 3, Prentice Hall: Englewood Cliffs, New Jersey.

Sun, Q., Ye, Z.-S., Revie, M., and Walls, L. (2019), "Reliability modelling of infrastructure load-sharing systems with workload adjustment," *IEEE Transactions on Reliability*, to appear.

Tseng, S.-T. and Wen, Z.-C. (2000), "Step-stress accelerated degradation analysis for highly reliable products," *Journal of Quality Technology*, 32(3), 209–216.

van Ryzin, G. and Vulcano, G. (2017), "An expectation-maximization method to estimate a rank-based choice model of demand," *Operations Research*, 65(2), 396–407.

Wang, X. (2010), "Wiener processes with random effects for degradation data," *Journal of Multivariate Analysis*, 101(2), 340–351.

Wang, X., Balakrishnan, N., Guo, B., and Jiang, P. (2015), "Residual life estimation based on bivariate non-stationary gamma degradation process," *Journal of Statistical Computation and Simulation*, 85(2), 405–421.

Weaver, B. P., Meeker, W. Q., Escobar, L. A., and Wendelberger, J. (2013), "Methods for planning repeated measures degradation studies," *Technometrics*, 55(2), 122–134.

Whitmore, G. and Schenkelberg, F. (1997), "Modelling accelerated degradation data using Wiener diffusion with a time scale transformation," *Lifetime Data Analysis*, 3(1), 27–45.

Xu, Y., Nguyen, T. V., Reinhard, M., and Gin, K. Y.-H. (2011), "Photodegradation kinetics of p-tert-octylphenol, 4-tert-octylphenoxy-acetic acid and ibuprofen under simulated solar conditions in surface water," *Chemosphere*, 85(5), 790–796.

Yan, H., Liu, K., Zhang, X., and Shi, J. (2016), "Multiple sensor data fusion for degradation modeling and prognostics under multiple operational conditions," *IEEE Transactions on Reliability*, 65(3), 1416–1426.

Ye, Z.-S. and Sun, Q. (2019), "Discussion on 'Challenges and new methods for designing reliability experiments'," *Quality Engineering*, 31(1), 125–128.

Ye, Z.-S., Wang, Y., Tsui, K.-L., and Pecht, M. (2013), "Degradation data analysis using Wiener processes with measurement errors," *IEEE Transactions on Reliability*, 62(4), 772–780.

Zeng, L., Deng, X., and Yang, J. (2016), "Constrained hierarchical modeling of degradation data in tissue-engineered scaffold fabrication," *IIE Transactions*, 48(1), 16–33.

Zhai, Q. and Ye, Z.-S. (2018), "Degradation in common dynamic environments," *Technometrics*, 60(4), 461–471.