IIE Transactions

ISSN: 0740-817X (Print) 1545-8830 (Online) Journal homepage: http://www.tandfonline.com/loi/uiie20

Constrained hierarchical modeling of degradation data in tissue-engineered scaffold fabrication

Li Zeng, Xinwei Deng & Jian Yang

To cite this article: Li Zeng, Xinwei Deng & Jian Yang (2016) Constrained hierarchical modeling of degradation data in tissue-engineered scaffold fabrication, IIE Transactions, 48:1, 16-33, DOI: 10.1080/0740817X.2015.1019164

To link to this article: http://dx.doi.org/10.1080/0740817X.2015.1019164

Accepted author version posted online: 15 Apr 2015.
Published online: 15 Apr 2015.

IIE TRANSACTIONS
2016, VOL. 48, NO. 1, 16–33
http://dx.doi.org/10.1080/0740817X.2015.1019164

Taylor & Francis
Taylor & Francis Group

Constrained hierarchical modeling of degradation data in tissue-engineered scaffold fabrication

Li Zeng[a], Xinwei Deng[b] and Jian Yang[c]

[a]The University of Texas at Arlington, Industrial and Manufacturing Systems Engineering, Arlington, VA, USA; [b]Virginia Tech, Statistics, Blacksburg, VA, USA; [c]Pennsylvania State University, Bioengineering, University Park, State College, PA, USA

ABSTRACT

In tissue-engineered scaffold fabrication, the degradation of scaffolds is a critical issue because it needs to match with the rate of new tissue formation in the human body. However, scaffold degradation is a very complicated process, making degradation regulation a challenging task. To provide a scientific understanding on the degradation of scaffolds, we propose a novel constrained hierarchical model (CHM) for the degradation data. The proposed model has two levels, with the first level characterizing scaffold degradation profiles and the second level characterizing the effect of process parameters on the degradation. Moreover, it can incorporate expert knowledge in the modeling through meaningful constraints, leading to insightful inference on scaffold degradation. Bayesian methods are used for parameter estimation and model comparison. In the case study, the proposed method is illustrated and compared with existing methods using data from a novel tissue-engineered scaffold fabrication process. A numerical study is conducted to examine the effect of sample size on model estimation.

ARTICLE HISTORY
Received 17 October 2014
Accepted 22 January 2015

KEYWORDS
Bayesian inference; Bayes factor; constrained hierarchical model; Gibbs sampling; tissue-engineered scaffolds

1. Introduction

The enormous need for tissue/organ grafts and the limitations of current therapies such as the shortage of organ donors have motivated the emergence of tissue engineering (TE) in recent years. TE is an interdisciplinary field, which combines the principles of polymer chemistry, engineering, and biological sciences in an effort to develop biological substitutes to repair/replace failing tissues and organs (Ikada, 2006; Fisher *et al.*, 2007). The basic concept of TE is illustrated in the left panel of Fig. 1, which involves growing relevant cells *in vitro* into a three-dimensional tissue/organ such as bone, skin, and heart valves prior to implantation. As cells alone lack the ability to grow in favored orientations similar to the native tissues, tissue formation is achieved by seeding the cells onto porous matrices known as *scaffolds* (Ma and Elisseeff, 2006; Laurencin and Nair, 2008). Biological molecules such as adhesive proteins are usually added to stimulate cellular growth. The scaffolds play a critical role in the development of engineered tissues/organs as a temporary substrate and microenvironment for cell accommodation and proliferation. For this reason, scaffold fabrication has become a major research area in biomanufacturing (Chu and Liu, 2008; Tateishi, 2008).

The scaffold fabrication process, shown in Fig. 1, consists of two steps: first, the scaffold material is synthesized. The material must be *biodegradable* so that the scaffold will eventually be replaced by the new tissues/organs in the human body (Fisher *et al.*, 2007). Then pores are constructed on the material to provide space for cell growth and flow transport of nutrients and metabolic wastes. Various techniques have been developed for pore construction, such as the widely used particulate leaching method (Liao *et al.*, 2002) as illustrated in Fig. 1; the polymer solution is first mixed with the porogen, usually sodium chloride (salt), to form a slurry, and then the slurry is cast into a mold and kept in a heated oven for solvent evaporation and polymerization. Finally, the dried composite is immersed in distilled water where the salt particles leach out, leaving behind a porous structure. The properties of scaffold products are determined by the setting of the parameters in the fabrication process such as the ratio of compositions in material synthesis, and the particle size of the porogen and polymerization conditions in pore construction, etc.

There are many considerations in scaffold fabrication to yield satisfactory products. A critical concern is the *degradation* of scaffolds within the human body (Buchanan, 2008). Scaffold degradation is a very complicated process caused by hydrolysis as illustrated in Fig. 2. In the human body, water molecules penetrate into the scaffold matrix, causing swelling of the matrix and triggering the breakdown of chemical chains that leads to weight loss of the scaffold. This process continues until the scaffold is completely dissolved. The rate of degradation is typically measured by degradation profiles shown in the right panel of Fig. 2, where the x-axis denotes the duration of the scaffold in the human body, and the y-axis denotes the corresponding percentage loss in scaffold properties (e.g., weight, mechanical strength, etc.). The degradation rate is crucial for tissue formation; fast degradation may cause fragile support to the cells, whereas slow degradation may impede the growth of new tissue. Therefore, it is important to appropriately regulate the degradation profile to match with the rate of new tissue formation in specific applications (Tran *et al.*, 2011). In the current scaffold

CONTACT Li Zeng lzeng@uta.edu

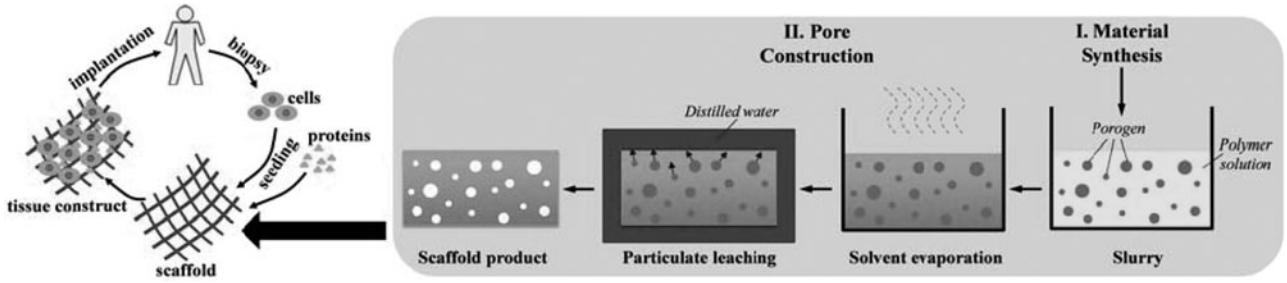

Figure 1. The concept of TE and scaffold fabrication process.

fabrication practice, however, degradation regulation is mainly done by trial-and-error experiments.

The key to the regulation of the degradation of a scaffold is to model the relationship between degradation profiles and process parameters. Such a model can be used to adjust the setting of process parameters to produce scaffold products with desirable degradation properties. There are two critical considerations in seeking proper models: first, the estimated model should be easily *interpretable* in terms of the intrinsic effect of process parameters on scaffold degradation. It can help researchers better understand the creation of scaffold products and their fabrication process. Since the physiochemical mechanisms of scaffold degradation are very complex, such insight is very important for conducting research on tissue-engineered scaffolds and developing novel biomaterials for scaffold fabrication. Second, expert knowledge on scaffold degradation (e.g., reasonable shapes of the degradation profile) should be taken into account in the modeling to make the results of any analysis practically meaningful.

Generally, there are two types of methods for modeling the scaffold degradation data: *analytical methods* based on physiochemical models of the degradation process and *data-driven methods* based on statistical models of degradation measurements. In the first direction, Chen *et al.* (2011) develop mechanistic modeling methods based on differential equations governing the underlying physiochemical processes of degradation. Metters (2001) proposes probabilistic modeling methods that describe state changes in polymer chains using stochastic models. However, these methods assume that the degradation mechanisms are known, which is not true in many applications (Buchanan, 2008).

Among data-driven methods, the following approaches can be potentially useful for modeling the degradation data:

(1) One approach is based on profile models in quality control (Woodall *et al.*, 2004; Noorossana *et al.*, 2011) since the degradation profile can be considered as a special type of quality profile data. However, the existing profile models are mostly one-dimensional, and can only characterize the relationship between degradation measurements and a single factor (e.g., time).

(2) Another approach is to treat process parameters and the degradation time as explanatory variables in surrogate models such as Gaussian Process (GP) models and Artificial Neural Networks (ANNs; Chen *et al.* (2006)). However, the resultant models may not be easy to interpret and it is difficult to incorporate expert knowledge.

(3) The third approach is the so-called Two-Step Regression (TSR) method for modeling dynamic responses in designed experiments (Govaerts and Noël, 2005; Goh, 2014; He *et al.*, 2015). It first builds a model for each observed profile and then obtains a model between the estimated parameters and design factors. Such a method has an intuitive interpretation, but it may introduce additional noise by using the estimated parameters as response in the second step.

In this work, we develop a constrained hierarchical modeling method to quantify the effect of process parameters on the degradation profiles of scaffold products. The proposed Constrained Hierarchical Model (CHM) has two levels, with the first level modeling degradation profiles and the second level modeling the effect of process parameters on the parameters of level-1 models. Expert knowledge is incorporated in the model through constraints. With limited data available on the scaffold fabrication process, we adopt the Bayesian framework (Daniels and Gatsonis, 1999; Gelman and Hill, 2007) for model estimation and model comparison. Specifically, Markov Chain Monte Carlo (MCMC) sampling procedures are used for estimating the model, and a Bayes factor method is developed for model comparison. In a similar spirit to the TSR method, the proposed method gives interpretable results. Moreover, it is easy to implement, flexible to incorporate expert knowledge, and has good prediction performance.

Hierarchical modeling, also called multilevel modeling, is a powerful tool to model data with complex structures. Special

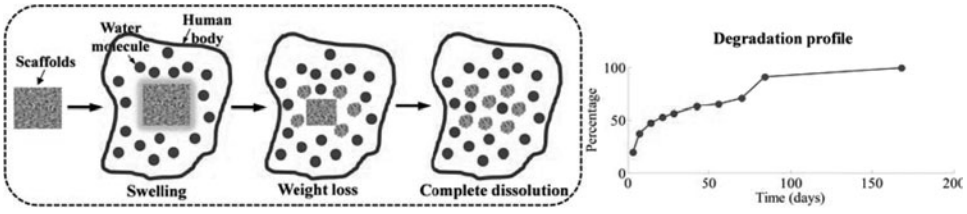

Figure 2. Scaffold degradation and example of degradation profile.

types of such models—e.g., mixed-effects models—have gained considerable popularity recently in social sciences and medical research (Verbeke and Molenberghs, 2000; Gelman and Hill, 2007; Jang, 2007). However, the models in existing work are unconstrained. This study will contribute by developing a systematic method to build constrained models including constraint formation based on expert knowledge and parameter estimation using Bayesian methods. Moreover, our study will also develop convenient Bayesian procedures for model comparison/selection of hierarchical models, which has not been considered in the literature.

The remainder of this article is organized as follows. Section 2 presents the proposed model. Section 3 formulates the tasks involved in building the model and then describes the Bayesian methods for parameter estimation and model comparison. Section 4 reports results of a case study, where the proposed method is applied to data from a novel tissue-engineered scaffold fabrication process. We will illustrate the model building steps using the proposed method and compare its prediction performance with other methods, including the TSR method, GP models, and ANNs. A numerical study is conducted in Section 5 to examine the effect of sample size on model estimation. Section 6 concludes the article and discusses open issues.

2. The proposed CHM

2.1. General form of the proposed model

Let y be the degradation measurement, t be the degradation time, and z be the process parameter. For convenience, z is assumed to be univariate in this study; the proposed method can be easily extended to cases with multiple process parameters. The basic idea of the proposed CHM is illustrated in Fig. 3. The model consists of three elements:

- *Level-1 Model*: A function $h(t; \boldsymbol{\beta})$ is used to characterize the relationship between degradation (y) and time (t), where $\boldsymbol{\beta}$ is the parameter vector of this function.
- *Level-2 Model*: A function $g(z; \mathbf{c})$ is used to characterize the relationship between the parameter of the level-1 model ($\boldsymbol{\beta}$) and the process parameter (z).
- *Constraints*: The models are subject to several constraints that represent expert knowledge on scaffold degradation (Buchanan, 2008). For example, an obvious constraint is that the degradation must be monotonically increasing over time or, in other words, the degradation rate is always positive. For some types of scaffolds, the degradation rate is supposed to be decreasing over time, or the acceleration must be always negative. Examples of violations are given in the right panel of Fig. 3.

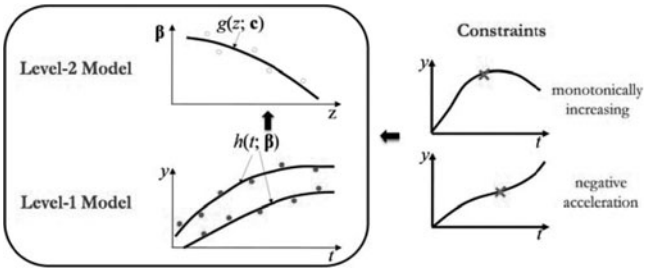

Figure 3. The basic idea of the proposed CHM. I

The mathematical expression of the proposed CHM can be written as follows:

$$\begin{aligned} &\textit{Level-1 Model} \\ &\qquad y = h(t; \boldsymbol{\beta}) + \varepsilon, \varepsilon \sim N(0, \sigma^2) \\ &\textit{Level-2 Model} \\ &\qquad \beta_i = g_i(z; \mathbf{c}_i), i = 0, \ldots, p-1, \\ &\qquad\quad \sigma^2 = g_\sigma(z; \mathbf{c}_\sigma) \\ &\textit{Constraints} \\ &\qquad \frac{dh}{dt} > 0, \frac{dh^2}{dt^2} < 0, \end{aligned} \tag{1}$$

where $\boldsymbol{\beta} = [\beta_0, \beta_1, \ldots, \beta_{p-1}]'$ is a p-dimensional coefficient vector of the level-1 model and ε is the error term, which represents the overall effect of measurement errors and other random errors in the scaffold fabrication process. Without loss of generality we assume that ε follows a normal distribution with zero mean and variance σ^2 that may depend on the process parameter z. Note that there is no random error in the level-2 models. The reason for this setup is given in Section 2.3. In the "constraints" part, the two aforementioned constraints are used for illustration. Other constraints may be used depending on intended application.

The proposed CHM has two key features that make it appropriate for scaffold degradation data. First, the two-level model structure separates the modeling of degradation profiles and the modeling of the effect of process parameters on degradation, which simplifies the modeling complexity as well as enables an easy and meaningful interpretation. Second, the constraints provide a convenient mechanism to accommodate expert knowledge on scaffold degradation.

2.2. Level-1 and level-2 models

Appropriate model forms should be chosen for the level-1 and level-2 models in Equation (1) to fit given data. In this work, we use polynomial models for them; that is,

$$h(t; \boldsymbol{\beta}) = \beta_{p-1} t^{p-1} + \cdots + \beta_0, \tag{2}$$

$$g_i(z; \mathbf{c}_i) = c_{i,q_i-1} z^{q_i-1} + \cdots + c_{i,0}, \tag{3}$$

$$g_\sigma(z; \mathbf{c}_\sigma) \text{or} \log[g_\sigma(z; \mathbf{c}_\sigma)] = c_{q_\sigma-1} z^{q_\sigma-1} + \cdots + c_0, \tag{4}$$

where Equation (2) has order $p-1$, Equation (3) has order $q_i - 1$ and $\mathbf{c}_i = [c_{i,q_i-1}, c_{i,q_i-2}, \ldots, c_{i,0}]'$, $i = 0, \ldots, p-1$, and Equation (4) has order $q_\sigma - 1$ and $\mathbf{c}_\sigma = [c_{q_\sigma-1}, c_{q_\sigma-2}, \ldots, c_0]'$. Logarithmic transformation of σ^2 may be used in the variance model for better fit.

Polynomial models are widely used in practice due to their convenience in data fitting, parameter estimation, and model interpretation. Their parsimonious model forms are especially suitable for the small-sample case of our problem as there are only limited data available in scaffold degradation studies. This situation is due to the lengthy degradation process (which may take months or years in some cases) and costly measurement. Nonetheless, it is worth pointing out that although we adopt the models in Equations (2) to (4) in this work, the proposed method is based on generic ideas and can be applied to other

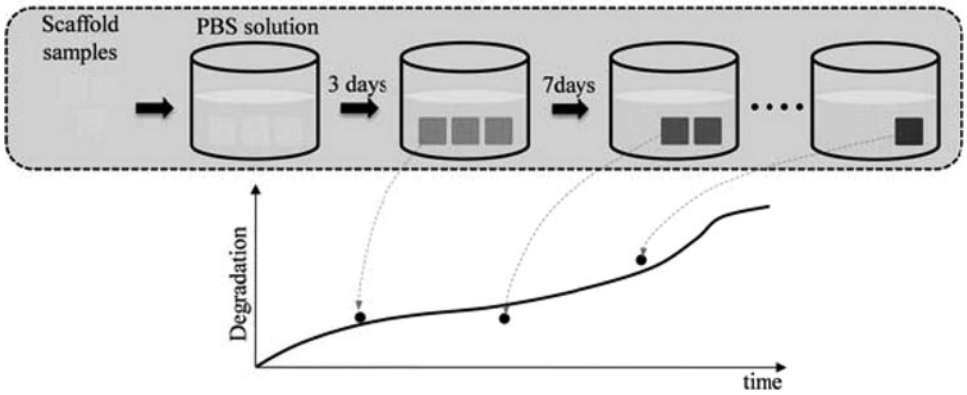

Figure 4. Data collection procedure in scaffold degradation studies.

types of model forms such as spline models and nonlinear models.

2.3. *Comparison with similar methods*

From the perspective of two-level modeling, the proposed CHM has some similarities to the TSR method mentioned in Section 1 and mixed-effect models (Jiang, 2007). However, it has the following unique characteristics.

First, the CHM simultaneously estimates parameters at the two levels in orderto achieve best overall fitting, whereas the TSR separately estimates model parameters at the two levels. That is, it first finds $\beta_i^{\wedge}$ at level 1 by regressing y on t and then finds $c_{0,0} = 0$ at level 2 by regressing $\beta_i^{\wedge}$ on z. Such separate estimation deteriorates the performance of the overall fitting. As the estimate of β_i obtained in the first step is reused as a response in the second step, it may also cause more uncertainty in the estimation. Second, the CHM has a more general form of random errors, which can be either homoscedastic or heteroscedastic (i.e., independent of or dependent on process parameters), whereas the TSR assumes a constant variance for random errors. Third, the proposed CHM can flexibly incorporate constraints, which is not straightforward for the TSR method or mixed-effect models.

Compared with mixed-effect models, the CHM has a special feature: level-2 models are free of random errors. There are two reasons for such a setup. First, having random errors at level 2 would make it difficult to impose constraints as the term $h(t; \boldsymbol{\beta})$ in level-1 models will be random. Second, with random errors at level 2, the CHM will be non-identifiable using data collected in scaffold studies. Figure 4 illustrates the data collection procedure in scaffold degradation studies. Typically, scaffold specimens are put in phosphate-buffered saline (PBS; i.e., salt) solution and have their weight loss measured at predetermined time points (Yang *et al.*, 2004). For each measurement, a specimen needs to be taken out of the solution and dried. It is then discarded, meaning that we can only obtain one observation from each specimen instead of the multiple observations required in mixed-effect models. In such a case, random errors at level 2 cannot be distinguished from those at level 1 (Frees, 2004).

3. Bayesian methods for model estimation and comparison

There are two tasks that need to be performed to build the CHM in Equation (1) under the models in Equations (2) to (4):

Task I. Parameter estimation: Note that we only need to estimate the level-2 parameters, $\{\mathbf{c}_0, \mathbf{c}_1, \ldots, \mathbf{c}_{p-1}, \mathbf{c}_\sigma\}$. Once they are determined, the level-1 parameters will also be determined.

Task II. Model comparison: To find a good model, we need to address a few considerations:

(Q1) Is logarithmic transformation on the variance needed?
(Q2) Does the level-1 model have zero intercept; i.e., $\beta_0 = 0$ in Equation (2)?
(Q3) Is the effect of the process parameter insignificant; i.e., $\mathbf{c}_i = \mathbf{0}$ in Equation (3)?
(Q4) Is the variance homoscedastic; i.e., $\mathbf{c}_\sigma = \mathbf{0}$ in Equation (4)?

Answers to the above questions will not only help find an optimal model but will also provide useful guidance to scaffold degradation research. For example, the value $\beta_0 = 0$ in (Q2) implies that the degradation behavior in the observed period may be consistent with that in the initial period, which is typically not observed. Also, $\mathbf{c}_i = \mathbf{0}$ in (Q3) suggests that the process parameter does not have significant effect on scaffold degradation.

Section 3.1 will detail the proposed Bayesian method for Task I. The considerations in Task II can be formulated as model comparison problems in the Bayesian framework. For example, (Q1) can be solved by comparing the model with $g_\sigma(z;\ \mathbf{c}_\sigma)$ and that with $\log[g_\sigma(z;\ \mathbf{c}_\sigma)]$. Bayesian methods for such comparisons will be given in Section 3.2.

3.1. *Bayesian method for parameter estimation*

Assume that the dataset contains m independent observations $\{(y_j, z_j, t_j), j = 1, \ldots, m\}$, where y_j is the degradation measured at time t_j under the value, z_j, of the process parameter. To avoid risks of inappropriate priors, flat (i.e., noninformative) priors are used in our study. In Bayesian statistics, the central task in model estimation is to find the posterior of the parameters. For the proposed CHM in Equations (1) to (4), the joint posterior distribution of parameters is

$$f(\mathbf{c}_0, \mathbf{c}_1, \ldots, \mathbf{c}_{p-1}, \mathbf{c}_\sigma | \text{data}). \tag{5}$$

As this distribution is analytically intractable, MCMC algorithms (Robert and Casella, 2004) will be used to sample from it (called *posterior sampling* later). Point estimates of the parameters can be obtained using location estimates (e.g., mean) of the

posterior sample. Prediction can also be found based on the posterior sample. In the following, the method for posterior sampling will be described first, followed by the method for getting point estimates and prediction.

3.1.1. *Posterior sampling*

In general, sampling from posteriors in a hierarchical model is a challenging task due to the high dimension of the parameters. In this study, a hybrid approach will be used, which integrates two MCMC methods: the *Gibbs sampler* and the *slice sampler*. The Gibbs sampler is a popular method for high-dimensional posterior sampling (Gelman *et al.*, 2004). Its basic idea is to divide the parameters into subgroups and sample from the conditional posterior of each subgroup given others. As the subgroups have smaller dimensions and their conditional posteriors are simpler than the joint posterior, the sampling can be more easily implemented. This method is especially suitable for hierarchical models, as the hierarchy in the models provides a natural way to divide parameters (Gilks, 2005). In sampling the conditional posteriors of subgroups, there are two cases: for standard conditional posteriors such as a normal distribution, the sampling can be done using random generators in software; for nonstandard ones, the sampling will be done using the slice sampler. The slice sampler is a novel MCMC algorithm that is a powerful method to sample nonstandard posteriors (Neal, 2003). One important advantage of this method lies in its convenience in implementation: it only needs the posterior to be sampled from and initial values as inputs.

Figure 5 shows two strategies to divide parameters in Gibbs sampling, where the dashed arrows indicate the order of sampling. In the first strategy, $\mathbf{c}_\sigma$ is one group and all of the other parameters form another group, while in the second strategy, each parameter forms a group. Generally, the first strategy can be used in small-dimension cases, and the second strategy should be used otherwise.

Posterior sampling procedures under the two strategies are given below. Let

$$\begin{aligned}
\mathbf{z} &= [z_1, \ldots, z_m]',\ \mathbf{y} = [y_1, \ldots, y_m]',\ \mathbf{t} = [t_1, \ldots, t_m]',\\
\boldsymbol{\varepsilon} &= [\varepsilon_1, \ldots, \varepsilon_m]',\ \mathbf{z}^{q_i-1} = [z_1^{q_i-1}, \ldots, z_m^{q_i-1}]',\\
\mathbf{Z}_i &= [\mathbf{z}^{q_i-1}, \mathbf{z}^{q_i-2}, \ldots, \mathbf{1}]',\ \mathbf{Z}_\sigma = [\mathbf{z}^{q_\sigma-1}, \mathbf{z}^{q_\sigma-2}, \ldots, \mathbf{1}]',\\
g_i(\mathbf{z};\mathbf{c}_i) &= c_{i,q_i-1}\mathbf{z}^{q_i-1} + \cdots + c_{i,0} = \mathbf{Z}_i\mathbf{c}_i,\\
g_\sigma(\mathbf{z};\mathbf{c}_\sigma) &= c_{q_\sigma-1}\mathbf{z}^{q_\sigma-1} + \cdots + c_0 = \mathbf{Z}_\sigma\mathbf{c}_\sigma,\\
\mathbf{V}_i(\mathbf{z};\mathbf{c}_\sigma) &= \mathrm{diag}(g_\sigma(\mathbf{z};\mathbf{c}_\sigma) \bullet /\mathbf{t}^{2i}),
\end{aligned}$$

where "diag(v)" denotes the diagonal matrix with the vector v being the diagonal elements, "•" is the dot product operator of two vectors, and "•/" is the dot divide operator.

Posterior Sampling Procedure under Strategy I

Pre-step: Specify the initial value of $\mathbf{c}_\sigma$, denoted by $\mathbf{c}_\sigma^{(0)}$.

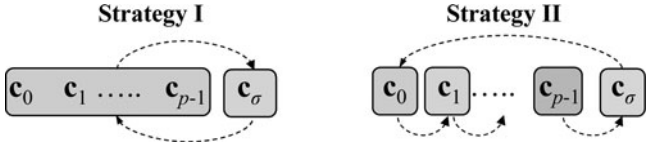

Figure 5. Two strategies to divide parameters in Gibbs sampling.

Step 1: Generate a sample $\{\mathbf{c}_0^{(1)}, \ldots, \mathbf{c}_{p-1}^{(1)}\}$ from the conditional posterior of $\{\mathbf{c}_0, \ldots, \mathbf{c}_{p-1}\}$ given

$$\begin{aligned}
&f\left(\mathbf{c}_0, \mathbf{c}_1, \ldots, \mathbf{c}_{p-1} \middle| \mathbf{c}_\sigma^{(0)}, \mathbf{y}\right)\\
&\quad \propto N\left(\mathbf{y} \middle| \sum_{i=0}^{p-1} g_i(\mathbf{z};\mathbf{c}_i) \bullet \mathbf{t}^i, \mathbf{V}_0\left(\mathbf{z};\mathbf{c}_\sigma^{(0)}\right)\right). \qquad (6)
\end{aligned}$$

Step 2: Generate a sample $\mathbf{c}_\sigma^{(1)}$ from the conditional posterior of $\mathbf{c}_\sigma$ given $\{\mathbf{c}_0^{(1)}, \ldots, \mathbf{c}_{p-1}^{(1)}\}$

$$\begin{aligned}
&f\left(\mathbf{c}_\sigma \middle| \mathbf{c}_0^{(1)}, \mathbf{c}_1^{(1)}, \ldots, \mathbf{c}_{p-1}^{(1)}, \mathbf{y}\right)\\
&\quad \propto N\left(\mathbf{y} \middle| \sum_{i=0}^{p-1} g_i\left(\mathbf{z};\mathbf{c}_i^{(1)}\right) \bullet \mathbf{t}^i, \mathbf{V}_0(\mathbf{z};\mathbf{c}_\sigma)\right). \qquad (7)
\end{aligned}$$

This will yield one sampling point $\{\mathbf{c}_0^{(1)}, \ldots, \mathbf{c}_{p-1}^{(1)}, \mathbf{c}_\sigma^{(1)}\}$. More sampling points can be generated by iterating Steps 1 and 2.

Posterior Sampling Procedure under Strategy II

Pre-step: Specify the initial values of the parameters, $\{\mathbf{c}_0^{(0)}, \ldots, \mathbf{c}_{p-1}^{(0)}, \mathbf{c}_\sigma^{(0)}\}$.

Step 1: Generate a sample $\mathbf{c}_0^{(1)}$ from the conditional posterior of $\mathbf{c}_0$ given $\{\mathbf{c}_1^{(1)}, \ldots, \mathbf{c}_{p-1}^{(1)}, \mathbf{c}_\sigma^{(0)}\}$:

$$\begin{aligned}
&\mathbf{c}_0|\mathbf{c}_1^{(0)}, \ldots, \mathbf{c}_{p-1}^{(0)}, \mathbf{c}_\sigma^{(0)},\\
&\quad \mathbf{y} \sim N\left(\left(\mathbf{Z}_0'\left(\mathbf{V}_0\left(\mathbf{z};\mathbf{c}_\sigma^{(0)}\right)\right)^{-1}\mathbf{Z}_0\right)^{-1}\mathbf{Z}_0'\left(\mathbf{V}_0\left(\mathbf{z};\mathbf{c}_\sigma^{(0)}\right)\right)^{-1}\mathbf{u}_0,\right.\\
&\quad \left.\times\left(\mathbf{Z}_0'\left(\mathbf{V}_0\left(\mathbf{z};\mathbf{c}_\sigma^{(0)}\right)\right)^{-1}\mathbf{Z}_0\right)^{-1}\right), \qquad (8)
\end{aligned}$$

where

$$\mathbf{u}_0 \equiv \mathbf{y} - \sum_{i=1}^{p-1} g_i\left(\mathbf{z};\mathbf{c}_i^{(0)}\right) \bullet \mathbf{t}^i.$$

$\{\mathbf{c}_1^{(1)}, \ldots, \mathbf{c}_{p-1}^{(1)}\}$ will be generated in sequence for $1 \le k \le p-1$.

Step $k+1$: Generate a sample $\mathbf{c}_k^{(1)}$ from the conditional posterior of $\mathbf{c}_k$:

$$\begin{aligned}
&\mathbf{c}_k\left|\mathbf{c}_1^{(1)}, \ldots, \mathbf{c}_{k-1}^{(1)}, \mathbf{c}_{k+1}^{(0)}, \ldots, \mathbf{c}_{p-1}^{(0)}, \mathbf{c}_\sigma^{(0)}, \mathbf{y}\right.\\
&\quad \sim N\left(\left(\mathbf{Z}_k'\left(\mathbf{V}_k\left(\mathbf{z};\mathbf{c}_\sigma^{(0)}\right)\right)^{-1}\mathbf{Z}_k\right)^{-1}\mathbf{Z}_k'\left(\mathbf{V}_k\left(\mathbf{z};\mathbf{c}_\sigma^{(0)}\right)\right)^{-1}\mathbf{u}_k,\right.\\
&\quad \left.\times\left(\mathbf{Z}_k'\left(\mathbf{V}_k\left(\mathbf{z};\mathbf{c}_\sigma^{(0)}\right)\right)^{-1}\mathbf{Z}_k\right)^{-1}\right), \qquad (9)
\end{aligned}$$

where

$$\mathbf{u}_k \equiv \left[\mathbf{y} - \sum_{i=0}^{k-1} g_i\left(\mathbf{z};\mathbf{c}_i^{(1)}\right) \bullet \mathbf{t}^i - \sum_{i=k+1}^{p-1} g_i\left(\mathbf{z};\mathbf{c}_i^{(0)}\right) \bullet \mathbf{t}^i\right] \bullet \Big/ \mathbf{t}^k.$$

Step $p+1$: Generate a sample $\mathbf{c}_\sigma^{(1)}$ from the conditional posterior of $\mathbf{c}_\sigma$ in Equation (7):

$$\begin{aligned}
&f\left(\mathbf{c}_\sigma \middle| \mathbf{c}_0^{(1)}, \mathbf{c}_1^{(1)}, \ldots, \mathbf{c}_{p-1}^{(1)}, \mathbf{y}\right)\\
&\quad \propto N\left(\mathbf{y} \middle| \sum_{i=0}^{p-1} g_i\left(\mathbf{z};\mathbf{c}_i^{(1)}\right) \bullet \mathbf{t}^i, \mathbf{V}_0(\mathbf{z};\mathbf{c}_\sigma)\right).
\end{aligned}$$

In implementing the above procedures, initial values of the parameters can be found in the following way: separately fit a polynomial model to data under each setting of the process parameter and then fit polynomial models using the parameter estimates of those models as a response. The conditional posteriors in Equations (6) and (7), which are proportional to the

normal likelihoods, are nonstandard distributions and will be sampled using the slice sampler. Since the conditional posteriors in Equations (8) and (9) are multivariate normal distributions, the sampling can be done using built-in routines in the software. Proofs of Equations (6) and (7) are given in Appendix A, and those of Equations (8) and (9) are given in Appendix B.

Note that there are several constraints in the proposed CHM in Equation (1). Incorporating constraints in model estimation is not trivial using conventional methods for estimation. Under the Bayesian framework, such difficulty can be easily conquered by discarding posterior samples that fall out of the constrained space (Geweke, 1986; Gelfand *et al.*, 1992). Specifically, when each sample is drawn from the posterior, the constraints will be checked. If the sample satisfies the constraints, it will be retained; otherwise, it will be discarded. In this way, the output of the posterior sampling will be within the constrained space.

3.1.2. *Point estimates and prediction*

Let $\{\mathbf{c}_0^{(w)}, \ldots, \mathbf{c}_{p-1}^{(w)}, \mathbf{c}_\sigma^{(w)}\}$, $w = 1, \ldots, W$, be the generated sample from the posterior sampling. When the posterior is symmetric, the sample mean can be taken as a point estimate of the parameters. To be more robust, the median or mode should be used. The sample mode can be obtained by first estimating the posterior density through nonparametric methods, such as kernel smoothing (Horova *et al.*, 2012) and then finding the mode of the estimated density through optimization.

From Equation (1), we can get $y \sim N(h(t;\ \boldsymbol{\beta}), \sigma^2)$. Prediction of degradation $\hat{y}^*$ under a given setting of the process parameter z^* and time point t^* can be obtained based on this expression. Specifically, for each item in the posterior sample, draw $y^{*(w)}$ from $N(h(t^*; \hat{\beta}^{*(w)}), \hat{\sigma}^{2*(w)})$, where $\hat{\boldsymbol{\beta}}^{*(w)} = [\hat{\beta}_0^{*(w)}, \hat{\beta}_1^{*(w)}, \ldots, \hat{\beta}_{p-1}^{*(w)}]'$, $\hat{\beta}_i^{*(w)} = g_i(z^*;\ \mathbf{c}_i^{(w)})$, $i = 0, \ldots, p-1$, and $\hat{\sigma}^{2*(w)} = g_\sigma(z;\ \mathbf{c}_\sigma^{(w)})$. This will produce a posterior sample of the prediction $\{y^{*(1)}, \ldots, y^{*(w)}\}$. The median of this sample can be used as the prediction $\hat{y}^*$.

3.2 *Bayesian method for model comparison*

A popular Bayesian tool for model comparison is the Bayes Factor (BF), which can compare any two models regardless of their forms (Kass and Raftery, 1995). The computation of BFs often encounters challenges in cases of high-dimensional parameters. In this study, we propose a new method to calculate the BF. This section will first briefly review the BF and existing computation methods and then describe the proposed method.

3.2.1. *The BF and computation methods*

The BF of two models, M_1 and M_2, is defined by (Gelman *et al.*, 2004)

$$BF_{12} = \frac{P(\mathbf{y}|M_1)}{P(\mathbf{y}|M_2)} = \frac{\int \pi(\boldsymbol{\theta}_1|M_1) f(\mathbf{y}|\boldsymbol{\theta}_1, M_1) d\boldsymbol{\theta}_1}{\int \pi(\boldsymbol{\theta}_2|M_2) f(\mathbf{y}|\boldsymbol{\theta}_2, M_2) d\boldsymbol{\theta}_2}, \tag{10}$$

where $\boldsymbol{\theta}_k$ is the parameter vector of model M_k, $k = 1, 2$. For example, if M_1 is the proposed model in Equations (1) to (4), then the parameter $\boldsymbol{\theta}_1 = [\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_{p-1}, \mathbf{c}_\sigma]'$. Here $\pi(\theta_k|M_k)$ denotes the prior under M_k, $f(\mathbf{y}|\boldsymbol{\theta}_k, M_k)$ is the likelihood, and $P(\mathbf{y}|M_k)$ is the Marginal Likelihood (ML) under this model. Being the ratio of the MLs, the BF has an intuitive interpretation. For example, $BF_{12} = 3$ means that M_1 is three times more plausible than M_2 according to the data. The following are widely used cutoff values for the Bayes factor (Kass and Raftery, 1995):

BF	2log	Evidence against M_j
$1 \sim 3$	$0 \sim 2$	Barely worth mentioning
$3 \sim 20$	26	Positive
$20 \sim 150$	$6 \sim 10$	Strong
>150	>10	Very strong

Usually the BF is obtained by separately estimating the ML of each of the competing models and then calculating their logarithmic difference (Han and Carlin, 2001):

$$\log(BF_{12}) = \log(P(\mathbf{y}|M_1)) - \log(P(\mathbf{y}|M_2)).$$

Since the marginal likelihood under model M can be written as

$$P(\mathbf{y}|M) = \int f(\mathbf{y}|\boldsymbol{\theta}, M)\pi(\boldsymbol{\theta}|M)d\boldsymbol{\theta} = E_{\pi(\boldsymbol{\theta}|M)}\left[f(\mathbf{y}|\boldsymbol{\theta}, M)\right];$$

i.e., the average likelihood over the prior, the simplest estimator of the ML is

$$\hat{P}(\mathbf{y}|M) = \frac{1}{W}\sum_{w=1}^{W} f(\mathbf{y}|\boldsymbol{\theta}^{(w)}, M),$$

where $\{\boldsymbol{\theta}^{(w)}: w = 1, \ldots, W\}$ is a sample from the prior $\pi(\theta|M)$. When a flat prior is used, most $\theta^{(w)}$ have very small likelihood values if the posterior is concentrated relative to the prior, so that the sampling process is inefficient and the estimator has large variance. Some estimators based on posterior sampling have been proposed for better estimation, including the harmonic mean estimator (Kass and Raftery, 1995):

$$\hat{P}(\mathbf{y}|M) = \frac{1}{W}\left\{\sum_{w=1}^{W}[f(\mathbf{y}|\boldsymbol{\theta}^{(w)}, M)]^{-1}\right\}^{-1}, \tag{11}$$

where $\{\boldsymbol{\theta}^{(w)}: w = 1, \ldots, W\}$ is a sample from the posterior $f(\theta|\mathbf{y}, M)$.

For the proposed CHM, there are two types of model comparison problems: non-nested and nested comparisons. Two models are called nested if one model can be obtained when some parameters of the other model take certain values. For the first type of problems, the harmonic mean estimator in Equation (11) will be used to calculate the BF. For the second type, we propose a new method to efficiently calculate the BF, which is now described.

3.2.2. *The proposed method for calculating BFs*

Let the two models to compare be M_l and M_s with parameters $\boldsymbol{\theta}_l$ and $\boldsymbol{\theta}_s$, and M_s is nested with M_l. Suppose $\boldsymbol{\theta}_l = [\boldsymbol{\theta}_s\ \boldsymbol{\xi}]'$ with unknown ξ, and $\xi = \xi_0$ in M_s. For simplicity, M_l is called the larger model and M_s the smaller model. In other words, the comparison is equivalent to determining if $\xi = \xi_0$. Theorem 1 gives a simple formula for the BF.

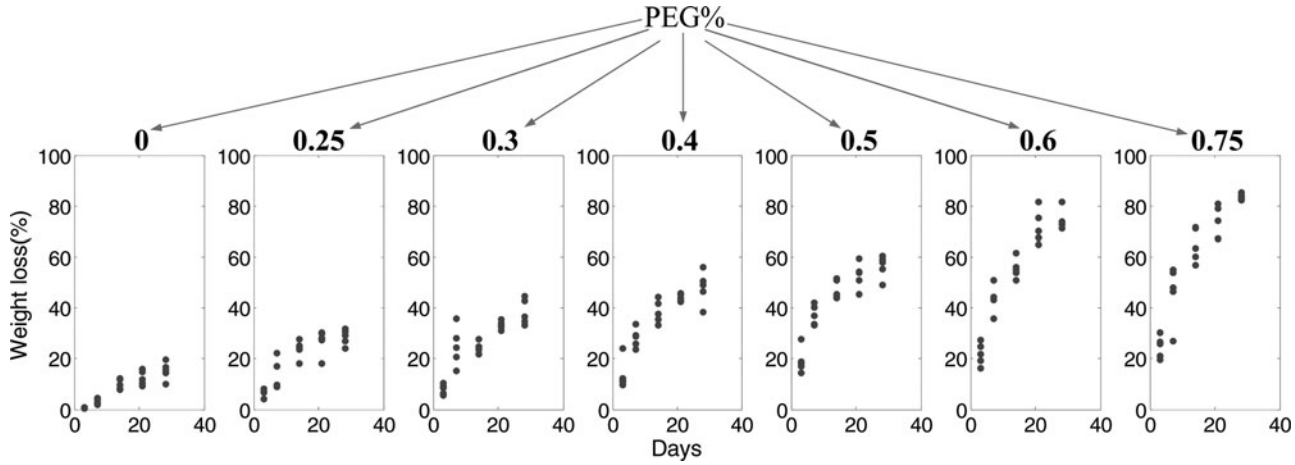

Figure 6. Data used in the case study: scaffold weight loss (y) under different settings of PEG percentage (z) measured at specified time points (t).

Theorem 1. *Assuming (i) the priors of $\boldsymbol{\theta}_s$ and $\boldsymbol{\xi}$ under M_l are independent; i.e., $\pi(\theta_l|M_l) = \pi(\theta_s|M_l)\pi(\xi|M_l)$; (ii) the prior of θ_s under M_l is the same as that under M_s; i.e., $\pi(\theta_s|M_l) = \pi(\theta_s|M_s)$; and (iii) ξ follows a flat prior within [a, b] under M_l, the BFr comparing M_l to M_s is*

$$BF_{ls} = \frac{1}{(b-a)\, f(\boldsymbol{\xi}_0|\mathbf{y}, M_l)}.$$

A proof of the theorem is given in Appendix C. Note that the three assumptions in the theorem are standard settings of priors in Bayesian analysis (Gelman *et al.*, 2004).

This result is intuitive: $f(\xi_0|\mathbf{y}, M_l)$ in the denominator is the marginal posterior density of ξ at ξ_0, which represents evidence for $\xi = \xi_0$; i.e., the smaller model. Since BF_{ls} measures evidence of the larger model against the smaller model, it has an inverse relationship with $f(\xi_0|\mathbf{y},M_l)$. A large $f(\boldsymbol{\xi}_0|\mathbf{y}, M_l)$ means that there is strong evidence for $\boldsymbol{\xi} = \boldsymbol{\xi}_0$ and thus the smaller model should be chosen. Consistently, the BF is small, indicating that the larger model is less supported by data and thus the smaller model is preferred.

Based on Theorem 1, we obtain the following estimator of the Bayes factor:

$$\hat{BF}_{ls} = \frac{1}{\left[\max_{1\le w\le W}(\boldsymbol{\xi}^{(w)}) - \min_{1\le w\le W}(\boldsymbol{\xi}^{(w)})\right] \times \hat{f}(\boldsymbol{\xi}|\mathbf{y}, M_l)\big|_{\boldsymbol{\xi}=\boldsymbol{\xi}_0}}, \quad (12)$$

where $\{\xi^{(w)}: w = 1, \ldots, W\}$ is a sample of ξ from the posterior under M_l. The maximum and minimum of this sample are used to estimate $(b-a)$; $\hat{f}(\boldsymbol{\xi}|\mathbf{y}, M_l)$ can be obtained by fitting this sample using kernel smoothing methods; then $f(\xi_0|\mathbf{y}, M_l)$ can be estimated by the value of this density function at $\xi = \xi_0$.

Compared with existing estimators, the new estimator in Equation (12) provides considerable convenience in computation; the existing estimators calculate the ML under each model to obtain the BF, which means that posterior sampling needs to be done under both models. In contrast, the new estimator only requires posterior sampling from the larger model. This simplification is made possible by the use of the nested structure of the models under comparison. Moreover, as the new estimator only depends on one posterior sample, it is expected to have a smaller variance than the existing estimators, which involve two samples. This is validated by the results of the case study in Section 4.3.

4. Case study

We apply the proposed method to a dataset from a novel tissue-engineered scaffold fabrication process using urethaned-doped polyester elastomers (CUPEs). The CUPEs is a new class of biomaterials for scaffold fabrication. Unlike conventional scaffold materials that are either stiff and incompliant or soft but weak, the CUPEs is fully elastic and sufficiently strong, making them potential scaffold materials to develop soft tissues such as cardiac tissues and blood vessels (Yang *et al.*, 2004; Dey *et al.*, 2008). Figure 6 displays the collected data containing scaffold weight loss percentages (y) under one key process parameter (z), the percentage of Polyethylene glycol (PEG) in scaffold material synthesis. There are seven settings of z at 0, 0.25, 0.3, 0.4, 0.5, 0.6, and 0.75. The weight loss was measured at five time points with $t = 3$, 7, 14, 21, 28 days under each setting of z. Five replicates were measured each time.

The level-1 and level-2 models in the CHM in Equation (1) should satisfy the following requirements:

(R1) *Good fitting of data*. The level-1 models should fit the degradation data well.
(R2) *Good fitting of level-1 parameter estimates*. The level-2 models should fit parameter estimates of level-1 models well.
(R3) *Easy interpretation*. As mentioned in the Introduction, models with easy interpretation are preferred. This is especially needed for level-2 models, which represent the effect of process parameters.

Results on building/validating the proposed CHM are reported as follows. Section 4.1 conducts some preliminary analysis to determine the appropriate forms for the level-1 and level-2 models based on the above requirements. With the chosen model forms, Section 4.2 applies the proposed method in Section 3.1 to estimate parameters of the CHM. The estimated model is compared with results from the unconstrained version

Table 1. Parameter setting in the candidate model forms of the CHM.

	Model																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
y	I	I	I	I	I	II	II	II	II	II	II	III	III	III	III	III	III	III	III	III
p−1	1	2	2	3	3	1	1	2	2	3	3	1	1	1	2	2	2	3	3	3
q−1	1	1	2	1	2	1	2	1	2	1	2	1	2	3	1	2	3	1	2	3

of the CHM (called UnCHM later) and the TSR method. Section 4.3 applies the method in Section 3.2 to compare possible variants of the CHM. Section 4.4 checks residuals to validate model assumptions. Section 4.5 compares the prediction performance of the CHM with UnCHM, TSR, GP, and ANNs through cross-validation. Section 4.6 summarizes the findings in the case study.

4.1. *Preliminary analysis*

Since degradation measurements are in percentages, logarithmic and logit transformation can be applied to the data, leading to the following possible model forms:

Level-1 Model

$$\text{(I) } y = \beta_{p-1}t^{p-1} + \cdots + \beta_0 + \varepsilon$$
$$\text{(II) } \log(y) = \beta_{p-1}t^{p-1} + \cdots + \beta_0 + \varepsilon \Rightarrow y = e^{\beta_{p-1}t^{p-1}+\cdots+\beta_0+\varepsilon}$$
$$\text{(III) } \operatorname{logit}(y) = \log\left(\frac{y}{1-y}\right) = \beta_{p-1}t^{p-1} + \cdots + \beta_0 + \varepsilon \Rightarrow y = \frac{e^{\beta_{p-1}t^{p-1}+\cdots+\beta_0+\varepsilon}}{1+e^{\beta_{p-1}t^{p-1}+\cdots+\beta_0+\varepsilon}}.$$

Level-2 Model

$$\beta_i = c_{i,q-1}z^{q-1} + \cdots + c_{i,0},\ \sigma^2 = c_{q-1}z^{q-1} + \cdots + c_0.$$

Here the level-1 model can take three forms: regular polynomials (I), exponential polynomials (II), and logit polynomials (III). The order of the polynomials is $p-1$. For convenience, polynomial models of order $q-1$ are used at level 2. In addition, to improve numerical stability in model fitting, values of z are scaled by 20 (i.e., $z/20$ is used). Considering different settings of polynomial orders at the two levels, 20 candidate model forms of the CHM as listed in Table 1 are fitted to the data in Fig. 6 using least squares methods. Specifically, the level-1 model is first estimated for data under each value of z, and then the level-2 model is fitted using level-1 parameter estimates as responses. The candidate models are compared in terms of the three abovementioned requirements.

The performance of candidate models in fitting the data (i.e., fitting at Level 1) is measured by Mean Squared Error (MSE), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). The MSE is an estimate of the random error variance (i.e., σ^2) with a smaller value indicating better fitting; the AIC/BIC measures trade-off between the goodness-of-fit of the model and its complexity, with a smaller value indicating better performance. Figure 7 shows the results of MSE and AIC under each value of z (legends in the plots are model indices). The results of BIC have similar patterns as AIC. Some models share one curve in the figure as they use same level-1 models. Note that the AIC values of the three groups (i.e., form I, II, and III) are not comparable as their responses (i.e., y, $\log(y)$, $\operatorname{logit}(y)$) are different. Considering both performance measures, models 2, 3, 4, 5 work similarly well in the first group. Since the level-1 model of models 2 and 3 has a lower polynomial order, they are chosen as the best in the first group. Models 10 and 11 are the best in the second group, and models 18, 19, 20 are the best in the third group. The best models in the three groups have similar MSEs. Thus, *in terms*

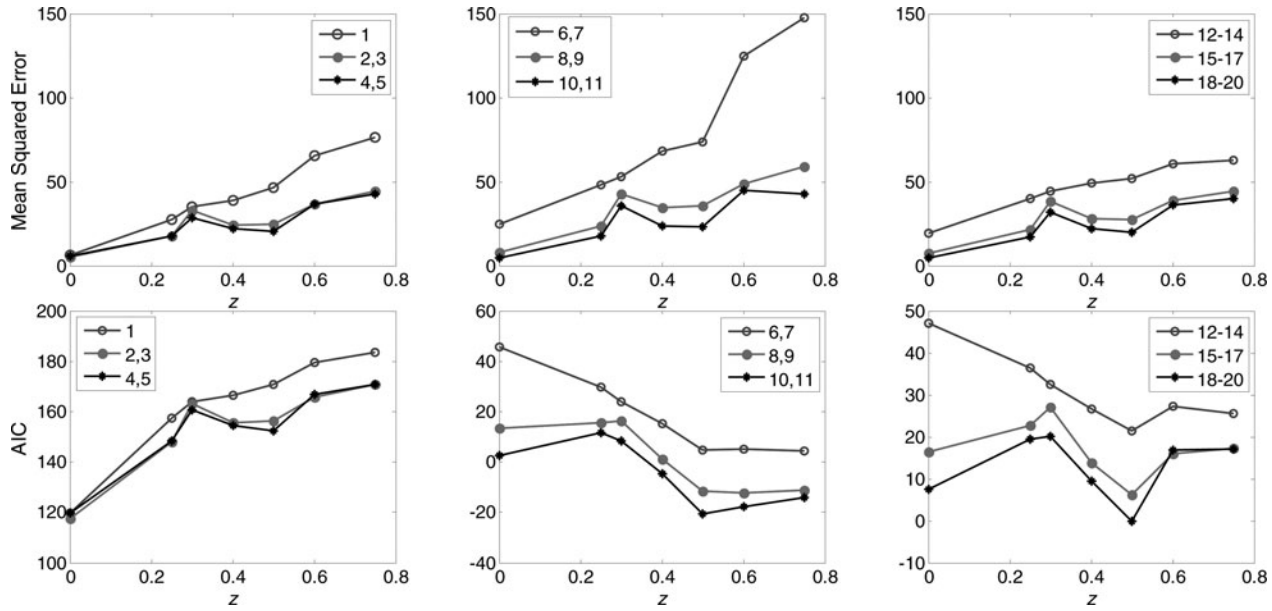

Figure 7. Performance of level-1 models under form I (left), II (middle), and III (right).

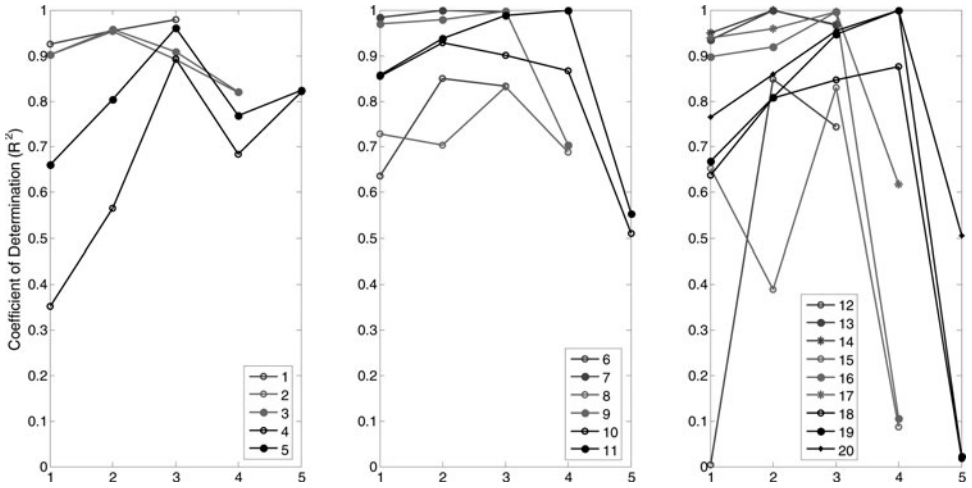

Figure 8. Performance of level-2 models in fitting parameter estimates of level-1 models.

of requirement (*R1*), *models 2, 3, 10, 11, 18, 19, 20 are good models*.

The performance of candidate model forms in fitting level-1 parameter estimates is measured by the coefficient of determination (i.e., R^2). Figure 8 shows results of this measure for the three groups. Note that each point in the plot represents the R^2 value of one model at level 2; if the candidate model contains an order-1 model at level 1 (e.g., model 1), there will be three level-2 models (β_1, β_0, and σ^2) and thus three points on the curve. In the first group, models 1, 2, 3 have good fitting ($R^2 > 0.8$) in all of the level-2 models, whereas models 4 and 5 are relatively poorly fitted ($R^2 < 0.7$) in some level-2 models; in the second group, model 7 seems to be over-fitted in all level-2 models ($R^2 > 0.98$), whereas other models are relatively poorly fitted in some level-2 models ($R^2 < 0.7$); in the third group, models 13 and 14 are well fitted in some models and over-fitted in one model, whereas other models are poorly fitted in some level-2 models ($R^2 < 0.5$). Thus, *in terms of requirement* (*R2*), *models 1, 2, 3 are good models*.

Considering the above two aspects, models 2 and 3 perform the best. *Since model 2 has simpler level-2 models than model 3, it is chosen as the best model form for the CHM*. This choice also satisfies requirement (R3), as level-2 models in model 2 can be easily interpreted; the process parameter has a linear effect on the shape and variability of scaffold degradation profiles. Figure 9 shows the fitted models using the chosen model form, where the level-1 models (upper) are quadratic and the level-2 models (lower) are linear. Note that some level-1 models exhibit a

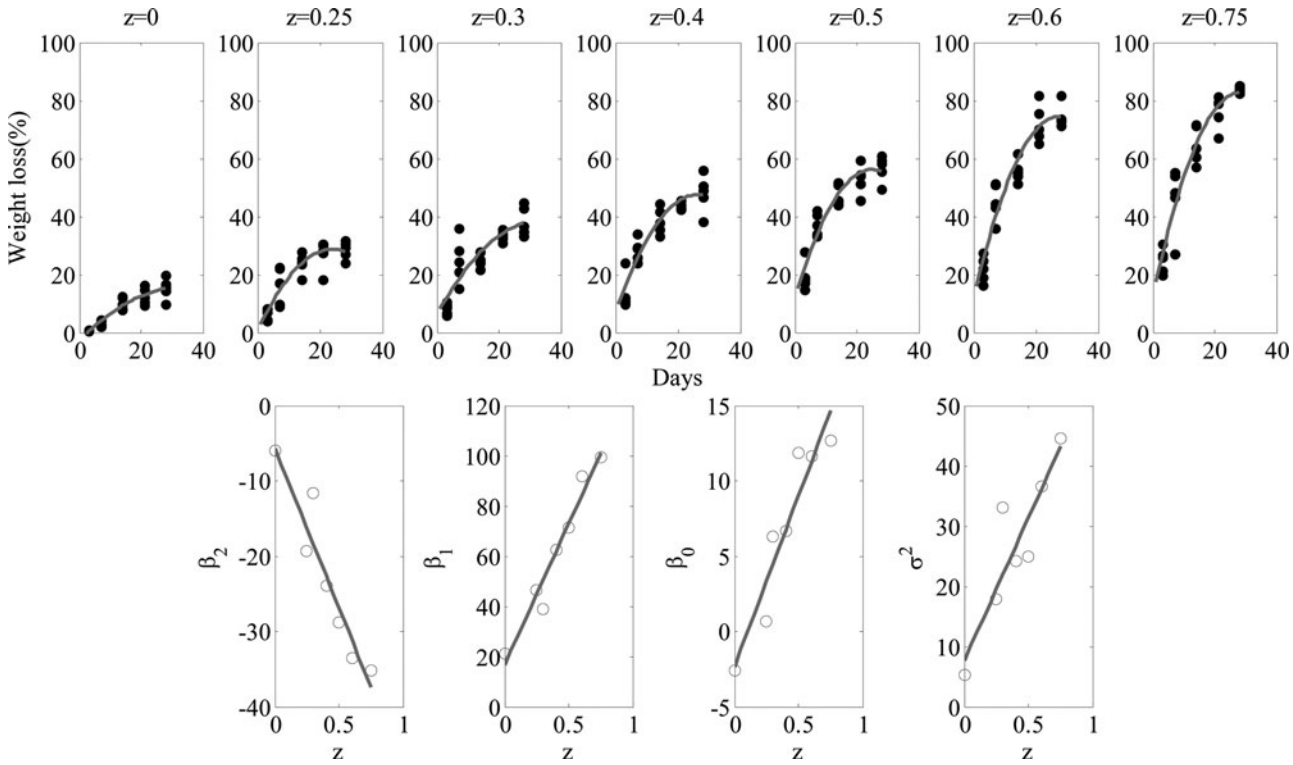

Figure 9. The chosen model forms: level-1 models (upper) and level-2 models (lower).

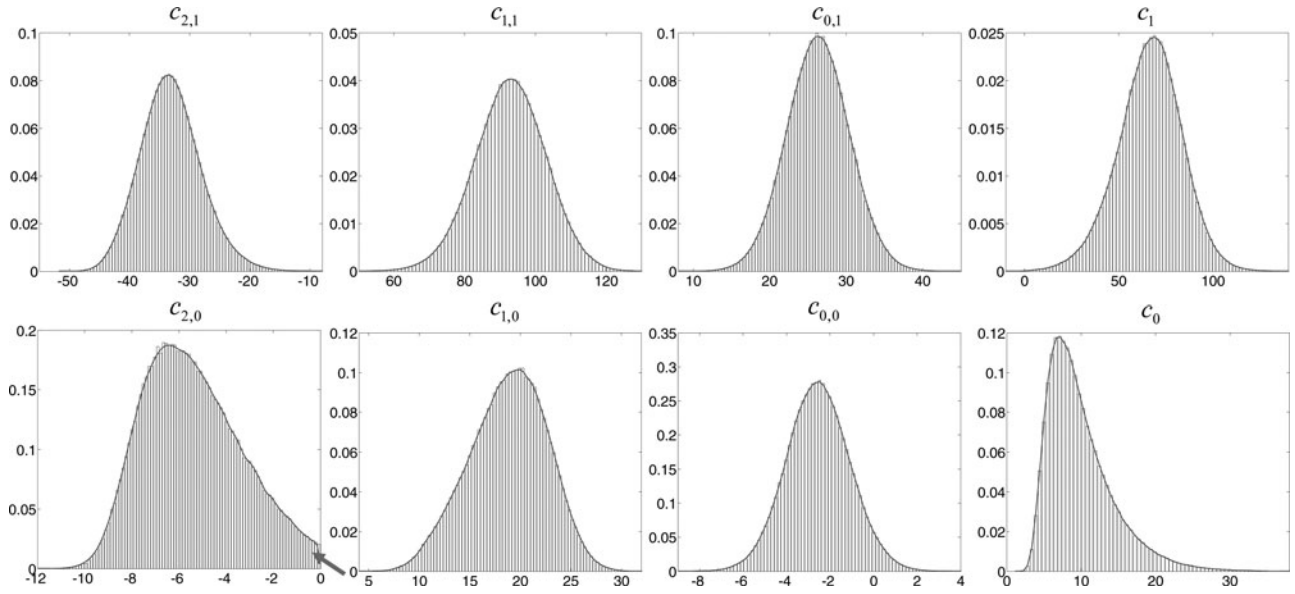

Figure 10. Normalized histograms and estimated density functions of the posterior sample.

decreasing trend at the right end. This will be corrected when constraints are incorporated in model fitting as shown in Section 4.2.

4.2. *Parameter estimation*

Based on results of the preliminary analysis and expert knowledge on the studied scaffold material, the CHM for the given dataset is as follows:

$$\begin{aligned}
&\textit{Level-1 Model}\\
&y = \beta_2 t^2 + \beta_1 t + \beta_0 + \varepsilon,\ \varepsilon \sim N(0, \sigma^2)\\
&\textit{Level-2 Model}\\
&\beta_2 = c_{2,1} z + c_{2,0},\ \beta_1 = c_{1,1} z + c_{1,0},\ \beta_0 = c_{0,1} z + c_{0,0}\\
&\sigma^2 = c_1 z + c_0\\
&\textit{Constraints}\\
&\frac{dy}{dt} = 2\beta_2 t + \beta_1 > 0,\ \frac{dy^2}{dt^2} = \beta_2 < 0,\ c_1 z + c_0 > 0,
\end{aligned} \tag{13}$$

where the three constraints represent positive degradation rate, negative acceleration, and positive variance, respectively. The parameters of the model are $\{c_{2,1}, c_{2,0}, c_{1,1}, c_{1,0}, c_{0,1}, c_{0,0}, c_1, c_0\}$.

Since the polynomial orders in Equation (13) are moderate, we can use either of the sampling procedures in Section 3.1.1 for parameter estimation. Point estimates of parameters are obtained using the mean, median, and mode of the posterior sample. This analysis is performed in Matlab: the *slicesample* function with 500 000 iterations is used in posterior sampling, the *ksdensity* function is used in posterior density estimation, and the *fmincon* function is used to find the mode of the estimated posterior density. Note that in imposing the first constraint, t takes values in the range of 0–28 days, which is of interest in this study. Since the first-order derivative is linear, we actually only need to check $t = 0$ and $t = 28$.

Figure 10 shows the normalized histograms and corresponding estimated densities of the posterior sample. Clearly, the histograms and the estimated densities closely match. The posteriors of $c_{2,0}$, $c_{1,0}$, c_1, and c_0 exhibit some degree of skewness, so median or mode estimates should be used. Also note that the posterior of $c_{2,0}$ is right-truncated (marked by the arrow in Fig. 10) due to the effect of the constraints. Table 2 lists the point estimates for each parameter. As expected, the estimates are very close for $c_{2,1}$, $c_{1,1}$, $c_{0,1}$, and $c_{1,0}$ and differ considerably for the skewed $c_{2,0}$, $c_{1,0}$, c_1, and c_0.

To evaluate the fitting performance of the CHM to the data, Fig. 11 displays the fitted level-1 models from the CHM, UnCHM, and TSR. It should be noted that the original TSR method in the literature (He *et al.*, 2015) assumes a constant variance for random errors; the TSR method used here is actually a modified version that includes a linear model of the variance at level 2, as in the CHM, to better fit the data. Given the limited data available, the fitting performance of the CHM is acceptable. From the lower panel, we can see that the two unconstrained methods (i.e., UnCHM and TSR) have a similar fitting performance in most cases, but the TSR behaves significantly worse under $z = 0$. The fitting performance of the CHM is comparable to the UnCHM, but the decreasing trend of the UnCHM under some settings of z (e.g., 0.5, 0.75) is corrected by the constraints in CHM. One point worth mentioning is that the fitting of CHM in Fig. 11 appears to be not as good as that in the upper panel of Fig. 9. The explanation for this observation is that the fitted models in Fig. 9 are from separate regressions at level 1 and level 2. In contrast, the fitting of CHM in Fig. 11 is from the overall estimation considering regression at both levels. Since the CHM assumes linear relationships free of random errors at level 2, it is reasonable that the level-1 fitting is not as good as in the separate regression.

Table 2. Point estimates of the parameters obtained from the posterior sample.

	Parameter							
	$c_{2,1}$	$c_{2,0}$	$c_{1,1}$	$c_{1,0}$	$c_{0,1}$	$c_{0,0}$	c_1	c_0
Mean	−33.12	−5.40	92.91	18.77	26.33	−2.55	66.45	9.97
Median	−33.34	−5.62	93.02	18.99	26.32	−2.56	67.10	8.95
Mode	−33.48	−6.45	93.06	19.90	26.28	−2.50	68.82	6.97

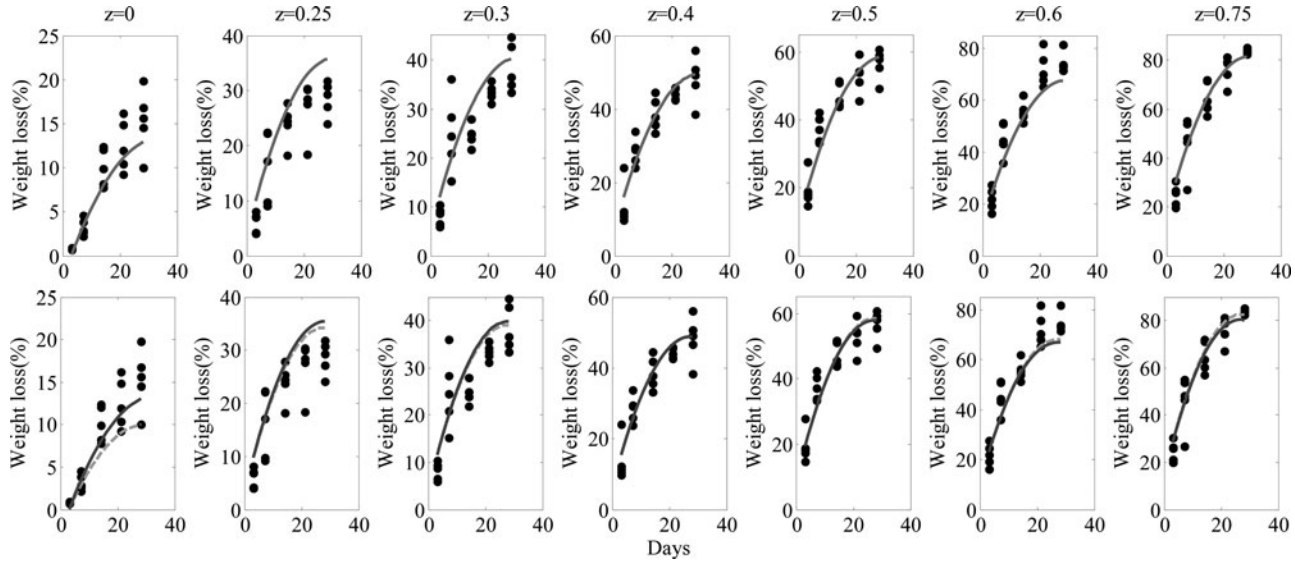

Figure 11. Fitted level-1 models: CHM (upper), UnCHM (lower, solid), and TSR (lower, dash).

4.3. *Comparison of the models*

Based on the posterior histograms in Fig. 10, the following concerns need to be addressed:

(Q1) Is logarithmic transformation on the variance needed?

(Q2) Does the level-1 model have zero intercept; i.e., $c_{0,1} = c_{0,0} = 0$?

For (Q2), since there is compelling evidence that $c_{0,1} > 0$ (all sampled values are positive in Fig. 10), we only need to determine if $c_{0,0} = 0$. The above concerns correspond to the following model comparisons:

(Q1) M_1: $\sigma^2 = c_1 z + c_0$ vs M_2: $\log(\sigma^2) = c_1 z + c_0$

(Q2) M_1: $c_{0,0} \in (-\infty, \infty)$ vs M_2: $c_{0,0} = 0$

In this study, (Q1) is solved first, and then (Q2) is solved based on the result of (Q1). Since the models in (Q1) are not nested, the harmonic mean estimator in Equation (11) is used to calculate the BF. For the two nested models in (Q2), both the harmonic mean estimator and the proposed estimator in Equation (12) are used. To show the variation in BF estimation, 25 samples are generated in the posterior sampling, each containing 20 000 iterations, and the BF for each sample is obtained. The results are given in the left panel of Fig. 12.

Clearly, the BFs for (Q1) show considerable variation over the samples, with an average of 5.16 (1.64 in log scale). This means that M_1 is about five times more plausible than M_2. According to the guidelines in Section 3.2.1, the evidence for M_1 is positive. Thus, we declare that the logarithmic transformation is not needed. For (Q2), the proposed estimator of the BF has a much smaller variance than the harmonic mean estimator, with an average of 1.32 (0.28 in log scale), which indicates barely worth mentioning evidence for M_1. This is consistent with the results on the Credible Intervals (CIs), which is the Bayesian version of confidence intervals. Using the posterior sample shown

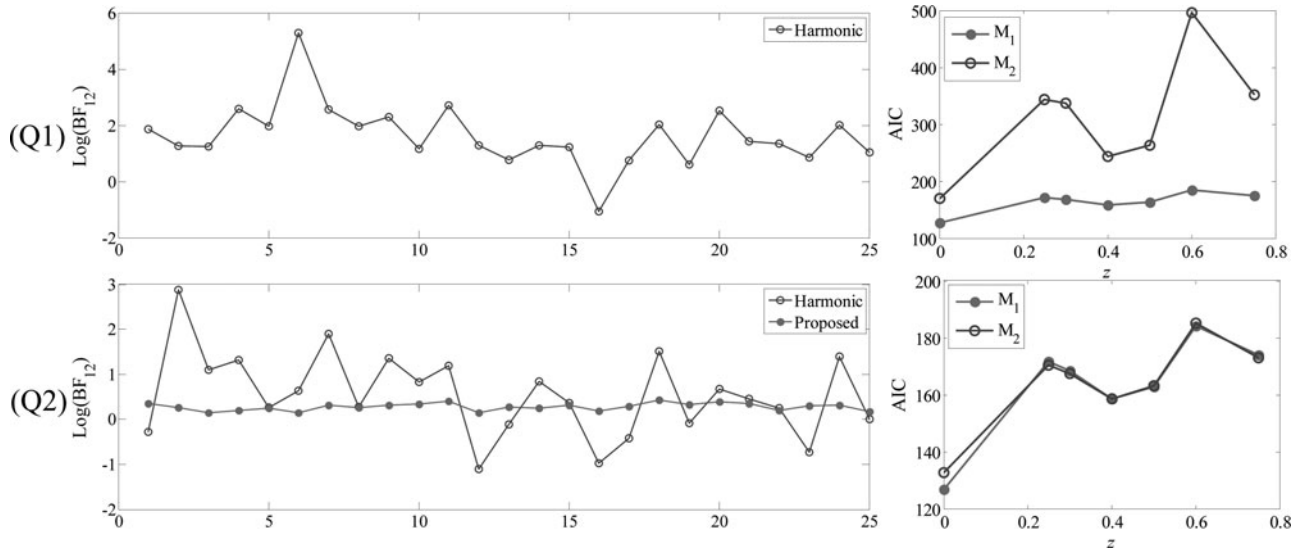

Figure 12. Calculated BFs and AIC values in model comparison.

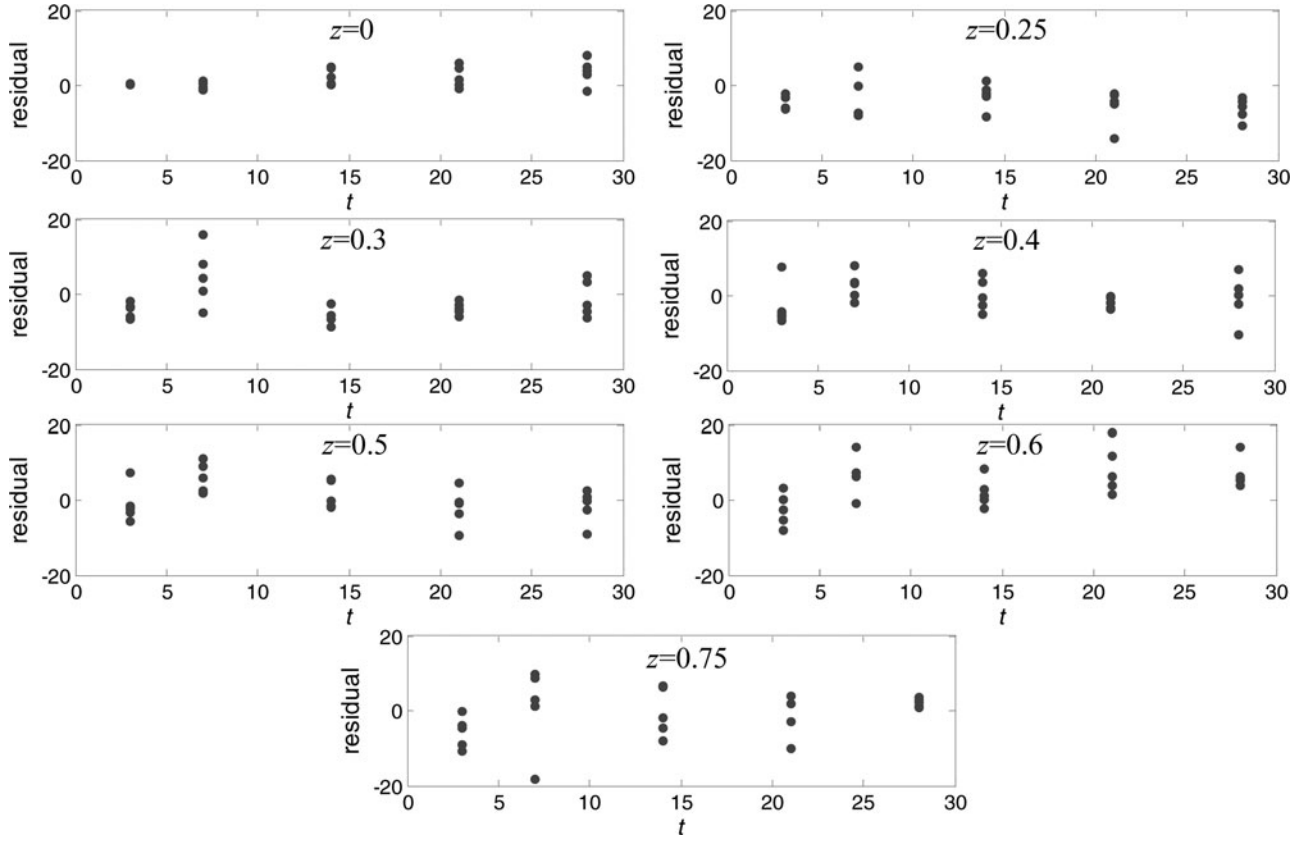

Figure 13. Residual plots under each value of the process parameter.

in Fig. 10, it is found that the 95% CI of $c_{0,0}$ is $[-5.39, 0.36]$, which covers zero, whereas the 90% CI is $[-4.92, -0.13]$, which does not cover zero. This means that the evidence for $c_{0,0} = 0$ (i.e., against M_1) can be significant or insignificant depending on the level of confidence. To be conservative, we choose M_1 and declare that $c_{0,0} \neq 0$. This analysis confirms that the proposed CHM in Equation (13) is the best model for the data.

The AIC is also calculated for the two models in (Q1) and (Q2). The results under each value of z are shown in the right panel of Fig. 12. We can see that the models in (Q1) have apparently different AIC values, and the AIC of M_1 is considerably smaller than that of M_2 in all cases. Thus, M_1 should be chosen, this is consistent with the decision based on the BF. For (Q2), the two models have very similar AIC values except that M_1 has a slightly smaller AIC under $z = 0$. M_1 should be chosen in this case, which is, again, consistent with the decision based on the BF.

4.4. *Residual analysis*

To check the adequacy of the fitted CHM, residual plots under each value of the process parameter are shown in Fig. 13. The mode estimates of parameters in Table 2 are used to calculate the residuals. There is no common pattern in the plots, indicating that the proposed model is reasonable for the data. Also, the variances of the residuals under different values of z are similar, which validates the adequacy of the variance model. Figure 14 shows the Q-Q (quantile-quantile) plots of the residuals. Clearly, the deviations of residuals from normal distribution are trivial, which verifies the normality assumption.

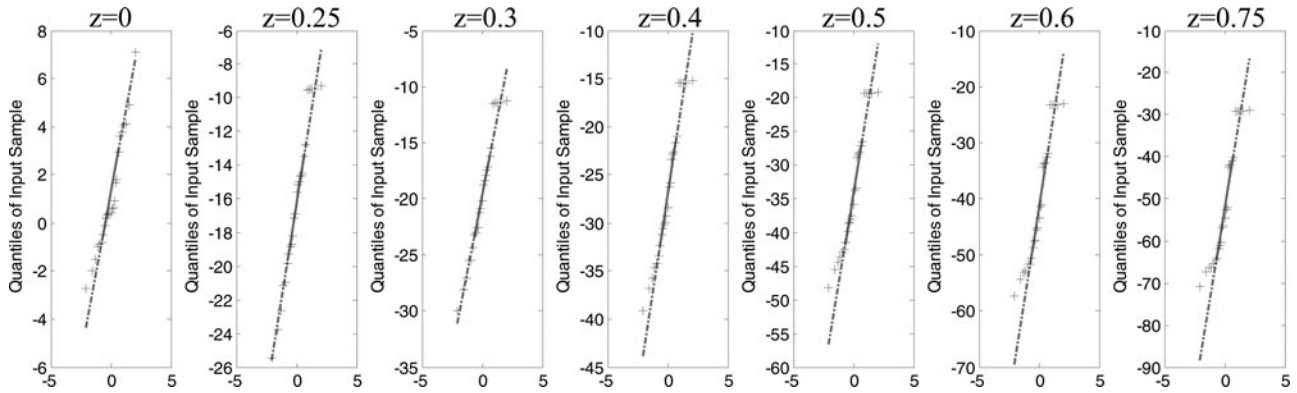

Figure 14. Q-Q plots of residuals under each value of the process parameter.

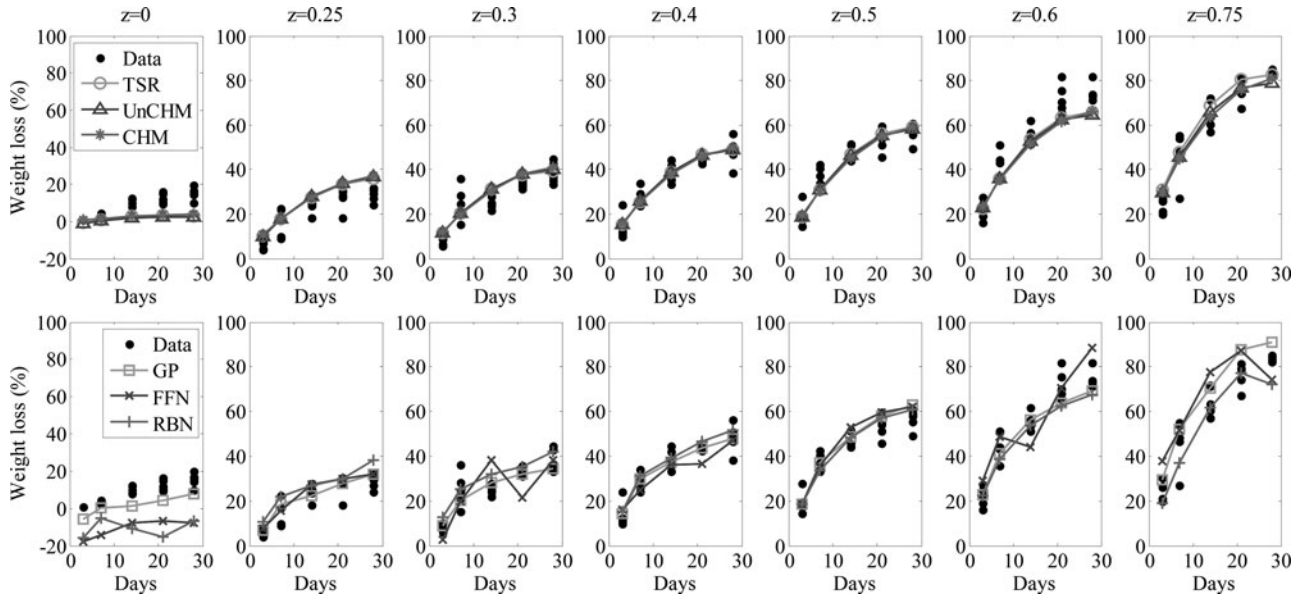

Figure 15. Degradation predictions from the six methods using original data.

4.5. *Prediction*

To evaluate the prediction performance of the proposed CHM, we adopt the leave-one-out cross-validation approach with respect to values of the process parameter ($z = 0$, 0.25, 0.3, 0.4, 0.5, 0.6, 0.75). Specifically, for each value of z, data under other values are used for model estimation, and data under this value are used for prediction of degradation at the five specified time points (i.e., $t = 3$, 7, 14, 21, 28 days). It should be noted that the predictions under $z = 0$ and $z = 0.75$ are extrapolations, whereas those under other values of z are interpolations.

We compare the prediction performance of the CHM with five other methods, including the UnCHM, the TSR, the GP model, a Feed-Forward Neural Network (FFNN), and a Radial Basis Neural Network (RBNN). The GP model used in this analysis follows a typical setup; i.e., with a constant mean, a Gaussian correlation function, and a normally distributed random error term with zero mean and constant variance (Ranjan *et al.*, 2011). The maximum likelihood estimates of parameters are obtained by genetic algorithm for the GP model. For the ANNs, there are two key parameters: the number of neurons (#neurons) for FFNN and the upper bound of MSE in training (trainMSE) for RBNN. In this study, different settings of these parameters (#neurons = 3–20, trainMSE = 8–40) are considered, and the settings that lead to best performance in training (for FFNN) or convergence of prediction error (for RBNN) are chosen.

The predicted degradations from the six methods are displayed in Fig. 15, with the three two-level models in the upper panel and the three surrogate models in the lower panel. Table 3 reports the corresponding Root Mean Squared Prediction Errors (RMSPEs). The following results are obtained.

(1) In the cases of interpolation (i.e., $z = 0.25$–0.6), the performance of the proposed CHM is comparable to other models except the GP model.
(2) In the cases of extrapolation (i.e., $z = 0$ and 0.75), the CHM gives the best prediction among the six methods. The prediction performance of the two ANNs are especially not promising in the cases of extrapolation (e.g., $z = 0$).
(3) Comparing the two groups, the two-level models generally perform better than the ANNs in most cases. For the GP model in comparison with two-level models, it has better prediction performance under $z = 0.25$ and 0.6 and comparable performance under $z = 0$, 0.3, 0.4, and 0.5 but worse performance under $z = 0.75$.

It is worth pointing out that all of the six methods do not produce good predictions for the case of $z = 0.6$. One potential reason lies in the large variability in the degradation data under large values of z (recall that the random error variance σ^2 increases with z as shown in Fig. 9). The prediction under $z = 0.75$ is the most challenging case, due to the large variability and inherent difficulty with extrapolation, whereas the CHM has the best prediction performance in this case. Another advantage of the CHM is that the predicted degradations are meaningful in all cases. From Fig. 15, it is easy to see that the surrogate models can yield predictions that are not practically reasonable, such as negative values (e.g., all of them under $z = 0$) and non-monotonic

Table 3. RMSPEs of the six methods using original data.

Methods		$z = 0$	$z = 0.25$	$z = 0.3$	$z = 0.4$	$z = 0.5$	$z = 0.6$	$z = 0.75$
Two-level models	TSR	8.0	6.3	6.0	4.7	5.3	8.4	7.1
	UnCHM	8.7	6.6	6.0	4.6	5.1	8.9	6.7
	CHM	7.4	6.7	6.0	4.7	5.3	9.0	6.6
Surrogate models	GP	7.4	4.3	5.1	4.4	5.4	6.7	9.5
	FFNN	19.4	4.6	9.5	5.9	6.2	10.0	12.2
	RBNN	20.3	7.0	5.8	4.8	5.1	7.7	9.3

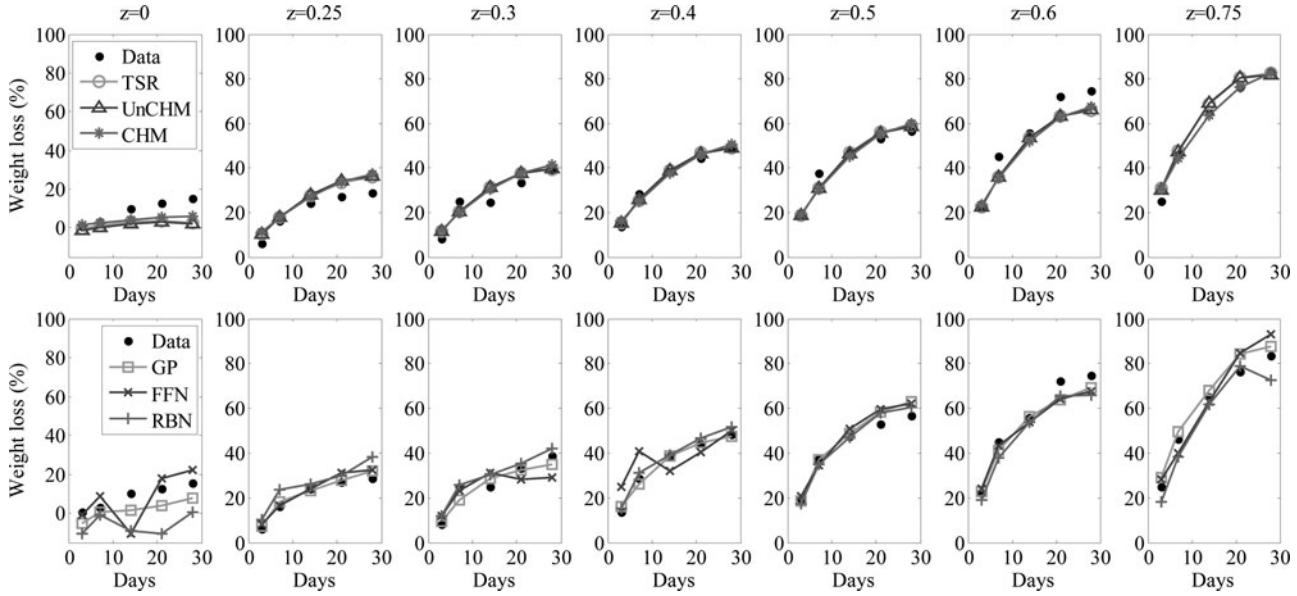

Figure 16. Degradation predictions from the six methods using average data.

values (e.g., FFNN under $z = 0.3, 0.6, 0.75$; RBNN under $z = 0.75$). Such an issue also exists for the TSR and UnCHM (e.g., under $z = 0$, their predictions at $t = 3$ days are negative and the prediction of TSR at $t = 28$ days is smaller than that at $t = 21$ days).

Note that there are five replicates at each time point in the original data. In practice, bioengineers often average the replicates for data analysis. To evaluate how this averaging procedure affects the performance of these methods, we repeat the above cross-validation study using the average data. The resulting predictions are shown in Fig. 16, and the corresponding RMSPEs are listed in Table 4. We can see that the results have similar patterns to those in Fig. 15, except that the prediction errors become smaller due to the averaging. Again, the CHM performs the best in the cases of extrapolation, and its prediction error under $z = 0.75$ is particularly small. Also, predictions with negative values and non-monotonic patterns are observed from other methods (e.g., TSR and UnCHM under $z = 0$, FFNN under $z = 0, 0.3, 0.4$, and RBNN under $z = 0, 0.75$).

4.6. *Summary of findings*

The case study validates the following advantages of the proposed CHM method.

(1) *Ability to incorporate constraints*. The CHM is the only method among the six methods compared in Section 4.5 that is able to take constraints into account. The inability to incorporate constraints may make the results not meaningful as shown in Figs. 15 and 16.

(2) *Easy interpretability*. The estimated CHM in the case study indicates that the process parameter has a linear effect on scaffold degradation profiles; and this effect is on both the shape and variability of the degradation. This is insightful information to scaffold researchers. In contrast, no such information can be obtained from the surrogate models.

(3) *Efficient model estimation*. As shown in Fig. 11, the CHM that simultaneously estimates the parameters at the two levels is better fit to data than the TSR.

(4) *Easy to implement*. The CHM method is very easy to implement as its model building (including model estimation and comparison) and prediction are all done using posterior samples that can be automatically generated. However, the use of the surrogate models has some intrinsic issues: the ANNs require the specification of some parameters, the appropriate values of which need to be found through trial-and-error; the GP method is likely to suffer numerical issues and become unstable when replicates are used in the analysis. For example, in the prediction study using original data in Section 4.5, the algorithms for model estimation do not converge 55% of cases.

Table 4. RMSPEs of the six methods using average data.

Methods		$z = 0$	$z = 0.25$	$z = 0.3$	$z = 0.4$	$z = 0.5$	$z = 0.6$	$z = 0.75$
Two-level models	TSR	7.7	5.0	4.5	1.9	3.5	7.0	3.9
	UnCHM	8.2	5.4	4.4	1.8	3.3	6.9	3.7
	CHM	5.7	5.6	4.4	2.0	3.7	7.1	2.7
Surrogate models	GP	7.1	2.0	3.7	1.4	3.7	4.8	6.0
	FFNN	10.3	2.7	5.9	8.2	4.4	5.1	6.7
	RBNN	15.9	5.9	3.6	2.3	3.1	6.1	7.0

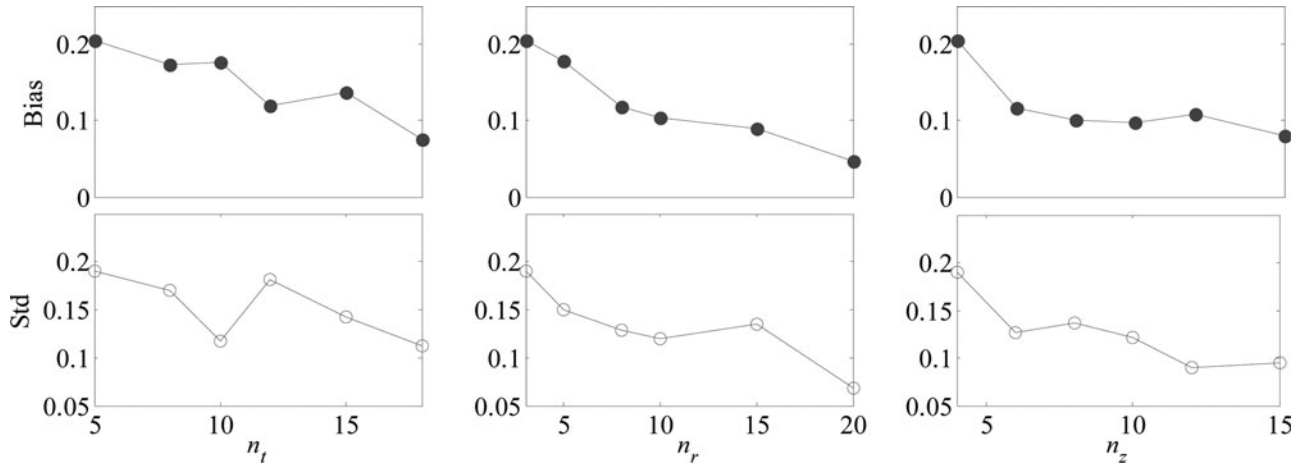

Figure 17. Performance of parameter estimation on the simulated datasets.

(5) *Good prediction*. The CHM has the best prediction performance in cases involving extrapolation, and its predictions are comparable to those advanced surrogate models in other cases.

5. Numerical study

We note that the performance of the proposed method in parameter estimation depends on the sample size of the data, which is determined by three design parameters in the data collection: the number of time points to measure degradation (n_t), the number of replicates measured at each time point (n_r), and the number of different settings of the process parameter (n_z). To further evaluate the performance of the proposed parameter estimation method, a series of simulations are conducted by varying these three parameters.

Let the range of t and z be the same as the data used in the case study; i.e., $t \in [3, 28]$, $z \in [0, 0.75]$. Three settings of the design parameters are considered: $n_t = 5, 8, 10, 12, 15, 18$ ($n_r = 3$, $n_z = 4$); $n_r = 3, 5, 8, 10, 15, 20$ ($n_t = 5$, $n_z = 4$); and $n_z = 4, 6, 8, 10, 12, 15$ ($n_t = 5$, $n_r = 3$). Under each setting, 30 datasets are simulated from the CHM in Equation (13) with the (true) parameters taking the values of the mode estimates in Table 2. Then parameter estimation is done for each simulated dataset using the proposed method in Section 3.1 (median estimates are calculated for convenience). A total of 4000 iterations are generated in the posterior sampling to make simulation time affordable.

Results of the estimation for $c_{2,1}$ are shown in Fig. 17 as an example. Two measures are used to evaluate the performance of the estimation, the relative bias (i.e., $|c_{2,1} - \hat{c}_{2,1}|/|c_{2,1}|$) and the relative standard deviation (i.e., $std(\hat{c}_{2,1})/|c_{2,1}|$). We can see that as the design parameters get larger, the bias and variance of the estimation get smaller, and the three design parameters exhibit similar patterns. This suggests that the accuracy of parameter estimation can be improved by increasing the number of scaffold specimens used in the degradation study.

6. Conclusions

This study develops a novel statistical modeling method for degradation data of tissue-engineered scaffolds. It can be used to identify significant process parameters in the scaffold fabrication process, predict scaffold degradation under given settings of the process parameters, and guide the design of experimental studies in practice. Results in the case study validate that the proposed method is able to provide interpretable and meaningful information regarding the effect of process parameters on scaffold degradation. The comparative study with existing methods shows that the proposed method has a prediction performance that is comparable to advanced surrogate models. The numerical study suggests that the accuracy of model estimation depends on sample size.

As an initial exploration on the degradation regulation problem in scaffold fabrication, this study has left open many issues that will be considered in our future research. First, in the posterior sampling, the constraints are applied by simply discarding samples that fall outside of the constrained space. This simple method is not applicable when the constrained space has little overlap with the high-density region of the posterior and thus most samples will be discarded. Advanced sampling methods for Bayesian-constrained estimation will be developed to solve this problem. Second, when the shape of the degradation profiles is complex, polynomial models may not be adequate to characterize the profiles. One idea is to use the extended version of polynomial models, the Piecewise Polynomial Models (PPMs; Denison *et al.* (1998)), as a complex profile can always be broken into a number of pieces, in which the local behavior can be approximated by a simple polynomial model. The intrinsic connection between the PPMs and the polynomial models will bring allow the extension of the Bayesian estimation methods developed in this study to the new CHM using PPMs. Third, this study separately fits a model for each parameter of the level-1 model, which may miss some common information on the parameters. A multivariate model will be explored in our future study to enhance the models at level 2 of the hierarchical modeling. Finally, models with multiple process parameters will also be considered when required data become available.

Funding

Li Zeng gratefully acknowledges financial support from the National Science Foundation under grant CMMI-1266225. Jian Yang acknowledges support from the National Science Foundation under grant CMMI-1266116.

Notes on contributors

Li Zeng is an Assistant Professor in the Department of Industrial and Manufacturing Systems Engineering at the University of Texas at Arlington. She received her B.S. degree in Precision Instruments and M.S. degree in Optical Engineering from Tsinghua University and Ph.D. in Industrial Engineering and M.S. degree in Statistics from the University of Wisconsin–Madison. Her research interest is process monitoring and control in complex manufacturing and healthcare delivery systems. She is a member of INFORMS and IIE.

Xinwei Deng is an Assistant Professor in the Department of Statistics at Virginia Tech. He received his Ph.D. degree in Industrial Engineering from Georgia Tech and his bachelor's degree in Mathematics from Nanjing University, China. His research interests are in statistical modeling and analysis of massive data, including high-dimensional classification, graphical model estimation, interface between experimental design and machine learning, and statistical approaches to nanotechnology. He is a member of INFORMS and ASA.

Jian Yang is an Associate Professor of Biomedical Engineering at the Pennsylvania State University. He is known as the inventor for citrate-based biomaterials for tissue engineering and medical devices. He has published 71 journal articles with many shown in prestigious journals such as *PNAS*, *Advanced Materials*, and *ACS Nano*. He has also received eight issued patents for his inventions in citrate polymers and their applications. He was a recipient of an NSF CAREER Award (2010) and Outstanding Young Faculty Award of College of Engineering at the University of Texas Arlington (2011). He serves as an Associate Editor for *Frontiers in Biomaterials* and on the editorial board for a number of journals in his field.

References

Buchanan, F. (ed). (2008) *Degradation Rate of Bioresorbable Materials: Prediction and Evaluation*, CRC, Boca Raton, FL.

Chen, V.C.P., Tsui, K.-L., Barton, R.R. and Meckesheimer, M. (2006) A review on design, modeling and applications of computer experiments. *IIE Transactions*, **38**(4), 273–291.

Chen, Y., Zhou, S. and Li, Q. (2011) Mathematical modeling of degradation for bulk-erosive polymers: applications in tissue engineering scaffolds and drug delivery systems. *Acta Biomaterialis*, **7**(3), 1140–1149.

Chu, P.K. and Liu, X. (2008) *Biomaterials Fabrication and Processing Handbook*, Taylor & Francis, Boca Raton, FL.

Daniels, M.J. and Gatsonis, C. (1999) Hierarchical generalized linear models in the analysis of variations in health care utilization. *Journal of the American Statistical Association*, **94**(445), 29–42.

Denison, D.G.T., Mallick, B.K. and Smith, A.F.M. (1998) Automatic Bayesian curve fitting. *Journal of the Royal Statistical Society B*, **60**(2), 333–350.

Dey, J., Xu, H., Shen, J., Thevenot, P., Gondi, S.R., Nguyen, K.T., Sumerlin, B.S., Tang, L. and Yang, J. (2008) Development of biodegradable crosslinked urethane-doped polyester elastomers. *Biomaterials*, **29**(35), 4637–4649.

Fisher, J.P., Mikos, A.G. and Bronzino, J.D. (eds). (2007) *Tissue Engineering*, Taylor & Francis, Boca Raton, FL.

Frees, W.E. (2004) *Longitudinal and Panel Data*, Cambridge University Press, New York, NY.

Gelfand, A.E., Smith, A.F.M. and Lee, T.M. (1992) Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling. *Journal of the American Statistical Association*, **87**(418), 523–532.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2004) *Bayesian Data Analysis*, second edition, Chapman & Hall/CRC, Boca Raton, FL.

Gelman, A. and Hill, J. (2007) *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge University Press, New York, NY.

Geweke, J. (1986) Exact inference in the inequality constrained normal linear regression model. *Journal of Applied Econometrics*, **1**(2), 127–141.

Gilks, W.R. (2005) Markov chain Monte Carlo, in *Encyclopedia of Biostatistics*, John Wiley & Sons, Hoboken, NJ.

Goh, T.N. (2014) Management of parametric response in design of experiments. *Quality and Reliability Engineering International*, **30**(8), 1191–1195.

Govaerts, B. and Noël, J. (2005) Analysing the results of a designed experiment when the response is a curve: methodology and application in metal injection moulding. *Quality and Reliability Engineering International*, **21**(5), 509–520.

Han, C. and Carlin, B.P. (2001) Markov chain Monte Carlo methods for computing Bayes factors: a comparative review. *Journal of the American Statistical Association*, **96**(455), 1122–1132.

He, Z., Zhou, P., Zhang, M. and Goh, T.N. (2015) A review of analysis of dynamic response in design of experiments. *Quality and Reliability Engineering International*, **31**(4), 535–542.

Horova, I., Kolacek, J. and Zelinka, J. (2012) *Kernel Smoothing in Matlab: Theory and Practice of Kernel Smoothing*, World Scientific Publishing Company, Singapore.

Ikada, Y. (2006) *Tissue Engineering: Fundamentals and Applications*, Elsevier, San Diego, CA.

Jiang, J. (2007) *Linear and Generalized Linear Mixed Models and Their Applications*, Springer, New York, NY.

Kass, R.E. and Raftery, A.E. (1995) Bayes factors. *Journal of the American Statistical Association*, **90**(430), 773–795.

Laurencin, C.T. and Nair, L.S. (2008) *Nanotechnology and Tissue Engineering: the Scaffold*, Taylor & Francis, Boca Raton, FL.

Liao, C.-J., Chen, C.-F., Chen, J.-H., Chiang, S.-F., Lin, Y.-J. and Chang, K.Y. (2002) Fabrication of porous biodegradable polymer scaffolds using a solvent merging/particulate leaching method. *Journal of Biomedical Materials Research*, **59**(4), 676–681.

Ma, P.X. and Elisseeff, J. (eds). (2006) *Scaffolding in Tissue Engineering*, Taylor & Francis, Boca Raton, FL.

Metters, A.T., Anseth, K.S. and Bowman, C.N. (2001) A statistical kinetic model for the bulk degradation of PLA-b-PEG-b-PLA hydrogel networks: incorporating network non-idealities. *Journal of Physical Chemistry B*, **105**(34), 8069–8076.

Neal, R.M. (2003) Slice sampling. *The Annals of Statistics*, **31**(3), 705–741.

Noorossana, R., Saghaei, A. and Amiri, A. (2011) *Statistical Analysis of Profile Monitoring*, John Wiley & Sons, Hoboken, NJ.

Ranjan, P., Haynes, R. and Karsten, R. (2011) A computationally stable approach to Gaussian process interpolation of deterministic computer simulation data. *Technometrics*, **53**(4), 366–378.

Robert, C.P. and Casella, G. (2004) *Monte Carlo Statistical Methods, second edition*, Springer, New York, NY.

Tateishi, T. (ed). (2008) *Biomaterials in Asia*, World Scientific Publishing Co., Singapore.

Tran, R., Naseri, E., Kolasnikov, A., Bai, X. and Yang, J. (2011) A new generation of sodium chloride porogen for tissue engineering. *Biotechnology and Applied Biochemistry*, **58**(5), 335–344.

Verbeke, G. and Molenberghs, G. (2000) *Linear Mixed Models for Longitudinal Data*, Springer, New York, NY.

Woodall, W.H., Spitzner, D.J., Montgomery, D.C. and Gupta, S. (2004) Using control charts to monitor process and product quality profiles. *Journal of Quality Technology*, **36**(3), 309–320.

Yang, J., Webb, A.R. and Ameer, G.A. (2004) Novel citric acid-based biodegradable elastomers for tissue engineering. *Advanced Materials*, **16**(6), 511–516.

Appendices

Appendix A. proofs of equations (6) and (7)

By Equations (1) to (4):

$$\mathbf{y} = \sum_{i=0}^{p-1} g_i(\mathbf{z};\ \mathbf{c}_i) \bullet \mathbf{t}^i + \boldsymbol{\varepsilon},\ \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{V}_0(\mathbf{z};\ \mathbf{c}_\sigma)), \tag{A.1}$$

thus the joint posterior of $\{\mathbf{c}_0, \ldots, \mathbf{c}_{p-1}, \mathbf{c}_\sigma\}$ is

$$f(\mathbf{c}_0, \mathbf{c}_1, \ldots, \mathbf{c}_{p-1}, \mathbf{c}_\sigma | \mathbf{y}) \propto \pi(\mathbf{c}_0, \mathbf{c}_1, \ldots, \mathbf{c}_{p-1}, \mathbf{c}_\sigma) \cdot N\left(\mathbf{y} | \sum_{i=0}^{p-1} g_i(\mathbf{z};\ \mathbf{c}_i) \bullet \mathbf{t}^i, \mathbf{V}_0(\mathbf{z};\ \mathbf{c}_\sigma)\right),$$

where $\pi(\cdot)$ is the prior. Since flat priors are used for all of the parameters:

$$f(\mathbf{c}_0, \mathbf{c}_1, \ldots, \mathbf{c}_{p-1}, \mathbf{c}_\sigma|\mathbf{y}) \propto N\left(\mathbf{y}\left|\sum_{i=0}^{p-1} g_i(\mathbf{z};\ \mathbf{c}_i) \bullet \mathbf{t}^i, \mathbf{V}_0(\mathbf{z};\ \mathbf{c}_\sigma)\right.\right).$$

The conditional posteriors in Equations (6) and (7) can be obtained from the joint posterior by conditioning on the given quantities. □

Appendix B. proof of equations (8) and (9)

By Equation (A1):

$$\mathbf{y} = \sum_{i=1}^{p-1} g_i(\mathbf{z};\ \mathbf{c}_i) \bullet \mathbf{t}^i + g_0(\mathbf{z};\ \mathbf{c}_0) + \boldsymbol{\varepsilon},\ \boldsymbol{\varepsilon} \sim N(\mathbf{0},\ \mathbf{V}_0(\mathbf{z};\ \mathbf{c}_\sigma)).$$

Given $\{\mathbf{c}_1, \ldots, \mathbf{c}_{p-1}, \mathbf{c}_\sigma\}$, the first term is a constant, so can be moved to the left. Let

$$\mathbf{u}_0 \equiv \mathbf{y} - \sum_{i=1}^{p-1} g_i(\mathbf{z};\ \mathbf{c}_i) \bullet \mathbf{t}^i;$$

thus,

$$\mathbf{u}_0 = g_0(\mathbf{z};\ \mathbf{c}_0) + \boldsymbol{\varepsilon} = \mathbf{Z}_0\mathbf{c}_0 + \boldsymbol{\varepsilon},\ \boldsymbol{\varepsilon} \sim N(\mathbf{0},\ \mathbf{V}_0(\mathbf{z};\ \mathbf{c}_\sigma)). \quad \text{(A.2)}$$

Equation (A2) can be viewed as a polynomial model of $\mathbf{z}$ with coefficient $\mathbf{c}_0$ and known variance–covariance matrix $\mathbf{V}_0(\mathbf{z};\ \mathbf{c}_\sigma)$. According to Gelman *et al.* (2004, p. 374), under flat priors, the conditional posterior of the coefficient of a linear model given variance–covariance matrix is multivariate normal:

$$\mathbf{c}_0|\mathbf{c}_1, \ldots, \mathbf{c}_{p-1}, \mathbf{c}_\sigma, \mathbf{y} \sim N\Big(\big(\mathbf{Z}_0'(\mathbf{V}_0(\mathbf{z};\ \mathbf{c}_\sigma))^{-1}\mathbf{Z}_0\big)^{-1}\mathbf{Z}_0'\big(\mathbf{V}_0\big(\mathbf{z};\ \mathbf{c}_\sigma\big)\big)^{-1}\mathbf{u}_0, \times \big(\mathbf{Z}_0'\big(\mathbf{V}_0\big(\mathbf{z};\ \mathbf{c}_\sigma\big)\big)^{-1}\mathbf{Z}_0\big)^{-1}\Big),$$

which gives Equation (8).

The conditional posterior of $\mathbf{c}_k$ given $\{\mathbf{c}_1, \ldots, \mathbf{c}_{k-1}, \mathbf{c}_{k+1}, \ldots, \mathbf{c}_{p-1}, \mathbf{c}_\sigma\}$, $1 \le k \le p-1$, can be proved in a similar way. Specifically, by Equation (A1):

$$\mathbf{y} = \sum_{i=0}^{k-1} g_i(\mathbf{z};\ \mathbf{c}_i) \bullet \mathbf{t}^i + \sum_{i=k+1}^{p-1} g_i(\mathbf{z};\ \mathbf{c}_i) \bullet \mathbf{t}^i + g_k(\mathbf{z};\ \mathbf{c}_k) \bullet \mathbf{t}^k + \boldsymbol{\varepsilon}.$$

Let

$$\mathbf{u}_k \equiv \left[\mathbf{y} - \sum_{i=0}^{k-1} g_i(\mathbf{z};\ \mathbf{c}_i) \bullet \mathbf{t}^i - \sum_{i=k+1}^{p-1} g_i(\mathbf{z};\ \mathbf{c}_i) \bullet \mathbf{t}^i\right] \bullet \Big/ \mathbf{t}^k,$$

then

$$\mathbf{u}_k = g_k(\mathbf{z};\ \mathbf{c}_k) + \boldsymbol{\varepsilon} \bullet /\mathbf{t}^k = \mathbf{Z}_k\mathbf{c}_k + \boldsymbol{\varepsilon} \bullet /\mathbf{t}^k. \quad \text{(A.3)}$$

Since $\boldsymbol{\varepsilon} \sim N(\mathbf{0},\ \mathbf{V}_0(\mathbf{z};\ \mathbf{c}_\sigma)) = N(\mathbf{0},\ diag(g_\sigma(\mathbf{z};\ \mathbf{c}_\sigma)))$:

$$\boldsymbol{\varepsilon} \bullet /\mathbf{t}^k \sim N(\mathbf{0},\ diag(g_\sigma(\mathbf{z};\ \mathbf{c}_\sigma) \bullet /\mathbf{t}^{2k})) = N(\mathbf{0},\ \mathbf{V}_k(\mathbf{z};\ \mathbf{c}_\sigma)).$$

Thus, Equation (A3) can be viewed as a polynomial model of $\mathbf{z}$ with coefficient $\mathbf{c}_k$ and known variance–covariance matrix $\mathbf{V}_k(\mathbf{z};\ \mathbf{c}_\sigma)$. Thus,

$$\mathbf{c}_k|\mathbf{c}_1, \ldots, \mathbf{c}_{k-1}, \mathbf{c}_{k+1}, \ldots, \mathbf{c}_{p-1}, \mathbf{c}_\sigma, \mathbf{y} \sim N\big(\big(\mathbf{Z}_k'(\mathbf{V}_k(\mathbf{z};\ \mathbf{c}_\sigma))^{-1}\mathbf{Z}_k\big)^{-1}\mathbf{Z}_k'\big(\mathbf{V}_k\big(\mathbf{z};\ \mathbf{c}_\sigma\big)\big)^{-1}\mathbf{u}_k, \times \big(\mathbf{Z}_k'(\mathbf{V}_k\big(\mathbf{z};\ \mathbf{c}_\sigma\big)\big)^{-1}\mathbf{Z}_k\big)^{-1}\big),$$

which gives Equation (9). □

Appendix C. proof of theorem 1.

From the definition of the BF in Equation (10):

$$BF_{ls} = \frac{P(\mathbf{y}|M_l)}{P(\mathbf{y}|M_s)} = \frac{\int \pi(\boldsymbol{\theta}_l|M_l) f(\mathbf{y}|\boldsymbol{\theta}_l, M_l) d\boldsymbol{\theta}_l}{\int \pi(\boldsymbol{\theta}_s|M_s) f(\mathbf{y}|\boldsymbol{\theta}_s, M_s) d\boldsymbol{\theta}_s}.$$

By condition (i):

$$BF_{ls} = \frac{\int\int \pi(\boldsymbol{\theta}_s|M_l)\pi(\xi|M_l) f(\mathbf{y}|\boldsymbol{\theta}_l, M_l) d\boldsymbol{\theta}_s d\boldsymbol{\xi}}{\int \pi(\boldsymbol{\theta}_s|M_s) f(\mathbf{y}|\boldsymbol{\theta}_s, M_s) d\boldsymbol{\theta}_s} = \int \left[\frac{\int \pi(\boldsymbol{\theta}_s|M_l) f(\mathbf{y}|\boldsymbol{\theta}_l, M_l) d\boldsymbol{\theta}_s}{\int \pi(\boldsymbol{\theta}_s|M_s) f(\mathbf{y}|\boldsymbol{\theta}_s, M_s) d\boldsymbol{\theta}_s}\right] \pi(\boldsymbol{\xi}|M_l) d\boldsymbol{\xi}.$$

Note that the denominator in the bracket—i.e., $\int \pi(\boldsymbol{\theta}_s|M_s) f(\mathbf{y}|\boldsymbol{\theta}_s, M_s) d\boldsymbol{\theta}_s$—is a constant since the integrand only depends on $\boldsymbol{\theta}_s$. The numerator $\int \pi(\boldsymbol{\theta}_s|M_l) f(\mathbf{y}|\boldsymbol{\theta}_l, M_l) d\boldsymbol{\theta}_s$ depends on ξ since the integrand, $\pi(\boldsymbol{\theta}_s|M_l) f(\mathbf{y}|\boldsymbol{\theta}_l, M_l)$, involves $\boldsymbol{\theta}_s$ and $\boldsymbol{\xi}$ (recall $\theta_l = [\theta_s\ \xi]'$). Thus, the quantity in the bracket is a function of ξ.

Let

$$\psi(\boldsymbol{\xi}) \equiv \left[\frac{\int \pi(\boldsymbol{\theta}_s|M_l) f(\mathbf{y}|\boldsymbol{\theta}_l, M_l) d\boldsymbol{\theta}_s}{\int \pi(\boldsymbol{\theta}_s|M_s) f(\mathbf{y}|\boldsymbol{\theta}_s, M_s) d\boldsymbol{\theta}_s}\right],$$

then

$$BF_{ls} = \int \psi(\boldsymbol{\xi}) \pi(\boldsymbol{\xi}|M_l) d\boldsymbol{\xi}, \quad \text{(A.4)}$$

which means that the BF is an average of $\psi(\xi)$ over the prior of ξ under M_l. By condition (iii):

$$\pi(\boldsymbol{\xi}|M_l) = \frac{1}{b-a}. \quad \text{(A.5)}$$

The joint posterior with this prior is

$$f(\boldsymbol{\theta}_s, \boldsymbol{\xi}|\mathbf{y}, M_l) = \frac{\pi(\boldsymbol{\theta}_s|M_l)\pi(\boldsymbol{\xi}|M_l) f(\mathbf{y}|\boldsymbol{\theta}_s, \boldsymbol{\xi}, M_l)}{f(\mathbf{y}|M_l)},$$

which gives the marginal posterior of ξ:

$$f(\boldsymbol{\xi}|\mathbf{y}, M_l) = \int f(\boldsymbol{\theta}_s, \boldsymbol{\xi}|\mathbf{y}, M_l) d\boldsymbol{\theta}_s = \frac{\int \pi(\boldsymbol{\theta}_s|M_l)\pi(\boldsymbol{\xi}|M_l) f(\mathbf{y}|\boldsymbol{\theta}_s, \boldsymbol{\xi}, M_l) d\boldsymbol{\theta}_s}{f(\mathbf{y}|M_l)}.$$

Consequently,

$$f(\boldsymbol{\xi}_0|\mathbf{y}, M_l) = \int f(\boldsymbol{\theta}_s, \boldsymbol{\xi}_0|\mathbf{y}, M_l) d\boldsymbol{\theta}_s = \frac{\int \pi(\boldsymbol{\theta}_s|M_l)\pi(\boldsymbol{\xi}_0|M_l) f(\mathbf{y}|\boldsymbol{\theta}_s, \boldsymbol{\xi}_0, M_l) d\boldsymbol{\theta}_s}{f(\mathbf{y}|M_l)}$$

$$= \frac{\int \pi(\boldsymbol{\theta}_s|M_s)\pi(\boldsymbol{\xi}_0|M_l) f(\mathbf{y}|\boldsymbol{\theta}_s, M_s) d\boldsymbol{\theta}_s}{f(\mathbf{y}|M_l)}.$$

Thus,

$$\frac{f(\boldsymbol{\xi}|\mathbf{y}, M_l)}{f(\boldsymbol{\xi}_0|\mathbf{y}, M_l)} = \frac{\int \pi(\boldsymbol{\theta}_s|M_l)\pi(\boldsymbol{\xi}|M_l) f(\mathbf{y}|\boldsymbol{\theta}_s, \boldsymbol{\xi}, M_l) d\boldsymbol{\theta}_s}{\int \pi(\boldsymbol{\theta}_s|M_s)\pi(\boldsymbol{\xi}_0|M_l) f(\mathbf{y}|\boldsymbol{\theta}_s, M_s) d\boldsymbol{\theta}_s}$$
$$= \frac{\int \pi(\boldsymbol{\theta}_s|M_l) f(\mathbf{y}|\boldsymbol{\theta}_s, \boldsymbol{\xi}, M_l) d\boldsymbol{\theta}_s}{\int \pi(\boldsymbol{\theta}_s|M_s) f(\mathbf{y}|\boldsymbol{\theta}_s, M_s) d\boldsymbol{\theta}_s}.$$

By the definition of $\psi(\boldsymbol{\xi})$, we can get

$$\psi(\boldsymbol{\xi}) = \frac{f(\boldsymbol{\xi}|\mathbf{y}, M_l)}{f(\boldsymbol{\xi}_0|\mathbf{y}, M_l)}. \tag{A.6}$$

Plugging Equations (A5) and (A6) to Equation (A4) gives

$$BF_{ls} = \int \frac{f(\boldsymbol{\xi}|\mathbf{y}, M_l)}{f(\boldsymbol{\xi}_0|\mathbf{y}, M_l)} \cdot \frac{1}{(b-a)} d\boldsymbol{\xi}$$
$$= \frac{\int f(\boldsymbol{\xi}|\mathbf{y}, M_l) d\boldsymbol{\xi}}{(b-a) f(\boldsymbol{\xi}_0|\mathbf{y}, M_l)} = \frac{1}{(b-a) f(\boldsymbol{\xi}_0|\mathbf{y}, M_l)},$$

so the theorem holds. □