IISE Transactions

Taylor & Francis
Taylor & Francis Group

ISSN: 2472-5854 (Print) 2472-5862 (Online) Journal homepage: http://www.tandfonline.com/loi/uiie21

Constrained Gaussian Process with Application in Tissue-engineering Scaffold Biodegradation

Li Zeng , Xinwei Deng & Jian Yang

To cite this article: Li Zeng , Xinwei Deng & Jian Yang (2017): Constrained Gaussian Process with Application in Tissue-engineering Scaffold Biodegradation, IISE Transactions, DOI: 10.1080/24725854.2017.1414973

To link to this article: https://doi.org/10.1080/24725854.2017.1414973

Accepted author version posted online: 11 Dec 2017.

Constrained Gaussian Process with Application in Tissue-engineering Scaffold Biodegradation

Authors' Biographies

Li Zeng is an Assistant Professor in the Department of Industrial and Systems Engineering at Texas A&M University. She received her B.S. degree in Precision Instruments and M.S. degree in Optical Engineering from Tsinghua University, China, and Ph.D. in Industrial Engineering and M.S. degree in Statistics from the University of Wisconsin-Madison. Her research interests are systems informatics and process control in complex manufacturing and healthcare delivery systems. She is a member of INFORMS and IISE.

Xinwei Deng is an Associate 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 a 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 also received 8 issued patents for his inventions in citrate polymers and their applications. Dr. Yang was a recipient of NSF CAREER Award (2010) and Outstanding Young Faculty Award of College of Engineering at University of Texas Arlington (2011). Dr. Yang serves as an Associate Editor for Frontiers in Biomaterials and on the editorial board for a number of journals in his field.

Abstract

In many biomanufacturing areas such as tissue-engineering scaffold fabrication, the biodegradation performance of products is a key to produce products with desirable properties. Prediction of biodegradation often encounters the challenge on how to incorporate expert knowledge appropriately. This paper proposes a constrained Gaussian process (CGP) method for predictive modeling with application to scaffold biodegradation. It provides a unified framework of using proper constraints to accommodate various types of expert knowledge in predictive modeling, including censoring, monotonicity, and bounds requirements. Efficient Bayesian sampling procedures for prediction are also developed. The performance of the proposed method is demonstrated in a case study from a novel scaffold fabrication process. Compared with the unconstrained GP and artificial neural networks, the

proposed method can provide more accurate and meaningful prediction. A simulation study is also conducted to further reveal the properties of the CGP.

Keywords: biomanufacturing, biodegradation, constrained Gaussian process (CGP), censoring, monotonicity, predictive modeling

ACCEPTED MANUSCRIPT

1. Introduction

Biomanufacturing is an emerging area and getting rapid growth in recent years (Grant and Settles, 2009). In biomanufacturing, *biodegradation* is an important performance aspect of products (Buchanan, 2008), especially for those integrated into human systems and made by degradable biomaterials. Figure 1 depicts the complicated biodegradation process caused by hydrolysis: in the human body environment, water molecules penetrate into the matrix of the product, causing it to swell. This triggers the breakdown of chemical chains, leading to weight loss, which continues until complete dissolution of the product.

For many biomanufacturing processes, the biodegradation rate of products needs to be designed to meet requirements in specific applications. One typical example is the *scaffold* fabrication in tissue engineering, as illustrated in Figure 2, in an attempt to develop biological substitutes for failing tissues/organs (Chu and Liu, 2008; Fisher *et al*., 2007; Sultana, 2013). First, relevant cells are grown in vitro into a three-dimensional tissue/organ. To enable the cells to grow in favored orientations similar to the native tissue, the cells are seeded onto the scaffold, which is a highly porous matrix made by degradable biomaterials. The pores on the scaffold provide space for flow transport of nutrients and metabolic wastes, thus forming a temporary substrate and microenvironment for cells. Then the cell-scaffold composite is implanted into the human body, where the scaffold eventually degrades, leaving only the new tissue/organ. As the scaffold plays a critical role to the success of this development, it is very crucial to match the degradation rate of scaffolds to the cell growth rate in the application of interest (Burdick and Mauck, 2011). If the degradation rate is too fast, there would be insufficient support to the cells, while if the degradation rate is too slow, the scaffold may impede the growth of new tissues.

In scaffold fabrication, the biodegradation performance of products is usually characterized by an experiment setup as shown in Figure 3 (Dey *et al*., 2008; Henry *et al*., 2007; Nicodemus and Bryant, 2008). Scaffold specimens are incubated in phosphate buffered saline (PBS, i.e., salt solution, used to mimic the human body environment) for a period of time; at each predetermined time point, one specimen is taken out, dried, measured by weight loss, and discarded after that. Scaffold products with desired biodegradation performance can be obtained by adjusting process variables in scaffold fabrication such as those in material synthesis (e.g., compositions of the biomaterial and their percentages) and those in pore construction (e.g., pore size and processing conditions) (Liao *et al*., 2002; Cui *et al*., 2015).

Achieving desired biodegradation performance is, however, challenging due to the lack of an understanding of the relationship between process variables and biodegradation performance of products. Analytical models of the relationship are often not available since the effects of process variables on scaffold biodegradation are very complicated. As a result, the trial-and-error approach is predominant in this field currently (Burdick and Mauck, 2011). In this study, we focus on data-driven methods for predictive modeling of scaffold biodegradation. The objective is to establish an empirical model for biodegradation prediction in scaffold fabrication such that it will enable process optimization to produce scaffolds with required biodegradation performance.

It is worth pointing out that the problem considered in this work is different from degradation modeling in reliability studies (e.g., Bian and Gebraeel, 2014; Chen and Tsui, 2013; Ranfiee *et al*., 2014). First, the biodegradation of scaffolds is treated as a controllable performance aspect of products in this study, which bears a different nature than the degradation of engineering components as a reliability concern. Second, the modeling of biodegradation is to characterize the relationship between product biodegradation performance and process variables in scaffold fabrication, while the degradation modeling in reliability studies is to characterize the time evolution of degradation. Finally, the data used in this study are scaffold biodegradation measurements (as shown in Figure 3) under different settings of process variables. The measurements are often collected at a small number (e.g., 5 in the case study) of time points as quantification of biodegradation is very time-consuming (taking months or years in some cases). In contrast, classical time-series data are usually used in reliability studies.

Surrogate models are common methods for predictive modeling of complex relationships between process variables/design parameters (predictors) and product performance (response) in manufacturing applications (Arendt *et al*., 2015; Chen *et al*., 2006; Tsai *et al*., 2012). Among various surrogate models, Gaussian process (GP) and artificial neural networks (ANNs) are two popular ones widely used in similar problems as ours. For example, the GP modeling is used in the prediction of product mechanical performance in nanomanufacturing (Pourhabib, *et al*., 2015) and of wafer geometric quality in semiconductor manufacturing (Jin *et al*., 2012). The ANNs are used in the prediction of surface roughness and other quality measures in machining (Feng and Wang, 2003, 2004; Feng *et al*., 2006).

However, the aforementioned surrogate methods may not work well for modeling scaffold biodegradation because expert knowledge needs to be incorporated to ensure meaningful prediction. Such knowledge includes: (*i*) *Full-degradation censoring*. Once the full degradation is reached, the weight loss

measurement will be a constant 100%. (*ii*) *Monotonicity*. Intrinsically, the biodegradation of scaffolds is monotonically increasing with respect to time and to some process variables in scaffold fabrication. (*iii*) *Bounds of weight loss*. The percentage of weight loss is bounded between 0% and 100%. Without guidelines from expert knowledge, those methods are likely to result in poor predictions and interpretation. It calls for a novel modeling method that is able to accommodate these three types of expert knowledge.

Zeng *et al*. (2016) develop a constrained hierarchical model for the scaffold biodegradation modeling problem, where one type of expert knowledge is incorporated as constraints on model parameters. This approach is easy to implement with good interpretation, but it only works for monotonicity constraints. In the literature, some nonparametric constrained modeling methods, such as shape-constrained function estimation methods (e.g., Chatterjee *et al*., 2015; Shively *et al*., 2011; Wang and Ghosh, 2012), may be useful for this problem. However, they can only deal with monotonicity constraints too. Moreover, they are designed for the one-dimensional case (i.e., a single predictor) as opposed to the multi-dimensional case (i.e., more than one predictors) assumed in this study. There are also some work considering shape constrains in the GP modeling (Lenk and Choi, 2017; Lin and Dunson, 2014; Riihimäki and Vehtari, 2010; Wang, 2012; Wang and Berger, 2016). These methods can incorporate constraints conveniently and work for multi-dimensional cases, but, again, they are limited to monotonicity and other shape constraints.

In this work, we propose a constrained GP (CGP) method for the predictive modeling of scaffold biodegradation. It provides a unified framework to accommodate the aforementioned three types of expert knowledge in the form of constraints. Efficient Bayesian algorithms are also developed for model estimation and prediction. The algorithms address several issues in the implementation of the CGP such as the identification of constrained locations and sampling of posteriors. The contribution of this work lies in three aspects: First, the CGP method introduces a novel, convenient way to accommodate expert knowledge in predictive modeling of product performance. The GP is known as a flexible method for predictive modeling (Rasmussen and Williams, 2006). The formation of expert knowledge as proper constraints on the GP modeling makes its intrinsic flexibility nimble for accurate prediction with meaningful interpretation. Moreover, the proposed method can provide useful inference for scaffold fabrication such as estimate of the time to reach full degradation. Second, although the proposed method is illustrated using the scaffold biodegradation problem in this work, it has broad applicability in other

manufacturing processes such as biomaterial-based additive manufacturing (Wei *et al*., 2015) and for other types of product performance such as mechanical and swelling performances (Wang, et al., 2015; Wei *et al*., 2015), where the three types of expert knowledge also apply. Third, unlike most existing studies that utilize simulated or observational data of large sample sizes, this study demonstrates a case of experimental data with limited samples and the advantages of the proposed CGP in prediction are validated in comparison with GP and ANNs. A simulation study is also conducted to reveal important properties of the CGP.

The remainder of this paper is organized as follows. Section 2 reviews the basics of the Gaussian process model and presents definition of the scaffold biodegradation modeling problem. Section 3 describes the proposed CGP method to impose each type of constraints. Some related problems are discussed in Section 4. Results of the case study are given in Section 5. Section 6 presents two numerical examples. Finally, Section 7 concludes the paper and discusses future work. Bayesian sampling procedures to implement the CGP method are summarized in Appendix III to provide convenience for practitioners.

2. Background and Problem Definition

In this section, we will briefly review the basics of GP modeling and prediction. Then we define the scaffold biodegradation modeling problem.

2.1 Gaussian process model

Suppose the observed data are $(\mathbf{x}_i, y_i)$, $i=1,\ldots,n$, where $\mathbf{x}_i = [x_{i1},..., x_{id}]'$ is the ith realization of the d-dimensional predictor and y_i is the corresponding response. Following the GP literature (Santner *et al*. 2003; Fang *et al*., 2005), we call $\mathbf{x}_1,\ldots,\mathbf{x}_n$ as locations. To model the relationship between the response and predictors, the GP modeling considers

$$y_i = \mu + f(\mathbf{x}_i) + \varepsilon_i, \tag{1}$$

where μ is the mean, $f(\mathbf{x}_i)$ is a random function of $\mathbf{x}_i$, and $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ is the random error, called nugget effect, which is independent of $f(\mathbf{x}_i)$. Here the random function $f(\mathbf{x})$ follows a Gaussian process with zero mean and covariance function $\sigma_f^2 R(\mathbf{x})$. That is, the vector$[f(\mathbf{x}_1), f(\mathbf{x}_2),..., f(\mathbf{x}_n)]'$ follows a multivariate normal distribution with $f(\mathbf{x}_i) \sim N(0, \sigma_f^2)$ and $\text{cov}(f(\mathbf{x}_i), f(\mathbf{x}_j)) = \sigma_f^2 R_{ij}$ for $i \neq j$. A popular choice of the correlation function is the Gaussian correlation function (Rasmussen and Williams, 2006) such that

$$R_{ij} = \prod_{w=1}^{d} \exp[-\theta_w (x_{iw} - x_{jw})^2], \tag{2}$$

where $\boldsymbol{\theta} = [\theta_1, ..., \theta_d]'$ are scale parameters on each dimension of the predictor. It means that the correlation of $f(\mathbf{x}_i)$ and $f(\mathbf{x}_j)$ depends on the distance between the two locations $\mathbf{x}_i$ and $\mathbf{x}_j$.

By denoting $\mathbf{y} = [y_1, y_2, ..., y_n]'$, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n]'$, $\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2, ..., \varepsilon_n]'$, it is easy to see that

$$\mathbf{y} \sim N(\mu \mathbf{1}_n, \sigma_f^2 \mathbf{R}(\mathbf{X}, \mathbf{X}) + \sigma_\varepsilon^2 \mathbf{I}_n), \tag{3}$$

where $\mathbf{1}_n$ is an n-dimensional column vector of 1s and $\mathbf{R}(\mathbf{X},\mathbf{X})$ is the correlation matrix of $\mathbf{f}(\mathbf{X}) = [f(\mathbf{x}_1), f(\mathbf{x}_2), ..., f(\mathbf{x}_n)]'$ with the (i, j)th entry R_{ij}, and $\mathbf{I}_n$ is the $n \times n$ identity matrix. For notational convenience, we denote the covariance function by

$$\mathbf{K}^{00}(\mathbf{X}, \mathbf{X}) = \sigma_f^2 \mathbf{R}(\mathbf{X}, \mathbf{X}) = (K^{00}(\mathbf{x}_i, \mathbf{x}_j))_{n \times n}, \tag{4}$$

where $K^{00}(\mathbf{x}_i, \mathbf{x}_j) = \text{cov}(f(\mathbf{x}_i), f(\mathbf{x}_j)) = \sigma_f^2 R_{ij}$. Here the superscript "00" is to distinguish the covariance function of the GP from other covariance functions that will be given in Section 3.2.

The predication based on the GP modeling is straightforward as follows. Let $\mathbf{X}^* = [\mathbf{x}_1^*, \mathbf{x}_2^*, ..., \mathbf{x}_{n^*}^*]'$ be the vector of locations for prediction, and $\mathbf{f}(\mathbf{X}^*) = [f(\mathbf{x}_1^*), f(\mathbf{x}_2^*), ..., f(\mathbf{x}_{n^*}^*)]'$ be the function values at these locations. Since $f(\mathbf{x})$ follows a Gaussian process, it is easy to find that $(\mathbf{f}(\mathbf{X})', \mathbf{f}(\mathbf{X}^*)')'$ follows multivariate normal, and $\mathbf{f}(\mathbf{X}^*)$ given $\mathbf{y}$ also follows a multivariate normal with mean and variance-covariance matrix as follows (Schabenberger and Gotway, 2005)

$$\begin{aligned} \text{E}[\mathbf{f}(\mathbf{X}^*) \mid \mathbf{y}, \boldsymbol{\psi}] &= \mu \mathbf{1}_{n^*} + \mathbf{K}^{00}(\mathbf{X}^*, \mathbf{X})[\mathbf{K}^{00}(\mathbf{X}, \mathbf{X}) + \sigma_\varepsilon^2 \mathbf{I}_n]^{-1}(\mathbf{y} - \mu \mathbf{1}_n), \\ \text{cov}[\mathbf{f}(\mathbf{X}^*) \mid \mathbf{y}, \boldsymbol{\psi}] &= \mathbf{K}^{00}(\mathbf{X}^*, \mathbf{X}^*) - \mathbf{K}^{00}(\mathbf{X}^*, \mathbf{X})[\mathbf{K}^{00}(\mathbf{X}, \mathbf{X}) + \sigma_\varepsilon^2 \mathbf{I}_n]^{-1} \mathbf{K}^{00}(\mathbf{X}, \mathbf{X}^*), \end{aligned} \tag{5}$$

where $\boldsymbol{\psi} = [\mu, \boldsymbol{\theta}, \sigma_f^2, \sigma_\varepsilon^2]$ is the parameters of the GP, $\mathbf{K}^{00}(\mathbf{X}^*, \mathbf{X}) = \text{cov}(\mathbf{f}(\mathbf{X}^*), \mathbf{f}(\mathbf{X})) = \sigma_f^2 \mathbf{R}(\mathbf{X}^*, \mathbf{X})$ and $\mathbf{K}^{00}(\mathbf{X}^*, \mathbf{X}^*)$ is similarly defined. Thus, the conditional mean $\text{E}[\mathbf{f}(\mathbf{X}^*) \mid \mathbf{y}, \boldsymbol{\psi}]$ in Eq. (5) can be used as a best linear unbiased predictor of $\mathbf{f}(\mathbf{X}^*)$.

It needs to mention that the estimation and prediction based on the GP modeling is often described in the Bayesian framework. In this framework, the distributions based on the basic setup of GP, e.g., Eq. (3), are called prior distributions, while those updated distributions given data, e.g., Eq. (5), are called posterior distributions. This framework will be followed in this paper.

2.2. Predictive modeling of scaffold biodegradation

There are two sets of predictors to scaffold biodegradation: time and process variables in scaffold fabrication. For convenience, we will use the two-dimensional case, with time *t* and a process variable *z* as predictors, to illustrate the biodegradation modeling problem, though the proposed CGP is a generic method for multi-dimensional cases. Suppose the biodegradation experiment is conducted at a grid of n_t time points $(t_1 < t_2 < ... < t_{n_t})$ and n_z values of the process variable $(z_1 < z_2 < ... < z_{n_z})$. We denote $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ as these $n=n_t \times n_z$ different settings of the predictors. The corresponding responses are the weight loss measurements $\{y_1, \ldots, y_n\}$. Figure 4 illustrates the structure of the data (not real data), where each stream of solid dots represent biodegradation measurements under the same value of the process variable. The objective of this work is to model the relationship between the response and the predictors, and thus enable to predict the weight loss at new locations $\mathbf{x}_1^*, \mathbf{x}_2^*, ..., \mathbf{x}_{n^*}^*$. In the modeling, the three types of expert knowledge described in the Introduction (i.e., censoring, monotonicity and bounds requirements) will be taken into account.

It is worth pointing out that a new location may represent an unsampled time point under a sampled value of *z* (e.g., a time point between t_1 and t_2 under $z = z_1$), or a sampled time point under an unsampled value of *z* (e.g., t_2 under a *z* value between z_1 and z_2), or an unsampled time point under an unsampled value of *z* (e.g., a time point between t_1 and t_2 under a *z* value between z_1 and z_2). In addition, the *t* value and/or *z* value of a new location can be within or out of the observed data region; estimation of weight loss in these two cases are called interpolation and extrapolation, respectively. For simplicity, both cases are called "prediction" in this paper, and the performance of the proposed method in each case will be investigated in the case study.

3. The Proposed Method

This section describes the proposed CGP method for scaffold biodegradation modeling, where the three types of expert knowledge will be incorporated in the form of constraints, referred to as *censoring constraint*, *monotonicity constraint* and *bound constraint*. It is designed to simultaneously impose multiple constraints of these three types. To facilitate the understanding, how to impose each type of constraints will be presented in the following.

3.1 Imposing censoring constraint

Censored measurement is often encountered in scaffold biodegradation experiments, as shown in Figure 5 where the censored measurement occurs at $\mathbf{x}_n = [t_{n_t}, z_{n_z}]'$. This measurement indicates that full degradation reached at or before t_{n_t} under $z = z_{n_z}$. Obviously, the measured value "100%" cannot be

directly used in the modeling and prediction as the response at $\mathbf{x}_n$. Here we propose a novel method to take this censored measurement into account.

The idea is as follows: let the dash line in Figure 5 represent the true biodegradation trajectory during the measurement period. The piece of the line above "100%" is not realizable due to the full-degradation of scaffold; it is simply an extension of the realizable biodegradation trajectory. Let y_n be the weight loss at $\mathbf{x}_n$, denoted by the dot circle in Figure 5. Note that y_n is not an actual measurement but an imagined quantity. Also, y_n is a random variable which cannot be lower than 100% according to the actual measurement at this location. Consequently, we can impose a constraint "$y_n \geq 100$" in the modeling. The CGP method following this idea is described below.

Let us first consider the case with one censored measurement as shown in Figure 5. Denoting the uncensored data as $\mathbf{y}^{(n-1)} = [y_1, y_2, ..., y_{n-1}]'$, $\mathbf{X}^{(n-1)} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{n-1}]'$, the whole dataset is $\mathbf{y} = [\mathbf{y}^{(n-1)}; y_n]$, $\mathbf{X} = [\mathbf{X}^{(n-1)}; \mathbf{x}_n]$. From Eq. (3), the prior of y_n is a truncated Gaussian

$$\pi(y_n) = N\left(\mu, \sigma_f^2 + \sigma_\varepsilon^2\right) \cdot I(y_n \geq 100), \tag{6}$$

where $I(.)$ is an indicator function. Given this prior, the joint (conditional) posterior of the prediction $\mathbf{f}(\mathbf{X}^*)$ and y_n can be found, as stated below.

Proposition 1: *Given the prior in Eq. (6), the joint posterior distribution of ($\mathbf{f}(\mathbf{X}^*)$, y_n) is*

$$\begin{aligned} P(\mathbf{f}(\mathbf{X}^*), y_n \mid \mathbf{y}^{(n-1)}, \boldsymbol{\psi}) &= P(\mathbf{f}(\mathbf{X}^*) \mid \mathbf{y}^{(n-1)}, y_n, \boldsymbol{\psi}) \cdot P(y_n \mid \mathbf{y}^{(n-1)}, \boldsymbol{\psi}) \\ &= P(\mathbf{f}(\mathbf{X}^*) \mid \mathbf{y}, \boldsymbol{\psi}) \cdot P(y_n \mid \mathbf{y}^{(n-1)}, \boldsymbol{\psi}) \end{aligned},$$

where

$$y_n \mid \mathbf{y}^{(n-1)}, \boldsymbol{\psi} \sim N\left(m(\mathbf{x}_n), V(\mathbf{x}_n)\right) \cdot I(y_n \geq 100), \tag{7}$$

$$\mathbf{f}(\mathbf{X}^*) \mid \mathbf{y}, \boldsymbol{\psi} \sim N\left(\mathbf{m}(\mathbf{X}^*), \mathbf{V}(\mathbf{X}^*)\right), \tag{8}$$

with

$$m(\mathbf{x}_n) = \mu + \mathbf{K}^{00}(\mathbf{x}_n, \mathbf{X}^{(n-1)})[\mathbf{K}^{00}(\mathbf{X}^{(n-1)}, \mathbf{X}^{(n-1)}) + \sigma_\varepsilon^2 \mathbf{I}_{n-1}]^{-1}(\mathbf{y}^{(n-1)} - \mu \mathbf{1}_{n-1}),$$

$$V(\mathbf{x}_n) = K^{00}(\mathbf{x}_n, \mathbf{x}_n) - \mathbf{K}^{00}(\mathbf{x}_n, \mathbf{X}^{(n-1)})[\mathbf{K}^{00}(\mathbf{X}^{(n-1)}, \mathbf{X}^{(n-1)}) + \sigma_\varepsilon^2 \mathbf{I}_{n-1}]^{-1}\mathbf{K}^{00}(\mathbf{X}^{(n-1)}, \mathbf{x}_n),$$

$$\mathbf{m}(\mathbf{X}^*) = \mu \mathbf{1}_{n^*} + \mathbf{K}^{00}(\mathbf{X}^*, \mathbf{X})[\mathbf{K}^{00}(\mathbf{X}, \mathbf{X}) + \sigma_\varepsilon^2 \mathbf{I}_n]^{-1}(\mathbf{y} - \mu \mathbf{1}_n),$$

$$\mathbf{V}(\mathbf{X}^*) = \mathbf{K}^{00}(\mathbf{X}^*, \mathbf{X}^*) - \mathbf{K}^{00}(\mathbf{X}^*, \mathbf{X})[\mathbf{K}^{00}(\mathbf{X}, \mathbf{X}) + \sigma_\varepsilon^2 \mathbf{I}_n]^{-1}\mathbf{K}^{00}(\mathbf{X}, \mathbf{X}^*).$$

The proof is given in Appendix I. The above result indicates that the joint posterior of $\mathbf{f}(\mathbf{X}^*)$ and y_n given the uncensored data $\mathbf{y}^{(n-1)}$ and the GP parameter $\boldsymbol{\psi}$ can be decomposed into two parts: the conditional posterior of y_n given $\mathbf{y}^{(n-1)}$, and the conditional posterior of $\mathbf{f}(\mathbf{X}^*)$ given $\mathbf{y}^{(n-1)}$ and y_n. The specific forms of the two conditional posteriors are given in Eqs. (7) and (8). Specifically, the conditional

posterior of y_n is a truncated normal distribution, while the conditional posterior of $\mathbf{f}(\mathbf{X}^*)$ is a multivariate normal distribution. Consequently, an estimate for $\mathbf{f}(\mathbf{X}^*)$ can be found by sampling from the joint posterior of $\mathbf{f}(\mathbf{X}^*)$ and y_n in two steps: first draw a sample of y_n from the truncated normal distribution in Eq. (7), and then given that value, draw a sample of $\mathbf{f}(\mathbf{X}^*)$ from the multivariate normal distribution in Eq. (8).

The above method can be extended straightforwardly to deal with cases of more than one censored measurements. Assume there are n_c censored measurements, with locations $\mathbf{X}^{(c)} = [\mathbf{x}^{(1)},..., \mathbf{x}^{(n_c)}]'$ and imagined responses $\mathbf{y}^{(c)} = [y^{(1)},..., y^{(n_c)}]'$. Then the prior in Eq. (6) becomes

$$\pi(\mathbf{y}^{(c)}) = N\left(\mu\mathbf{1}_{n_c}, \mathbf{K}^{00}(\mathbf{X}^{(c)}, \mathbf{X}^{(c)}) + \sigma_\varepsilon^2\mathbf{I}_{n_c}\right) \cdot I(y^{(1)} \geq 100,..., y^{(n_c)} \geq 100).$$

Given this prior, the posteriors in Eqs. (7) and (8) still apply except that the scalar terms are replaced by their vector counterparts.

3.2 Imposing monotonicity constraint

As mentioned in the Introduction, scaffold biodegradation follows some intrinsic monotonicity properties with respect to time or certain process variable according to expert knowledge. As shown in Figure 3, each scaffold specimen only yields one data point and it will be destructed during the weight loss measurement and discarded after that. In other words, the data points on the observed biodegradation profile are obtained from different specimens rather than from the same specimen over time. As a result, it is possible that some later measurements are smaller than earlier measurements (e.g., the 2nd data point in the lower of Figure 3) due to sample uncertainty among scaffold specimens. Consequently, predictions violating the intrinsic monotonicity may occur. Therefore, monotonicity constraints are necessary in modeling and prediction of the scaffold biodegradation data in practice.

One advantage of adopting the GP modeling is that *the derivative process of the GP is also a GP* (Rasmussen and Williams, 2006), thus making it convenient to impose monotonicity constraints. Hereafter the two processes will be referred to as the *original GP* and the *derivative GP*. For easy understanding, we will first consider monotonicity with respect to a single predictor, and then generalize to both predictors.

Case I: Monotonicity with respect to a single predictor

Let us consider imposing monotonicity constraints with respect to time. We first need to specify a set of locations, called *constrained set*, where monotonicity is required. Let $\mathbf{X}^\Delta = [\mathbf{x}_1^\Delta, \mathbf{x}_2^\Delta,..., \mathbf{x}_m^\Delta]'$ be the vector of m constrained locations, and $\mathbf{f}'(\mathbf{X}^\Delta) = [f'(\mathbf{x}_1^\Delta), f'(\mathbf{x}_2^\Delta),..., f'(\mathbf{x}_m^\Delta)]'$ be the first derivatives with respect to

time at these locations. Thus, the monotonicity constraints are $f'(\mathbf{x}_1^{\Delta}) \geq 0,..., f'(\mathbf{x}_m^{\Delta}) \geq 0$, as illustrated in Figure 6. Note that the locations in the constrained set do not have to be the observed ones $\{\mathbf{x}_1,\ldots,\mathbf{x}_n\}$. The covariance of the derivative GP and the covariance of the derivative GP and the original GP are

$$\mathbf{K}^{01}(\mathbf{X},\mathbf{X}^{\Delta}) = (K^{01}(\mathbf{x}_i,\mathbf{x}_j^{\Delta}))_{n\times m}, \text{ where } K^{01}(\mathbf{x}_i,\mathbf{x}_j^{\Delta}) = 2\sigma_f^2 R_{ij}\theta_1(x_{i1} - x_{j1}^{\Delta}),$$
$$\mathbf{K}^{10}(\mathbf{X}^{\Delta},\mathbf{X}) = (K^{10}(\mathbf{x}_j^{\Delta},\mathbf{x}_i))_{m\times n} = (K^{01}(\mathbf{x}_i,\mathbf{x}_j^{\Delta}))_{n\times m},$$
$$\mathbf{K}^{11}(\mathbf{X}^{\Delta},\mathbf{X}^{\Delta}) = (K^{11}(\mathbf{x}_j^{\Delta},\mathbf{x}_k^{\Delta}))_{m\times m}, \text{ where } K^{11}(\mathbf{x}_j^{\Delta},\mathbf{x}_k^{\Delta}) = 2\sigma_f^2 R_{jk}\theta_1[(1-2\theta_1(x_{j1}^{\Delta} - x_{k1}^{\Delta})^2],$$

(9)

where $K^{11}(\cdot\ ,\cdot)$ is the covariance function of the derivative GP, and $K^{01}(\cdot\ ,\cdot)/K^{10}(\cdot\ ,\cdot)$ is the covariance function of the original GP and the derivative GP. In the above formulas, x_{i1}, x_{j1}^{Δ} and x_{k1}^{Δ} are the time values (among $t_1,...,t_{n_t}$) of locations $\mathbf{x}_i$, $\mathbf{x}_j^{\Delta}$ and $\mathbf{x}_k^{\Delta}$, R_{ij} is the correlation function defined in Eq. (2) between locations $\mathbf{x}_i$ and $\mathbf{x}_j^{\Delta}$, and θ_1 is the scale parameter for time in the correlation function. Derivations of Eq. (9) are provided in Appendix II.

Given the monotonicity constraints, the prior of $\mathbf{f}'(\mathbf{X}^{\Delta})$ is

$$\pi(\mathbf{f}'(\mathbf{X}^{\Delta})) = N\left(\mathbf{0},\mathbf{K}^{11}(\mathbf{X}^{\Delta},\mathbf{X}^{\Delta})\right)\cdot I(f'(\mathbf{x}_1^{\Delta}) \geq 0,..., f'(\mathbf{x}_m^{\Delta}) \geq 0). \quad (10)$$

The joint posterior of $\mathbf{f}(\mathbf{X}^*)$ and $\mathbf{f}'(\mathbf{X}^{\Delta})$ can be derived in a similar way as Proposition 1.

Proposition 2: *Given the prior in Eq. (9), the joint posterior distribution of* $(\mathbf{f}(\mathbf{X}^*),\mathbf{f}'(\mathbf{X}^{\Delta}))$ *is*

$$P(\mathbf{f}(\mathbf{X}^*),\mathbf{f}'(\mathbf{X}^{\Delta})\,|\,\mathbf{y},\boldsymbol{\psi}) = P(\mathbf{f}(\mathbf{X}^*)\,|\,\mathbf{f}'(\mathbf{X}^{\Delta}),\mathbf{y},\boldsymbol{\psi})\cdot P(\mathbf{f}'(\mathbf{X}^{\Delta})\,|\,\mathbf{y},\boldsymbol{\psi}),$$

where

$$\mathbf{f}'(\mathbf{X}^{\Delta})\,|\,\mathbf{y},\boldsymbol{\psi} \sim N(\mathbf{m}(\mathbf{X}^{\Delta}),\mathbf{V}(\mathbf{X}^{\Delta}))\cdot I(f'(\mathbf{x}_1^{\Delta}) \geq 0,..., f'(\mathbf{x}_m^{\Delta}) \geq 0),$$

(11)

$$\mathbf{f}(\mathbf{X}^*\,|\,\mathbf{f}'(\mathbf{X}^{\Delta}),\mathbf{y},\boldsymbol{\psi} \sim N\left(\mathbf{m}(\mathbf{X}^*),\mathbf{V}(\mathbf{X}^*)\right),$$

(12)

with

$$\mathbf{m}(\mathbf{X}^{\Delta}) = \mathbf{K}^{10}(\mathbf{X}^{\Delta},\mathbf{X})[\sigma_\varepsilon^2\mathbf{I} + \mathbf{K}^{00}(\mathbf{X},\mathbf{X})]^{-1}(\mathbf{y} - \mu\mathbf{1}_n),$$
$$\mathbf{V}(\mathbf{X}^{\Delta}) = \mathbf{K}^{11}(\mathbf{X}^{\Delta},\mathbf{X}^{\Delta}) - \mathbf{K}^{10}(\mathbf{X}^{\Delta},\mathbf{X})[\sigma_\varepsilon^2\mathbf{I} + \mathbf{K}^{00}(\mathbf{X},\mathbf{X})]^{-1}\mathbf{K}^{01}(\mathbf{X},\mathbf{X}^{\Delta}),$$
$$\mathbf{m}(\mathbf{X}^*) = \mu\mathbf{1}_{n^*} + \mathbf{A}^{-1}(\mathbf{A}_1^T\mathbf{B}_1^{-1}(\mathbf{y} - \mu\mathbf{1}_n) + \mathbf{B}_2^{-1}\mathbf{A}_2\mathbf{f}'(\mathbf{X}^{\Delta})),$$
$$\mathbf{V}(\mathbf{X}^*) = \mathbf{A}^{-1}.$$

In the above terms, the matrix $\mathbf{A} = \mathbf{A}_1^T\mathbf{B}_1^{-1}\mathbf{A}_1 + \mathbf{B}_2^{-1}$ with $\mathbf{A}_1 = \mathbf{K}^{00}(\mathbf{X},\mathbf{X}^*)[\mathbf{K}^{00}(\mathbf{X}^*,\mathbf{X}^*)]^{-1}$, $\mathbf{A}_2 = \mathbf{K}^{01}(\mathbf{X}^*,\mathbf{X}^{\Delta})[\mathbf{K}^{11}(\mathbf{X}^{\Delta},\mathbf{X}^{\Delta})]^{-1}$, $\mathbf{B}_1 = \sigma_\varepsilon^2\mathbf{I} + \mathbf{K}^{00}(\mathbf{X},\mathbf{X}) - \mathbf{K}^{00}(\mathbf{X},\mathbf{X}^*)[\mathbf{K}^{00}(\mathbf{X}^*,\mathbf{X}^*)]^{-1}\mathbf{K}^{00}(\mathbf{X}^*,\mathbf{X})$, and $\mathbf{B}_2 = \mathbf{K}^{00}(\mathbf{X}^*,\mathbf{X}^*) - \mathbf{K}^{01}(\mathbf{X}^*,\mathbf{X}^{\Delta})[\mathbf{K}^{11}(\mathbf{X}^{\Delta},\mathbf{X}^{\Delta})]^{-1}\mathbf{K}^{10}(\mathbf{X}^{\Delta},\mathbf{X}^*)$. This result is similar to Lemma 3.1 in the

study of Wang (2012). It has a similar interpretation as Proposition 1, that is, we can find estimate for $\mathbf{f}(\mathbf{X}^*)$ by sampling from the joint posterior of $\mathbf{f}(\mathbf{X}^*)$ and $\mathbf{f}'(\mathbf{X}^\Delta)$: first draw a sample of $\mathbf{f}'(\mathbf{X}^\Delta)$ from the truncated multivariate normal distribution in Eq. (11), and then draw a sample of $\mathbf{f}(\mathbf{X}^*)$ from the multivariate normal distribution in Eq. (12).

If the biodegradation increases monotonically as the process variable *z* increases according to expert knowledge, monotonicity constraints with respect to *z* need to be imposed. All the formulas will take the same form as above except that θ_1 and $x_{i1} - x_{j1}^\Delta$ are replaced by θ_2 and $x_{i2} - x_{j2}^\Delta$, respectively, in Eq. (9).

Case II: Monotonicity with respect to both predictors

When the monotonicity constraints are applicable for both of the two predictors, we can specify a general constrained set

$$\mathbf{X}^\Delta = [\mathbf{x}_1^\Delta, ..., \mathbf{x}_{m_1}^\Delta, \mathbf{x}_{m_1+1}^\Delta, ..., \mathbf{x}_m^\Delta]',$$

with their first derivatives

$$\mathbf{f}'(\mathbf{X}^\Delta) = [f'^t(\mathbf{x}_1^\Delta), ..., f'^t(\mathbf{x}_{m_1}^\Delta), f'^z(\mathbf{x}_{m_1+1}^\Delta), ..., f'^z(\mathbf{x}_m^\Delta)]',$$

where $f'^t(\cdot) = \partial f(\cdot)/\partial t$, $f'^z(\cdot) = \partial f(\cdot)/\partial z$. The constraints to impose are

$$f'^t(\mathbf{x}_1^\Delta) > 0, ..., f'^t(\mathbf{x}_{m_1}^\Delta) > 0, f'^z(\mathbf{x}_{m_1+1}^\Delta) > 0, ..., f'^z(\mathbf{x}_m^\Delta) > 0.$$

That is, there are monotonicity constraints with respect to time at m_1 locations $\{\mathbf{x}_1^\Delta, ..., \mathbf{x}_{m_1}^\Delta\}$ and those with respect to the process variable at $m-m_1$ locations $\{\mathbf{x}_{m_1+1}^\Delta, ..., \mathbf{x}_m^\Delta\}$, as illustrated in Figure 7.

In this case, Proposition 2 still applies, except that the covariance functions, i.e., $\mathbf{K}^{01}(\mathbf{X}, \mathbf{X}^\Delta)$ and $\mathbf{K}^{11}(\mathbf{X}^\Delta, \mathbf{X}^\Delta)$ in Eqs. (11)-(12), are replaced by

$$\mathbf{K}^{10}(\mathbf{X}^\Delta, \mathbf{X}) = [\mathbf{K}^{t0}(\mathbf{X}_{\mathrm{I}}^\Delta, \mathbf{X}) \quad \mathbf{K}^{z0}(\mathbf{X}_{\mathrm{II}}^\Delta, \mathbf{X})],$$
$$\mathbf{K}^{11}(\mathbf{X}^\Delta, \mathbf{X}^\Delta) = \begin{bmatrix} \mathbf{K}^{tt}(\mathbf{X}_{\mathrm{I}}^\Delta, \mathbf{X}_{\mathrm{I}}^\Delta) & \mathbf{K}^{tz}(\mathbf{X}_{\mathrm{I}}^\Delta, \mathbf{X}_{\mathrm{II}}^\Delta) \\ \mathbf{K}^{zt}(\mathbf{X}_{\mathrm{II}}^\Delta, \mathbf{X}_{\mathrm{I}}^\Delta) & \mathbf{K}^{zz}(\mathbf{X}_{\mathrm{II}}^\Delta, \mathbf{X}_{\mathrm{II}}^\Delta) \end{bmatrix}, \quad (13)$$

where $\mathbf{X}_{\mathrm{I}}^\Delta = [\mathbf{x}_1^\Delta, ..., \mathbf{x}_{m_1}^\Delta]'$, $\mathbf{X}_{\mathrm{II}}^\Delta = [\mathbf{x}_{m_1+1}^\Delta, ..., \mathbf{x}_m^\Delta]'$, and $\mathbf{X}^\Delta = [\mathbf{X}_{\mathrm{I}}^\Delta, \mathbf{X}_{\mathrm{II}}^\Delta]'$. In the superscripts of the terms at the right side, "0" indicates the original GP, "*t*" indicates the first derivative with respect to time, and "*z*" indicates the first derivative with respect to the process variable. Specific formulas of the covariance functions and derivations are given in Appendix II.

3.3 Imposing bound constraint

Bound constraints may exist at one or several locations for prediction. Let us consider a general case where the predictions must satisfy $f(\mathbf{x}_1^*) \in U_1, ..., f(\mathbf{x}_{n^*}^*) \in U_{n^*}$, with $U_1, ..., U_{n^*}$ being the bounds of weight loss according to expert knowledge. This equates to a prior for the predictions

$$\pi(\mathbf{f}(\mathbf{X}^*)) = N\left(\mathbf{0}, \mathbf{K}^{00}(\mathbf{X}^*, \mathbf{X}^*)\right) \cdot \mathbf{1}_{\{f(\mathbf{x}_1^*) \in U_1, ..., f(\mathbf{x}_{n^*}^*) \in U_{n^*}\}}. \quad (14)$$

The resulting posterior of $\mathbf{f}(\mathbf{X}^*)$ is given below.

Proposition 3: *Given the prior in Eq. (14), the posterior of* $\mathbf{f}(\mathbf{X}^*)$ *is*

$$\mathbf{f}(\mathbf{X}^*) \mid \mathbf{y}, \boldsymbol{\psi} \sim N\left(\mathbf{m}(\mathbf{X}^*), \mathbf{V}(\mathbf{X}^*)\right) \cdot \mathbf{1}_{\{f(\mathbf{x}_1^*) \in U_1, ..., f(\mathbf{x}_{n^*}^*) \in U_{n^*}\}}, \quad (15)$$

where the mean and variance-covariance matrices of the normal distribution are as defined in Eq. (8).

This result is natural based on the results in Propositions 1 and 2. To obtain samples of $\mathbf{f}(\mathbf{X}^*)$, we need to draw from the truncated multivariate normal distribution in Eq. (15).

4. Model Estimation and Inference

Note that the posterior distributions in Propositions 1-3 are conditional on the parameters of the original GP model $\boldsymbol{\psi} = [\mu, \boldsymbol{\theta}, \sigma_f^2, \sigma_\varepsilon^2]$, where μ is the mean, $\boldsymbol{\theta} = [\theta_1, ..., \theta_d]'$ are parameters of the correlation function, σ_f^2 is the process variance, and σ_ε^2 is the random error variance. Thus, $\boldsymbol{\psi}$ needs to be estimated from data to generate predictions. Basically, the implementation of the proposed CGP method involves two steps: estimation of $\boldsymbol{\psi}$ and prediction of $\mathbf{f}(\mathbf{X}^*)$ given the estimate of $\boldsymbol{\psi}$. A fully Bayesian approach treats $\boldsymbol{\psi}$ as a random vector like $\mathbf{f}(\mathbf{X}^*)$ and conducts the two steps simultaneously by finding the joint posterior of $\boldsymbol{\psi}$ and $\mathbf{f}(\mathbf{X}^*)$. However, this contains challenging issues, e.g., specifying priors for components of $\boldsymbol{\psi}$ which are intrinsically correlated (Wang, 2012) and sampling of the high-dimensional posterior. In this study, we adopt the idea of empirical Bayesian methods (Robert, 2007) and use the maximum likelihood estimate (MLE) of $\boldsymbol{\psi}$. Then we predict $\mathbf{f}(\mathbf{X}^*)$ by sampling from its conditional posterior given the MLE of $\boldsymbol{\psi}$. Details of the two steps are given in this section. Some other related issues will also be discussed.

4.1 Parameter estimation

A commonly used method to find the MLE of $\boldsymbol{\psi}$ (Ranjan *et al*., 2011) is briefly described here. Defining $\delta = \sigma_\varepsilon^2 / \sigma_f^2$, the closed form of the MLEs of μ and σ_f^2 given $\boldsymbol{\theta}$ and δ are

$$\hat{\mu}(\boldsymbol{\theta}, \delta) = [\mathbf{1}_n^T (\mathbf{R}(\mathbf{X}, \mathbf{X}) + \delta \mathbf{I}_n)^{-1} \mathbf{1}_n]^{-1} \mathbf{1}_n^T (\mathbf{R}(\mathbf{X}, \mathbf{X}) + \delta \mathbf{I}_n)^{-1} \mathbf{y},$$

$$\hat{\sigma}_f^2(\boldsymbol{\theta}, \delta) = \frac{\left(\mathbf{y} - \hat{\mu}(\boldsymbol{\theta}, \delta)\mathbf{1}_n\right)^T (\mathbf{R}(\mathbf{X}, \mathbf{X}) + \delta \mathbf{I}_n)^{-1} \left(\mathbf{y} - \hat{\mu}(\boldsymbol{\theta}, \delta)\mathbf{1}_n\right)}{n},$$

that is, the MLEs of μ and σ_f^2 are functions of other parameters (i.e., $\boldsymbol{\theta}$ and δ). Consequently, the MLEs of $\boldsymbol{\theta}$ and δ can be found by minimizing the negative profile log-likelihood

$$-2\log L \propto \log|\mathbf{R}(\mathbf{X},\mathbf{X})+\delta\mathbf{I}_n| + n\log[(\mathbf{y}-\mathbf{1}_n\hat{\mu}(\boldsymbol{\theta},\delta))^T(\mathbf{R}(\mathbf{X},\mathbf{X})+\delta\mathbf{I})^{-1}](\mathbf{y}-\mathbf{1}_n\hat{\mu}(\boldsymbol{\theta},\delta)),$$

where $|\cdot|$ is the determinant of a matrix. To address possible issues with bumpy likelihood surface near boundaries of the parameter space, a new parameterization for $\boldsymbol{\theta}$ (Butler *et al.*, 2014) can be used

$$\omega_w = \log_{10}(\theta_w) \quad w = 1,...,d$$

With this parameterization, peaks and dips of the likelihood surface will appear in the middle of the parameter space to facilitate a thorough search for MLEs of $\boldsymbol{\theta}$ and δ.

4.2 Posterior sampling

Sampling from the posteriors given in Propositions 1-3 is challenging due to their high dimension and need for truncation. We propose the following strategy for the posterior sampling which integrates built-in random generators in software (e.g., MATLAB) and Markov chain Monte Carlo (MCMC) algorithms (Robert and Casella, 2004).

- The posterior of y_n in Eq. (7), which is a truncated univariate normal distribution, can be sampled simply by drawing from the normal distribution and discarding samples that do not satisfy the constraint. Alternatively, to enhance the efficiency of sampling, we can use popular MCMC algorithms which directly draw from nonstandard distributions such as truncated normal. One good choice is the slice sampler (Neal, 2003), which is both powerful and convenient to use as it only needs the posterior to be sampled from and a set of casually picked initial values as inputs.
- The posterior of $\mathbf{f}'(\mathbf{X}^{\Delta})$ in Eq. (11), a truncated multivariate normal distribution, is more complex to sample. In this case, generating all elements of $\mathbf{f}'(\mathbf{X}^{\Delta})$ simultaneously from the distribution has many issues including the inefficiency in multivariate truncation. A better method is the Gibbs sampler which generates each element of $\mathbf{f}'(\mathbf{X}^{\Delta})$, i.e., $f''(\mathbf{x}_1^{\Delta}),...,f'^{z}(\mathbf{x}_m^{\Delta})$, separately from its conditional posterior given other elements (Gelfand *et al.*, 1992). This method will be used in this study for all samplings from truncated multivariate normal.
- The posteriors of $\mathbf{f}(\mathbf{X}^*)$ in Eqs. (8) and (12) are multivariate normal distributions, which can be sampled using software. The posterior of $\mathbf{f}(\mathbf{X}^*)$ in Eq. (15) is a truncated multivariate distribution, which will be sampled using the abovementioned Gibbs sampler. Note that in calculating the variance matrix $\mathbf{V}(\mathbf{X}^*)$ in these equations, a simplification based on the Sherman–Morrison–Woodbury

formula, i.e., $\mathbf{A}^{-1} = \mathbf{B}_2 - \mathbf{B}_2\mathbf{A}_1^T(\mathbf{B}_1 + \mathbf{A}_1\mathbf{B}_2\mathbf{A}_1^T)^{-1}\mathbf{A}_1\mathbf{B}_2$, can be used for faster and more stable computation.

There is also a short note on the order of sampling when multiple types of constraints are imposed: whenever the censoring constraint is considered, the sampling of y_n in Eq. (7) should be done the first; whenever the bound constraint is considered, the sampling of $\mathbf{f}(\mathbf{X}^*)$ in Eq. (15) should be done at the last. Detailed procedures of the posterior sampling in each case are summarized in Appendix III.

4.3 Identifying constrained set

When imposing monotonicity constraints, we need to specify the constrained set $\{\mathbf{x}_1^\Delta,...,\mathbf{x}_m^\Delta\}$. A straightforward method is to manually identify a set of locations where the data exhibit a violating trend, but this may miss some locations that should be constrained. Alternatively, one can use all the training and prediction locations which may include many unnecessary locations. But then problems may occur in model estimation and prediction because the inverse of the covariance matrix of GP becomes more difficult to compute as more locations and/or more nearby locations are involved. Wang (2012) proposes a rigorous procedure to decide the minimal constrained set, which is useful when large samples exist and/or locations are densely distributed. In studies of scaffold biodegradation, limited data are typically available and locations are sparsely distributed, so we provide a simplified, easy-to-implement procedure below to find a reasonable constrained set with a small number of locations (assuming predictions are required to be monotonically increasing).

Step 1: Identify a set of candidate locations $\{\mathbf{x}_1^c,\mathbf{x}_2^c,...\}$ among the training locations manually. To be conservative, we can just use the whole training set as the candidate set.

Step 2: For each location $\mathbf{x}^c$ in the candidate set, since the posterior of the first-derivative at this location is a normal distribution, i.e., $f'(\mathbf{s}^c)\,|\,\mathbf{y},\boldsymbol{\psi} \sim N(m(\mathbf{x}^c),V(\mathbf{x}^c))$, according to Eq. (11), we can find the probability of negative first-derivative at this location,

$$p_{Neg} = P(f'(\mathbf{x}^c) < 0\,|\,\mathbf{y},\boldsymbol{\psi}) = \Phi\left(-\frac{m(\mathbf{x}^c)}{\sqrt{V(\mathbf{x}^c)}}\right). \quad (16)$$

Location(s) with a large p_{Neg} should be constrained.

Step 3: To make sure the predictions at given locations $\mathbf{X}^*$ satisfy the monotonicity requirement, those locations should also be constrained. This leads to a constrained set as follows

$$\mathbf{X}^{\Delta} = \underset{\mathbf{x}^c \in \{\mathbf{x}_1^c, \mathbf{x}_2^c, \ldots\}}{\arg} \left[p_{Neg}(\mathbf{x}^c) \geq 0.5 \right] \cup \{\mathbf{x}_1^*, \ldots, \mathbf{x}_{n^*}^*\}.$$

4.4 Estimating full-degradation time

In the case with full degradation censoring in Section 3.1, scaffold researchers are interested in the *full-degradation time*, which is the time point when full degradation is reached. Let the full-degradation time under the process setting that yields the censored data (i.e., $z = z_{n_z}$ in Figure 5) be t_F. An estimate of this time can be obtained as a by-product of the prediction. According to Eqs. (8), (12) and (15), for any $\mathbf{x}^* = [t, z_{n_z}]'$, $f(\mathbf{x}^*)|\cdot \sim N(m(\mathbf{x}^*), V(\mathbf{x}^*))$, where "$\cdot$" is the conditioning set in those equations. Thus, the probability of full degradation at $\mathbf{x}^*$ is

$$p_{Full}(\mathbf{x}^*) = P(f(\mathbf{x}^*) \geq 100 \,|\, \cdot) = 1 - \Phi\left(\frac{100 - m(\mathbf{x}^*)}{\sqrt{V(\mathbf{x}^*)}} \right),$$

(17)

where Φ is the cumulative distribution function of the standard normal distribution. Based on this result, we can define the estimate of the full-degradation time as

$$\hat{t}_F = \underset{t_1 < t < t_{n_t}}{\arg} \left[p_{Full}(\mathbf{x}^*) = 0.5 \right] = \underset{t_1 < t < t_{n_t}}{\arg} \left[E((f(\mathbf{x}^*)) = 100 \right].$$

(18)

Here this estimate has two interpretations: as the time point with 50% chance of full degradation, or as the time point where the expected weight loss is 100%. Since $f(\mathbf{x}^*)|\cdot$ follows a normal distribution which is symmetric, these two are equivalent. Note that the percentage "50%" in the first definition can be replaced by a higher value, e.g., 60% or 80%, depending on the concern/preference in the specific application.

5. Case Study

In this study, the proposed CGP method is applied to data from a novel tissue-engineering scaffold fabrication process (Dey *et al*., 2008; Yang *et al*., 2004). This process uses a new class of biomaterials called urethaned-doped polyester elastomers (CUPEs) to fabricate scaffolds. Unlike conventional biomaterials 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. Figure 8 displays a dataset from the process, with scaffold weight loss percentages (y) under different settings of a critical process variable (z), the percentage of *Polyethylene glycol* (PEG) in

scaffold material synthesis. Figures 8(a)-(c) contain three streams of data under z=75%, 25% and 0%, which will be used to demonstrate the proposed method in one-dimensional cases. Figure 8(d) contains data at t=3, 7, 14, 21, 28 days under z=0%, 25%, 30%, 40%, 50%, 60%, and 75%, which will be used to demonstrate the application in a two-dimensional case. Each data point in the figures is the average of five replicates.

Section 5.1 reports results of CGP prediction with each of the three types of constraints described in Sections 3.1~3.3 using the three one-dimensional data streams in Figures 8(a)-(c). It needs to remark that the one-dimensional case often occurs in practice, which is to address prediction under an interested setting of process variables. Here these cases are used to illustrate details of imposing constraints. Section 5.2 shows prediction results on the two-dimensional data in Figure 8(d). We first demonstrate the use and performance of CGP at unsampled time points and/or settings of the process variable (Section 5.2.1). Then a comparative study of CGP, GP and ANNs is conducted through leave-one-out cross validation to identify the advantages of CGP over other methods (Section 5.2.2). In the posterior sampling for each prediction, 20000 samples are generated using Matlab functions with 2000 burn-ins.

5.1 One-dimensional prediction

The data in Figure 8(a) contain a censoring measurement (at t=42 days) and thus will be used to demonstrate prediction with censoring constraint; the data in Figure 8(b) have an "abnormal" measurement (at t=14 days) which may lead to predictions violating the monotonicity requirement, and thus will be used to demonstrate prediction with monotonicity constraint; and the data in Figure 8(c) will be used to demonstrate the case with bound constraint. Procedures and results of the predictions are given as follows.

Let us first consider the case in Figure 8(a). Predictions are made at t=5, 8, 10, 15, 20, 25, 30, 32, 35, 38, 40 days using two methods: the (unconstrained) GP based on all the measurements (i.e., y_6=100%), and the proposed constrained GP with a censoring constraint at t=42 days. Figure 9(a) shows the resulting predictions. We can see that the two methods yield similar predictions during the uncensored period (3~28 days), whereas their behaviors during the censored period (28~42 days) are clearly different. Specifically, the predictions from the CGP is accurate and adequate, which implies that full degradation occurred before the last time point (i.e., 42 days). The 95% credible bands of the constrained prediction are narrow during the uncensored period and become wider after that, reflecting the increase of

uncertainty as time goes on. In contrast, the predictions from the unconstrained GP show a linear trend toward 100% in this period, which is expected since the censored measurement "100%" was used directly in this method. These results validate the importance of taking the censoring information into account for meaningful predictions.

We can also obtain the full-degradation probability, p_{Full} in Eq. (17), from the two methods for comparison. Figure 9(b) shows their estimates of the full-degradation probability at t=4, 5, …, 45 days. Again, the proposed CGP method yields reasonable results, with $p_{Full}(42)=0.86$, which is consistent with the fact (i.e., the censoring observed at this time point). Since $p_{Full}(38)=0.46$ and $p_{Full}(39)=0.57$, a simple estimate of the full-degradation time is 38.5 days based on the definition in Eq. (18). The unconstrained results appear not reasonable, where the estimated probability of full-degradation at 42 days is relatively small (around 50%), though censoring was actually observed. This suggests that imposing censoring constraint makes sense not only to prediction, but also to other inferences on scaffold degradation.

In the case of Figure 8(b), the weight loss measurement at t=14 days (y_3=24.7%) appears "abnormal", which is slightly smaller than the measurement at t=7 days (y_2=25.0%). Before applying the proposed method, the unconstrained GP is tried to see if there is a need for imposing monotonicity constraints. First, the MLEs of the GP parameters are obtained: $\hat{\mu}=25.4921$, $\hat{\theta}=0.2658$, $\hat{\sigma}_z=10.3558$, $\hat{\sigma}_\varepsilon=6.3509$. Then prediction of weight loss is made at t=5, 8, 10, 12, 15, 17, 20, 22, 25 days. The results, as shown in Figure 10(a), exhibit an increasing trend, indicating that the unconstrained GP yields reasonable predictions.

To further understand the data, we calculate the probability of negative first-derivative, i.e., p_{Neg} in Eq. (16), during the observed period (3~28 days), which is shown in Figure 10(b). All the probabilities are small (<0.5), meaning that monotonicity constraints are probably not needed. One explanation is that despite the mild outlier at t=14 days, the whole dataset has an increasing trend overall. In this case, due to the smoothing effect of the random error in the GP model, the predictions will follow the overall trend of data and will not be affected much by the outlier.

In the case of Figure 8(c), the MLEs of the GP parameters are $\hat{\mu}=5.8871$, $\hat{\theta}=0.1244$, $\hat{\sigma}_z=9.0483$, $\hat{\sigma}_\varepsilon=0.9289$. We compare predictions at t=1, 5, 8, 10, 12, 15, 17, 20, 22, 25 days using the unconstrained GP and the proposed CGP with a bound constraint "$f(t)>0$" at t=1 day. Note that t=1 day is not in the observed time range, so the prediction at this time point is an extrapolation. According to the results

shown in Figure 11, the unconstrained prediction at t=1 day is negative (–0.4%), which is not meaningful. Such problems are not unexpected for extrapolations. In contrast, the constrained prediction at this time point is positive (0.58%), and the 95% credible bands are also positive.

5.2 Two-dimensional prediction

5.2.1 Prediction using the proposed CGP

Now we apply the CGP method to the two-dimensional dataset in Figure 8(d) to demonstrate its use and performance in a general multi-dimensional case. Unlike in the one-dimensional case where the three types of constraints are imposed separately, here we will focus on the monotonicity constraints and bound constraints, and all constraints will be imposed simultaneously when needed. Also, we assume the degradation is monotonically increasing with respect to both time and the process variable according to expert knowledge. We will consider the two situations of prediction mentioned in Section 2.2, i.e., *interpolations* where prediction is made within the observed data region, and *extrapolations* where prediction is made out of the region. It is well known that the latter situation is challenging in general, and the unconstrained GP tends not to work well in this situation due to its flexibility.

We first predict weight loss at the observed time points (i.e., t=3, 7, 14, 21, 28 days) under three new settings of the process variable: z=15%, 32% and 70%, which represent the situation of interpolations. Noticing the "abnormal" data point in the dataset (t=14 days, z=30%), monotonicity is our concern here. In the first step, the MLEs of the GP parameters are found: $\hat{\mu} = 23.6849$, $\hat{\theta}_1 = 0.4451$, $\hat{\theta}_2 = 1.8604$, $\hat{\sigma}_z = 29.3117$, $\hat{\sigma}_\varepsilon = 2.3743$. Then, we check the probability of negative first-derivative with respect to time and the process variable to see if there is a need to impose constraints. The results are shown in Figures 12(a)-(b) respectively. In Figure 12(a), the probability is close to zero everywhere within the data region except the small neighborhood around the abnormal data point; but even in that neighborhood, the peak is lower than 0.2. This means that the predictions will probably satisfy the monotonicity with respect to t automatically and constraints are not needed. The probability of negative first-derivative in Figure 12(b) shows similar patterns, which is near zero in most part of the data region, with small values in the margins, meaning that monotonicity constraints with respect to z are not needed either. Therefore, the unconstrained GP is used for the prediction. The results are shown in Figure 12(c), which are all meaningful, as predicted by the probabilities of negative first-derivative in Figures 12(a)-(b). Especially, it seems that the predictions under z=32% are not affected by the abnormal data point under z=30%.

For extrapolations, two cases are considered. In the first case, we focus on predicting weight loss at some unobserved time points, t=30, 35, 40, 42 days under z=75%. According to Figure 12(a), the probability of negative first-derivative at the last observed time point (i.e., t=28 days) is rather high, especially under z=75%. So monotonicity constraints with respect to time are imposed at all the time points to predict. The constrained GP predictions are given in Figure 13(a), and the corresponding unconstrained predictions are also given in the figure for comparison. Clearly, the constraints are necessary to produce meaningful predictions. It deserves to point out that the predicted weight loss at t=42 days under z=75% is 100%, i.e., full degradation has reached, which is consistent with the actual measurement under this setting of z as shown in Figure 8(a). In the second case, we predict weight loss under two unobserved settings of the process variable, z=80% and 90%. As implied by Figure 12(b), the probability of negative first-derivative tends to getting high around z=75%, so monotonicity constraints with respect to z are imposed at the locations to predict. The results of constrained GP and unconstrained GP are given in Figure 13(b). The unconstrained predictions are, again, not meaningful: values under z=90% are lower than those under z=80% at some time points, and they are similar to the observations under z=75%. These results suggest that the constrained GP method is particularly useful in extrapolations to enable meaningful predictions.

5.2.2 Comparative study

To further understand the advantages of the CGP method, we compare its prediction performance with the two popular surrogate models mentioned in the Introduction, i.e., the unconstrained GP and ANNs, using data in Figure 8(d). Specifically, two powerful ANN methods are considered, the feed-forward neural network (FNN) and the radial basis neural network (RNN). There are two key parameters of the ANN methods: the number of neurons (#neurons) in FNN, and the upper bound of mean squared error in training (trainMSE) for RNN. In this study, these parameters are tuned by considering different settings of them (#neurons=3~20, trainMSE=8~40), and the settings that lead to the best prediction performance are chosen.

To assess the prediction performance of the methods, we adopt the leave-one-out cross validation with respect to values of the process variable (i.e., z=0%, 25%, 30%, 40%, 50%, 60%, 75%). Specifically, for each value of z, data under other values are used for model training, and data under this value are used for weight loss prediction at the five observed time points (i.e., t=3, 7, 14, 21, 28 days). Note that the predictions under z=0% and z=75% are *extrapolations*, while those under other values of z are

interpolations. Like in Section 5.2.1, performance of the four methods in these two situations is discussed below.

The predictions in interpolations (i.e., z=25%, 30%, 40%, 50% and 60%) are shown in Figure 14, and the corresponding root mean squared prediction errors (RMSPEs) are summarized in Table 1. In these cases, the predictions from the unconstrained GP are meaningful (i.e., monotonically increasing with time), so constraints are not needed; in other words, the CGP will produce the same results as the GP, as shown in Figure 14. In terms of prediction accuracy, the GP and CGP methods perform quite well: better than FNN and RNN in the cases of z = 20%, 40%, 60%, as shown in Table 1. In terms of prediction meaningfulness, both GP and CGP give meaningful predictions, while the FNN predictions violate monotonicity in three cases (z=0%, 30%, 40%), and the RNN predictions violate in two cases (z=0%, 75%). In summary, as we have seen in Section 5.2.1, both GP and CGP give promising prediction performance in interpolations.

The prediction results in extrapolations (i.e., z=0%, 75%) are given in Figure 15 and the corresponding RMSPEs are summarized in Table 1. One can see that constraints are not active for the CGP method under z=75%, leading to the same performance for both GP and CGP. The results under z=0% are interesting, where the four methods perform dramatically differently. In fact, prediction in this case is very challenging due to the lack of training data around z=0% (the closest are those under z=25%). The proposed CGP used in this case imposes bound constraints "$f(t)>0$" and monotonicity constraints with respect to time at all the locations to predict. In terms of prediction accuracy, the CGP gives the best prediction accuracy and substantially outperforms the GP, FNN and RNN methods. In terms of prediction meaningfulness, only the CGP predictions are always meaningful; GP gives negative prediction values at the beginning time points, while the two neural networks produce negative prediction at some time points and not monotonically increasing. These results suggest, again, that the proposed CGP is especially useful in extrapolations to provide meaningful predictions. Moreover, it can also improve the prediction accuracy.

6. Numerical Study

To further reveal properties of the proposed CGP method, we conduct two simulation studies by generating new datasets through slightly modifying the original datasets used in Section 5. Since the

CGP's advantages on extrapolations have been well demonstrated in the case study, here we will create cases of interpolations where the monotonicity is violated and imposing constraints are necessary.

6.1 One-dimensional example

In the one-dimensional data in Figure 8(b), we have noted the abnormal data point at t=14 days, and found that since it does not affect the overall monotonic trend of data, constraints are not needed, as shown in Figure 10. Now we replace this data point with a smaller value (e.g., 10%), while keeping other data points as their original values. The new dataset is shown in Figure 16(a), where the modified data point appears to be a serious outlier and has decisive influence on the overall trend of data. As a result, the probability of negative first-derivative given in Figure 16(b) becomes very different from that in Figure 10(b), with large values at some time points, e.g., t=11, 12 days, indicating that monotonicity constraints must be imposed. This is validated by the unconstrained GP predictions given in Figure 16(a), which exhibits a dramatically downward trend around the modified data point.

To apply the constrained GP method, a constrained set is first identified following the procedure in Section 4.3. The set is found to be {3, 5, 8, 9, 10, 11, 12, 13, 14, 15, 17, 20, 21, 22, 23, 24, 25, 28} days, at each of which monotonicity constraint is imposed. The resulting predictions are shown in Figure 16(a), which have an increasing trend and seem not to be affected much by the outlier. Another interesting result is that the 95% credible bands of the constrained GP is much narrower than the 95% confidence bands of the unconstrained GP, which indicates a lower level of uncertainty. These results suggest that the CGP will be useful in interpolations when the data set contains influential outliers.

6.2 Two-dimensional example

The two-dimensional data in Figure 8(d) are similarly modified by replacing the third data point (at t=14 days) under each setting of z with a value that is a little smaller than the second data point (at t= 7 days). Figure 17(a) shows the new dataset, where the modified data points substantially changed the overall trend of data. In Figure 17(b), the probability of negative first-derivative with respect to time now looks very different from the one in Figure 12(a), with a high peak (around 0.9) in the neighborhood around the third data points across all values of z. Correspondingly, unlike in Figure 12(c), the unconstrained predictions under z=15%, 32% and 70% violate the monotonicity. In contrast, when monotonicity constraints on related locations are imposed, the predictions become monotonic as shown in Figure 17(a). This validates what we found in the one-dimensional examples, i.e., the CGP is needed to produce meaningful predictions when substantial outliers exist.

A simulation following the similar strategy of data generation is also conducted to find the advantage of the CGP on prediction accuracy in cases of outliers. In the simulation, the data point at t=14 days is set to be proportional to that at t=7 days under the same setting of z, i.e., $y_3/y_2=r$, where $0<r\leq 1$ is an outlying factor, a smaller value of which indicates more serious outlying level of the dataset. r=1, 0.95, 0.9, 0.85 are considered, and given each of these values, the root mean squared prediction errors of GP and CGP in leave-one-out cross validation (like that reported in Section 5.2.2) are found. The results under z=25% and 60% are given in Figure 18. The benefit of CGP in prediction accuracy is clear, which increases approximately linearly as the outlying level increases (i.e., r gets smaller).

7. Discussion

Biodegradation prediction is a key problem in tissue-engineering scaffold fabrication. This study proposes a constrained Gaussian process method to solve this problem, which is able to incorporate various types of expert knowledge such as full-degradation censoring, monotonicity and bounds requirements in the prediction. According to the case study, the CGP method can yield meaningful and more accurate predictions when the regular GP fails, and it performs better than popular ANN methods. Especially, it has promising performance in extrapolations as well as interpolations with influential outliers, where prediction is usually very difficult.

Another interesting and useful finding that deserves mentioning is that the random error ε has an effect on the prediction when monotonicity constraints are imposed. Some researchers point out that including the random error in the GP model may introduce unnecessary over-smoothing, and thus make efforts to minimize over-smoothing (Ranjan *et al*., 2011). However, this is true only in contexts with intrinsically deterministic responses, such as computer experiments, and the purpose of having the random error in the model is mainly to solve computational issues related with ill-conditioned matrices in the likelihood function. Our application is a different case where the random error is used to characterize the substantial randomness contained in scaffold biodegradation measurements. In fact, the smoothing effect of the random error may be even beneficial in that it enables the predictions from the GP model to satisfy the monotonicity requirement automatically. The case study shows such examples (Figures 10 and 12) where the predictions from unconstrained GP are satisfactory and thus monotonicity constraints are not needed.

Due to the flexibility of GP and the wide existence of expert knowledge as considered in this study, the proposed CGP method can be useful in many applications. In our future research, we will extend the current methodology to handle special problems in practice. Three possible directions of study are as follows. First, in the current study, we assume there are an equal number of measurements under each setting of the process variable and there are no missing data. Given the high-demanding biodegradation measurement process in scaffold fabrication as mentioned in the Introduction, unbalanced design and/or missing data are likely to exist. We will modify this method to cover such situations. Second, the GP model used in the current work is an ordinary version of GP with a constant mean (i.e., μ in Eq. (1)). In the universal version of GP, the mean part takes a more complicated form usually as a function of the predictors. A natural question is how to incorporate those constraints in this case. One idea is that we can model the mean part using shape-constrained splines. Finally, we will also extend the current framework to impose other types of constraints in scaffold fabrication and other biomanufacturing or manufacturing applications.

Appendices

Appendix I: Proof of Proposition 1

The following is a well-known result on multivariate normal distribution that will be used in the proof: given two vectors $\mathbf{X}_1$ and $\mathbf{X}_2$ following multivariate normal distribution

$$\begin{bmatrix}\mathbf{X}_1\\ \mathbf{X}_2\end{bmatrix} \sim N\left(\begin{bmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{bmatrix}\right),$$

then $\mathbf{X}_2$ given $\mathbf{X}_1=\mathbf{x}$ follows a normal distribution with

$$\begin{aligned}\mathrm{E}(\mathbf{X}_2 \mid \mathbf{X}_1=\mathbf{x}) &= \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{x}-\boldsymbol{\mu}_1)\\ \mathrm{cov}(\mathbf{X}_2 \mid \mathbf{X}_1=\mathbf{x}) &= \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\end{aligned}.$$

(A1)

From the GP model in Eq. (1),

$$\begin{bmatrix}\mathbf{y}^{(n-1)}\\ y_n\end{bmatrix} \sim N\left(\begin{bmatrix}\mu\mathbf{1}_{n-1}\\ \mu\end{bmatrix}, \begin{bmatrix}\mathbf{K}^{00}(\mathbf{X}^{(n-1)},\mathbf{X}^{(n-1)})+\sigma_\varepsilon^2\mathbf{I}_{n-1} & \mathbf{K}^{00}(\mathbf{X}^{(n-1)},\mathbf{x}_n)\\ \mathbf{K}^{00}(\mathbf{x}_n,\mathbf{X}^{(n-1)}) & \sigma_f^2+\sigma_\varepsilon^2\end{bmatrix}\right).$$

Plugging in the terms in the above into (A1) leads to $m(\mathbf{x}_n)$ and $V(\mathbf{x}_n)$ in Eq. (7). Considering the constraint $y_n \geq 100$, the truncated normal distribution in Eq. (7) will be obtained. Similarly, according to the joint distribution of $\mathbf{f}(\mathbf{X}^*)$ and $\mathbf{y}$,

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}(\mathbf{X}^*) \end{bmatrix} \sim N\left(\begin{bmatrix} \mu \mathbf{1}_n \\ \mu \mathbf{1}_{n^*} \end{bmatrix}, \begin{bmatrix} \mathbf{K}^{00}(\mathbf{X},\mathbf{X}) + \sigma_\varepsilon^2 \mathbf{I}_n & \mathbf{K}^{00}(\mathbf{X},\mathbf{X}^*) \\ \mathbf{K}^{00}(\mathbf{X}^*,\mathbf{X}) & \mathbf{K}^{00}(\mathbf{X}^*,\mathbf{X}^*) \end{bmatrix} \right),$$

and thus $\mathbf{m}(\mathbf{X}^*)$ and $\mathbf{V}(\mathbf{X}^*)$ in Eq. (7) will be obtained.

Appendix II: Derivation of Eq. (9) and (13)

According to the definitions in Rasmussen and Williams (2006), we can find the covariance of the derivative GP and the covariance of the derivative GP and the original GP:

Case I: Monotonicity with respect to *t* (Eq. (9))

$$K^{01}(\mathbf{x}_i, \mathbf{x}_j^\Delta) = \text{cov}(f(\mathbf{x}_i), f'(\mathbf{x}_j^\Delta)) = \frac{\partial \,\text{cov}(f(\mathbf{x}_i), f(\mathbf{x}_j^\Delta))}{\partial t} = \sigma_f^2 R_{ij} \cdot 2\theta_1 (x_{i1} - x_{j1}^\Delta),$$

$$K^{10}(\mathbf{x}_j^\Delta, \mathbf{x}_i) = \text{cov}(f'(\mathbf{x}_j^\Delta), f(\mathbf{x}_i)) = \frac{\partial \,\text{cov}(f(\mathbf{x}_j^\Delta), f(\mathbf{x}_i))}{\partial t} = \sigma_f^2 R_{ij} \cdot 2\theta_1 (x_{i1} - x_{j1}^\Delta),$$

$$K^{11}(\mathbf{x}_j^\Delta, \mathbf{x}_k^\Delta) = \text{cov}(f'(\mathbf{x}_j^\Delta), f'(\mathbf{x}_k^\Delta)) = \frac{\partial^2 \,\text{cov}(f(\mathbf{x}_j^\Delta), f(\mathbf{x}_k^\Delta))}{\partial t^2} = \sigma_f^2 R_{jk} \cdot 2\theta_1 [(1 - 2\theta_1 (x_{j1}^\Delta - x_{k1}^\Delta)^2].$$

Case II: Monotonicity with respect to both *t* and *z* (Eq. (13))

$$K^{0t}(\mathbf{x}_i, \mathbf{x}_j^\Delta) = K^{t0}(\mathbf{x}_j^\Delta, \mathbf{x}_i) = \text{cov}\left[f(\mathbf{x}_i), \frac{\partial f(\mathbf{x}_j^\Delta)}{\partial t} \right] = \frac{\partial \,\text{cov}[f(\mathbf{x}_i), f(\mathbf{x}_j^\Delta)]}{\partial t} = \frac{\partial [\sigma_f^2 e^{-\theta_1 (x_{i1} - x_{j1}^\Delta)^2 - \theta_2 (x_{i2} - x_{j2}^\Delta)^2}]}{\partial t}$$
$$= 2\sigma_f^2 R_{ij} \theta_1 (x_{i1} - x_{j1}^\Delta),$$

$$K^{0z}(\mathbf{x}_i, \mathbf{x}_j^\Delta) = K^{z0}(\mathbf{x}_j^\Delta, \mathbf{x}_i) = \text{cov}\left[f(\mathbf{x}_i), \frac{\partial f(\mathbf{x}_j^\Delta)}{\partial z} \right] = \frac{\partial \,\text{cov}[f(\mathbf{x}_i), f(\mathbf{x}_j^\Delta)]}{\partial z} = \frac{\partial [\sigma_f^2 e^{-\theta_1 (x_{i1} - x_{j1}^\Delta)^2 - \theta_2 (x_{i2} - x_{j2}^\Delta)^2}]}{\partial z}$$
$$= 2\sigma_f^2 R_{ij} \theta_2 (x_{i2} - x_{j2}^\Delta),$$

$$K^{tt}(\mathbf{x}_j^\Delta, \mathbf{x}_k^\Delta) = \text{cov}\left[\frac{\partial f(\mathbf{x}_j^\Delta)}{\partial t}, \frac{\partial f(\mathbf{x}_k^\Delta)}{\partial t} \right] = \frac{\partial^2 \,\text{cov}[f(\mathbf{x}_j^\Delta), f(\mathbf{x}_k^\Delta)]}{\partial t^2} = \frac{\partial^2 [\sigma_f^2 e^{-\theta_1 (x_{j1}^\Delta - x_{k1}^\Delta)^2 - \theta_2 (x_{j2}^\Delta - x_{k2}^\Delta)^2}]}{\partial t^2}$$
$$= 2\sigma_f^2 R_{jk} \theta_1 [1 - 2\theta_1 (x_{j1}^\Delta - x_{k1}^\Delta)^2],$$

$$K^{tz}(\mathbf{x}_j^\Delta, \mathbf{x}_k^\Delta) = \text{cov}\left[\frac{\partial f(\mathbf{x}_j^\Delta)}{\partial t}, \frac{\partial f(\mathbf{x}_k^\Delta)}{\partial z} \right] = \frac{\partial^2 \,\text{cov}[f(\mathbf{x}_j^\Delta), f(\mathbf{x}_k^\Delta)]}{\partial t \partial z} = \frac{\partial^2 [\sigma_f^2 e^{-\theta_1 (x_{j1}^\Delta - x_{k1}^\Delta)^2 - \theta_2 (x_{j2}^\Delta - x_{k2}^\Delta)^2}]}{\partial t \partial z}$$
$$= -4\sigma_f^2 R_{jk} \theta_1 \theta_2 (x_{j1}^\Delta - x_{k1}^\Delta)(x_{j2}^\Delta - x_{k2}^\Delta),$$

$$K^{zt}(\mathbf{x}_j^\Delta, \mathbf{x}_k^\Delta) = \text{cov}\left[\frac{\partial f(\mathbf{x}_j^\Delta)}{\partial z}, \frac{\partial f(\mathbf{x}_k^\Delta)}{\partial t} \right] = \frac{\partial^2 \,\text{cov}[f(\mathbf{x}_j^\Delta), f(\mathbf{x}_k^\Delta)]}{\partial z \partial t} = \frac{\partial^2 [\sigma_f^2 e^{-\theta_1 (x_{j1}^\Delta - x_{k1}^\Delta)^2 - \theta_2 (x_{j2}^\Delta - x_{k2}^\Delta)^2}]}{\partial z \partial t}$$
$$= -4\sigma_f^2 R_{jk} \theta_1 \theta_2 (x_{j1}^\Delta - x_{k1}^\Delta)(x_{j2}^\Delta - x_{k2}^\Delta),$$

$$K^{zz}(\mathbf{x}_j^\Delta, \mathbf{x}_k^\Delta) = \text{cov}\left[\frac{\partial f(\mathbf{x}_j^\Delta)}{\partial z}, \frac{\partial f(\mathbf{x}_k^\Delta)}{\partial z} \right] = \frac{\partial^2 \,\text{cov}[f(\mathbf{x}_j^\Delta), f(\mathbf{x}_k^\Delta)]}{\partial z^2} = \frac{\partial^2 [\sigma_f^2 e^{-\theta_1 (x_{j1}^\Delta - x_{k1}^\Delta)^2 - \theta_2 (x_{j2}^\Delta - x_{k2}^\Delta)^2}]}{\partial z^2}$$

$$= 2\sigma_f^2 R_{jk}\theta_2[1-2\theta_2(x_{j2}^{\Delta}-x_{k2}^{\Delta})^2].$$

Appendix III: Sampling procedures for the CGP method described in Sections 3.1~3.3

Section 3.1: prediction with censoring constraint

Step 1: Find MLEs of the GP parameters $\hat{\boldsymbol{\psi}}$ using the method described in Section 3.4.1.

Step 2: Draw $y_n^{(b)}$ from Eq. (7). If $y_n^{(b)}$<100, redraw the sample until the constraint is satisfied.

Step 3: Given $y_n^{(b)}$, draw $\mathbf{f}(\mathbf{X}^*)^{(b)}$ from Eq. (8).

Repeat Steps 2 and 3 to obtain a stream of the posterior samples $\{\mathbf{f}(\mathbf{X}^*)^{(b)} : b = 1, 2, ...\}$.

Section 3.2: prediction with monotonicity constraint

The following is the conditional distribution of each variable in a multivariate normal distribution, which is the basis for sampling from a truncated multivariate normal distribution.

Let $\mathbf{X} = [X_1, ..., X_m]' \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\mathbf{X}_{-j}$ is $\mathbf{X}$ excluding X_j, j=1,…,m, then

$$\begin{bmatrix} \mathbf{X}_{-j} \\ X_j \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\mu}_{-j} \\ \mu_j \end{bmatrix}, \begin{bmatrix} \mathrm{V}(\mathbf{X}_{-j}) & \mathrm{cov}(\mathbf{X}_{-j}, X_j) \\ \mathrm{cov}(X_j, \mathbf{X}_{-j}) & \mathrm{V}(X_j) \end{bmatrix} \right).$$

By (A1), the conditional distribution of X_j is

$$X_j \mid \mathbf{X}_{-j} \sim N(m_j, v_j),$$

where

$$m_j = \mu_j + \mathrm{cov}(X_j, \mathbf{X}_{-j})[\mathrm{V}(\mathbf{X}_{-j})]^{-1}(\mathbf{X}_{-j} - \boldsymbol{\mu}_{-j}),$$
$$v_j = V(X_j) - \mathrm{cov}(X_j, \mathbf{X}_{-j})[V(\mathbf{X}_{-j})]^{-1}\mathrm{cov}(\mathbf{X}_{-j}, X_j).$$

Step 1: Find MLEs of the GP parameters $\hat{\boldsymbol{\psi}}$ using the method described in Section 3.4.1.

Step 2: Draw $\mathbf{f}'(\mathbf{X}^{\Delta})^{(b)}$ from Eq. (11): for j=1,…,m, draw a sample from

$$f'(\mathbf{x}_j^{\Delta})^{(b)} \mid \mathbf{f}'(\mathbf{X}_{-j}^{\Delta})^{(b-1)}, \mathbf{y}, \hat{\boldsymbol{\psi}} \sim N(\mu_j^{(b-1)}, v_j) \text{ until } f'(\mathbf{x}_j^{\Delta})^{(b)} > 0,$$

where $\mathbf{X}_{-j}^{\Delta}$ is $\mathbf{X}^{\Delta}$ excluding $\mathbf{x}_j^{\Delta}$, $\mathbf{f}'(\mathbf{X}_{-j}^{\Delta})^{(b-1)} = [f'(\mathbf{x}_1^{\Delta})^{(b)}, ..., f'(\mathbf{x}_{j-1}^{\Delta})^{(b)}, f'(\mathbf{x}_{j+1}^{\Delta})^{(b-1)}, ..., f'(\mathbf{x}_m^{\Delta})^{(b-1)}]'$, and the parameters of the univariate normal distribution are

$$\mu_j^{(b-1)} = m(\mathbf{x}_j^{\Delta}) + \Omega(\mathbf{x}_j^{\Delta}, \mathbf{X}_{-j}^{\Delta})\Omega^{-1}(\mathbf{X}_{-j}^{\Delta}, \mathbf{X}_{-j}^{\Delta})(\mathbf{f}'(\mathbf{X}_{-j}^{\Delta})^{(b-1)} - \mathbf{m}(\mathbf{X}_{-j}^{\Delta})),$$
$$v_j = \Omega(\mathbf{x}_j^{\Delta}, \mathbf{x}_j^{\Delta}) - \Omega(\mathbf{x}_j^{\Delta}, \mathbf{X}_{-j}^{\Delta})\Omega^{-1}(\mathbf{X}_{-j}^{\Delta}, \mathbf{X}_{-j}^{\Delta})\Omega(\mathbf{X}_{-j}^{\Delta}, \mathbf{x}_j^{\Delta}),$$

where $m(\mathbf{x}_j^{\Delta})$ is the j^{th} element of $\mathbf{m}(\mathbf{X}^{\Delta})$ in Eq. (11), $\mathbf{m}(\mathbf{X}_{-j}^{\Delta})$ is $\mathbf{m}(\mathbf{x}^{\Delta})$ excluding the j^{th} element, $\Omega(\mathbf{x}_j^{\Delta}, \mathbf{x}_j^{\Delta})$ is the (j, j)$^{\text{th}}$ element of $\mathbf{V}(\mathbf{X}^{\Delta})$ in Eq. (10), $\Omega(\mathbf{x}_j^{\Delta}, \mathbf{X}_{-j}^{\Delta})$ is the j^{th} row of $\mathbf{V}(\mathbf{X}^{\Delta})$ excluding the entry from the j^{th} column, $\Omega(\mathbf{X}_{-j}^{\Delta}, \mathbf{X}_{-j}^{\Delta})$ is $\mathbf{V}(\mathbf{X}^{\Delta})$ excluding the j^{th} row and j^{th} column, and $\Omega(\mathbf{X}_{-j}^{\Delta}, \mathbf{x}_j^{\Delta})$ is the j^{th} column of $\mathbf{V}(\mathbf{X}^{\Delta})$ excluding the entry from the j^{th} row.

Step 3: Given $\mathbf{f}'(\mathbf{X}^{\Delta})^{(b)}$, draw $\mathbf{f}(\mathbf{X}^*)^{(b)}$ from Eq. (12).

Repeat Steps 2 and 3 to obtain a stream of posterior samples $\{\mathbf{f}(\mathbf{X}^*)^{(b)} : b = 1, 2, ...\}$.

Section 3.3: prediction with bound constraint

The sampling involves drawing from a truncated multivariate normal distribution (Eq. (15)). This will be done following the method in Step 2 of Case II.

Step 1: Find MLEs of the GP parameters $\hat{\boldsymbol{\psi}}$ using the method described in Section 3.4.1.

Step 2: Draw $\mathbf{f}(\mathbf{X}^*)^{(b)}$ from Eq. (15) following Step 2 in Case II, except that the generated samples are screened by the bound constraints $f(\mathbf{x}_1^*) \in U_1, ..., f(\mathbf{x}_{n^*}^*) \in U_{n^*}$.

Repeat Step 2 to obtain a stream of posterior samples $\{\mathbf{f}(\mathbf{X}^*)^{(b)} : b = 1, 2, ...\}$.

Acknowledgement

Li Zeng gratefully acknowledges financial support from the National Science Foundation under grant CMMI-1649009.

References

Adler, R. J. (1981) *The Geometry of Random Fields*, John Wiley & Sons, New York, NY.

Arendt, P. D., Apley, D. W., and Chen, W. (2015) "A Preposterior Analysis to Predict Identifiability in the Experimental Calibration of Computer Models", *IIE Transactions*, **48**(1):75-88.

Bian, L., and Gebraeel, N. (2014) "Stochastic Modeling and Real-time Prognostics for Multi-component Systems with Degradation Rate Interactions", *IIE Transactions*, **46**:470-482.

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

Burdick, J. A., and Mauck, R. L. (ed.) (2011) *Biomaterials for Tissue Engineering Applications – A Review of the Past and Future Trends*, Springer-Verlag/Wien, New York, NY.

Butler, A., Haynes, R. D., Humphries, T. D., and Ranjan, P. (2014) "Efficient Optimization of the Likelihood Function in Gaussian Process Modeling", *Computational Statistics and Data Analysis*, **73**:40-52.

Chatterjee, S., Guntuboyina, A., and Sen, B. (2015) "On Risk Bounds in Isotonic and Other Shape Restricted Regression Problems", *The Annals of Statistics*, **43**(4):1774-1800.

Chen, N., and Tsui, K.-L. (2013) "Condition Monitoring and Remaining Useful Life Prediction Using Degradation Signals: Revisited", *IIE Transactions*, **45**:939-952.

Chen, V. C. P., Tsui, K.-L., Barton, R. R., and Mechesheimer, M. (2006) "A Review on Design, Modeling and Applications of Computer Experiments", *IIE Transactions*, **38**:273-291.

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

Cui, L., Zhang, N., Cui, W., Zhang, P., and Chen, X. (2015) "A Novel Nano/Micro-Fibrous Scaffold by Melt-Spinning Method for Bone Tissue Engineering", *Journal of Bionic Engineering*, **12**(1):117-128.

Dey, J., Xu, H., Shen, J., Thevenot, P., Gondi, S. R., Nguyen, K. T., Sumerlin, B. S., Tang, L., Yang, J. (2008) "Development of Biodegradable Crosslinked Urethane-doped Polyester Elastomers", *Biomaterials*, **29**(35):4637-4649.

Fang, K. T., Li, R., and Sudjianto, A. (2005) *Design and Modeling for Computer Experiments*, Chapman&Hall/CRC Press, New York, NY.

Feng, C., and Wang, X. (2003) "Surface Roughness Predictive Modeling: Neural Networks versus Regression", *IIE Transactions*, **35**:11-27.

Feng, C., and Wang, X. (2004) "Data Mining Techniques Applied to Predictive Modeling of the Knurling Process", *IIE Transactions*, **36**:253-263.

Feng, C., Yu, Z., and Kusiak, A. (2006) "Selection and Validation of Predictive Regression and Neural Network Models based on Designed Experiments", *IIE Transactions*, **38**(1):13-23.

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

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.

Grant, H., and Settles, S. (2009) "A Survey of Issues in Biomanufacturing Research", *IIE Transactions*, **41**(6):537-545.

Henry, J. A., Simonet, M., Pandit, A., and Neuenschwander, P. (2007) "Characterization of a Slowly Degrading Biodegradable Polyesterurethane for Tissue Engineering Scaffolds", *Journal of Biomedical Materials Research Part A*, **82**(3):669-679.

Jin, R., Chang, C.-J., and Shi, J. (2012) "Sequential Measurement Strategy for Water Geometric Profile Estimation", *IIE Transactions*, **44**:1-12.

Lavine, M., and Mockus, A. (1995) "A Nonparametric Bayes Method for Isotonic Regression", *Journal of Statistical Planning and Inference*, **46**:235-248.

Lenk, P. J., and Choi, T. (2017) "Bayesian Analysis of Shape-restricted Functions Using Gaussian Process Priors", *Statistica Sinica*, **27**:43-69.

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.

Lin, L., and Dunson, D. B. (2014) "Bayesian Monotone Regression Using Gaussian Process Projection", *Biometrika*, **101**(2):303-317.

Neal, R. M. (2003) "Slice Sampling", *The Annals of Statistics*, **31**(3):705-741.

Nicodemus, G. D., and Bryant, S. J. (2008) "Cell Encapsulation in Biodegradable Hydrogels for Tissue Engineering Applications", *Tissue Engineering: Part B*, **14**(2):149-165.

Pourhabib, A., Huang, J., Wang, K., Zhang, C., Wang, B., and Ding, Y. (2015) "Modulus Prediction of Buckypaper based on Multi-fidelity Analysis Involving Latent Variables", *IIE Transactions*, **47**:141-152.

Ranfiee, K., Feng, Q., and Coit, D. W. (2014) "Reliability Modeling for Dependent Competing Failure Processes with Changing Degradation Rate", *IIE Transactions*, **46**:483-496.

Ranjan, P., Haynes, R., and Karsten, R. (2011) "A Computationally Stable Approach to Gaussian Process Interpolation of Deterministic Computer Simulation Data", *Technometrics*, **53**(4):366-377.

Rasmussen, C. E., and Williams, C. K. I. (2006) *Gaussian Processes for Machine Learning*, The MIT Press, Cambridge, MA.

Riihimäki, J., and Vehtari, A. (2010) "Gaussian Processes with Monotonicity Information", *Journal of Machine Learning Research*, **9**:645-652.

Robert, C. P. (2007) *The Bayesian Choice*, second edition, Springer, New York, NY.

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

Santner, T.J., Williams, B. J., and Notz, W. I. (2003) *The Design and Analysis of Computer Experiments*, New York: Springer.

Schabenberger, O., and Gotway, C. A. (2005) *Statistical Methods for Spatial Data Analysis*, Chapman &Hall/CRC, Boca Raton, FL.

Shively, T. S., Walker, S. G., and Damien, P. (2011) "Nonparametric Function Estimation subject to Monotonicity, Convexity and Other Shape Constraints", *Journal of Econometrics*, **161**:166-181.

Sultana, N. (2013) *Biodegradable Polymer-Based Scaffolds for Bone Tissue Engineering*, Springer, New York, NY.

Tsai, C.-H., Chang, C.-J., Wang, K., Zhang, C., Liang, Z., and Wang, B. (2012) "Predictive Model for Carbon Nanotube-reinforced Nanocomposite Modulus Driven by Micromechanical Modeling and Physical Experiments", *IIE Transactions*, **44**(7):590-602.

Wang, J., and Ghosh, S. K. (2012) "Shape Restricted Nonparametric Regression with Bernstein Polynomials", *Computational Statistics and Data Analysis*, **56**(9):2729-2741.

Wang, X. (2012) *Bayesian Modeling Using Latent Structures*, PhD Thesis, Duke University.

Wang, X., and Berger, J. O. (2016) "Estimating Shape Constrained Functions Using Gaussian Processes", *Uncertainty Quantification*, **4**:1-25.

Wang, J., Wei, J., Su, S., Qiu, J., and Wang, S. (2015) "Ion-linked Double-network Hydrogel with High Toughness and Stiffness", *Journal of Materials Science*, **50**(16):5458-5465.

Wei, J., Wang, J., Su, S., Wang, S., and Qiu, J. (2015) "Tough and Fully Recoverable Hydrogels", *Journal of Materials Chemistry B*, **3**:5284-5290.

Wei, J., Wang, J., Su, S., Wang, S., Qiu, J., Zhang, Z., Christopher, G., Ning, F., and Cong, W. (2015) "3D Printing of An Extremely Tough Hydrogel", *RSC Advances*, **5**:81324-81329.

Yang, J., Webb, A. R., Ameer, G. A. (2004) "Novel Citric Acid-based Biodegradable Elastomers for Tissue Engineering", *Advanced Materials*, **16**(6):511-516.

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.

Table 1. Root mean squared prediction errors (RMSPE) of the four methods

Methods	0%	25%	30%	40%	50%	60%	75%	Average
GP	7.1	2.0	3.7	1.4	3.7	4.8	6.0	4.1
CGP	**1.9**	**2.0**	3.7	**1.4**	3.7	**4.8**	**6.0**	**3.4**
FNN	10.3	2.7	5.9	8.2	4.4	5.1	6.7	6.2
RNN	15.9	5.9	**3.6**	2.3	**3.1**	6.1	7.0	6.3

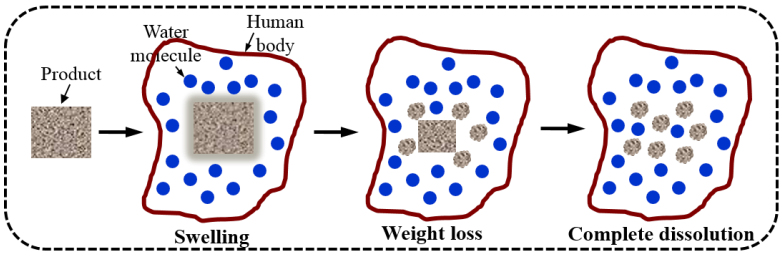


Figure 1. An illustration of biodegradation of products made by degradable biomaterials

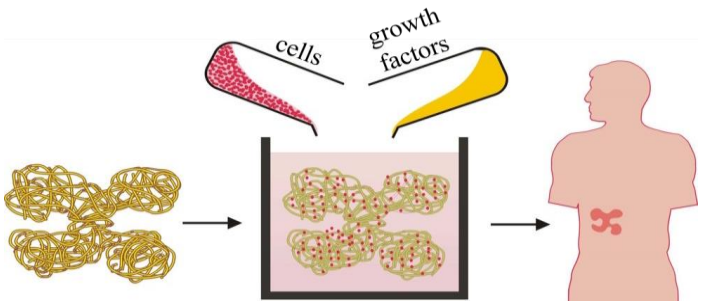


Figure 2. The development of biological substitutes for failing tissues/organs in tissue engineering

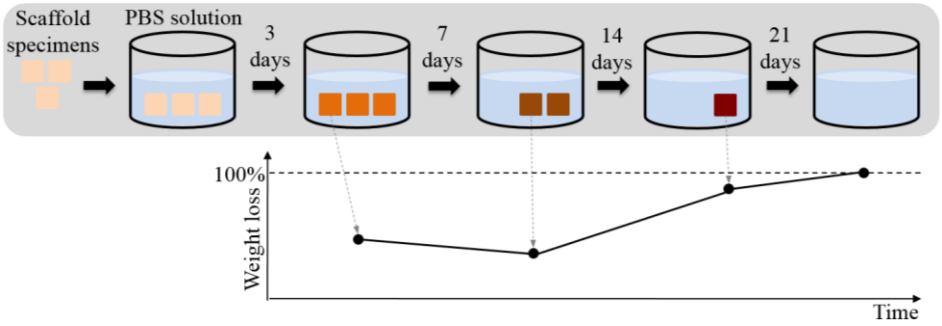


Figure 3. Experimental setup to characterize biodegradation in scaffold fabrication

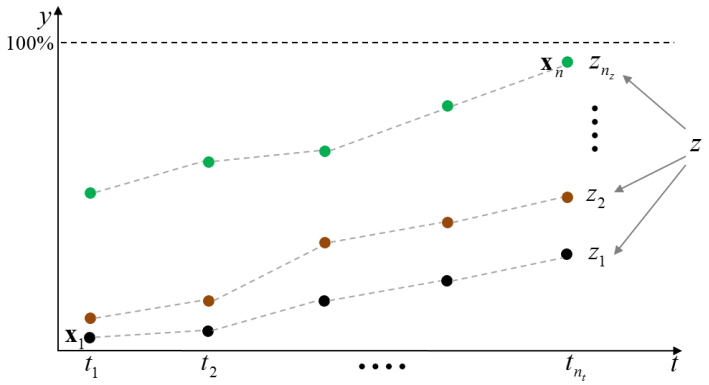


Figure 4. Illustration of the biodegradation data used in this study

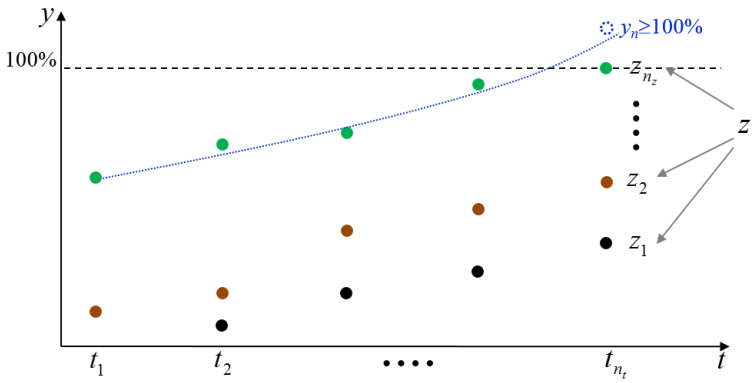


Figure 5. Illustration of the censoring constraint

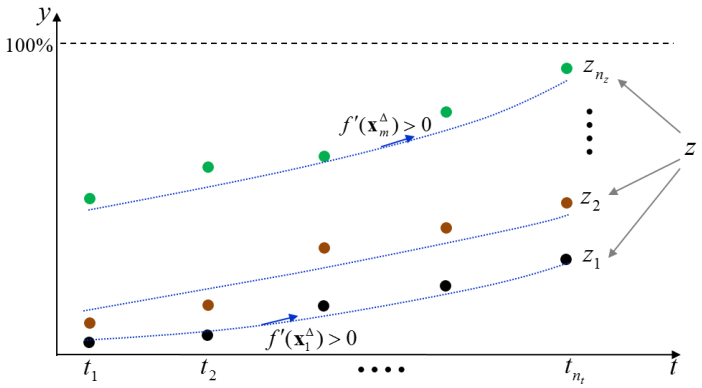


Figure 6. Illustration of monotonicity constraints with respect to time

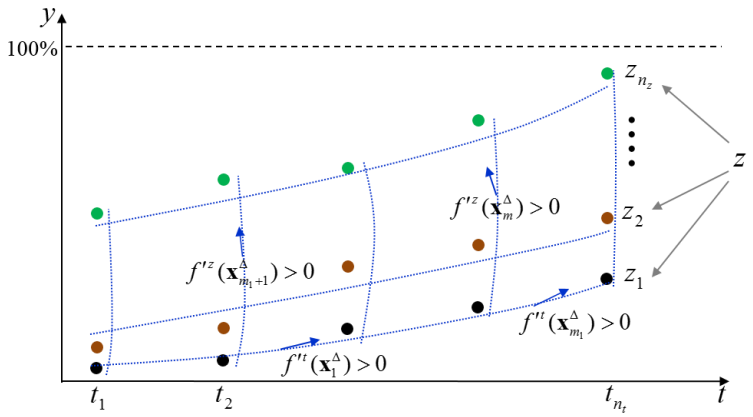


Figure 7. Illustration of monotonicity constraints with respect to both predictors

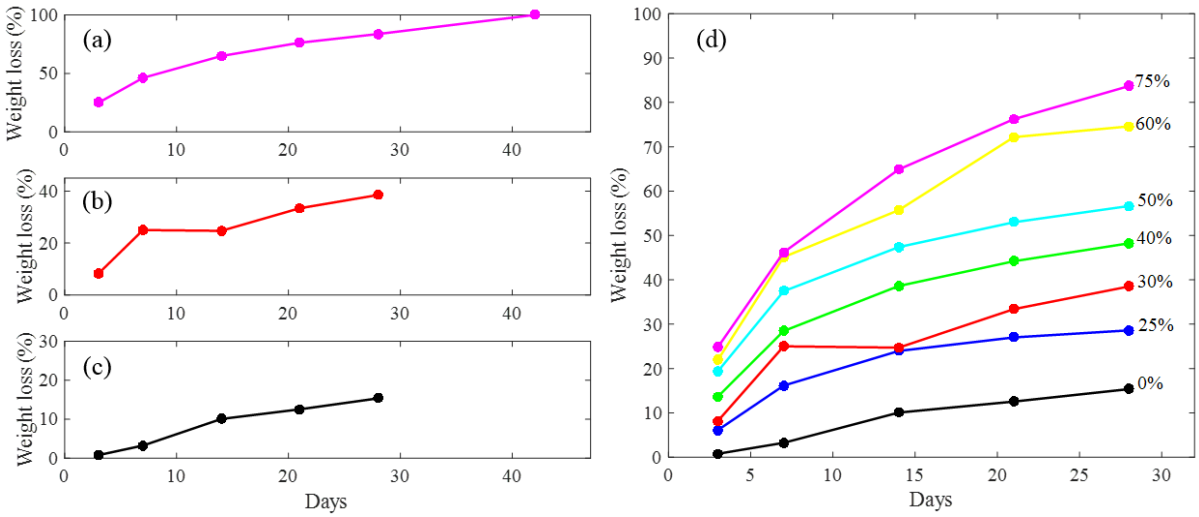
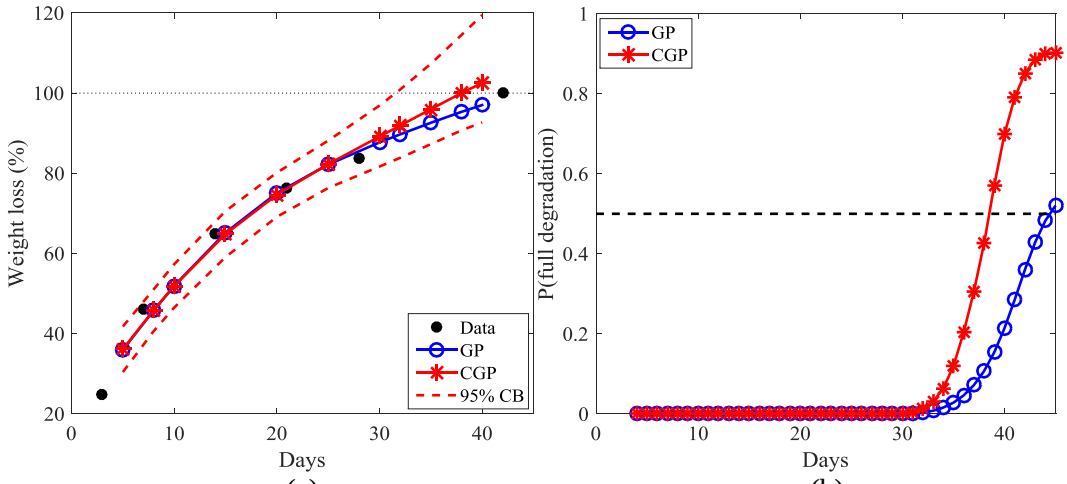
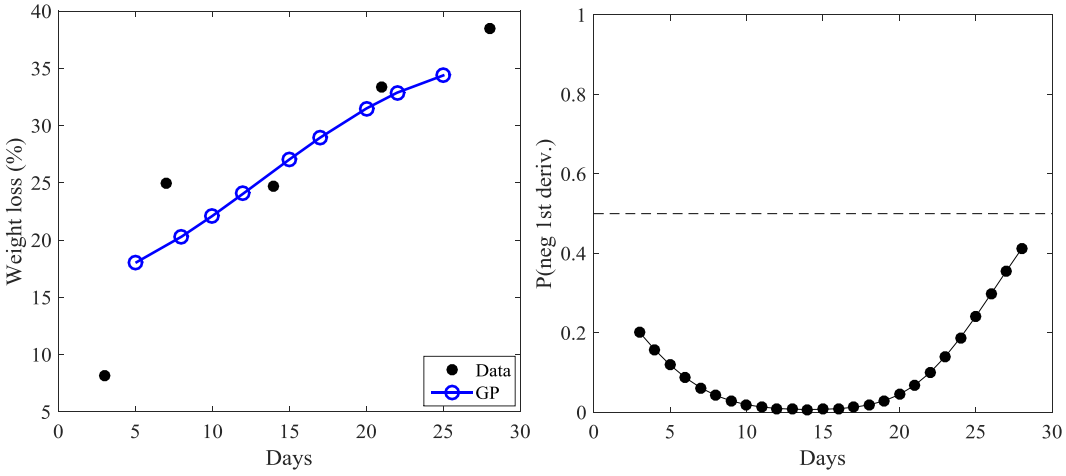


Figure 8. Degradation measurements used in the case study: (a) one-dimensional data with z=75%, (b) one-dimensional data with z=30%, (c) one-dimensional data with z=0%, and (d) two-dimensional data



(a) (b)

Figure 9. One-dimensional case in Figure 8(a): (a) predictions and (b) full-degradation probability



(a) (b)

Figure 10. One-dimensional case in Figure 8(b): (a) predictions and (b) probability of negative derivative

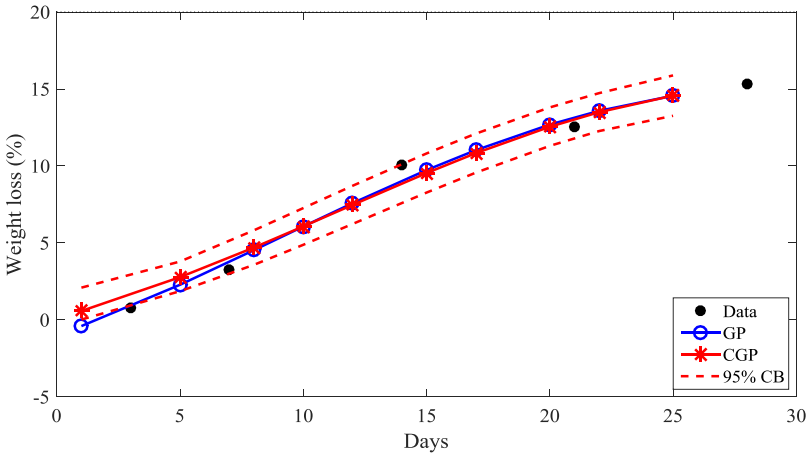


Figure 11. Prediction in one-dimensional case in Figure 8(c)

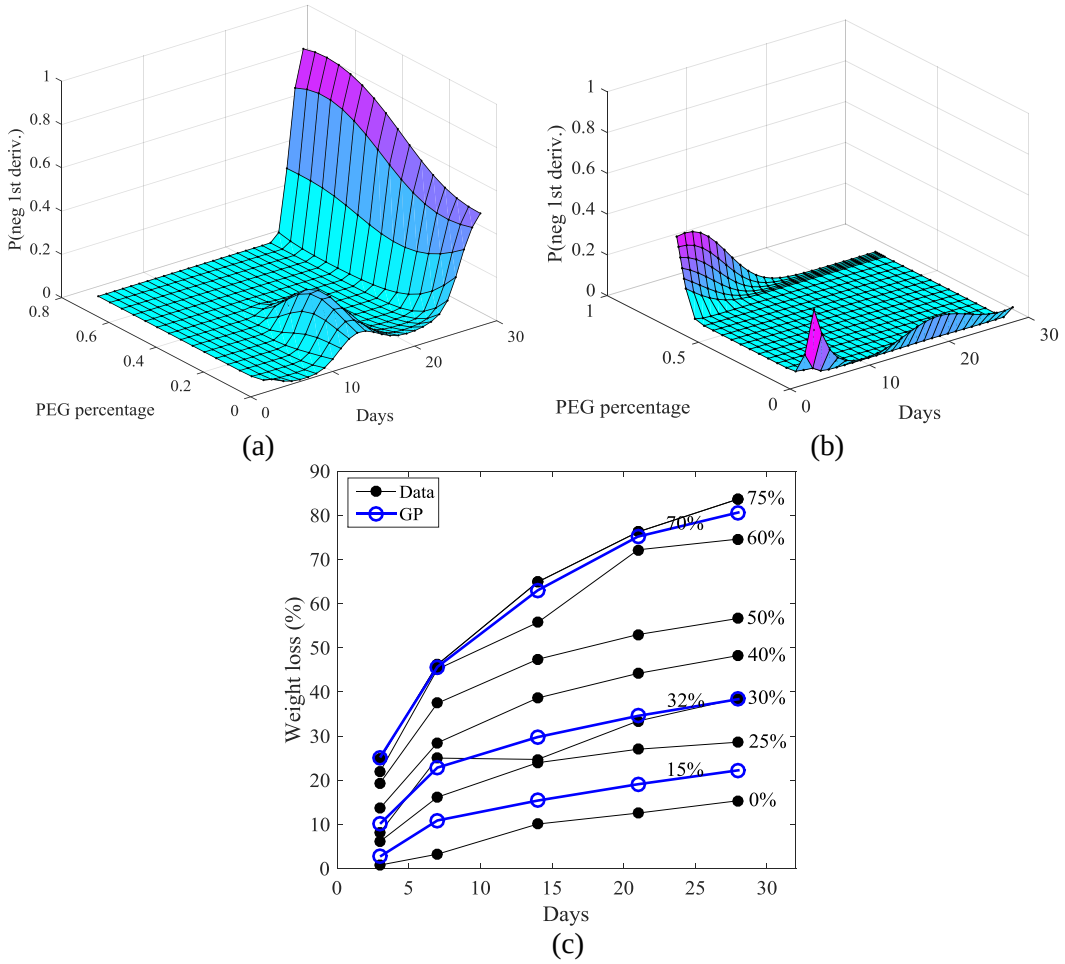


Figure 12. Prediction in the two-dimensional case in Figure 8(d): (a) probability of negative first-derivative with respect to *t*, (b) probability of negative first-derivative with respect to *z*, (c) interpolations under new settings of z = 15%, 32%, 70%

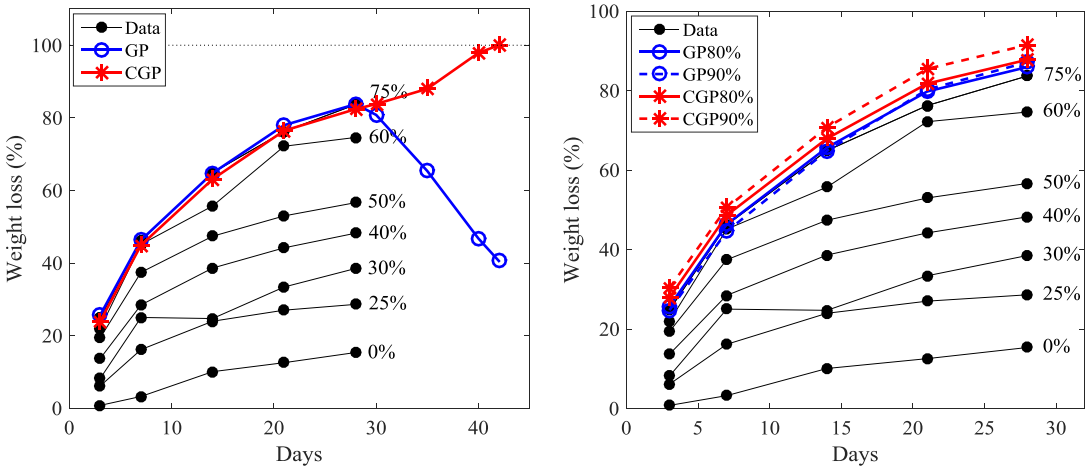


Figure 13. Prediction in the two-dimensional case in Figure 8(d): (a) extrapolations at new time points t = 30, 35, 40, 42 days; (b) extrapolations under new settings of z = 80%, 90%

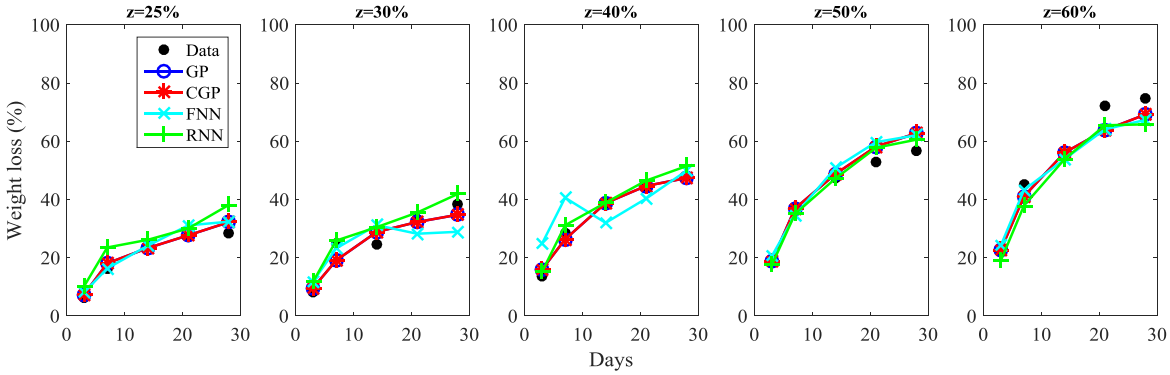


Figure 14. Leave-one-out predictions of the four methods in interpolations

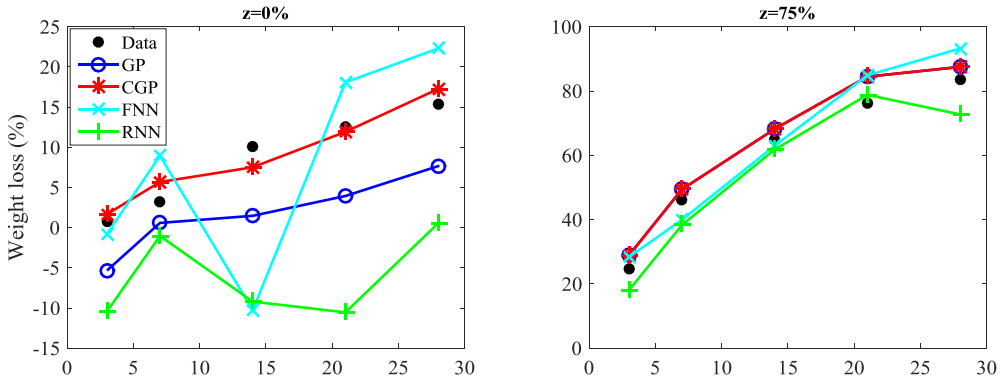
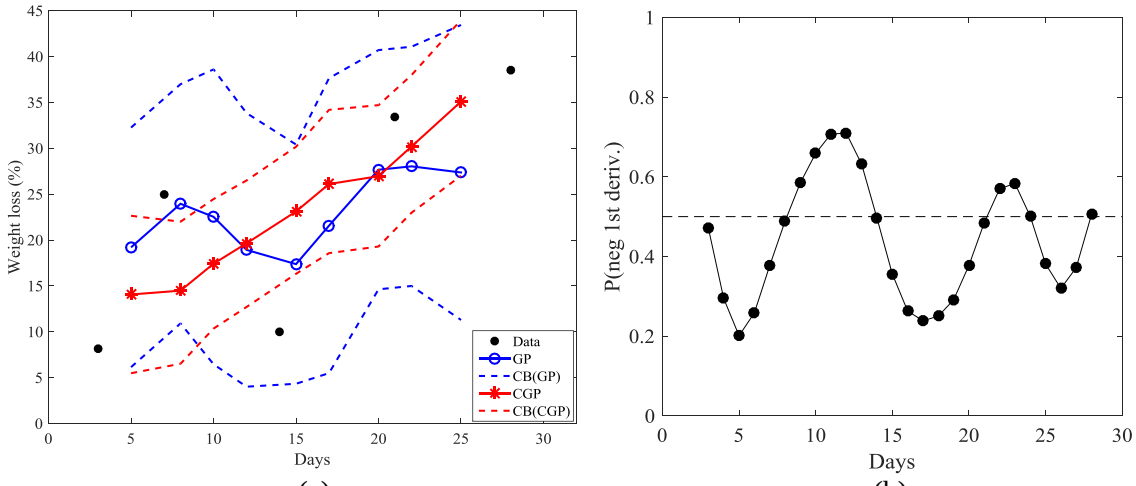
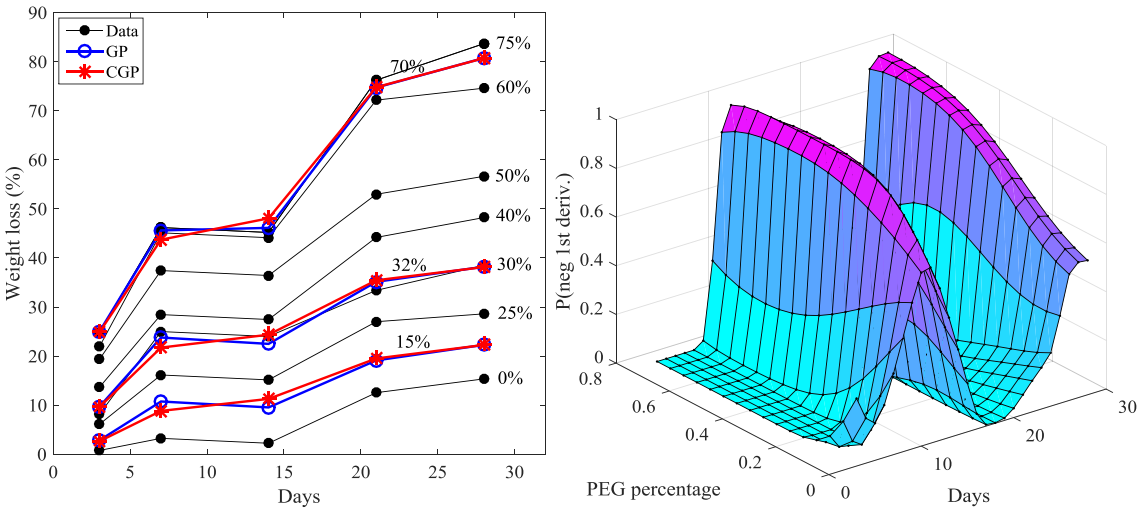


Figure 15. Leave-one-out predictions of the four methods in extrapolations



(a) (b)

Figure 16. One-dim example: (a) data and predictions, and (b) probability of negative first-derivative



(a) (b)

Figure 17. Two-dim example: (a) data and predictions, and (b) probability of negative first-derivative

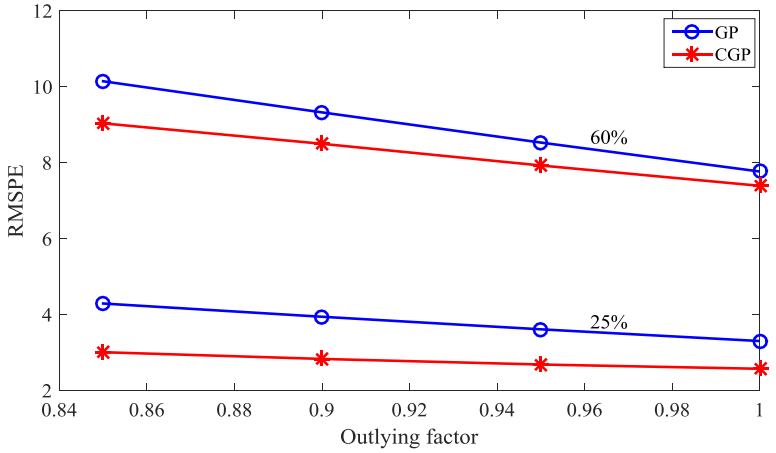


Figure 18. Prediction errors under different outlying levels of the modified dataset