MODELING OF QUALITY PROFILE DATA WITH APPLICATION IN

MANUFACTURING AND BIOMEDICAL ENGINEERING

A Dissertation

by

QIAN WU

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Li Zeng
Committee Members,	Yu Ding
	Shiren Wang
	Suhasini Subba Rao
Head of Department,	Mark Lawley

August 2019

Major Subject: Industrial Engineering

ABSTRACT

The quality of the output of a complex system is often recorded as multi-dimensional profile data with panel structure. In such structure, the quality of each individual in the output is measured repeatedly based on time or other variables. In this dissertation, the quality profile data are modeled to address two types of problems: (a) to explore the underlying relationship between the parameter of interest in the complex system and the resulting quality under the condition that the principal mechanism is not fully known and (b) to quantify the uncertainties among the output. For the first type of problem, we consider a constrained semiparametric varying coefficient model. The system parameter of interest is treated as a covariate whose effect upon the resulting quality is modeled nonparametrically as a function of time. Any existing physicochemical knowledge related to other factors in the system that affect the resulting output quality is modeled parametrically as an additive term in the model. In the situation that expert knowledge about the effect of the parameter is available, some constraints can be incorporated in the model such that the estimated effect aligns with the given knowledge. For the second type of problem, mixed-effect model is developed to quantify the uncertainties among output using random effects. These random effects can be utilized for anomaly detection or for variation quantification where deviation among individuals is of interest depending on the context of the data. Three case studies from manufacturing and biomedical engineering domains are presented in the dissertation where the above two types of problems are discussed.

DEDICATION

I dedicate my dissertation work to my family and friends. A feeling of gratitude to my loving parents who support me all the way with endless love and encouragement. With the unconditional love, understanding and trust from my mom, Yunping Hong and my dad, Yueqin Wu, I am able to pursue my goal.

This dissertation is also dedicated to my friends for being with me throughout the entire process. I will always appreciate all the patience, kindness and happiness they gave me during the doctorate program. All of them have been my cheerleaders.

ACKNOWLEDGEMENTS

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professor Li Zeng, Professor Yu Ding and Professor Shiren Wang in the Department of Industrial & Systems Engineering and Professor Suhasini Subba Rao in the Department of Statistics.

The data analyzed in Chapter II was provided by Professor Shiren Wang from his work published in 2015. The data analyzed in Chapter III was provided by Professor Hanli Liu in the Department of Bioengineering at the University of Texas at Arlington from her work published in 2017.

Funding Sources

TABLE OF CONTENTS

LIST OF FIGURES

ix

LIST OF TABLES

CHAPTER I

INTRODUCTION

I.1. General introduction to engineering system

An engineering system is usually composed of multiple components that regularly

interact with each other to yield product or service. The complexity of system increases as

the product or service becomes more integrated. To fulfill quality requirements on system

output, two critical issues need to be solved: (1) understand the relationship between

quality of output and components of the system, and (2) quantify the uncertainties among

the output. The two issues interact with each other. Large variation among the individuals

of output motivates the optimization of the system to produce satisfying product or service,

which can be enabled by a better understanding of how the quality of output is related to

components of the system. Such improvement also helps to diagnose root causes of

changes and failures in the system that result in unsatisfactory output. The detection of

such changes and failures makes it possible to adjust the system to maintain and improve

its stability, leading to the decrease of uncertainties among output products and service.

Due to the complexity of systems, it is difficult to analyze a system based on domain

knowledge only. Statistical/data analytics methods along with quality data have been

introduced and become more and more important in this field.

I.2. Quality profile data

Profile data is a new class of quality data that exist in many applications. Compared

to univariate quality characteristics that summarize the quality of a sample using one

single value, a profile measures the quality as a function between a response quality

characteristic and some explanatory variables. This allows direct observation and deeper understanding of the change in quality over factors of interest. A profile dataset typically consists of a collection of profiles, each for one individual sample. As a result, profile data reflect quality variation among samples as well as relationship between quality and relevant factors. Depending on the nature of different systems, profile data take different forms and the analysis of them varies.

I.3. Research tasks

In this dissertation, we address the above-mentioned two target problems with three studies in biomedical and manufacturing applications.

I.3.1. Modeling the relationship between parameter of interest in the system and the output quality using profile data

The problem is analyzed in Chapter II and III.

In Chapter II, we propose a method that models the relationship between the parameter of interest in the system and the resulting quality with accommodation of available expert knowledge. A semiparametric varying coefficient model is created. The parameter of interest is treated as a covariate in the model, whose effect upon the resulting quality is modeled nonparametrically as a function of time. Any existing physicochemical knowledge related to other factors in the system that affect the resulting output quality is modeled parametrically as an additive term in the model. In the situation that expert knowledge about the effect of the parameter to be explored is available, constraints are incorporated in the model such that the estimated effect aligns with the given knowledge.

In Chapter III, we consider the relationship between categorical parameters (treatment/placebo) and corresponding quality output. Hypothesis tests are first conducted to validate that the effect of such parameters is significant and varies over time. Then the system parameter effect is modeled as a parametric or nonparametric function of time.

I.3.2. Quantification of individual randomness using quality profile data

The problem is analyzed in Chapter III and IV.

The mixed-effect model is used here, where the individual quality profile is divided into the population effect and individual randomness. The former describes the portion representing the overall trend of the population, while the latter describes the uncertainty among individuals. In Chapter III, we focus on the progress of treatment quality over time and use parametric and functional mixed effect models to quantify the individual randomness. In contrast, Chapter IV aims at detecting out-of-control products in manufacturing where a nonlinear mixed effect model is developed to characterize the individual randomness. A T^2 chart is then constructed based on the estimated random effects for detection purpose.

CHAPTER II

CONSTRAINED VARYING COEFFICIENT MODEL FOR TIME-COURSE

EXPERIMENTS IN SOFT TISSUE FABRICATION

II.1. Introduction

Soft tissue injury widely occurs among elderly population due to long-term pressure combined with the decline of physiological functions. There is a substantial and increasing demand in soft tissue repair and replacement surgeries (Parker et al., 2015; World Health Organization, 2011; Ortman et al., 2014). Among various treatments of soft tissue injury, soft tissue grafts are one favored treatment in the long term. However, there is a big gap between the demand and supply of soft tissue grafts due to the shortage of donation (Weiss et al., 2017). Driven by the need to close this gap, fabrication of artificial (engineered) soft tissues has become an emerging research topic.

A key to the success of soft tissue fabrication lies in appropriate biomaterials. For given biomaterial types, the use of 3D printing allows fabrication of structures with complicated and customized shapes like human tissues for each individual patient (He et al., 2014). Figure II-1 illustrates the concept of 3D printing for fabricating soft tissue products using meniscus as an example. It starts from acquiring image of the tissue by a medical scanner such as computed tomography (CT) and magnetic resonance imaging (MRI); then the image is processed to create a computer-aided design (CAD) model, and the tissue is printed layer by layer based on the CAD model (Wei et al., 2015).
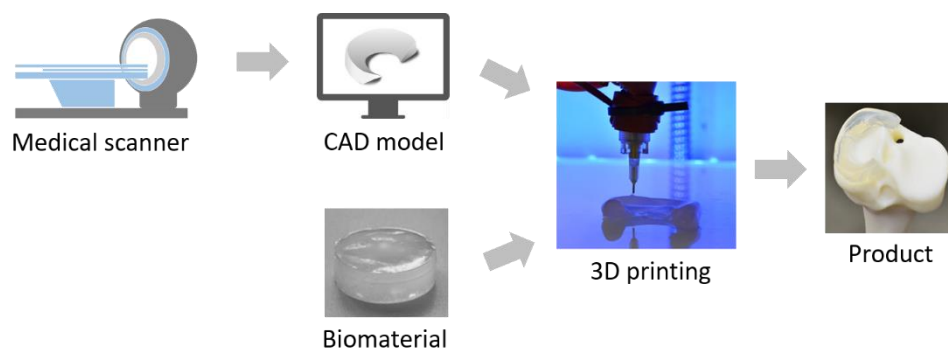
Figure II-1. An illustration of 3D printing for fabricating artificial meniscus

To serve proper functions of natural tissues, the fabricated product needs to meet certain requirements. Take artificial meniscus as an example. It must bear certain mechanical properties when exposed to shear, tension and compression forces (Bochynska et al., 2016) since meniscus acts as cushion to the knee joint to maintain its stability. Moreover, meniscus is lubricated so it is necessary to consider surface characteristics of the product (Fox et al., 2012). In addition, the biocompatibility of the product is crucial as it will be implanted to human body.

In biomaterial fabrication, satisfactory material properties depend on the setting of process parameters (e.g., percentages of ingredients, heating temperature). It is important to understand the relationship between process parameters and material properties, or, the effect of process parameters on material properties. Data-driven methods are preferred for modeling such effects since the fabrication involves a series of physicochemical mechanisms which are complex or unknown. This requires data of material properties resulted from different settings of process parameters. Also, as the effect of process parameters evolves with time, the dynamic trajectory of material properties is measured to characterize the stability of tissue products which is critical to serve their designed

functions. Such data are obtained by conducting time-course experiments illustrated in the upper panel of Figure 2, where y is the material property and x is the process parameter of interest. Within the budget of cost, m different settings of the process parameter, $x_1, \ldots x_m$, are considered. Under the ith setting, the material is fabricated and values of y at n time points, denoted as $y_i(t_j)$, $i = 1, \ldots, m$, $j = 1, \ldots, n$, are measured. The collected data form a two-way table as given in the lower panel of Figure 2. Graphically, they manifest as multiple time-course curves each from a setting of the process parameter.

Setting	t_1	t_2	...	t_n
x_1	$y_1(t_1)$	$y_1(t_2)$	...	$y_1(t_n)$
⋮	⋮	⋮	...	⋮
x_m	$y_m(t_1)$	$y_m(t_2)$	...	$y_m(t_n)$
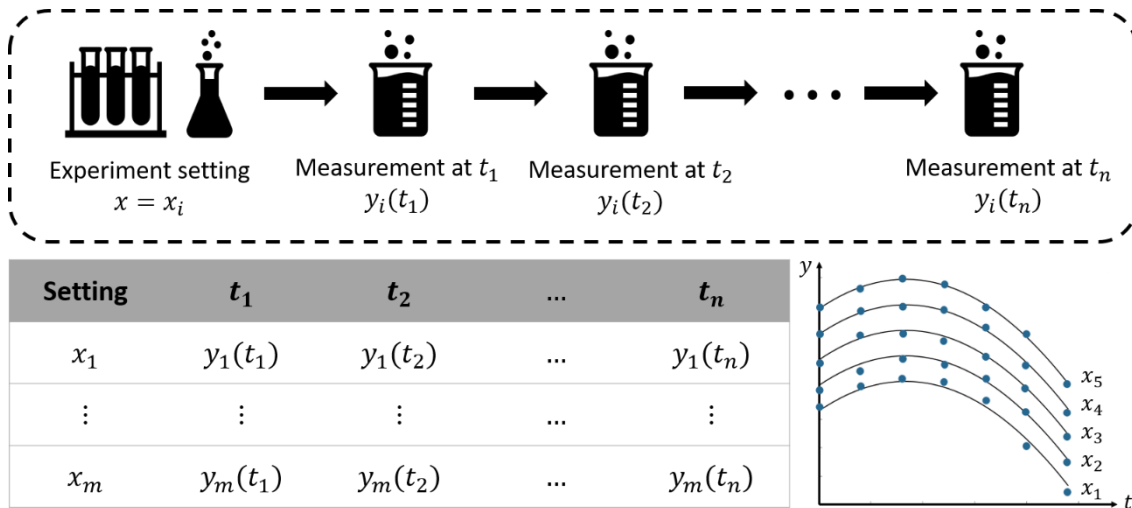
Figure II-2. Data collected in biomaterial experiments, where $y_i(t_j)$ is the measurement of material property at the process parameter value x_i and time t_j.

There are several considerations in modeling the data in Figure 2. First, the modeling is to provide better understanding of the complex material fabrication process hence a good interpretability is required. Second, expert or domain knowledge on the process should be accommodated to make the estimated model practically meaningful.

Finally, the model must be applicable to data of small size as experiments in the fabrication of soft tissues are costly and time-consuming.

For the aforementioned considerations, we take a modeling strategy based on varying coefficient models (VCMs) (Hastie and Tibshirani, 1993). The key idea of conventional VCMs is to use a linear (or nonlinear) function to model the relationship of response and covariates where the coefficients (including intercept) are non-parametric functions of time. Thus they can capture both the complex effects of covariates and dynamics of the effects. With a linear form, the model is also easy to understand and does not require large samples to fit. However, one limitation of such models is that they are not able to accommodate expert knowledge.

In this work, we propose a constrained VCM approach to model the data in Figure 2 with accommodation of expert knowledge in biomaterial fabrication. The proposed model is a linear function of covariate with *semi-parametric* coefficients: the intercept is a parametric function of time and the coefficient of covariate is a non-parametric function of time. Such a semi-parametric model structure allows expert knowledge to be accommodated in the form of constraints on the model coefficients. Specifically, knowledge on the baseline material behavior is incorporated by specifying an appropriate form for the parametric intercept function, while knowledge on the effect of process parameter on material property is incorporated by imposing constraints on the non-parametric coefficient function. For the latter, we consider a dynamic stability constraint that applies in many engineering applications. Model estimation methods based on

7

smoothing spline and weighted smoothing spline are also developed for the proposed model.

The contribution of this work lies in three aspects. First, it establishes an interpretable framework to model data from time-course experiments in soft tissue fabrication. The proposed methodology can be applied to broader areas, such as bio-manufacturing, with similar forms of experimental data, and it can produce a model with good interpretability to help researchers learning new knowledge on their processes. Second, we develop novel ways to accommodate different types of engineering expert knowledge in data analytics for better interpretation and prediction. Finally, the use of weighted smoothing spline in the proposed approach provides a new, convenient method for imposing constraints in non-parametric function estimation. While smoothing spline is powerful in non-parametric fitting, it is known to be sensitive to large random errors/outliers in real observations. By imposing constraints through a weighting scheme, we can regulate smoothing spline fitting and make it satisfy various requirements in specific applications.

The remainder of this paper is organized as follows. Section 2 presents a review of related literature. Section 3 describes the proposed modeling approach and explains the underlying idea. Section 4 gives the model estimation algorithms. Section 5 is the case study, where the proposed approach is applied to a dataset from artificial meniscus fabrication and its good interpretability and prediction performance are demonstrated. Section 6 provides a numerical example to illustrate the advantage of accommodating

expert knowledge in data analytics. Section 7 concludes this study and discusses future work.

II.2. Literature review

The time-course data in Figure 2 take a form similar to longitudinal/panel data that are popular in economics, social sciences and medical literature (Hsiao, 2003; Islam, 1995). There are three relevant modeling approaches in the literature: (1) *Marginal or population-averaged models* which assume the response (biomaterial property in our problem) is a linear model of covariates (process parameters in biomaterial fabrication) through some link function (Verbeke et al., 2008). Non-parametric intercept is used in some studies to improve model flexibility (Zeger and Diggle, 1994). The effects of covariates on the response are represented by their coefficients, but the dynamics of the effects is often not captured. Moreover, an adequate link function may not be available in practice to characterize the complex effects of covariates. (2) *Mixed effect models* which use a linear function to model the relationship of response and covariates with random effects for quantifying the randomness between samples (Laird and Ware, 1983). They are often used for panel data from individuals such as patients in a medical study, where between-individual variation is of interest. However, this is not the case in this study. (3) *Transition models* which account for the time dependence of the response using Markov models (Verbeke et al., 2008). The effect of a covariate is separated into two parts, as represented by its coefficient in the regression component of the model and by the influence of past values of the response on its present value. Such an approach is not

9

applicable for our problem since biomaterial fabrication requires a direct characterization of the effect of process parameters.

Imposing constraints in statistical modeling has been studied recently. For example, a general framework for shape constrained generalized additive models (SCAM) has been developed which can incorporate shape constraints such as monotonicity in generalized additive modeling (Pya and Wood, 2015). A constrained hierarchical model is proposed for modeling degradation of biomaterials, where monotonicity and concavity constraints are incorporated in the model (Zeng et al., 2016). A constrained Gaussian process is developed for the same application, considering bound, censoring and monotonicity constraints (Zeng et al., 2018). Splines with non-negativity constraint is used for analyzing wake effect in wind turbine power generation (Hwangbo et al., 2018). This work considers a *dynamic stability* constraint on the effect of process parameter which has not been considered yet in the literature.

II.3. The proposed model

Let y be the measurement of the biomaterial property of interest (e.g., toughness, viscosity, etc.), x be the value of a process parameter (e.g., percentage of certain ingredient, temperature, etc.) in biomaterial fabrication, and t be time. We consider the following varying coefficient model as

$$y(t) = \beta_0(t) + \beta_1(t)x + \epsilon, \ \epsilon \sim N(0, \sigma_\epsilon^2), \tag{II.1}$$

with $\beta_0(t; \theta)$ *to be a parametric function with parameter* θ,

$\beta_1(t)$ *to be a non-parametric function.*

Here ϵ is the time-independent error term which represents the overall effect of measurement errors and other random errors in the fabrication process. Without loss of generality the error term is assumed to be normally distributed with mean 0 and a constant variance σ_ϵ^2. The $\beta_0(t)$ is the intercept, and $\beta_1(t)$ is the coefficient of process parameter.

The proposed model in Eq. (II.1) follows the spirit of conventional VCMs that allows the coefficients to evolve with time to capture the dynamics of the effect of process parameter. It has a linear structure with respect to the process parameter, which does not require a large number of different settings of the process parameter to learn the model. This is especially favorable in the considered application where limited number of settings are affordable in experimental studies. The linear structure is also easy to interpret. The above aspects make this model suitable for the data in Figure II-2.

The core of the proposed model lies in the coefficients $\beta_0(t)$ and $\beta_1(t)$, which have good interpretations and provide basis for accommodating expert knowledge. Specifically, the intercept term $\beta_0(t)$ is to represent the material behavior under the null setting of process parameter (i.e., $x = 0$), or the baseline material behavior. We consider a *parametric* function of time for $\beta_0(t)$ since the baseline material behavior is often relatively simple and much knowledge about it is available, including physiochemical models of the material and historical data on material characterization. The slope term $\beta_1(t)$ is to represent the effect of process parameter on the material property. In general, this effect is complex and little knowledge about it is available, especially in the considered application where the fabricated biomaterials are novel and very different from traditional materials. Therefore, we assume $\beta_1(t)$ to be a *non-parametric* function of time.

The above setup makes it convenient to incorporate different types of expert knowledge in the form of constraints on the two coefficients. Specifically, expert knowledge on the baseline material behavior can be incorporated by constraining on the form of $\beta_0(t)$ (e.g., specifying certain parametric form for $\beta_0(t)$); and knowledge on the effect of process parameter can be incorporated by imposing constraints in the non-parametric fitting of $\beta_1(t)$. Examples of such constraints are sign constraints (e.g., the effect must be positive) and shape constraints (e.g., the effect must be monotonically non-decreasing over time).
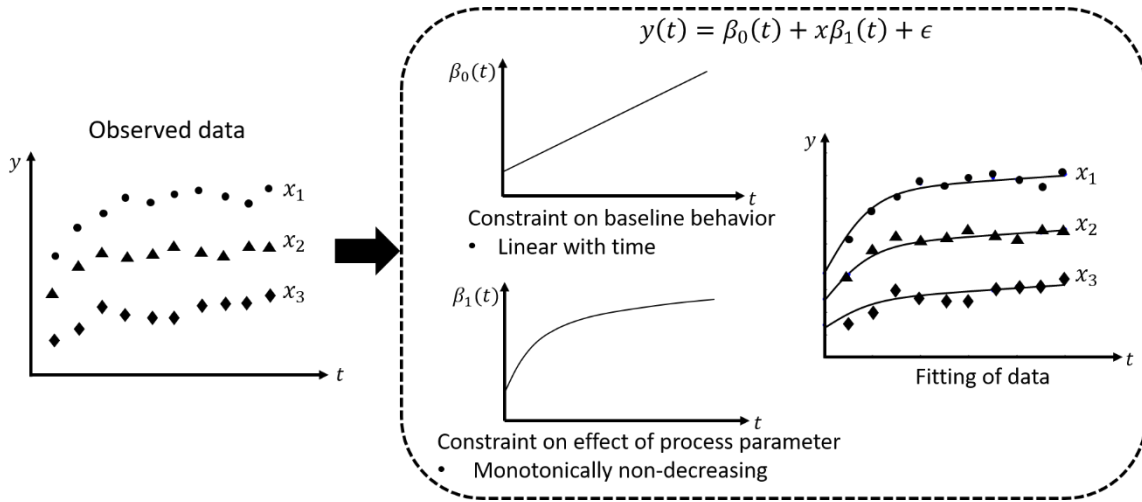
Figure II-3. The idea of expert knowledge accommodation in the proposed modeling approach.

The idea of expert knowledge accommodation is illustrated in Figure II-3 using a simple example. The observed data are shown in the left, and the coefficients and fitted values are in the right panel of the figure. Assume we have the following expert knowledge: Under the baseline setting, the material property approximately linearly

increases over time; and the effect of process parameter is monotonically non-decreasing over time. Accordingly, $\beta_0(t)$ is modeled as a linear function of time, while $\beta_1(t)$ is a non-parametric function of time with a monotonicity constraint. The modeling is expected to produce good fitting to observed data and estimates of the two coefficients that reveal baseline material behavior and the effect of process parameter on material property, as shown in the right panel of Figure II-3.

We would like to remark that the proposed model can be easily extended to more general cases such as non-continuous response under the generalized linear models. Also, although the model assumes a single covariate x, it is applicable to the case of multiple covariates by considering the model in a form of $y(t) = \beta_0(t) + x^T \beta_1(t) + \epsilon$, where $\beta_1(t)$ contains p components and each component is a non-parametric function of time.

II.4. Model estimation

Suppose the time-course data contain m settings of the process parameter, $x_1, x_2, \ldots, x_m$, and measurements of the property at n time points under each setting. Using i to index the process settings and j to index time points, the matrix form of the proposed model can be written as

$$Y(t_j) = \beta_0(t_j)1_{m \times 1} + \beta_1(t_j)X + \epsilon, \quad j = 1, \ldots, n, \tag{II.2}$$

where $Y(t_j) = (y_1(t_j), \ldots, y_m(t_j))'$, $1_{m \times 1} = (1, \ldots, 1)'$, $X = (x_1, \ldots, x_n)'$, and $\epsilon = (\epsilon_1, \ldots, \epsilon_n)'$.

For the model estimation, we consider two scenarios in practice:

Scenario 1: There is only expert knowledge on baseline material behavior.

In this scenario, the parametric form of $\beta_0(t; \theta)$ is determined by the expert knowledge on baseline material behavior, and $\beta_1(t)$ takes a general non-parametric form. A method to estimate the parameter θ of $\beta_0(t)$ and $\beta_1(t)$ is described in Section 4.1.

Scenario 2: There is expert knowledge on both baseline behavior and effect of process parameter.

In this scenario, in addition to a parametric form of $\beta_0(t; \theta)$ determined by the expert knowledge on baseline material behavior, some constraints are imposed on the non-parametric form of $\beta_1(t)$ to reflect the expert knowledge on the effect of process parameter on material property. In particular, we will investigate a *dynamic stability* constraint which is a characteristic feature of biomaterials and also a common phenomenon in many engineering applications. How to impose this constraint in the estimation of $\beta_1(t)$ is described in Section 4.2.

II.4.1. Estimation under scenario 1

In this scenario, we consider the smoothing spline as the non-parametric function for $\beta_1(t)$ since the process parameter usually has a continuous smooth effect over time. Thus the model estimation can be written as

$$\left(\hat{\theta}^*, \hat{\beta}_1^*(t)\right) = \arg\min \sum_{i=1}^m \sum_{j=1}^n \left\{ \left[y_i(t_j) - \beta_0(t; \theta) - x_i \beta_1(t_j) \right]^2 \right\} + \lambda \int \beta_1''(t) dt,$$

(II.3)

where λ is a smoothness parameter. As there are two unknowns, the parametric component θ (and $\beta_0(t; \theta)$) and the non-parametric component $\beta_1(t)$, the estimation will follow an iterative scheme in a similar fashion as the generalized semi-parametric VCMs (Qi et al.,

2017). To differentiate from the constrained estimate of $\beta_1(t)$ in Scenario 2, $\hat{\beta}_1^*(t)$ will be referred to as the unconstrained estimate of $\beta_1(t)$.
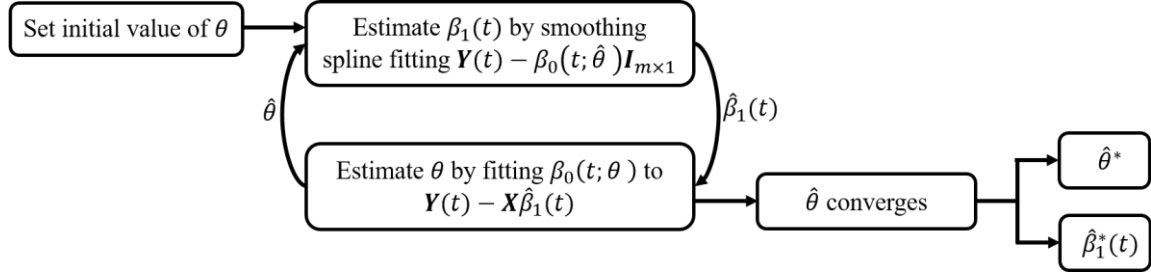
Figure II-4. An illustration of model estimation under Scenario 1

The idea of the iterative estimation procedure is illustrated in Figure II-4. The estimation contains two building blocks: the estimation of $\beta_1(t)$ by smoothing spline and the estimation of θ by parametric function fitting. It starts with assigning an initial value for θ. Given the value of θ (and thus $\beta_0(t; \theta)$), an estimate of $\beta_1(t)$ is obtained by fitting the adjusted part, $Y(t) - \beta_0(t; \hat{\theta})I_{m \times 1}$, using smoothing spline. Similarly, given the estimate of $\beta_1(t)$, an estimate of θ is obtained by fitting the function $\beta_0(t; \theta)$ to the adjusted part, $Y(t) - X\hat{\beta}_1(t)$. This process will iterate to update the estimates of θ and $\beta_1(t)$ until $\hat{\theta}$ converges. Steps of this method are summarized in Algorithm 4.1 given in the supplementary file.

Specifically, in the estimation of $\beta_1(t)$ given $\hat{\theta}$, the smoothing spline method for conventional VCMs will be used (Eubank et al., 2004). This method gives an analytical solution for

$$\hat{\beta}_1(t) = argmin \sum_{i=1}^{m} \sum_{j=1}^{n} \left\{ \left[\tilde{y}_i(t_j) - x_i \beta_1(t_j) \right]^2 \right\} + \lambda \int \beta_1''(t) dt, \qquad \text{(II.4)}$$

15

where $\tilde{y}_i(t_j) = y_i(t_j) - \beta_0(t; \hat{\theta})$. Details of the method can be found in the supplementary files. The objective function in Eq. (II.4) contains two parts: the residual sum of squares (RSS) that represents accuracy of fitting and the penalty to non-smoothness controlled by the smoothness parameter λ. Methods for automatic smoothing parameter selection are available in the literature, including prediction-based methods such as cross-validation (CV) (Wahba and Wold, 1975), generalized cross-validation (GCV) (Craven and Wahba, 1979), Mallow's C_p (Wakefield, 2013) and improved Akaike information criterion (Hurvich, et al., 1998), likelihood-based methods such as maximum likelihood (Anderssen and Bloomfield, 1974), generalized maximum likelihood (GML) (Wahba, 1985) and restricted maximum likelihood (REML) (Wood, 2011), and risk-based methods (Lee, 2003). In our study, GCV is used for smoothing parameter selection as it is recommended for small samples (Aydin and Tuzemen, 2012) which aligns well with the nature of our data.

In the estimation of θ given $\hat{\beta}_1(t)$, it becomes a nonlinear (linear if $\beta_0(t; \theta)$ is linear) least-squares curve fitting problem, which is

$$\hat{\theta} = argmin \sum_{i=1}^{m} \sum_{j=1}^{n} \left(r_i(t_j) - \beta_0(t_j; \theta) \right)^2, \tag{II.5}$$

where $r_i(t_j) = y_i(t_j) - x_i \hat{\beta}_1(t_j)$. The above optimization problem can be solved by either trust-region-reflective (Moré, 1978) or levenberg-marquardt algorithm (Coleman and Li, 1996).

II.4.2. Estimation under scenario 2

Dynamic stability refers to the phenomenon that the material property approaches to an equilibrium state as time goes on. In biomaterial fabrication, a common expert knowledge is that the material property under each setting of process parameter will stabilize along time, which implies that the effect of process parameter $\beta_1(t)$ will converge to a constant. Note that the data-driven approach may not follow such underlying mechanism since it relies only on data to reflect the pattern of the observed data. The smoothing spline method used in Section 4.1 thus could fit the experimental data well, but may not produce a $\hat{\beta}_1(t)$ with the desired shape, especially when the data suffer from large random errors and/or outliers, which often happens in biomaterial experiments.

To address this challenge, we consider a constrained smoothing spline for the estimation of $\beta_1(t)$ by incorporating the dynamic stability as a constraint as follows:

$$\hat{\beta}_1(t) = \begin{cases} argmin \sum_{i=1}^{m} \sum_{j=1}^{n} \left\{ \left[y_i(t_j) - \beta_0(t; \hat{\theta}^*) - x_i \beta_1(t_j) \right]^2 \right\} + \lambda \int \beta_1''(t) dt \\ s.t. \ \beta_1(t) \text{ converges as } t \to \infty \end{cases} \tag{II.6}$$

where $\hat{\theta}^*$ is the estimate of θ obtained from the unconstrained scenario in Section 4.1. Here θ is treated as a known plug-in parameter with the consideration that the parametric function (i.e., $\beta_0(t; \theta)$) fitting is not as flexible as the non-parametric fitting in general. If θ is treated as an unknown under the constrained scenario, it can be difficult to find an estimate of θ that both fits $\beta_0(t; \theta)$ adequately and satisfy the constraint. Thus, it would be better to obtain the estimate of θ under the simpler unconstrained scenario, and then apply the flexible non-parametric fitting of $\beta_1(t)$ to the adjusted data, $y_i(t_j) - \beta_0(t; \hat{\theta}^*)$.

Note that the formulation in Eq. (II.6) is not easy to solve because the constraint is rather implicit and nonlinear. One may think the constraint can be integrated in the objective function in a form similar to the smoothness penalty; however, it will be difficult to solve the resulting optimization problem since the objective function can be quite nonlinear. To circumvent this difficulty, we consider a weighted smoothing spline method and adjust the weights appropriately such that the estimate $\hat{\beta}_1(t)$ converges with time. The reasoning here is to allow more bias at some points through the weighting scheme, thus forcing the resultant estimate to satisfy the constraint. Specifically, we use the following weighted smoothing spline to obtain $\hat{\beta}_1(t)$ as

$$\hat{\beta}_1(t) = argmin \sum_{i=1}^{m} \sum_{j=1}^{n} w_{ij} \left\{ \left[y_i(t_j) - \beta_0(t; \hat{\theta}^*) - x_i \beta_1(t_j) \right]^2 \right\} + \lambda \int \beta_1''(t) dt, \text{ (II.7)}$$

where the weight w_{ij} is given to the squared error under the ith setting at time t_j, $i = 1, \dots, m, j = 1, \dots, n$. If all the weights are equal to 1, the weighted smoothing spline degenerates to smoothing spline; if a weight $w_{ij} < 1$, that means more bias is allowed in fitting the corresponding data point. The close-form solution of Eq. (II.7) and derivation can be found in the supplementary files.

It is worth pointing out that the weighted smoothing spline in the literature is usually used for improving the local performance of curve fitting. That is, weights are assigned to either the squared error term or the smoothness penalty term in Eq. (II.4) to reduce the bias in the area with large local variation (Davies and Meise, 2008; Maria and Malva, 2005). The use of weighted smoothing spline in this work is for an opposite purpose: allowing tolerable bias in the fitting when local variation is large in order to make the estimate $\hat{\beta}_1(t)$ satisfy the constraint.

Figure II-5 elaborates the proposed method for solving the constrained estimation problem in Eq. (II.6). Basically, as shown in the left and right panels of the figure, the constrained estimate of $\beta_1(t)$ is obtained by correcting the unconstrained estimate $\hat{\beta}_1^*(t)$ obtained in Section 4.1 so that the resulting estimate exhibits converging behavior. Specifically, starting with $\hat{\beta}_1^*(t)$, we update the weights w_{ij} in Eq. (II.7) and obtain a new estimate $\hat{\beta}_1(t)$ iteratively. In each iteration, the convergence performance of $\hat{\beta}_1(t)$ will be assessed, and accuracy of fitted values, i.e., $x\hat{\beta}_1(t)$, will be checked. The iterative procedure stops when the stopping rule (e.g., fitting accuracy becomes unacceptable) is reached, resulting in a number of $\hat{\beta}_1(t)$ and their corresponding convergence performance assessments. The estimate that has the best convergence performance will be selected as the constrained estimate of $\beta_1(t)$. This proposed estimation method is summarized in Algorithm 4.2 which is given in the supplementary files. A series of technical details involved in this method are given below.
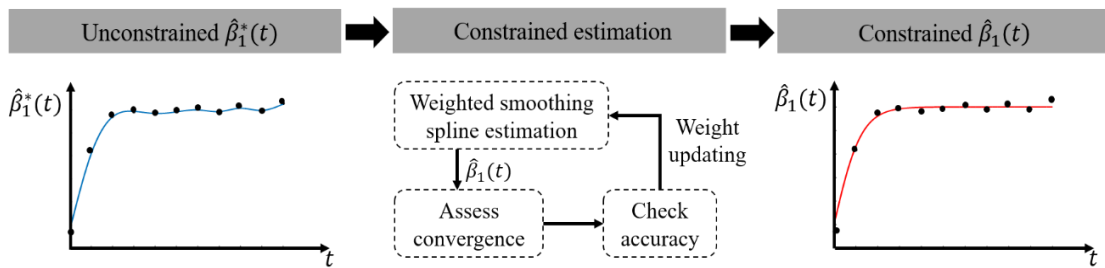
Figure II-5. The proposed method to incorporate dynamic stability constraint in the estimation of $\boldsymbol{\beta_1(t)}$.

II.4.2.1. When to impose constraint

The converging behavior of the effect of process parameter occurs at a certain stage during the observed period. Let $[t^*, t_n]$ be the converging stage, where t^* is the starting time point, $0 < t^* \le t_n$. The dynamic stability constraint should be imposed on this interval in the constrained estimation of $\beta_1(t)$. Usually there is no knowledge on the converging stage, and t^* needs to be determined from data. In this subsection we provide a procedure based on reverse cumulative average to estimate t^* automatically.

In the literature, the reverse cumulative average, i.e., cumulative average starting from the last time point, is used to reveal trend of a time-course stream of data (Yang et al., 2004). Following this idea, we first consider how to identify the converging stage of one time-course stream from the ith setting, $i = 1, \dots, m$, and then extend for all streams of data from m settings. For a single stream $y_i(t_j)$, $j = 1, \dots, n$, we can calculate the difference between $y_i(t_j)$ and its corresponding reverse cumulative average, i.e., $\frac{1}{n-j+1}\sum_{s=j}^{n} y_i(t_s) - y_i(t_j)$, denoted as $D_j(i)$. If t_j is within the converging stage, then $D_j(i)$ will have a small value. Therefore, an estimate of t^* can be obtained by finding the time point from which $D_j(i)$ starts to become small.

For multiple streams under the m settings, we can take the median of all $D_j(i)$, $i = 1, \dots, m$, denoted as D_j. Here the median estimate is used for robustness purpose. To determine the magnitude of D_j that signals converging, we can compare it with $h\hat{\sigma}$, where h is a pre-specified positive value (e.g., 1) and $\hat{\sigma}$ is the estimated standard deviation of data. A conventional estimate $\hat{\sigma}$ is obtained by finding the robust variance estimate of data

20

under each setting (Davies and Meise, 2008) and then taking the median of the m estimates.

Specifically, the procedure of finding the estimate of t^* consists of the following steps:

Step 1. At each time point t_j, calculate the difference of the observations and their corresponding reverse cumulative average:

$$D_j = \underset{i}{median} \left\{ \frac{1}{n-j+1} \sum_{s=j}^{n} y_i(t_s) - y_i(t_j): i = 1, \ldots, m \right\}, \; j = 1, \ldots n.$$

Step 2. Estimate the variance of data σ^2:

$$\hat{\sigma} = \underset{i}{median}\{\hat{\sigma}_i : i = 1, \ldots, m\},$$

where $\hat{\sigma}_i = \frac{1.48}{\sqrt{2}} median \{y_i(t_j) - y_i(t_{j-1}): j = 2, \ldots n\}$.

Step 3. In the sequence $\{D_1, D_2 \ldots, D_n\}$, find the longest subsequence $\{D_a, D_{a+1}, \ldots, D_n\}$, $a \in \{1, \ldots, n\}$, bounded by $\pm h\hat{\sigma}$. Then t_a is the estimate of t^*.
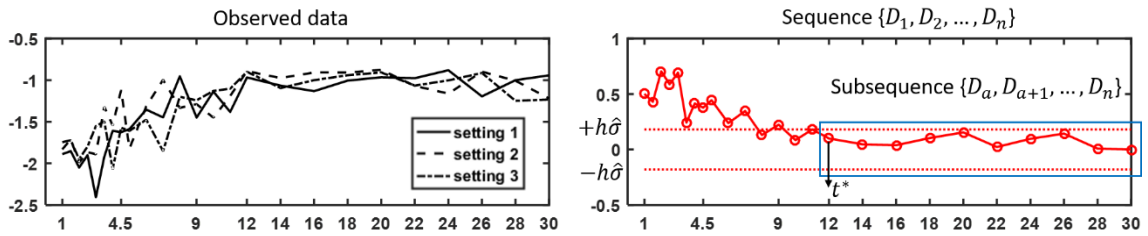
Figure II-6. An example of finding starting time point of converging stage.

Figure II-6 gives an example with $m = 3$ settings and $n = 25$ time points to illustrate the above procedure. Intuitively, when convergence behavior is present, the

21

difference D_j would be small, as shown in the sequence $\{D_1, D_2 \dots, D_n\}$ in the right panel. $h\hat{\sigma}$ is used as the threshold for the sequence to decide whether it is small enough to signal converging. The width of the signaling region depends on the value of h; a larger value of h tolerates more random errors. In the example given in Figure II-6, $h = 1$ is used. The 16th time point is selected as the starting point t^*, and thus the converging stage is $[t_{16}, t_n]$.

II.4.2.2. How to measure convergence performance of $\widehat{\boldsymbol{\beta}}_1(t)$

To assess the convergence performance of the estimate $\hat{\beta}_1(t)$, a mathematical definition of convergence of $\beta(t)$ as a function of time is needed. Formally, the convergence of $\beta(t)$ can be defined as follows: $|\beta(t) - G| < \delta$ for all $t \geq t_a$, where G is a constant, δ is a small value, and t_a is the estimate of the starting time point of the converging stage obtained in Section 4.2.1. This leads to

$$|\beta(t_a) - G| = |\beta(t_a) - \beta(t_{a+1}) + \beta(t_{a+1}) - \beta(t_{a+2}) + \cdots - G|$$

$$\leq \sum_{i=a} |\beta(t_i) - \beta(t_{i+1})|$$

$$\approx \sum_{i=a} |\Delta t_i \beta'(t_i)| \rightarrow \int_{t_a}^{t_n} |\beta'(t)| dt \leq \delta$$

According to the above inequalities, if $\int_{t_a}^{t_n} |\beta'(t)| dt \leq \delta$, then $\beta(t)$ is approaching G approximately. Therefore, we can use

$$C_{\hat{\beta}_1} = \int_{t_a}^{t_n} |\hat{\beta}_1'(t)| dt \qquad \qquad (\text{II.8})$$

as a measure of convergence for the estimate $\hat{\beta}_1(t)$. A smaller value of $C_{\hat{\beta}_1}$ represents better convergence performance of $\hat{\beta}_1(t)$.

II.4.2.3. How to ensure acceptable fitting accuracy

In addition to good converging behavior, there needs a way to evaluate fitting accuracy in the constrained estimation. Note that the weights in the weighted smoothing spline vary from iteration to iteration in the estimation. This means that fitting accuracy can also vary and the resultant estimate may not give satisfactory accuracy. Here we provide a simple method to ensure acceptable fitting accuracy.

In some studies (e.g., Davies and Meise, 2008), fitting accuracy is controlled by specifying a non-parametric confidence region based on residuals and the fitting lying in the region is selected. Inspired by this idea, we can guarantee fitting accuracy by defining a region of acceptable accuracy (RAA) for the fitted values $\hat{y}_i(t_j) = \beta_0(t; \hat{\theta}^*) + x_i \hat{\beta}_1(t_j), i = 1, \dots, m, j = 1, \dots, n$. The width of RAA depends on one's tolerance on fitting accuracy. Intuitively, if the width $\to \infty$, that means fitting accuracy is completely ignored in the constrained estimation, and the narrower the width, the higher the requirement on accuracy. A reasonable choice of the RAA is the confidence interval based on observed data

$$A_\alpha = y_i(t_j) \pm t_{1-\frac{\alpha}{2}, n-1} \times \frac{\hat{\sigma}}{\sqrt{n}}, \tag{II.9}$$

where α is a tuning parameter to control width of A_α and $\hat{\sigma}$ is the estimated standard deviation of data obtained in Step 2 of the procedure given in Section 4.2.1. The tuning parameter α follows the selection rule of significance level in computing confidence intervals and some common values are 0.001, 0.01 and 0.05. Practitioners can tune the value of α based on the required accuracy level.

23

The region of acceptable accuracy A_α will be used in the stopping rule of iterations in search for the constrained estimate $\hat{\beta}_1(t)$. Specifically, the iteration process will stop when the number of fitted values falling outside A_α increases. As the search starts with the unconstrained estimate $\hat{\beta}_1^*(t)$, this means that the fitting accuracy of the constrained estimate should not be considerably worse than the fitting when the convergence constraint is not considered.

II.4.2.4. How to update weights

Weight updating in the iterations is critical in the constrained estimation. Ideally, the updated weights should improve the convergence performance of the resulting $\hat{\beta}_1(t)$ so that the satisfactory estimate can be quickly achieved. In other words, the updating should be along the direction of improvement in convergence performance. Here we propose a weight updating method based on the current fitting performance. Specifically, the new weight is

$$w_{ij} = \frac{\min\limits_{i,j}|r_{ij}|}{|r_{ij}|}, \quad i = 1, \dots, m, \ j = a, \dots, n \qquad (II.10)$$

where $r_{ij} = y_i(t_j) - \hat{y}_i(t_j)$ is the residual of the current fitting and $\min\limits_{i,j}|r_{ij}|$ is the smallest absolute residual. Using the above weights, the point with the smallest residual (i.e., most accurate) will be assigned a weight of 1, while other points will be assigned a weight smaller than 1. The larger the residual (i.e., the less accurate the fitting), the smaller the weight. The idea here is similar to that of robust regression (Rousseeuw, 1987): If there exist any observations with large random errors or outliers, those points will be given small weights. Thus, large bias will be allowed in fitting them, or, in other words, they

will have little influence on the estimation. Since those points are the major cause for non-convergence as mentioned previously, this is expected to improve the convergence performance of $\hat{\beta}_1(t)$.

II.4.2.5. Choice of tuning parameters

The proposed constrained estimation involves two tuning parameters, λ that controls the smoothness level of $\hat{\beta}_1(t)$, and α that controls the width of acceptable accuracy region. For λ, the optimal value determined in Section 4.1 will be used in the constrained scenario to maintain the same level of smoothness as in the unconstrained scenario. For α, a smaller value corresponds to a tighter region of acceptable accuracy, which requires a larger number of iterations in the estimation. A moderate value $\alpha = 0.01$ is used in the case study to avoid too long searching time while achieving a reasonably good level of fitting accuracy.

II.5. Case study

In this section, we apply the proposed approach described in Sections 3 and 4 to a real dataset in a 3D printing of meniscus application. The dataset comes from the fabrication of a novel biomaterial called calcium-alginate/polyacrylamide (CA/PAAm) double-network (DN) hydrogel (Wang et al., 2015). Hydrogel is a popular class of materials constructed by a network of polymer chains. The CA/PAAm-DN hydrogel contains two networks: the alginate/polyacrylamide (A/PAAm) network and the network formed by calcium. With this special structure, it has good shape fidelity with mechanical properties similar to those of natural meniscus, and hence is a promising material for 3D printing of meniscus. As illustrated in Figure II-7 modified from scheme 1 in Wang et al.,

2015, one key step in synthesizing this type of hydrogel is to immerse the one-network A/PAAm hydrogel in $CaCl_2$ solution to build the second network through Ca^{2+} in the solution. The concentration of $CaCl_2$ solution is a critical process parameter in this step which determines properties of the resulting hydrogel. One property of interest is *swelling*, i.e., change in volume of hydrogel, which is important to geometric fidelity of 3D printed meniscus. Thus, experiments are done to explore how the concentration of $CaCl_2$ solution affects the swelling behavior of hydrogel.
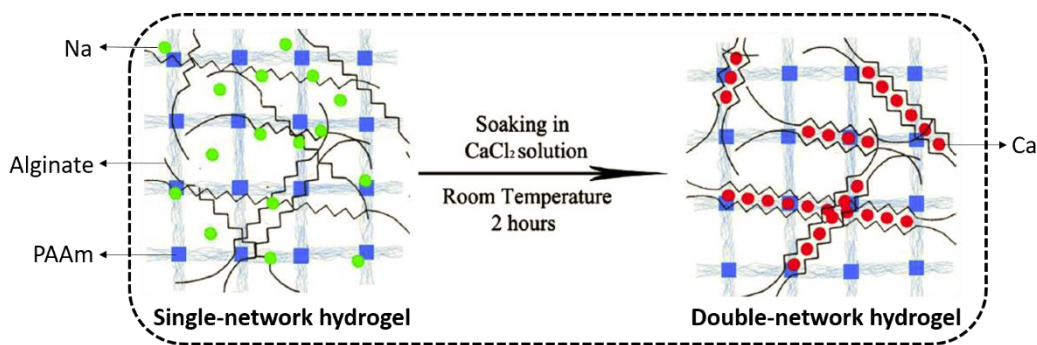
Figure II-7. Illustration of a key step in the fabrication of CA/PAAm hydrogel, modified from scheme 1, Wang et al., 2015

Details of the case study are reported in the following. Section 5.1 introduces the data used in the case study and pre-processing. Section 5.2 gives available expert knowledge on the baseline swelling behavior of the hydrogel material and the effect of process parameter on swelling. Section 5.3 shows the results of model estimation without and with consideration of expert knowledge on the effect of process parameter. Section 5.4 assesses the prediction performance of the proposed approach and compares it with

the conventional VCM. The data used and R codes can be found in the supplementary files to this article.
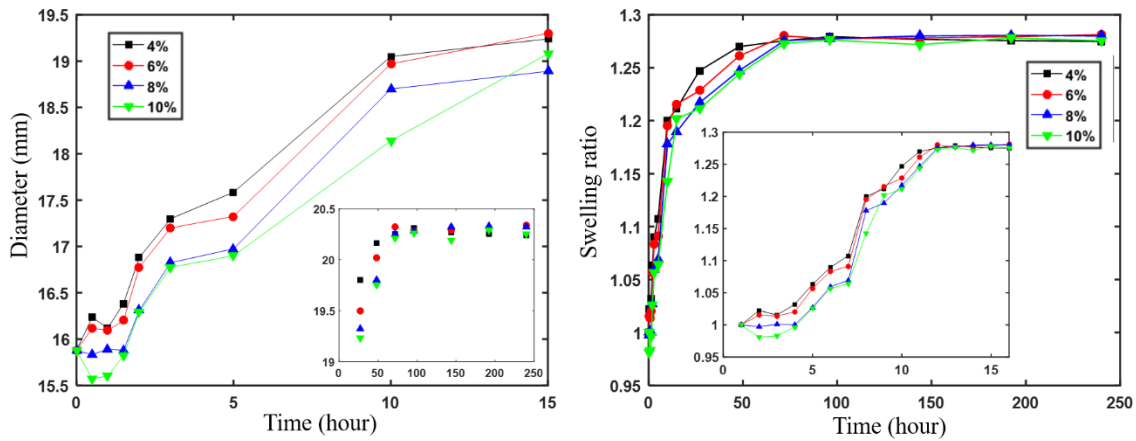
Figure II-8. Hydrogel swelling data used in the case study: original (left) and scaled (right).

II.5.1. Data and pre-processing

The CA/PAAm-DN hydrogel is made into cylinder-shaped samples under four values of the weight concentration of $CaCl_2$ solution: 4%, 6%, 8% and 10 %. The diameters of the samples are recorded initially, and then measured at 16 time points during a period of 10 days. The original data collected from the experiment are displayed in the left panel of Figure II-8, where the portion after two days are shown in a small plot. We scaled these data by the initial diameter to get rid of the effect of unit, so the response used in modeling is volumetric swelling ratio instead of the absolute measurement of diameter. The transformed data are displayed in the right panel of Figure II-8, where the small plot shows a more detailed view of the data arranged by time index (i.e., the 1st time point, …,

the 16th time point). We will fit the transformed data using the proposed model (Eq. (II.1)).

Here y is the volumetric swelling ratio of the fabricated material, x is the weight

concentration of $CaCl_2$ solution, $m=4$, and $n=16$.

II.5.2. Available expert knowledge

In the proposed model, the intercept $\beta_0(t)$ represents the swelling behavior of the

material when the concentration of $CaCl_2$ solution is 0% (i.e., when deionized water is

used in material fabrication). In this case, the resulting material is a regular type, single-

network hydrogel. $\beta_1(t)$ represents the effect of concentration of $CaCl_2$ solution on the

swelling behavior of the material. There is expert knowledge on these two quantities,

which is described below.

About swelling behavior of regular hydrogels (i.e., baseline swelling behavior of

the fabricated material), previous research finds that their volume swelling ratio is

proportional to mass swelling ratio, and the mass swelling ratio has a typical behavior

pattern that is monotonically non-decreasing and gradually flattening out when time goes

on (Ehrenhofer et al., 2018). In some literature, exponential functions are used to fit the

mass swelling ratio of various types of hydrogels (Sadeghi and Hosseinzadeh, 2013;

Slaughter et al., 2015). Thus, we assume that $\beta_0(t)$ has the following parametric form

$$\beta_0(t; \boldsymbol{\theta}) = \theta_1(1 - \theta_2 \cdot e^{-\theta_3 t}), \tag{II.10}$$

which is characterized by three parameters $\boldsymbol{\theta} = [\theta_1, \theta_2, \theta_3]$. The parameters have good

physical interpretations: θ_1 represents the plateau of equilibrium volume swelling ratio

(i.e., the converged value of volume swelling ratio when it stops changing over time), and

θ_2 and θ_3 together describe the increasing rate of volume swelling ratio before equilibrium is reached (Steiner et al., 2016).

About the effect of CaCl$_2$ concentration on hydrogel swelling, as the CA/PAAm-DN hydrogel is a new material under development, little understanding exists in the literature. From the experimental data in the right panel of Figure II-8, a rough impression is that higher concentration leads to smaller swelling ratio. In other words, the concentration seems to have a negative effect on swelling. But there lacks domain knowledge to support this. The dynamics of the effect is also unknown; it may monotonically increase or decrease with time, or oscillate in a complex way. Hence, we cannot constrain the sign/shape of $\beta_1(t)$ and should have it learned from data. The only knowledge we have is that $\beta_1(t)$ will converge with time. Intrinsically, as can be seen from the trend of data, swelling of each hydrogel sample will achieve a plateau finally, and thus the effect of process parameter will converge to a constant.

II.5.3. Model estimation

We obtain first the estimate of $\beta_0(t)$ by the method described in Section 4.1 (Algorithm 4.1) and then the estimate of $\beta_1(t)$ with dynamic stability constraint by the method described in Section 4.2 (Algorithm 4.2). Parameter setting in the estimation and results are given below:

Estimation of $\beta_0(t)$

Initial values of the parameters are set to be $\boldsymbol{\theta}^{(0)} = [\theta_1^{(0)}, \theta_2^{(0)}, \theta_3^{(0)}] = [1, 0.5, 1]$. The algorithm stops after 11 iterations and yields $\widehat{\boldsymbol{\theta}} = [1.28, 0.21, 0.12]$. Figure II-9 shows the updating process of the parameter estimation (1-norm of the three parameters,

i.e.,$|\theta_1| + |\theta_2| + |\theta_3|$, is plotted for convenience) and the resulting estimate of $\beta_0(t)$. As expected according to the above-mentioned expert knowledge, the estimated $\beta_0(t)$ is non-decreasing and flattening out after approximately 40 hours.
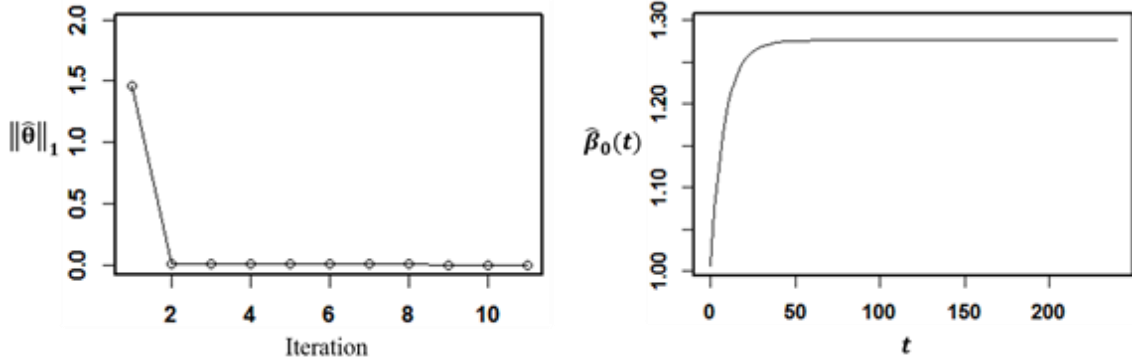
Figure II-9. Estimation of $\beta_0(t)$ norm of parameters in each iteration (left) and estimate of $\beta_0(t)$ (right).

Estimation of $\beta_1(t)$

From the results of Algorithm 4.1, we have the initial (unconstrained) estimate $\hat{\beta}_1^{(0)}(t) = \hat{\beta}_1^*(t)$. To incorporate the dynamic stability constraint, we first find out the starting point t^* of the converging stage using the procedure given in Section 4.2.1. The standard deviation of data is estimated to be $\sigma = 0.0137$, and $h = 1$ is used. The starting point is found to be $t^* = t_{12} = 72$ hours. Then we obtain the region of acceptable accuracy by Eq. (II.8) with $\alpha = 0.01$. Those regions associated with the four settings of process parameter are shown in Figure II-10. In each plot in the figure, the solid markers connected by solid lines represent the observed values, the hollow markers connected by dash lines

represent the fitted values, and the gray lines are the upper and lower bounds of the region of acceptable accuracy.
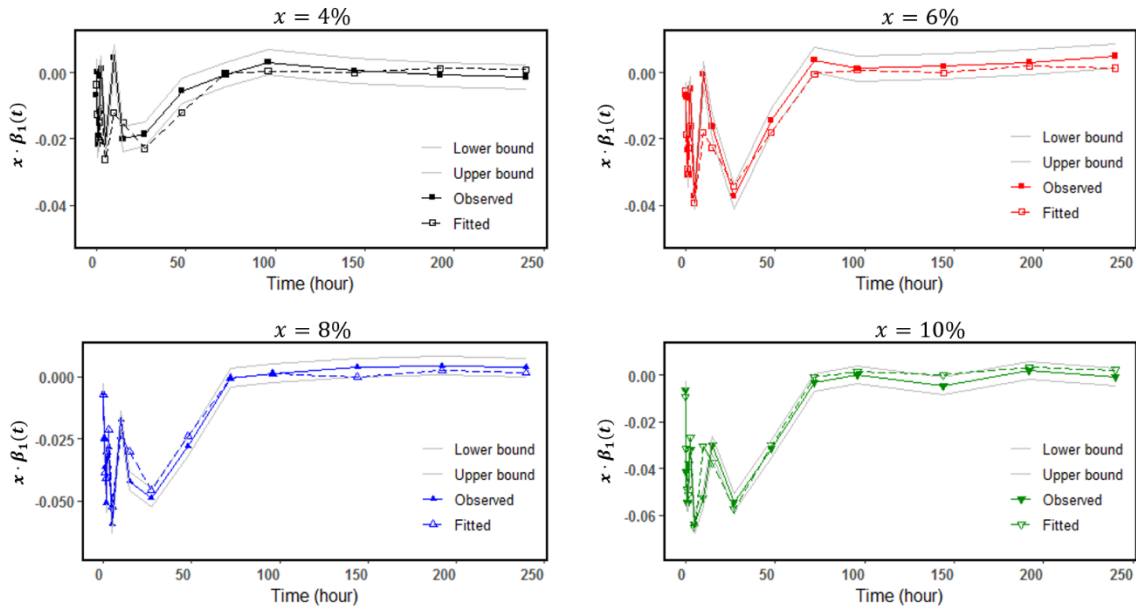
Figure II-10. Region of acceptable accuracy under each setting of process parameter.

Using Algorithm 4.2, we obtain the constrained $\hat{\beta}_1(t)$ with $C_{\hat{\beta}_1} = 0.0574$, which has a higher degree of convergence than the unconstrained estimate $\hat{\beta}_1^*(t)$ with $C_{\hat{\beta}_1^*} = 0.1184$. The two estimates are shown in Figure II-11 for comparison. It is clear that the constrained one is more stable in the tail part. Compared with $\hat{\beta}_0(t)$ shown in the right panel of Figure II-9, the estimated $\beta_1(t)$ has a complex shape during the period $t = [0, 50]$, telling us that the effect of concentration of $CaCl_2$ solution oscillates in the early stage. This is understandable because it is known that the swelling behavior is affected by

multiple forces (e.g., shrinkable elastic energy and swelling osmotic energy) which compete at the beginning before reaching a balance (Wang et al., 2015). Another finding is that this effect is negative, i.e., higher concentration of the solution tends to shrink the volume of hydrogel. These findings are useful to researchers in 3D printing of meniscus to understand the swelling behavior of the novel CA/PAAm-DN hydrogel.
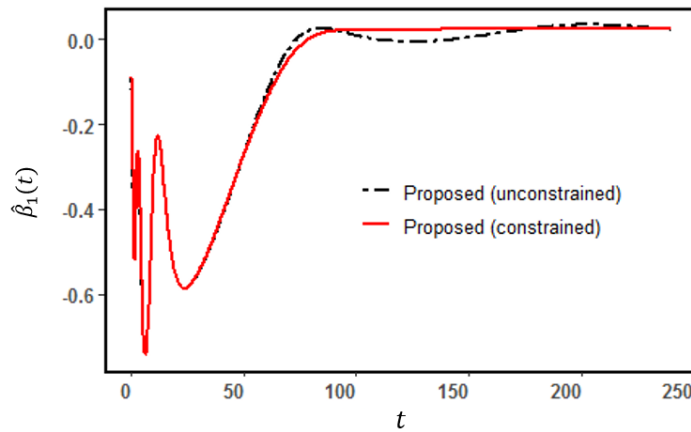
Figure II-11. Estimate of $\beta_1(t)$ with and without dynamic stability constraint.

Based on the estimate of $\beta_0(t)$ in Figure II-9 and the constrained estimate of $\beta_1(t)$ in Figure II-11, we obtain the fitted values of the original data on volume swelling ratio, which are displayed in the left panel of Figure II-12. The data are well fitted with a sum of squared residuals of 0.0032. The largest fitting error occurs under $x = 6\%$, $t = 27$ hour, and is only 1.5% of the actual measurement. The residual plot is given in the right panel of Figure II-12. All the residuals are up and down around the zero line and appear similar among the four settings of x, which is consistent with the assumptions on random errors in Eq. (II.1).
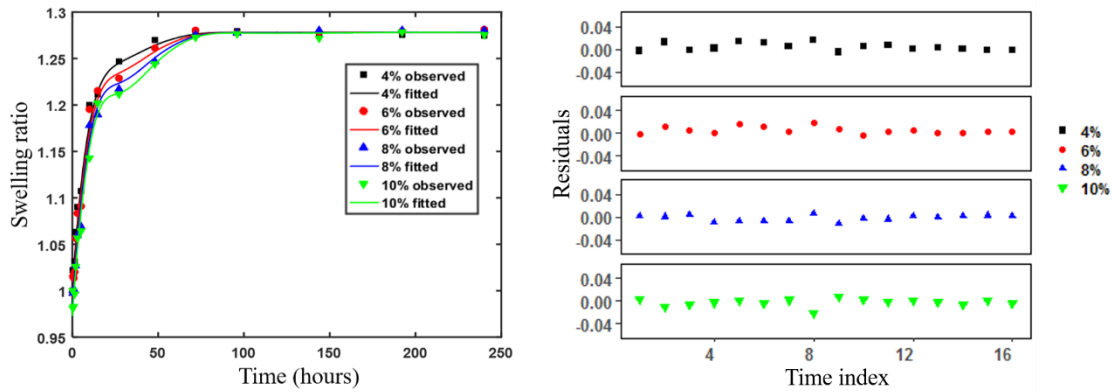
Figure II-12. Fitted values of the original data (left) and residuals (right).

II.5.4. Comparison with existing VCMs

It is interesting to compare the proposed approach with existing approaches to fit VCMs. A flexible and efficient approach in the literature is the previously mentioned SCAM method (Section 2), which can be applied to fit conventional VCMs where both $\beta_0(t)$ and $\beta_1(t)$ are non-parametric. The upper panel of Figure II-13 shows the estimates of $\beta_0(t)$ and $\beta_1(t)$ using the SCAM method without constraint and the proposed approach with dynamic stability constraint. It is seen that the non-parametric estimate of $\beta_0(t)$ is jagged with several local peaks and valleys before $t = 60$ hours, and the estimate of $\beta_1(t)$ also has an irregular shape with longer oscillation in the early stage and two humps in the later stage. Compared to their counterparts from the proposed approach, the estimates from the SCAM method are hard to interpret and not consistent with expert knowledge.

Since the SCAM method is able to incorporate shape constraints in the model, we impose a monotonicity constraint on $\beta_0(t)$ to improve the estimates of this method. The

updated fitting results are given in the lower panel of Figure II-13. Now the estimate of $\beta_0(t)$ has a simple shape that is expected according to the expert knowledge, meaning that the monotonicity constraint works on it. On the other hand, however, the estimate of $\beta_1(t)$ becomes very simple, too, to match with the change in the estimate of $\beta_0(t)$. Although dynamic stability in the later stage is satisfactory now, the oscillation in the early stage is masked, which is not consistent with the expert knowledge nor helpful in understanding the subtle details of the swelling behavior. The fitting performance of the resulting model also becomes worse than the case without the monotonicity constraint (detailed residual analysis is ignored here).
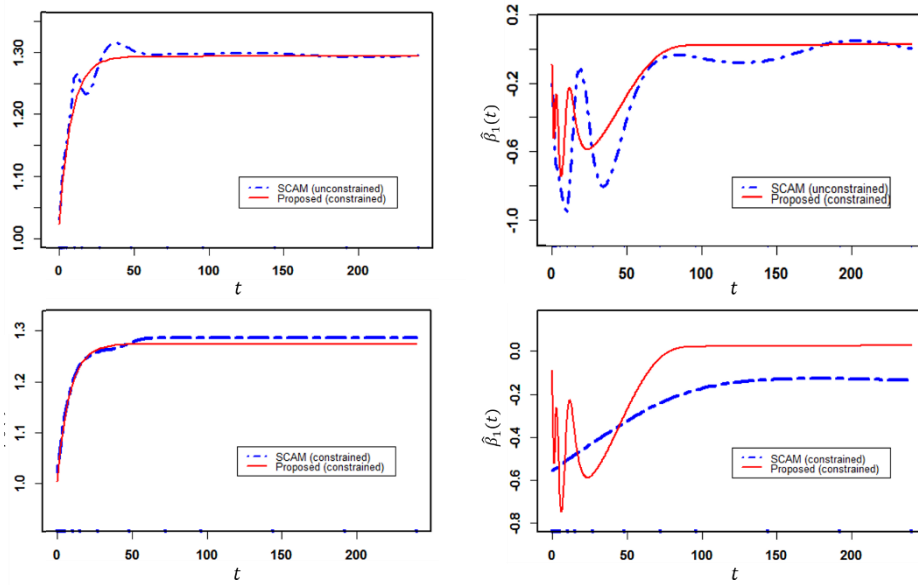
Figure II-13. Estimates of $\beta_0(t)$ and $\beta_1(t)$ of the proposed approach and SCAM without (upper) and with (lower) monotonicity constraint

Another important aspect of the comparison is the prediction performance of the two approaches. Due to the limited available data, we assess their prediction performance based on the leave-one-out cross validation (LOOCV) approach. Specifically, data at the jth time point, $j = 1, ..., n$, are used as test data (of size m), and the remaining data are used as training data (of size $m \times (n - 1)$). We first fit the model using the training data and then predict the response in the test data. The prediction performance at each data point is measured by the relative prediction error, i.e., $\frac{|y_i(t_j) - \hat{y}_i(t_j)|}{y_i(t_j)}$.
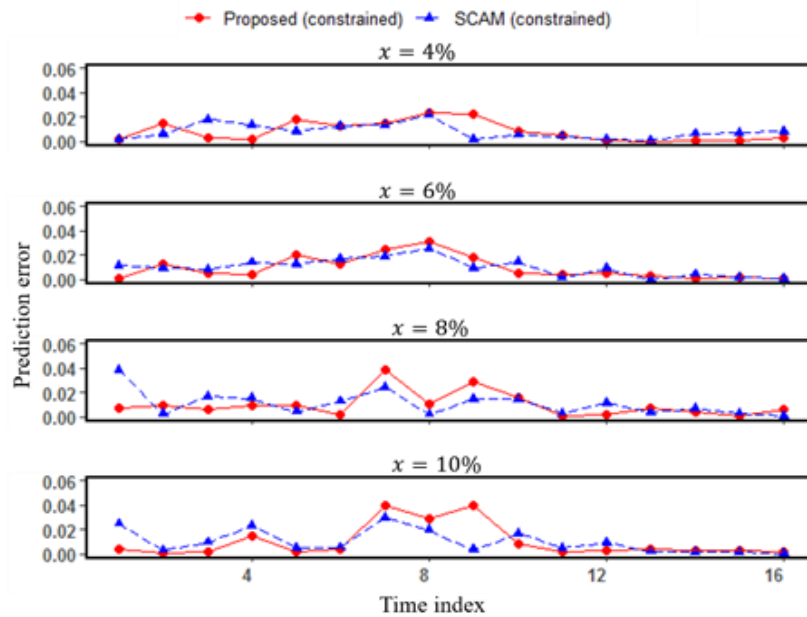
Figure II-14. Prediction performance of the proposed approach and constrained SCAM.

Figure II-14 shows prediction performance of the proposed approach and the SCAM method with monotonicity constraint. Roughly speaking, the prediction errors of

both methods are small (within 4%), indicating good prediction performance. Between the two, no one is uniformly better than the other, and their errors are similar in many cases. So the general conclusion is that the proposed model has comparable prediction performance to the conventional VCM with both $\beta_0(t)$ and $\beta_1(t)$ being non-parametric. Note that the original motivation of the proposed model is interpretability (by accommodating expert knowledge). Now this model is found to perform well in prediction as well and it is comparable to the more flexible conventional VCM in this aspect.

It deserves pointing out that the proposed model performs consistently better in prediction than the conventional VCM in the later stage, from t_{10} to t_{16}, thanks to the dynamic stability constraint. In addition, predictions at t_1 are actually extrapolations as they are made based on data at $t_2,\ldots, t_{16}$. The proposed model also has better performance in this case, which is probably owing to its full accommodation of available expert knowledge.

II.6. Numerical study

To further demonstrate the effectiveness and advantage of the proposed approach, a numerical example is presented in this section. The data generation model is as follows

$$Y(t_j) = \beta_0(t_j) + \beta_1(t_j)X + \epsilon, \ \ \beta_0(t) = t^{\theta_1}e^{-\frac{t}{\theta_2}} + \theta_3, \ \ \beta_1(t) = \frac{\phi_1 t}{\sqrt{1+\phi_2 t^2}},$$

where the parameters of $\beta_0(t)$ are $\theta_1 = 3, \theta_2 = 1.2,$ and $\theta_3 = 1.5,$ the parameters of $\beta_1(t)$ are $\phi_1 = 5,$ and $\phi_2 = 0.4, \ \ X = [x_1, \ldots, x_m]^T = [0.05, 0.19, 0.36, 0.72, 0.91]',\ \ \epsilon = [\epsilon_1, \ldots, \epsilon_m]^T$ with $\epsilon_i \sim N(0, 0.3^2), m = 5$ and $n = 30$. The simulated data are shown in Figure II-15, where the curves are the true functions (i.e., $\beta_0(t) + \beta_1(t)x_i, i = 1, \ldots, m$)

and the dots are actual observations. The true functions converge when t gets large. However, due to considerable random errors, the actual observations do not reflect this trend very well. This kind of situations is common in biomaterial experiments. We will use the proposed approach and the SCAM method to fit the data and compare their fitting performance. Note that there is no obvious shape requirement in this case, so no constraint is imposed in the SCAM method.
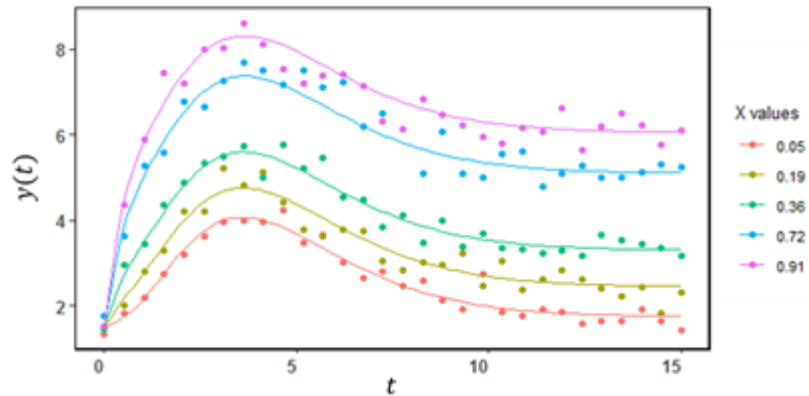
Figure II-15. The simulated dataset.

We first fit the data using the proposed approach without dynamic stability constraint and the SCAM method. The estimates of $\beta_0(t)$ and $\beta_1(t)$ from these two methods together with the true functions are given in the left panel of Figure II-16. The estimate of $\beta_0(t)$ from the proposed approach matches with the true curve very well, while that from the SCAM method has distinct deviations from the true curve, especially in the tail part. The estimates of $\beta_1(t)$ from the two methods exhibit similar shapes. Both of them have large fluctuations in the later stage and deviate substantially from the true curve.

These results show that without imposing the dynamic stability constraint, the proposed approach has advantage in capturing the shape of the true $\beta_0(t)$, but is not able to capture the shape of the true $\beta_1(t)$. Both methods are subject to the trend of data and affected significantly by individual data points. In other words, they are sensitive to measurement errors and outliers.

Then we add the dynamic stability constraint in the proposed approach. The updated fitting results are given in the right panel of Figure II-16. The starting time of convergence stage is estimated to be $t^* = 2.59$. The estimate of $\beta_0(t)$ from the proposed approach is still similar to the unconstrained case and close to the true curve. The estimate of $\beta_1(t)$ from the proposed approach now exhibits convergence behavior in the later stage and becomes close to the true function. The difference of these estimates from those from the SCAM method validates that incorporation of the dynamic stability constraint in the estimation of $\beta_1(t)$ can improve the fitting performance as well as produce interpretable results consistent with expert knowledge.
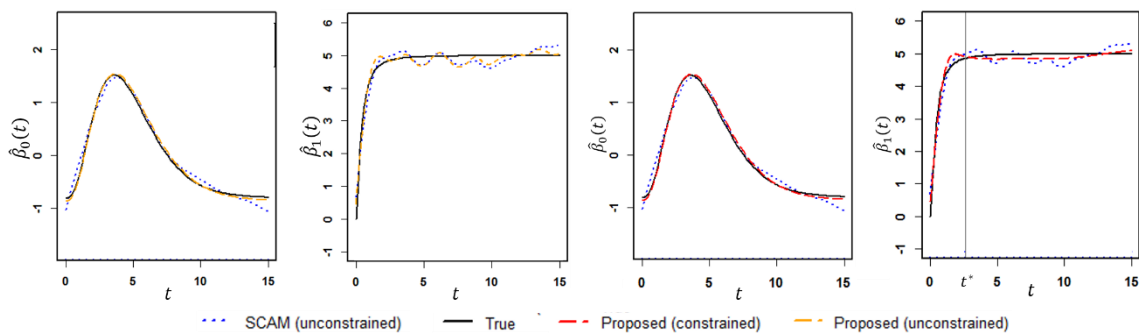
Figure II-16. Estimates of $\beta_0(t)$ and $\beta_1(t)$ from SCAM and the proposed approach with (left) and without (right) dynamic stability constraint

II.7. Conclusion and discussion

This study proposes a constrained VCM approach to model data from time-course experiments in the fabrication of artificial soft tissues. This approach is able to accommodate expert knowledge on the baseline material behavior and the effect of process parameter on material property. According to the case study, the proposed approach leads to results with good interpretability. It also has comparable prediction performance to the conventional VCM, and performs better in extrapolations.

There are some open issues in this study. First, though the semi-parametric model itself can be easily extended to the case of multiple covariates, the method for imposing the dynamic stability constraint in the modeling is designed for the univariate covariate case. When there are more than one process parameters, Algorithm 4.2 is not suitable any more. There is a need to make the constraint applicable in this case.

Second, in the weighted smoothing spline fitting, we update weights based on the residuals in previous iteration (Eq. (II.10)). Though $C_{\hat{\beta}_1}$ decreases as iteration increases (shown in the left panel of Figure II-9) in the case study, this method does not always guarantee such monotonicity. We have considered other ways for weight updating such as $w_{ij}^{(k)} = w_{ij}^{(k-1)} \cdot \frac{\min_j \Sigma_i |r_{ij}|}{\Sigma_i |r_{ij}|}$. This weight also depends on the fitting performance in previous iteration. However, it differentiates the weights on time points only but not on settings of process parameter, i.e., $w_{ij}^{(k)} = w_{i'j}^{(k)}, w_{ij}^{(k)} \neq w_{ij'}^{(k)}$ for $i \neq i'$ and $j \neq j'$, which provides less flexibility in searching for $\hat{\beta}_1(t)$ with small $C_{\hat{\beta}_1}$. Such a weight updating appears not performing as well as the proposed one in terms of the convergence

performance of $\hat{\beta}_1(t)$ in the case study. We will further explore other weight updating methods to find one that performs better than the proposed method.

Finally, this study considers a single constraint in the modeling. A more general scenario in practice is that multiple constraints need to be considered simultaneously. For example, a process parameter may affect the product property of interest in a monotonic way while at the same time its effect also follows a convergence pattern. In such a case, two constraints, monotonicity and stability, need to be incorporated, which is a challenging problem. It becomes more complex when multiple process parameters are involved. One can imagine the case where a monotonicity constraint applies to one process parameter while a stability constraint applies to another. These problems will be explored systematically in our future research.

II.8. References

Anderssen, R. S. and Bloomfield, P. (1974) Numerical Differentiation Procedures for Non-exact Data. *Numerische Mathematik*, **22**(3), 57-182.

Aydin, D. and Tuzemen, M. S. (2012) Smoothing Parameter Selection Problem in Nonparametric Regression Based on Smoothing Spline: A Simulation Study. *Journal of Applied Sciences*, **12**(7), 636-644.

Bochynska, A. I., Hannink, G., Grijpma, D. W. and Buma, P. (2016) Tissue Adhesives for Meniscus Tear Repair: An Overview of Current Advances and Prospects for Future Clinical Solutions. *The Journal of Materials Science: Materials in Medicine*, **27**(5), 85.

Coleman, T. and Li, Y. (1996) An Interior Trust Region Approach for Nonlinear Minimization Subject to Bounds. *SIAM Journal on Optimization*, **6**(2), 418-445.

Craven, P. and Wahba, G. (1979) Smoothing Noisy data with Spline Functions: Estimating the Correct Degree of Smoothing by the Method of Generalized Cross-Validation. *Numerical Mathematics*, **31**(4), 377-403.

Davies, P. L. and Meise, M. (2008) Approximating Data with Weighted Smoothing Splines. *Journal of Nonparametric Statistics*, **20**(3), 207-228.

Ehrenhofer, A., Elstner, M. and Wallmersperger, T. (2018) Normalization of Hydrogel Swelling Behavior for Sensoric and Actuatoric Applications. *Sensors and Actuators B: Chemical*, **255**, 1343-1353.

Eubank, R. L., Huang, C., Maldonado, Y. M., Wang, N., Wang, S. and Buchanan, R. J. (2004) Smoothing Spline Estimation in Varying-Coefficient Models. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, **66**(3), 653-667.

Fox, A. J. S., Bedi, A. and Rodeo, S. A. (2012) The Basic Science of Human Knee Menisci: Structure, Composition, and Function. *Sports Health*, **4**(4), 340-351.

Hastie, T. and Tibshirani, R. (1993) Varying-coefficient Models. *Journal of the Royal Statistical Society: Series B*, **55**(4), 757-796.

He, Y., Xue, G. H. and Fu, J. Z. (2014) Fabrication of Low Cost Soft Tissue Prostheses with the Desktop 3D Printer. *Scientific Reports*, **4**, 6973.

Hsiao, C. (2003) *Analysis of Panel Data*, Cambridge University Press, Cambridge, UK.

Hurvich, C. M., Simonoff, J. S., and Tsai, C.-L. (1998) Smoothing Parameter Selection in Nonparametric Regression Using an Improved Akaike Information Criterion. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, **60**(2), 271-293.

Hwangbo, H., Johnson, A. L. and Ding, Y. (2018) Spline Model for Wake Effect Analysis: Characteristics of A Single Wake and Its Impacts on Wind Turbine Power Generation. *IISE Transactions*, **50**(2), 112-125.

Islam, N. (1995) Growth Empirics: A Panel Data Approach. *The Quarterly Journal of Economics*, **110**(4), 1127-1170.

Laird, N. M. and Ware, J. H. (1983) Random-effects Models for Longitudinal Data. *Biometrics*, **38**(4), 963-974.

Lee, T. C. M. (2003) Smoothing Parameter Selection for Smoothing Splines: A Simulation Study. *Computational Statistics & Data Analysis*, **42**(1-2), 139-148.

Maria, L. and Malva, O. (2005) *Contour Smoothing Based on Weighted Smoothing Splines, Developments in Spatial Data Handling*, 125-136, Springer Berlin Heidelberg, Berlin, Heidelberg.

Moré, J. J. (1978) The Levenberg-Marquardt Algorithm: Implementation and Theory. *Numerical Analysis*, **630**, 105-116.

Ortman, J. M., Velkoff, V. A. and Hogan, H. (2014) An Aging Nation: The Older Population in the United States. U.S. Department of Commerce.

Parker, B. R., Hurwitz, S., Spang, J., Creighton, R. and Kamath, G. (2016). Surgical Trends in the Treatment of Meniscal Tears: Analysis of Data from the American Board

of Orthopaedic Surgery Certification Examination Database. *The American Journal of Sports Medicine*, **44**(7), 1717-1723.

Pya, N. and Wood, S. N. (2015). Shape Constrained Additive Models. *Statistics and Computing*, **25**(3), 543-559.

Qi, L., Sun, Y. and Gilbert, P. B. (2017) Generalized Semiparametric Varying-coefficient Model for Longitudinal Data with Applications to Adaptive Treatment Randomizations. *Biometrics*, **73**(2), 441-451.

Rousseeuw, P. (1987) *Robust Regression and Outlier Detection*. John Wiley & Sons, Hoboken, NJ.

Sadeghi, M. and Hosseinzadeh, H. (2013) Synthesis and Properties of Collagen-g-poly (Sodium Acrylate-co-2-hydroxyethylacrylate) Superabsorbent Hydrogels. *Brazilian Journal of Chemical Engineering*, **30**(2), 379-389.

Slaughter, B. V., Blanchard, A. T., Maass, K. F. and Peppas, N. A. (2015) Dynamic Swelling Behavior of Interpenetrating Polymer Networks in Response to Temperature and pH. *Journal of Applied Polymer Science*, **132**(24), 42076.

Steiner, S., Jensen, W. A., Grimshaw, S. D. and Espen, B. (2016) Nonlinear Profile Monitoring for Oven-Temperature Data. *Journal of Quality Technology*, **48**(1), 84-97.

Verbeke, G., Fitzmaurice, G. E., Davidian, M. and Molenberghs, G. (2008) *Longitudinal Data Analysis*, Chapman and Hall/CRC, New York, NY.

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.

Wahba, G. (1985) A Comparison of GCV and GML for Choosing the Smoothing Parameter in the Generalized Spline Smoothing Problem, *The Annals of Statistics*, **13**(4), 1378-1402.

Wahba, G. and Wold, S. (1975) A Completely Automatic French Curve: Fitting Spline Functions by Cross Validation. *Communications in Statistics*, **4**(1), 1-17.

Wakefield, J. (2013) *Bayesian and Frequentist Regression Methods*. Springer, New York, NY.

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.

Weiss, B., Dakkak, M., Rockl, G., Sukhu, B., Mohr, J., Maru, K., Eye and Tissue Data Committee. (2017) Development of National System Performance Metrics for Tissue Donation, Production, and Distribution Activity. *Cell and Tissue Banking*, **18**(3), 281-296.

Wood, S. N. (2011) Fast Stable Restricted Maximum Likelihood and Marginal Likelihood Estimation of Semiparametric Generalized Linear Models. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, **73**(1), 3-36.

World Health Organization (2011) *Global Health and Aging*, NIH Publication no. **11**-7737.

Yang, W., Bitetti-Putzer, R. and Karplus, M. (2004) Free Energy Simulations: Use of Reverse Cumulative Averaging to Determine the Equilibrated and the Time Required for Convergence. *The Journal of Chemical Physics*, **120**(6), 2618-2628.

Zeger, S. L. and Diggle, P. J. (1994) Semiparametric Models for Longitudinal Data with Application to CD4 Cell Numbers in HIV Seroconverters. *Biometrics*, **50**(3), 689-699.

Zeng, L., Deng, X. and Yang, J. (2018) Constrained Gaussian Process with Application in Tissue-engineering Scaffold Biodegradation, *IISE Transactions*, **50**(5), 431-447.

CHAPTER III

LEARNING HEMODYNAMIC EFFECT OF TRANSCRANIAL INFRARED LASER

STIMULATION: A LONGITUDINAL DATA ANALYSIS

III.1. Introduction

Photobiomodulation is a noninvasive treatment that uses low-level laser to achieve beneficial therapeutic outcomes (Bartos et al., 2016). Though photobiomodulation has been applied in clinical practices for over 40 years since the first observation of its medical benefits in 1967, it was recently found to be potentially useful for various brain disorders and neurological diseases such as stroke and Parkinson's disease (Hennessy and Hamblin, 2016; Karu, 2013; Naeser and Hamblin, 2011) . From then on, brain photobiomodulation has gained much attention. One form of brain photobiomodulation is transcranial infrared laser stimulation (TILS). Previous studies demonstrated that TILS can improve cognitive, emotional and executive functions and reduce depression symptoms (Barrett and Gonzales-Lima, 2013; Blanco et al., 2015; Disner et al., 2016; Hwang et al., 2016).

TILS is triggered by photon absorption occurred in the mitochondria in cells and followed by a chain of complex reactions which eventually lead to alterations in energy metabolism in brain (Wang et al. 2017). Though much preclinical work has been done to explore the underlying biochemical mechanism of TILS, its fundamental principle is still not fully clear so that skepticism about its clinical performance exists (Hamblin et al., 2018). Among the few clinical trials on TILS, Wang et al. conducted experiments investigating changes in hemodynamics caused by TILS (Wang et al. 2017). In their study, hemodynamic responses were measured at a number of time points from a placebo group

and a treatment group, and for each time point, the mean responses of the two groups were compared using two-sample t-test. The test results confirmed that significant hemodynamic changes were caused by TILS. This simple statistical analysis sheds light on the effectiveness of TILS treatment by establishing the significance of its population-level performance. However, the time effect of this treatment was ignored as the t-test was applied to each time point separately. In addition, only the mean of individual responses to the treatment was considered, while the variation among individuals, which is critical for assessing the treatment performance as well, is not clear.

In this study, we propose to learn the effect of TILS on hemodynamics of human brain through longitudinal data analysis. Our proposed methods model the time effect of TILS during and after treatment. Also, both the population-level performance and variation among individuals are considered in order to learn the individual-level uncertainty as well as the expected average therapeutic outcome of TILS. Results of this study will provide a deeper and more comprehensive understanding on the TILS effect. Specifically, three longitudinal data analysis methods are used. First, *repeated measures ANOVA* is applied to confirm the statistical significance of treatment-placebo difference, time effect and individual variation. Then two *parametric mixed-effect models* are proposed to characterize the population-level effect as a function of time and individual variation. Finally, a *non-parametric functional mixed-effect model* is proposed to address problems of the parametric mixed-effect models. Performance of the models will be compared and interpretation of the selected model will be given.

47

The remainder of this paper is organized as follows. Section 2 introduces data used in this study and reports the results of repeated measures ANOVA. Section 3 presents the proposed parametric mixed-effect models and estimation results. Section 4 presents the proposed functional mixed-effect model and estimation results. The proposed models will be compared in Section 5 in terms of fitting, prediction and variance pattern. Section 6 concludes this study and discusses future research.

III.2. Repeated measures ANOVA

III.2.1. Experiment and dataset

The data used in this study were collected from the experiments conducted by Wang et al. (Wang et al. 2017). The experimental setup is shown in the left panel of Figure III-1. 11 healthy individuals participated two separate experiments: the placebo treatment on their right forehead, followed by the TILS treatment on the same location. A 1064-nm continuous wave laser device was used for both types of treatments. The laser power was controlled at 3.4 W for the TILS treatment. For the placebo treatment, the power was tuned down to 0.1 W, and the laser aperture was further covered by a black cap to ensure that no laser was delivered to the subject. A single-channel broadband near-infrared spectroscopy (bb-NIRS) system was used to measure change in the concentration of oxygenated hemoglobin (ΔHbO) in brain during the experiments, which consists of a tungsten halogen lamp as light source, a miniature spectrometer as detector, an "I"-shaped probe holder to hold an optical fiber bundle relayed to the white light, an optical shutter to switch on and off the white light, and a laptop computer to acquire and save the experimental data.

Each experiment lasted a period of 15 minutes. As shown in the right panel of Figure III-1, this period was divided to three stages: pre-treatment (first 2 minutes), treatment (next 8 minutes) and recovery (last 5 minutes). In the pre-treatment stage, neither placebo nor TILS was applied so that the baseline of oxygenated hemoglobin (HbO) can be measured for each participant. The treatment was then applied during the treatment stage and discontinued in the recovery stage. Spectral data were collected every minute during treatment and recovery, and then converted to ΔHbO from the baseline (Wang et al. 2016).

Figure III-2 shows the ΔHbO measurements of the placebo group and the TILS group. We denote the data using the following notations. Let y_{hij} be the ΔHbO of participant i in group h at time t_j, where $h = 0$ (placebo) or 1 (TILS), $i = 1, ..., N_h$, $j = 1, ..., n$. In this dataset, the two groups have equal number of participants, $N_1 = N_2 = 11$, and the number of measured time points in each session is $n = 13$.
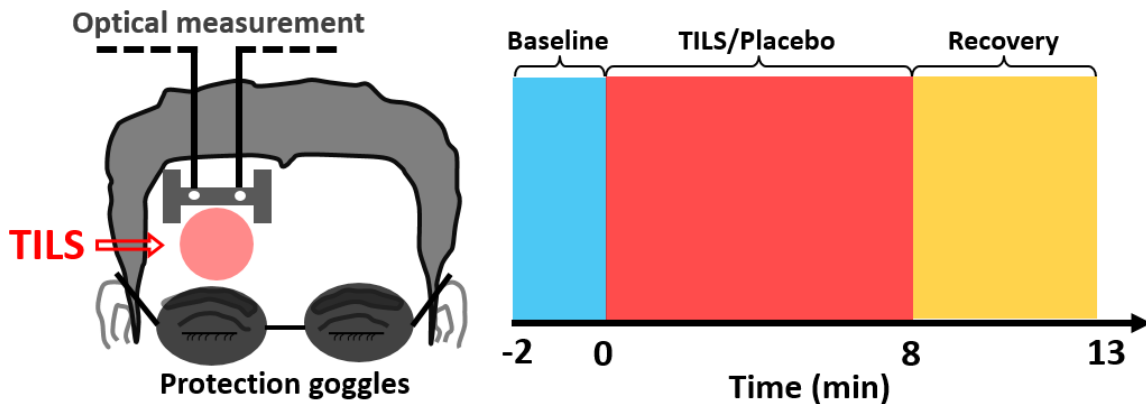
Figure III-1. Setup of the TILS/placebo treatment experiment (left) and data acquisition scheme (right).
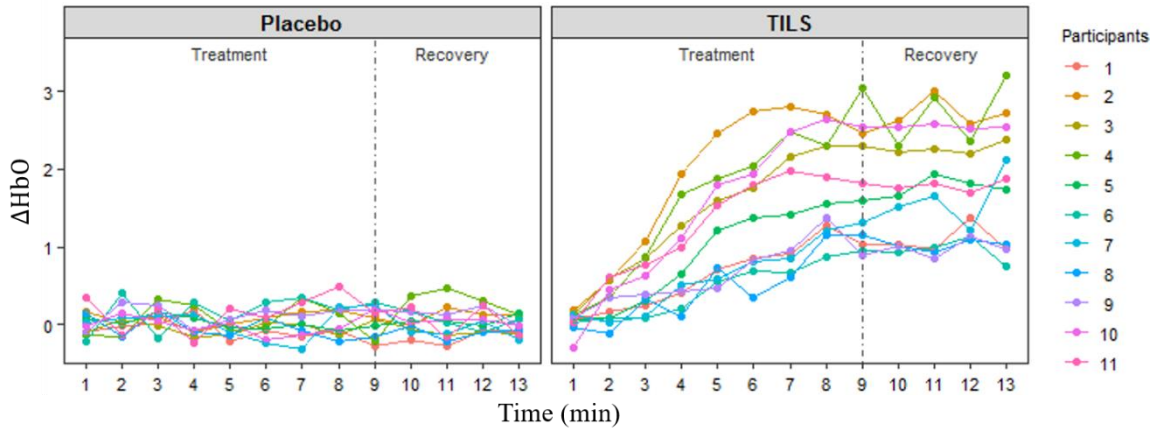
Figure III-2. Data collected from Placebo group (left) and TILS group (right)

III.2.2. Repeated measures ANOVA and hypothesis tests

We first test if group (TILS vs. placebo), time and individuals cause variation in the ΔHbO measurements in Figure III-2. This test is necessary to set a foundation for formal modeling of the TILS effect. If the difference between groups is negligible, the effectiveness of TILS at the population-level will be declined. This indicates that modeling the effect of TILS does not make much sense. In other words, modeling the TILS effect is only meaningful when the difference caused by treatment is significant. Time may also play a role in the variation of ΔHbO data. When the time effect is significant, it means that the effect of TILS is time-varying, and thus it should be modeled as a function of time. In addition, individual differences of the participants in their responses to the treatment may help explain the variation in ΔHbO too. If this difference is significant, it should be taken into account in the modeling. The above three aspects can be tested all at once using repeated measures ANOVA (Hedeker and Gibbons, 2006) given below.

50

Table III-1. Repeated measures ANOVA table

Source	Degrees of freedom	Sum-of-squares	Mean sum-of-squares
Group	$2-1$	$SS_G = n \sum_{h=0}^{1} N_h (\bar{y}_{h..} - \bar{y}_{...})^2$	$MS_G = \dfrac{SS_G}{2-1}$
Time	$n-1$	$SS_T = N(\sum_{j=1}^{n} (\bar{y}_{..j} - \bar{y}_{...})^2)$	$MS_T = \dfrac{SS_T}{n-1}$
Group $\times$ Time	$(2-1)$ $\times$ $(n-1)$	$SS_{GT} = n \sum_{h=0}^{1} \sum_{j=1}^{n} N_h (\bar{y}_{h.j} - \bar{y}_{h..} - \bar{y}_{..j} + \bar{y}_{...})^2$	MS_{GT} $= \dfrac{SS_{GT}}{(2-1)\times(n-1)}$
Individuals in Group	$N^* - 2$	$SS_{S(G)} = n \sum_{h=0}^{1} \sum_{i=1}^{N_h} (\bar{y}_{hi.} - \bar{y}_{h..})^2$	$MS_{S(G)} = \dfrac{SS_{S(G)}}{N-2}$
Residuals	$(N-2)$ $\times$ $(n-1)$	$SSE = \sum_{h=0}^{1} \sum_{i=1}^{N_h} \sum_{j=1}^{n} (y_{hij} - \bar{y}_{h.j} - \bar{y}_{hi.} + \bar{y}_{h..})^2$	MSE $= \dfrac{SSE}{(N-2)\times(n-1)}$
Total	$Nn - 1$	$SS_y = \sum_{h=0}^{1} \sum_{i=1}^{N_h} \sum_{j=1}^{n} (y_{hij} - \bar{y}_{...})^2$	$MS_y = \dfrac{SS_y}{Nn-1}$

*$\sum_{h=1}^{2} N_h = N$ denotes the total number of participants in the experiment.

The repeated measures ANOVA is based on the following point-wise model

$$y_{hij} = \mu + \gamma_h + \tau_j + (\gamma\tau)_{hj} + \pi_{i(h)} + e_{hij}, \qquad (III.1)$$

where μ is the grand mean, γ_h is the group effect, τ_j is the time effect, $(\gamma\tau)_{hj}$ is the interaction effect of group and time, $\pi_{i(h)}$ is the individual difference component for participant i in group h, and e_{hij} is the random error. In general, it is assumed that $\pi_{i(h)}$ and e_{hij} are independently normally distributed, $\pi_{i(h)} \sim N(0, \sigma_\pi^2)$, $e_{hij} \sim N(0, \sigma_e^2)$. It is also required that $\sum_{h=1}^{2} \gamma_h = 0, \sum_{j=1}^{n} \tau_j = 0, \sum_{h=1}^{2} \sum_{j=1}^{n} (\gamma\tau)_{hj} = 0$ to make all effects in the model differentiable [6]. The ANOVA table is shown in Table III-1. Using the data in Figure III-2, the mean sum-of squares are obtained: $MS_G = 107.9045$, $MS_T = 2.18616$, $MS_{GT} = 2.224202$, $MS_{S(G)} = 2.129356$, and $MSE = 0.05848785$.

III.2.2.1. Group by time interaction test

$$H_0: (\gamma\tau)_{01} = \cdots = (\gamma\tau)_{1n} = 0 \ vs. H_a: \text{not all } (\gamma\tau)_{hj} \text{ are zeros}, \ h = 0, 1, \ j = 1, \cdots, n$$

Under the null hypothesis, the group by time interaction is tested by an F-test with degree of freedom $n - 1$ and $(N - 2)(n - 1)$ with $F_0 = \frac{MS_{GT}}{MSE}$. The computed p-value in our case is much less than 0.0001, which gives a strong supportive evidence of the alternative hypothesis. The rejection of the test indicates that the group and time effects are significant however confounded. The difference between groups varies overtime and the curves not parallel.

III.2.2.2. Group effect hypothesis test

$$H_0: \gamma_0 = \gamma_1 = 0 \ vs. \ H_a: \text{not both } \gamma_0 \text{ and } \gamma_1 \text{ are zeros.}$$

Under the null hypothesis, $F_0 = \frac{MS_G}{MS_{S(G)}}$ follows F distribution with degrees of freedom 1 and $(N - 2)$. The computed p-value is smaller than 0.0001 so the hypothesis is rejected at typical levels of significance such as 0.01, 0.02 and 0.05. The test result indicates that the effect of TILS is significant in this experiment so modeling this effect is meaningful.

III.2.2.3. Time effect hypothesis test

$$H_0: \tau_1 = \cdots = \tau_n = 0 \ vs. \ H_a: \text{not all of } \tau_1, \dots, \tau_n \text{ are zeros.}$$

Under H_0, $F_0 = \frac{MS_T}{MSE}$ follows F distribution with degrees of freedom $(n - 1)$ and $(N - 2)(n - 1)$. The null hypothesis is rejected as the p-value is smaller than 0.0001. This indicates that the value of ΔHbO depends on time, which suggests to model the effect of TILS as a function of time.

III.2.2.4. Individual effect hypothesis test

$$H_0: \sigma_\pi^2 = 0 \quad vs. \quad H_a: \sigma_\pi^2 \neq 0$$

Under H_0, $F_0 = \frac{MS_{S(G)}}{MSE}$ follows F distribution with degrees of freedom $(N-2)$ and $(N-2)(n-1)$. Again, the null hypothesis is rejected as the p-value is smaller than 0.0001, which means that the participants respond to the treatment differently. This result necessitates the quantification of individual variation in the TILS effect.

In summary, the above test results motivate us to model the effect of TILS on the concentration of oxygenated hemoglobin in brain. In the modeling, time should be involved as this effect is found to be time-varying. The variation of individual responses should also be taken into account. The following sections provide two types of models that satisfy the requirements. Note that only the data of the TILS group (shown in the right panel of Figure III-2) will be used in the modeling of TILS effect, and the subscript h will be dropped in the notations hereafter.

III.3. Parametric mixed-effect model

As a starting point of the modeling, this section considers parametric mixed-effect models which is a popular category of models for longitudinal data analysis (Laird and Ware ,1982; Breslow and Clayton, 1993). Basics of such models will be given first and then two specific models will be proposed for the TILS data.

III.3.1. Basics

For interpretation convenience, a general parametric mixed effect model is written in the following hierarchical form (Lindstrom and Bates, 1990),

$$\text{Level-1:} \ y_{ij} = f(t_j; \theta_i) + e_{ij} \tag{III.2a}$$

53

$$\text{Level-2: } \theta_i = \beta + u_i \tag{III.2b}$$

The Level-1 model assumes that y_{ij}, the ΔHbO of participant i at time t_j, depends on a parametric function f of time with parameter θ_i and random error e_{ij}. e_{ij} is assumed to follow a normal distribution $N(0, \sigma_e^2)$. The Level-2 model further defines the parameter θ_i as combination of a constant, β, called fixed effect, and a random variable u_i, called random effect. It is assumed that $u_i \sim N(0, \sigma_u^2)$.

Note that the model in Eq. (III.2a) is a point-wise model. The continuous versions of the notations are $y_i(t)$, $f(t; \theta_i)$, etc. Intuitively, the responses of participant i at different time points form a curve $y_i(t)$ as shown in Figure III-2. The shape of the curve is determined by θ_i and it varies from one participant to another as clearly seen from the figure. Thus, β represents the population-level, or baseline, shape of the curve, while u_i represents the difference of participant i from the baseline.

Once the form of Leve-1 model is determined, the unknown parameters β, σ_e and σ_u can be estimated by maximizing the likelihood using the expectation-maximization (EM) algorithm [16, 17]. Then $f(t; \hat{\beta})$ can be interpreted as the population-level effect of TILS treatment, and $\hat{\sigma}_u^2$ indicates the individual variation of the TILS effect. It deserves to mention that the selection of the parametric function $f(t; \theta)$ is subjective and often relies on domain knowledge. Here we give two models with different forms of $f(t; \theta)$ and their corresponding estimation results.

III.3.2. Model I

From the right panel of Figure III-2, we observe that the ΔHbO shows an overall increasing trend for all participants during the treatment period, and a decreasing trend for

some participants (#1, #5, #6 and #10) during the recovery period. Therefore, we assume

the following model form

$$\text{Level-1: } f(t_j; \boldsymbol{\theta}_i) = \theta_{1i} t_j^2 + \theta_{2i} t_j + \theta_{3i} + e_{ij}, \tag{III.3a}$$

$$\text{Level-2: } \begin{cases} \theta_{1i} = \beta_1 + u_{1i}, \\ \theta_{2i} = \beta_2 + u_{2i}, \\ \theta_{3i} = \beta_3 + u_{3i}. \end{cases} \tag{III.3b}$$

In the Level-1 model, f is a quadratic function of time with three parameters θ_{1i},

θ_{2i}, θ_{3i}. When θ_{1i} is negative, the curve of participant i first increases monotonically when

$0 \leq t \leq -\frac{\theta_{2i}}{2\theta_{1i}}$ and then decreases, and the increasing/decreasing rate is controlled by θ_{1i}.

All the three parameters contribute to the highest value that the curve can achieve. In the

Level-2 model, each parameter contains a fixed effect and a random effect which allows

much flexibility to fit individual curves. The random effects are assumed to be normally

distributed with $u_{1i} \sim N(0, \sigma_{u_1}^2)$, $u_{2i} \sim N(0, \sigma_{u_2}^2)$, $u_{3i} \sim N(0, \sigma_{u_3}^2)$.

The fitted population-average curve (i.e., $f(t; \widehat{\boldsymbol{\beta}})$, where $\widehat{\boldsymbol{\beta}} = [\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3]$) and

individual curves (i.e., $f(t; \widehat{\boldsymbol{\theta}}_i)$, where $\widehat{\boldsymbol{\theta}}_i = [\hat{\theta}_{1i}, \hat{\theta}_{2i}, \hat{\theta}_{3i}]$) are shown in the left panel and

right panel, respectively, of Figure III-3 with the parameter estimates given in Appendix

E. The fitted population-average curve shows an overall increasing trend till around the

11th minute. The fitted individual curves show similar trend and match with the data well.

According to Appendix E, the estimated variances of random effects are considerably

large, indicating that individual variation of the TILS effect is significant. Consistent with

that, the shapes of individual curves in the right panel of Figure III-3 vary a lot from each

other.
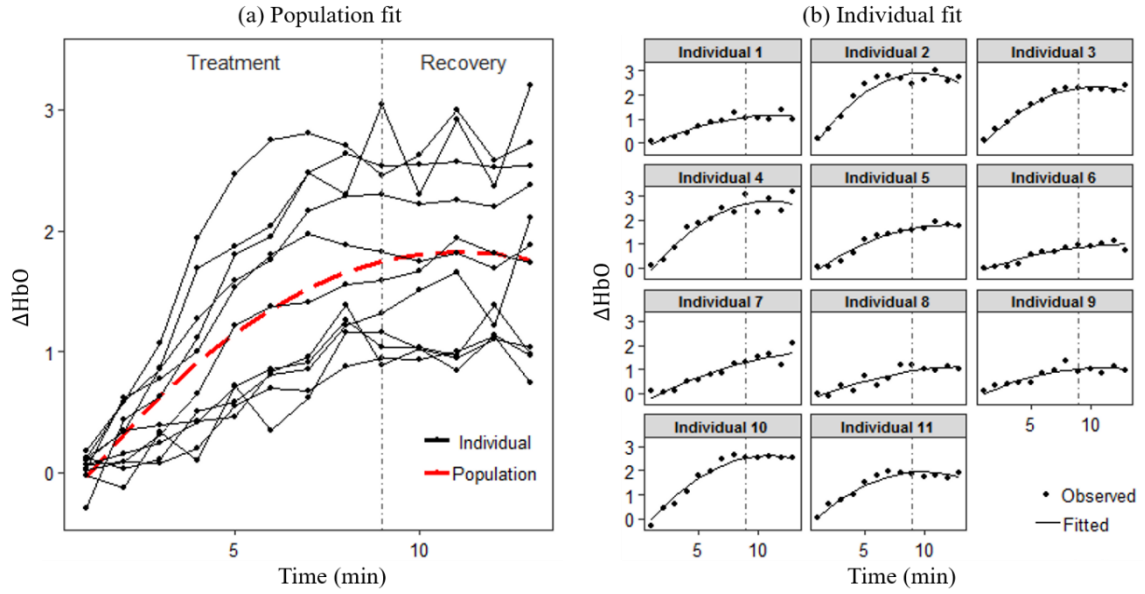
Figure III-3. Population and individual fitting of Model I.

The fitted Model I tells us that in general, the change in oxygenated hemoglobin concentration in brain increases as the TILS treatment continues. When the treatment stops, the change remains increasing and then starts to decrease, meaning that the effect of TILS lasts but gets weaker with time.

III.3.3. Model II

Alternatively, it is reasonable to assume that ΔHbO is non-decreasing as time goes on and achieves a plateau eventually. The corresponding model has the following form,

$$\text{Level-1: } f\left(t_j; \boldsymbol{\theta}_i\right) = \theta_{1i} + \theta_{2i}\left(1 - \exp\left(-\theta_{3i}(t_j - 1)\right)\right) + e_{ij}, \quad \text{(III.4a)}$$

$$\text{Level-2: } \begin{cases} \theta_{1i} = \beta_1 + u_{1i}\,, \\ \theta_{2i} = \beta_2 + u_{2i}\,, \\ \theta_{3i} = \beta_3 + u_{3i}\,. \end{cases} \quad \text{(III.4b)}$$

Figure III-4 shows an illustration of the function $f(t; \boldsymbol{\theta})$ in the Level-1 model in Eq. (III.4a). It has an increasing trend starting from θ_1 at $t_1 = 1$ and reaches a plateau of height $\theta_1 + \theta_2$ when t is large. θ_3 reflects the speed of increase before reaching the plateau. The larger the θ_3 is, the faster the increase is and the more likely that the curve flattens out at the end of the experiment period. The Level-2 model is the same as in Model I, where each parameter contains a fixed effect and a random effect.
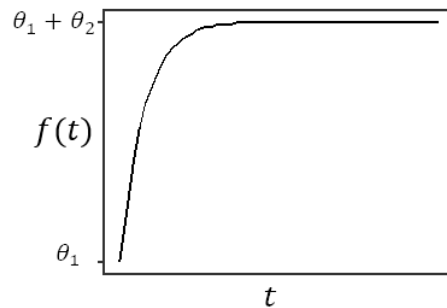
Figure III-4. Illustration of Level-1 model in Model II

The fitted population-average curve and individual curves are shown in Figure III-5 with the parameter estimates listed in Appendix E. The population-level effect keeps increasing without reaching a plateau during the whole experiment period. The fitted individual curves show similar trend and match with data well. According to Appendix E, the estimates of variance components are considerably large and similar to those of Model I.

Based on the fitted Model II, we can conclude that the change in oxygenated hemoglobin concentration in brain keeps increasing during the treatment stage and even

in the recovery stage, meaning that the TILS effect sustains for quite long time after the treatment stops.
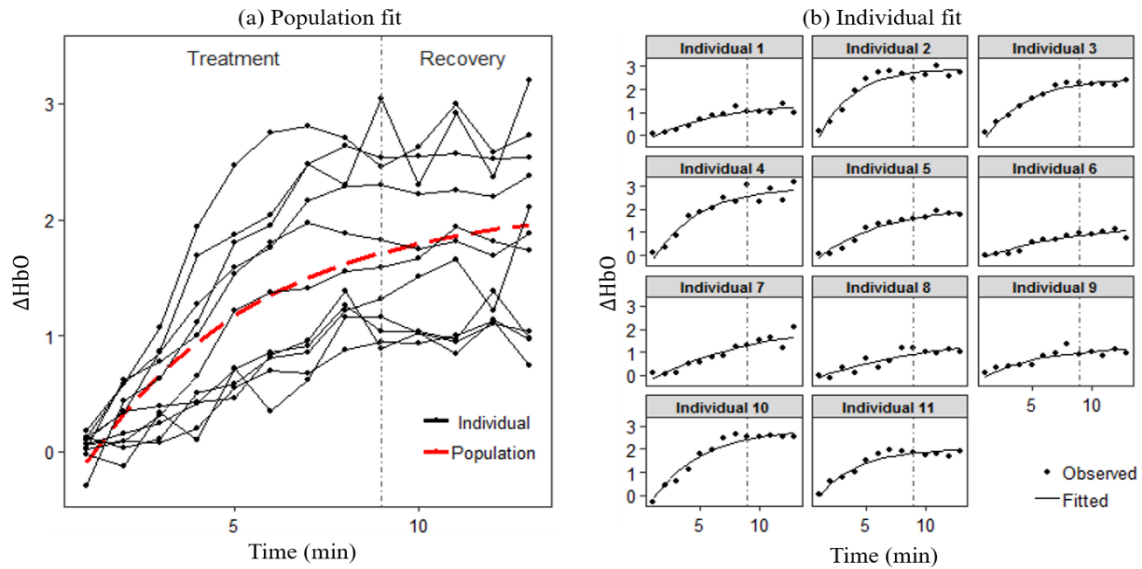
Figure III-5. Population and individual fitting of Model II

III.3.4. Limitation of parametric mixed-effect models

Both Model I and Model II provide estimates of the population-level effect of TILS treatment on hemodynamics, but their interpretations are not fully consistent. About the TILS effect during the treatment stage ($1 \leq t \leq 9$), both models indicate that ΔHbO increases over time. However, they differ regarding the TILS effect during the recovery stage: Model I suggests that ΔHbO first keeps increasing and then starts to decrease, while Model II says that ΔHbO keeps increasing during the whole recovery stage.

The above difference comes from the selection of parametric form for $f(t; \boldsymbol{\theta})$ in Eq. (III.2a), which pre-specifies the shape of the population-average curve and individual

curves subjectively. The risk associated with parametric models is twofold. First, the selected parametric form may not be consistent with the true underlying mechanism. As a result, the estimated population-average curve may lead to biased interpretation on the TILS effect and generate inaccurate predictions to unobserved data. Second, the parametric form together with the random effects in the Level-2 model also determines the variance structure of the response variable. It is possible that the fitting of population and individual curves is good, but the corresponding variance pattern does not match with the data.

III.4. Functional mixed-effect model

The issues with parametric mixed-effect models can be addressed by functional mixed-effect model which utilizes a nonparametric approach. In this section, we first present the basics of functional mixed-effect model and then give the estimation results of this model.

III.4.1. Basics

The functional mixed-effect model was first introduced in 2002 (Guo, 2002). For the considered data, the model can be written as below,

$$y_{ij} = \beta(t_j) + \alpha_i(t_j) + e_{ij},\tag{III.5}$$

where $e_{ij} \sim N(0, \sigma_e^2)$. Here we assume that both $\beta(t)$ and $\alpha_i(t)$ are smooth functions of t, and thus the estimated $\hat{y}_i(t)$ is also a smooth curve. The random errors e_{ij} are independent among individuals and time points. The variance-covariance matrix of the random effects $\{\alpha_i(t_1), \alpha_i(t_2), \dots, \alpha_i(t_n)\}$ is denoted by Σ_α. Intuitively, $\beta(t)$ represents

the population-average curve that tells the overall trend of the TILS effect, while $\alpha_i(t)$ is the individual-specific deviation from the population curve.

In this study, we model $\beta(t)$ and $\alpha_i(t)$ using smoothing splines. Specifically, for participant i, the estimated $\hat{\beta}(t)$ and $\hat{\alpha}_i(t)$ balance the fitting accuracy and smoothness by minimizing a penalized residual sum of squares,

$$\{\hat{\beta}(t), \hat{\alpha}_i(t)\} = argmin \sum_{j=1}^{n} \left[y_{ij} - \beta(t_j) - \alpha_i(t_j)\right]^2$$

$$+\lambda_\beta \int_{t_1}^{t_n} \beta''(t)^2 dt + \lambda_\alpha \int_{t_1}^{t_n} \alpha_i''(t)^2 dt. \tag{III.6}$$

In Eq. (III.6), λ_β and λ_α are tuning parameters that control the smoothness of the estimated curves. Details of the model estimation are given in Appendices F and G.

III.4.2. Fitting results

The estimated population-level effect $\hat{\beta}(t)$ and random effects $\hat{\alpha}_i(t)$ are given in Figure III-6. In Figure III-6(a), the population-level effect displays monotonic increase during the whole experiment period with a decreasing rate in the recovery stage. In Figure III-6(c), we could observe that for all participants, the observed values (black dots) fall close to the individual fits, which indicates good fitting performance. Also, the fitted individual curves do not have similar shapes but vary a lot from each other, which is different from those from parametric mixed-effect models shown in Figures III-3 and III-5. To provide some intuition on the fitting, Figure III-6(b) shows the population fit, individual fit and the estimated random effect of a single participant as an example. It can be seen clearly that the individual fit (black solid line) is a combination of the population effect and random effect.

60
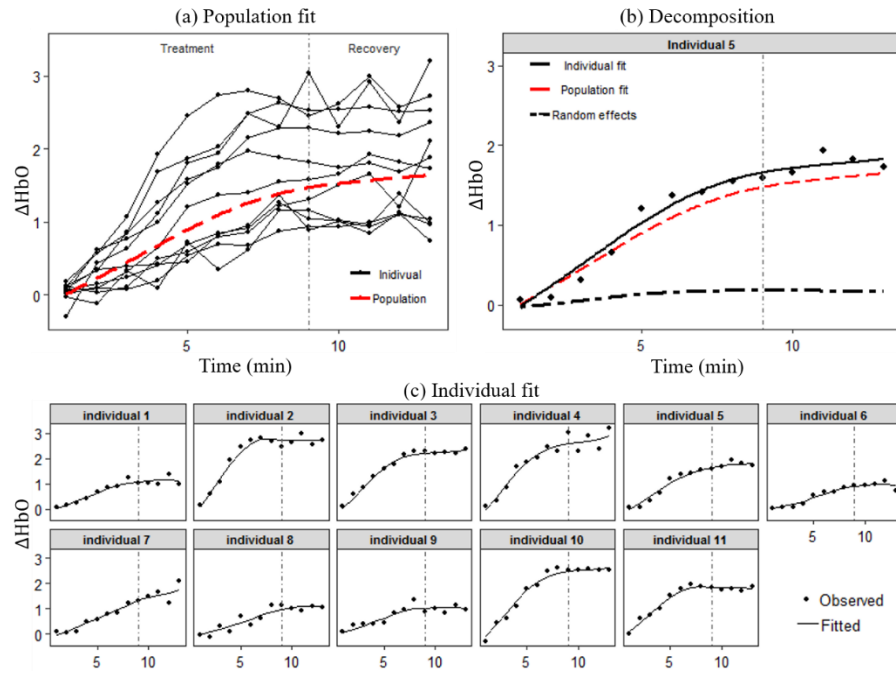
Figure III-6. Population and individual fitting of Model III.

The results tell us that the TILS treatment causes an increasing change in the oxygenated hemoglobin concentration in brain, and this effect sustains in the recovery stage after the treatment stops, though with a weakened rate. The individual variation among participants is significant as demonstrated by the fitted individual curves. Especially, their behaviors in the recovery stage are very different, some keeping increasing, some reaching a constant, and some going downward.

III.5. Model comparison

In this section we compare the three proposed models from three aspects: fitting performance, prediction performance and fitted variance pattern.

III.5.1. Fitting performance

Table III-2 lists the residual sum of squares and log likelihood of Models I, II, and III, which measure how well each model fits the TILS data. It is clear that Model III has the best fitting performance, with the smallest residual sum of squares and largest log likelihood. This is consistent with the visual impression of fitting performance in Figures III-3, III-5 and III-6. The two parametric models perform similarly in fitting.

Table III-2. Comparison of fitting performance of proposed models.

Model	Residual sum of squares	Log likelihood
Model I (Eq. (III.3))	4.508246	1.361531
Model II (Eq. (III.4))	4.977247	−3.189774
Model III (Eq. (III.5))	3.135987	147.9451

III.5.2. Prediction performance

In all the three models, individual variation is represented by the random effects and quantified by the estimated variances and covariances of the random effects. This makes the prediction of ΔHbO at individual level possible. Here we will compare the prediction performance of Models I, II and III through the following procedure. First, we split the data into a training set and a testing set. Specifically, the testing set is created by randomly selecting N_{test} observations among all participants and time points, and the training set is obtained after removing the testing set from the data. Each model is

estimated using the training set and then the estimated model is used to predict the values in the testing set. The performance of prediction is measured by the mean squared error (MSE), which is the mean of squared differences between the predicted values and actual observations.

To be robust to the effect of sample size and uncertainty in data splitting, five different sizes of the testing set $N_{test} = \{1, 5, 10, 20, 30\}$ are considered, and given each size, 100 runs are conducted and the average prediction error is computed. The results are shown in Figure III-7. Model III outperforms the other two models with the smallest average prediction error in all scenarios. Between the two parametric models, Model I exhibits relatively better prediction performance than Model II in most scenarios.
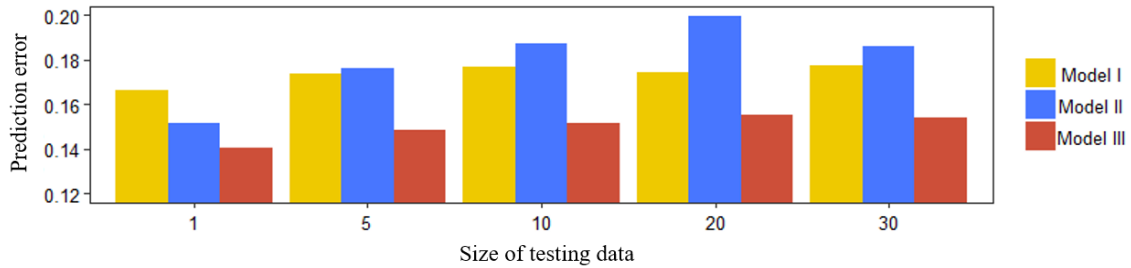
Figure III-7. Comparison of prediction performance of proposed models.

III.5.3. Variance pattern

We use $V(y)$ to denote the variance of ΔHbO. For parametric mixed-effect models, the variance is determined by both the Level-1 model and Level-2 model, and thus the resulting structure of variance is often complex. Here we derive the variance of Model I as an example for the purpose of comparing the fitted variance patterns of the proposed

63

models. Let $\rho(\cdot)$ be the correlation between two random effects. The variance of ΔHbO in

Model I is as below,

$$V(y) = Var\big(f(t; \boldsymbol{\theta})\big) + Var(e)$$

$$= t^3 \left(2\sigma_{u_1}^2 \sigma_{u_2}^2 \rho(u_1, u_2)\right) + t^2 \left(\sigma_{u_1}^2 + 2\sigma_{u_1}^2 \sigma_{u_3}^2 \rho(u_1, u_3)\right)$$

$$+t \left(\sigma_{u_2}^2 + 2\sigma_{u_2}^2 \sigma_{u_3}^2 \rho(u_2, u_3)\right) + \sigma_{u_3}^2 + \sigma_e^2 \qquad \text{(III.7)}$$

In contrast, the variance of Model III is determined by the random effect $\alpha(t)$ and random

error:

$$V(y) = Var\big(\alpha(t)\big) + Var(e) \qquad \text{(III.8)}$$
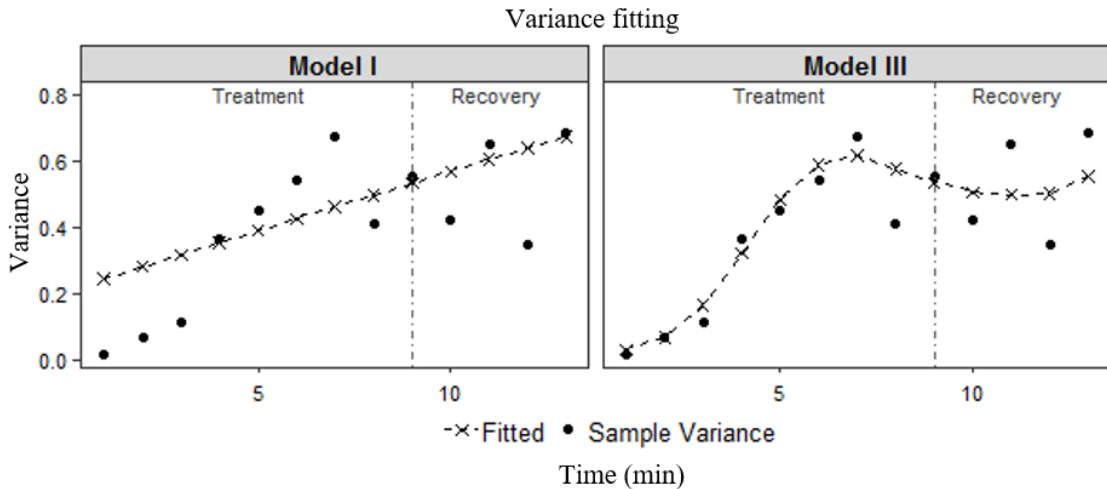
Figure III-8. Comparison of fitted variance patterns of proposed models.

Figure III-8 shows the variance pattern of the data and the fitted models. Sample

variance of all observations collected at each time point is used to represent the variance

of the data, and the fitted variance is obtained based on Eqs. (III.7) and (III.8). Again,

Model III performs better in fitting the variance pattern of the data. With small coefficients for the t^3 and t^2 terms in Eq. (III.7), the variance of Model I has a near-linear shape which does not match with the data.

III.6. Conclusion

In this study, we learn the effect of TILS treatment on the concentration of oxygenated hemoglobin in brain from experimental data by using three longitudinal data analysis methods. The result of repeated measures ANOVA validates that the TILS effect is significant. It also suggests that this effect should be modeled as a function of time and individual variation should be taken into account in the modeling. Following the idea, we propose two parametric mixed-effect models and a functional mixed-effect model to characterize the effect of TILS. By selecting appropriate parametric forms, the parametric mixed-effect models are able to provide desired approximation of the population-level effect of TILS and estimation of individual variation. However, the selection of function form in those models is subjective and could be misleading. In addition, such models lack the flexibility to fit the variance of the data. These issues are further addressed by the functional mixed-effect model. As a non-parametric method, it is free of the bias due to specification of model form and can let data speak.

According to the fitted functional mixed-effect model, the change in the concentration of oxygenated hemoglobin in brain caused by TILS increases as time goes on during the period when TILS is applied. Such an effect sustains even after TILS stops, though with a slower increasing rate. Another important fact is that large variation in the TILS effect exists among participants taking the treatment. These findings are useful

information to biomedical engineering researchers in the related field, which helps to reveal the underlying mechanism of TILS as a promising treatment and will direct future research toward deeper understanding of this treatment and better control of its clinical application.

There are a series of interesting topics for our future research on this topic. First, this study uses the concentration of oxygenated hemoglobin to illustrate the TILS effect on hemodynamics in brain. In the original experiment conducted (Wang et al. 2017), other response variables such as the concentration of deoxygenated hemoglobin and total hemoglobin are also measured. Similar approaches can be applied to those variables to obtain a comprehensive view about how TILS affects hemodynamics. Second, the effect of TILS on cerebral metabolism can also be assessed quantitatively as the change in cerebral concentration of oxidized cytochrome-c-oxidase (CCO), the terminal enzyme in the mitochondrial respiratory chain, are also measured in the experiment. In addition, we will explore the coupling effect between cerebral metabolism and blood oxygen supply via multivariate longitudinal data analysis.

III.7. References

Barrett, D. W. and Gonzales-Lima, F. (2013) Transcranial infrared laser stimulation produces beneficial cognitive and emotional effects in humans. *Neuroscience*, **230**, 13-23.

Bartos, A. et al. (2016) Pre-conditioning with near infrared photobiomodulation reduces inflammatory cytokines and markers of oxidative stress in cochlear hair cells. *Journal of Biophotonics*, **9**, 1125-1135.

Berk, M. (2018) sme: Smoothing-Splines Mixed-Effects Models. R package version 1.0.2.

Blanco, NJ. et al. (2015) Improving executive function using transcranial infrared laser stimulation. *Journal of Neuropsychology*, **11**(1), 14-25.

Breslow, N. E. and Clayton, D. G. (1993) Approximate inference in generalized linear mixed Models. *Journal of the American Statistical Association*, **88**, 9-25.

Disner, SG. et al. (2016) Transcranial laser stimulation as neuroenhancement for attention bias modification in adults with elevated depression symptoms. *Brain Stimulation*, **9**(5), 780-787.

Fitzmaurice, G. M. et al. (2008) *Longitudinal Data Analysis*, CRC Press.

Guo, W. (2002) Functional mixed effect models. *Biometrics*, **58**(1), 121-128.

Hamblin, M. R. et al. (2018) *Low-Level Light Therapy: Photobiomodulation.* The International Society for Optical Engineering.

Hedeker, D. and Gibbons, R. D. (2006) *Longitudinal data analysis*. Wiley-Interscience

Hennessy, M. and Hamblin,M. R. (2016) Photobiomodulation and the brain: a new paradigm. *Journal of Optics*, **19**(1).

Hwang, J. et al. (2016) Cognitive enhancement by transcranial laser stimulation and acute aerobic exercise. *Lasers in Medical Science*, **31**(6), 1151-1160.

Karu,T. (2013) Is it time to consider photobiomodulation as a drug equivalent? *Photomedicine and Laser Surgery*, **31**(5), 189-191.

Kuhn, E. and Lavielle, M. (2005) Maximum likelihood estimation in nonlinear mixed effects models. *Computational Statistics & Data Analysis*, **49**(4), 1020-1038.

Laird, N. M. and Ware, J. H. (1982) Random-effects models for longitudinal data. *Biometrics*, **38**(4), 963-974.

Lindstrom, M. J. and Bates, D. M. (1990) Nonlinear mixed effects models for repeated measures data. *Biometrics*, **46**(3), 673-687.

Naeser, M.A. and Hamblin. M. R. (2011) Potential for transcranial laser or LED therapy to treat stroke, traumatic brain injury and neurodegenerative disease. *Photomedicine and Laser Surgery*, **29**(7), 443-446.

Wang, X. et al. (2016) Interplay between up-regulation of cytochrome-c-oxidase and hemoglobin oxygenation induced by near-infrared laser. *Scientific Report*, **6**, 30540.

Wang, X. et al. (2017) Up-regulation of cerebral cytochrome-c-oxidase and hemodynamics by transcranial infrared laser stimulation: A broadband near-infrared spectroscopy study. *Journal of Cerebral Blood Flow & Metabolism*, **37**(12), 3789–3802.

CHAPTER IV

PHASE I MONITORING OF OPTICAL PRO-FILES WITH APPLICATION IN LOW-

EMITTANCE GLASS MANUFACTURING[1]

IV.1. Introduction

During the past two decades, concerns about energy and environment have led to many improvements in house and building design such as increased energy efficiency of windows. One critical technical innovation that makes this possible is the low-emittance (low-E) glass. This type of glass is manufactured through a coating process as illustrated in the upper panel of Figure IV-1, where in the coating chambers solid materials, e.g., metal, metal oxide and metal nitride, are deposited on the surface of flat glass ribbons. The coating will enhance the thermal/optical performance of the glass, making it to reflect up to 90% of infrared radiation while allowing visible light to enter (Arasteh et al., 2004; Carmody et al., 1996). Thus, low-E glass can reduce unwanted heat gain in summer and heat loss in winter. Due to its great benefit in energy savings, such glass has become more and more popular (Carmody and Haglund, 2012).

One quality concern in low-E glass manufacturing is the coating uniformity on the glass surface. In the current quality inspection practice illustrated in the lower panel of Figure IV-1, laser scanning is conducted on different locations of finished products to

69

measure optical properties such as reflectance, i.e., the percentage of light that reflects from the surface. The measurements at each location form an optical profile, where the x-axis denotes wavelength λ. Each profile is then summarized into two color scales, a^* (degree of redness/greenness) and b^* (degree of yellowness/blueness) (Hunter Lab, 2008), which are monitored by Shewhart control charts. The color scales are simple summaries of information contained in optical profiles and thus may not perform well in detecting changes. This study aims to advance quality control in low-E glass industry by using optical profiles directly. The coating uniformity of products consists of two aspects: the spatial uniformity of optical profiles at different locations of the same product and the between-product uniformity at the same location. We will focus on the latter in this study, i.e., we consider one profile from each product. Phase I analysis will be investigated, where a historical dataset is available and the purpose is to screen out outlying profiles in the dataset.
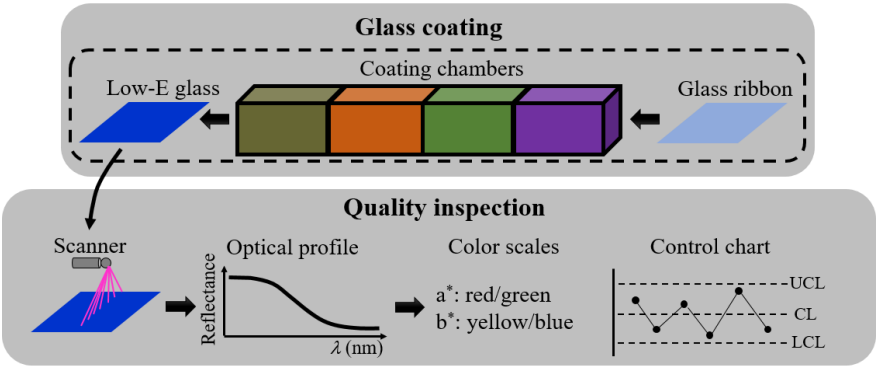
Figure IV-1. The coating process of low-E glass and quality inspection

Figure IV-2 shows a collection of 191 real optical profiles from low-E glass manufacturing, each containing 125 equally-spaced sampling points. The shape and variation of the data exhibit the following characteristics: (i) As shown in Fig. 2(a) and Fig. 2(b), the profile has a complex shape, with a high level of smoothness and small random noises, and (ii) the difference between profiles mainly lies in their shapes, as can be seen in Fig. 2(c). These characteristics need to be taken into account when looking for a suitable methodology for monitoring optical profile data.
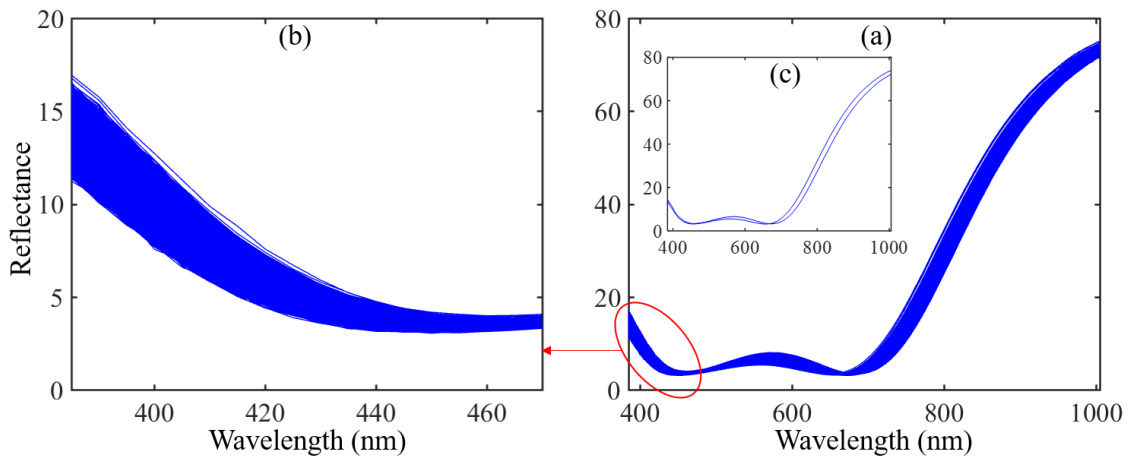
Figure IV-2. (a) A collection of 191 optical profiles, (b) zoom-in view of profiles at spectrum range 385-470 nm and (c) two single profiles in the dataset

The monitoring of quality profiles has been well studied recently; see the reviews of Woodall et al. (2004) and Noorossana et al. (2011). A large amount of work in this field deals with simple profiles that can be described by linear models (e.g., Kang and Albin, 2000; Mahmound, 2008) or polynomial models (e.g., Kazemzadeh et al., 2008, 2009). A few studies have been done on complex nonlinear profiles, which are potentially useful

for our problem. Specifically, we will take a look at those considering Phase I monitoring for the same purpose as this study. For Phase II monitoring of nonlinear profiles, see Zou et al. (2008) and Qiu et al. (2010).

The existing methods can be categorized into three groups. The first group treats each profile as a high-dimensional vector and uses dimension reduction methods such as principal component analysis and independent component analysis, and monitors the reduced data (e.g., Chang et al., 2012; Ding et al., 2006; Paynabar et al., 2013; Zhang and Albin, 2009; Zou, et al., 2014). The second group fits a nonparametric regression model such as B-spline to each profile, and monitors summary statistics of deviations from the fitted model (e.g., Chang and Yadama, 2010; Chou, et al., 2014). The third group fits a parametric or nonparametric regression model to the profiles, and monitors the estimates of parameters (e.g., Abdel-Salam et al., 2013; Jensen and Birch, 2009; Jensen et al. 2016; Mosesova et al. 2007; Williams et al., 2007a; Williams et al., 2007b). The first two groups are designed for a general category of nonlinear profiles which typically have irregular shapes and/or special features such as sharp peaks and jaggedness. In contrast, the optical profiles, shown in Figure IV-2, are very smooth and can be adequately characterized by regular regression models. Therefore, the third group is most relevant to our problem and will be reviewed in detail as follows.

For modeling the shape of profiles, Jensen and Birch (2009), Jensen et al. (2016), Williams et al. (2007a), and Williams et al. (2007b) choose parametric nonlinear regression models (e.g., four-parameter logistic model); Mosesova et al. (2007) provide a nonparametric regression model (B-spline) as an alternative; and Abdel-Salam et al.

(2013) consider a hybrid of parametric model and non-parametric model (p-spline) to be robust. For modeling the between-profile variation, Jensen et al. (2016) and Williams et al. (2007b) assume a normally distributed random error term to represent all variation in data; Williams et al. (2007a) build a regression model for the variance of data; and Abdel-Salam et al. (2013), Jensen and Birch (2009), and Mosesova et al. (2007) consider random effects of parameters to characterize the within-profile correlation in addition to the random error term. All those studies use a T^2 control chart to monitor the estimates of model parameters or random effects for change detection.

In this study, we propose a *piecewise polynomial random-coefficient (PRC) model*, called the *piecewise PRC model* later, for optical profiles. That is, the spectrum will be segmented into a number of pieces and a PRC model is fitted to the profile segments on each piece. A random-coefficient model is a special type of mixed-effect models where each coefficient has a random effect (West et al., 2015). For Phase I analysis, estimates of random effects in each piece will be monitored separately using a T^2 control chart as in previous work.

The proposed method uses a piecewise polynomial model to characterize the shape of optical profiles. There are two reasons for this. First, compared to the global modeling strategies (using parametric nonlinear models or nonparametric models) in most studies, monitoring a complex profile in segments will facilitate subsequent diagnosis efforts. This makes particular sense for the considered application. In low-E glass coating as illustrated in the upper panel of Figure IV-1, the deposition material in each chamber affects the resulting optical profile at certain range of the spectrum. Therefore, when change is

detected in a piece, engineers only need to examine related chambers to find root cause of the change. In fact, the segmentation strategy is also adopted by Jin and Shi (1999) and Chang and Yadama (2010), where profiles are divided either subjectively or by engineering knowledge. A heuristic procedure for profile segmentation based on data will be provided in this study which is useful when engineering knowledge is not available. Second, as shown in a proof-of-concept exploration on optical profiles (Zeng and Chen, 2015), a polynomial model is adequate for a profile segment. In addition, the simple polynomial model which is easy to estimate is beneficial for online monitoring. In low-E glass manufacturing, optical profiles are generated very fast (fifteen seconds between scannings). The computation problems in estimation of parametric nonlinear or nonparametric models (Jensen and Birch, 2009; Abdel-Salam et al. 2013) will pose challenges for monitoring large streams of optical profiles in practice.

As optical profiles have strong within-profile correlation, a mixed-effect model is used here. We choose a random-coefficient model specifically to reflect differences in the shape of profiles which accounts for the major part of the between-profile variation. Moreover, while most studies assume independent random effects, we consider correlated random effects and show its advantages in fitting. Furthermore, we investigate a potential problem caused by *highly correlated random effects* (HCRE). The rationale of this problem and its detrimental impacts on Phase I monitoring are revealed, and a remedy based on regressor transformation is provided. To the best of our knowledge, this work is the first attempt to understand and address problems associated with correlated random effects in model estimation and model-based monitoring.

74

It is also worth mentioning that optical profile data have been widely used in various applications, e.g., biomaterial fabrication (Fernandez-Oliveras et al., 2013), semiconductor manufacturing (Gaponenko, 1998), and food industry (Zude, 2009), to measure product quality. Although the proposed method is illustrated using data from low-E glass manufacturing, it can also be applied or easily extended to data from other applications.

The remainder of the paper is organized as follows. We will first present the proposed methodology for Phase I monitoring of optical profiles, including the definition of the piecewise PRC model, model construction and change detection. Then we will describe the HCRE problem and the proposed remedy. This is followed by a numerical study and a case study using real data from low-E glass manufacturing. Finally, we conclude this study and discuss future research.

IV.2. Proposed methodology for Phase I monitoring of optical profiles

IV.2.1. The piecewise PRC model

Suppose there are m optical profiles in a historical dataset, each containing N design points. We split the spectrum into a number of pieces and fit a PRC model to the profile segments on each piece. Given a piece with n design points, the segments are modeled by

$$\mathbf{y}_i = \mathbf{X}(\boldsymbol{\beta} + \mathbf{b}_i) + \boldsymbol{\varepsilon}_i \text{ for } i = 1, 2, \cdots m, \tag{IV.1}$$

where $\mathbf{y}_i = [y_{i1}, y_{i2}, \ldots, y_{in}]'$ is the $n \times 1$ vector of responses (e.g., reflectance of light) for the ith profile, $\mathbf{X} = [1, \mathbf{x}, \ldots, \mathbf{x}^p]$ is the $n \times (p + 1)$ matrix of polynomial regressors where $\mathbf{x}$ is the vector of design points on the spectrum, and $\boldsymbol{\varepsilon}_i \sim MN(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I})$ is the vector

of random errors where $\mathbf{0}$ is the vector of 0s, $\mathbf{I}$ is the identity matrix, and "MN" denotes multivariate normal distribution. Note that we assume the design points on each profile are the same, so $\mathbf{X}$ is constant for all profiles (thus the profile index is omitted). The coefficient of the regressors has two parts: the vector of fixed effects $\boldsymbol{\beta} = [\beta_0, \beta_1, \dots, \beta_p]'$ that is the same for all profiles, and the vector of random effects for the ith profile $\mathbf{b}_i = [b_{i0}, b_{i1}, \dots, b_{ip}]'$. For the random effects, we assume

$$\mathbf{b}_i \sim MN(\mathbf{0}, \mathbf{G}), \tag{IV.2}$$

where $\mathbf{G}$ is the variance-covariance matrix of $\mathbf{b}_i$, which has nonzero off-diagonal elements since correlated random effects are used.

The model in Eq. (IV.1) is a random-coefficient model as each coefficient has a random effect. Since coefficients characterize the shape of a profile, the random effects reflect the between-profile variation which mainly lies in the shape of profiles. The terms in the model have an easy interpretation: the fixed effects $\boldsymbol{\beta}$ represent the base-line regressor-response relationship, the random effects $\mathbf{b}_i$ represent the intrinsic variation in low-E glass manufacturing, and the random errors $\boldsymbol{\varepsilon}_i$ represent noises in the process and optical measurement. It is also worth mentioning that as optical profiles are very smooth, the random errors are small compared to the random effects.

IV.2.2. Profile segmentation

Segmenting profiles is similar to identifying knots in curve fitting studies (Hastie et al., 2009). However, our problem is more complex: curve fitting typically deals with a single curve (i.e., profile) and only considers the shape of the curve; process monitoring focuses on between-profile variations and thus takes both shapes and variation patterns of

the profiles into account. Here we propose a simple method for profile segmentation, which is described as follows.

We first point out some *basic* properties of the PRC model. From Eq. (IV.1), we can obtain the pointwise expression of the model

$$y_{ij} = X_j(\boldsymbol{\beta} + \mathbf{b}_i) + \varepsilon_{ij} \text{ for } j = 1, 2, \dots, n, \tag{IV.3}$$

where y_{ij} is the jth response on the ith profile, and X_j is the jth row of the design matrix $\mathbf{X}$. Consequently, the pointwise mean and variance are

$$\mu_{ij} = X_j\boldsymbol{\beta}, \quad v_{ij} = X_j\mathbf{G}X_j^T + \sigma_\varepsilon^2. \tag{IV.4}$$

Since μ_{ij} and v_{ij} are functions of x, we can call them the "mean profile" and "variance profile". By Eq. (IV.4), these two quantities have the following properties: (i) The mean profile depends on the fixed effects, while the variance profile depends on the variances/covariances of the random effects. In other words, differences in any two PRC models, whether in fixed effects or random effects, will be reflected by their mean and variance profiles. (ii) They are polynomial functions of x. These suggest that the segmentation can be made by dividing the (sample) mean profile and variance profile such that each piece of them can be modeled by a polynomial function.

This idea is illustrated in Figure IV-3, where plot (a) shows a set of $m = 200$ profiles simulated from PRC models, and plots (b) and (c) are their (sample) mean profile and variance profile respectively. The dataset contains two parts. The part on $x = [-1, 0]$ comes from a quadratic PRC model, and the part on $x = [0, 1]$ comes from another quadratic PRC model. We can see that the difference in the fixed effects of the two models

77

is reflected by shapes of the mean profile on the two intervals of x, while the difference in their variances/covariances of random effects is reflected by the variance profile similarly. Assuming the true structure of data is unknown, we can split the mean and variance profiles at their empirical stationary points (i.e., where the slope is zero, as marked in plots (b) and (c)), which results in segments from the two models.
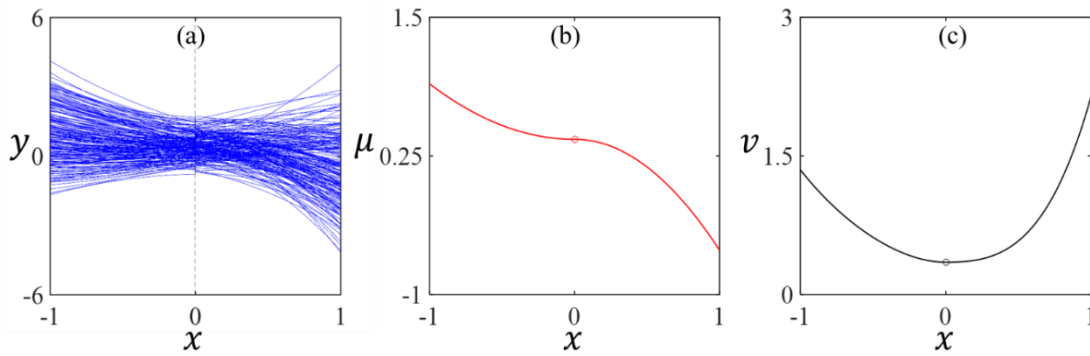
Figure IV-3. Simulated profiles and their mean profile and variance profile

Based on the above idea, the proposed procedure for profile segmentation is as follows:

Step 1. Obtain the mean profile and variance profile of the optical profiles.

Step 2. Identify the empirical stationary points, say $x_1^\mu, x_2^\mu, ...$ and $x_1^v, x_2^v, ...$, of the mean and variance profiles, and sort them in ascending order. These points are used as the initial breaking points, which tend to divide the mean and variance profiles into pieces of simple shapes that can be described adequately by polynomial models. Considering that the polynomial order of the variance profile is twice that of the mean profile (by Eq.

(IV.4)), we will limit to cubic or lower-order polynomial models in fitting each piece of the mean and variance profiles.

Step 3. Combine some adjacent pieces by removing their boundaries to minimize the number of pieces. Residual sum of squares (RSS) of fitting will be used as a stopping rule,

$$RSS^\mu = \Sigma_{j=1}^N \left(\bar{y}_j - \hat{\bar{y}}_j\right)^2, \ RSS^v = \Sigma_{j=1}^N \left(s_j^2 - \hat{s}_j^2\right)^2,$$

where $\bar{y}_j$ is the jth point of the mean profile, $\hat{\bar{y}}_j$ is the corresponding fitted value by a polynomial model, s_j^2 is the jth point of the variance profile, and $\hat{s}_j^2$ is its fitted value. A boundary can be removed only when it will not cause considerable increase (e.g., $\leq 1\%$) to RSS^μ or RSS^v.

Step 4. Adjust locations of the breaking points resulted from last step within a small neighborhood to improve the fitting. The adjustment stops when there is no considerable further reduction on RSS (e.g., $> 1\%$). The spectrum will be cut into pieces using the resulting breaking points.

IV.2.3. Model estimation

A PRC model will be fitted to the profile segments at each piece of spectrum. Let the variance-covariance matrix of random effects be

$$\mathbf{G} = \begin{bmatrix} \tau_0^2 & \tau_0\tau_1 r_{01} & \cdots & \tau_0\tau_p r_{0p} \\ \tau_1\tau_0 r_{01} & \tau_1^2 & \cdots & \tau_1\tau_p r_{1p} \\ \vdots & \vdots & \vdots & \vdots \\ \tau_p\tau_0 r_{0p} & \tau_p\tau_1 r_{1p} & \cdots & \tau_p^2 \end{bmatrix}, \quad\quad (IV.5)$$

where $\tau_0^2, \tau_1^2, \ldots, \tau_p^2$ are variances of the random effects $b_{i0}, b_{i1}, \ldots, b_{ip}$ ($\tau_0, \tau_1, \ldots, \tau_p$ are positive), and $r_{01}, r_{02}, \ldots r_{p-1,p}$ are the $p(p+1)/2$ correlation coefficients between the

random effects. Defining $\boldsymbol{\theta} = (\tau_0^2, \tau_1^2, \ldots, \tau_p^2, r_{01}, r_{02}, \ldots, r_{p-1,p}, \sigma_\varepsilon^2)$, the unknowns in the

PRC model include the fixed effects $\boldsymbol{\beta}$, the random effects $\mathbf{b}_i, i = 1, \ldots, m$, and $\boldsymbol{\theta}$. The

maximum likelihood (ML) estimator of the unknowns has a negative bias due to the loss

of degrees of freedom in estimating the fixed effects (Patterson and Tompson, 1971). The

restricted maximum likelihood (REML) method (Jiang, 2007) can solve this problem by

transforming the response to remove the fixed effects. This method will be used for the

estimation, the details of which are given below.

From Eq. (IV.1), the marginal distribution of the response is

$$\mathbf{y}_i \sim MN(\mathbf{X}\boldsymbol{\beta}, \mathbf{V}), \;\; \mathbf{V} = \mathbf{X}\mathbf{G}\mathbf{X}^T + \sigma_\varepsilon^2 \mathbf{I}. \tag{IV.6}$$

Let $\mathbf{A}$ be a matrix from the null space of $\mathbf{X}$ (i.e., $\mathbf{A}^T\mathbf{X} = \mathbf{0}$). The REML method

transforms the response to have zero mean and involve only $\mathbf{V}$,

$$\mathbf{A}^T\mathbf{y}_i \sim MN(\mathbf{0}, \mathbf{A}^T\mathbf{V}\mathbf{A}). \tag{IV.7}$$

Since $\mathbf{V}$ is a function of $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}}$ can be found by maximizing the likelihood of the

multivariate normal distribution in Eq. (IV.7). Then $\widehat{\boldsymbol{\beta}}$ and $\widehat{\mathbf{b}}_i$ can be obtained by

$$\widehat{\boldsymbol{\beta}} = \left(m\mathbf{X}^T\widehat{\mathbf{V}}^{-1}\mathbf{X}\right)^{-1}\left(\mathbf{X}^T\widehat{\mathbf{V}}^{-1}\sum_{i=1}^{m}\mathbf{y}_i\right), \tag{IV.8}$$

$$\widehat{\mathbf{b}}_i = \widehat{\mathbf{G}}\mathbf{X}^T\widehat{\mathbf{V}}^{-1}\left(\mathbf{y}_i - \mathbf{X}\widehat{\boldsymbol{\beta}}\right). \tag{IV.9}$$

Calculating $\widehat{\boldsymbol{\theta}}$ requires numerical optimization algorithms (Gumedze and Dunne,

2011). The stochastic approximation expectation maximization (SAEM) method (Delyon,

Lavielle and Moulines, 1999) is used here. For notational simplicity, we will consider

estimation using a single profile (i.e. $\mathbf{y}_i$); estimation using all the m profiles is similar

(note that $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m$ are independent samples from the same multivariate normal distribution).

SAEM follows a similar scheme as the traditional EM algorithm that performs E-step and M- step iteratively to update estimates. Treating the random effects in the PRC model as missing data, the complete data are $(\mathbf{y}_i, \mathbf{b}_i)$. Correspondingly, the log-likelihood function is

$$l(\boldsymbol{\theta}|\mathbf{y}_i, \mathbf{b}_i) = \log p(\mathbf{y}_i, \mathbf{b}_i|\boldsymbol{\theta}) = \log p(\mathbf{b}_i|\boldsymbol{\theta}) + \log p(\mathbf{y}_i|\boldsymbol{\theta}, \mathbf{b}_i), \qquad \text{(IV.10)}$$

where $p(\cdot)$ is the density function, and the two terms in the right are given by

$$\log p(\mathbf{b}_i|\boldsymbol{\theta}) = -\frac{1}{2}[\log(|\mathbf{G}|) + \mathbf{b}_i^T \mathbf{G}^{-1} \mathbf{b}_i], \qquad \text{(IV.11)}$$

$$\log p(\mathbf{y}_i|\boldsymbol{\theta}, \mathbf{b}_i) = -\frac{1}{2}\left[n\log(\sigma_\varepsilon^2) + \frac{(\mathbf{y}_i - \mathbf{X}\boldsymbol{\beta} - \mathbf{X}\mathbf{b}_i)^T(\mathbf{y}_i - \mathbf{X}\boldsymbol{\beta} - \mathbf{X}\mathbf{b}_i)}{\sigma_\varepsilon^2}\right] \qquad \text{(IV.12)}$$

<u>E-step</u>: Given the current estimate $\widehat{\boldsymbol{\theta}}^{(k)}$ in the kth iteration, EM calculates the expectation of the log-likelihood in Eq. (IV.10) with respect to $p(\mathbf{b}_i|\widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{y}_i)$ by

$$E_{\mathbf{b}_i|\widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{y}_i}[l(\boldsymbol{\theta}|\mathbf{y}_i, \mathbf{b}_i)] = -\frac{1}{2}\left[\log(|\mathbf{G}^{(k)}|) + tr\left(\mathbf{V}_b^{(k)} \mathbf{G}^{(k)-1}\right) + \boldsymbol{\mu}_b^{(k)T} \mathbf{G}^{(k)-1} \boldsymbol{\mu}_b^{(k)}\right]$$

$$-\frac{1}{2}\left[n\log(\sigma_\varepsilon^2) + \frac{tr\left(\mathbf{V}_\varepsilon^{(k)}\right) + \boldsymbol{\mu}_\varepsilon^{(k)T} \boldsymbol{\mu}_\varepsilon^{(k)}}{\sigma_\varepsilon^2}\right] \qquad \text{(IV.13)}$$

where

$$\boldsymbol{\mu}_b^{(k)} = \mathbf{G}^{(k)} \mathbf{X}^T \mathbf{P}^{(k)} \mathbf{y}_i,$$

$$\mathbf{V}_b^{(k)} = \mathbf{G}^{(k)} - \mathbf{G}^{(k)} \mathbf{X}^T \mathbf{P}^{(k)} \mathbf{X} \mathbf{G}^{(k)},$$

$$\boldsymbol{\mu}_\varepsilon^{(k)} = \sigma_\varepsilon^{2(k)} \mathbf{P}^{(k)} \mathbf{y}_i, \quad \mathbf{V}_\varepsilon^{(k)} = \sigma_\varepsilon^{2(k)} - \sigma_\varepsilon^{4(k)} \mathbf{P}^{(k)},$$

$$\mathbf{G}^{(k)} = \mathbf{G}\left(\widehat{\boldsymbol{\theta}}^{(k)}\right), \mathbf{V}^{(k)} = \mathbf{V}\left(\widehat{\boldsymbol{\theta}}^{(k)}\right),$$

$$\sigma_\varepsilon^{2(k)} = \sigma_\varepsilon^2\left(\widehat{\boldsymbol{\theta}}^{(k)}\right), \mathbf{P}^{(k)} = \mathbf{A}(\mathbf{A}^T \mathbf{V}^{(k)} \mathbf{A})^{-1} \mathbf{A}.$$

One drawback of the EM algorithm is that computing the expectation in Eq. (IV.13) may be difficult and time consuming. To alleviate this problem, the SAEM algorithm approximates the expectation in the kth iteration by sampling the missing data (i.e., $\mathbf{b}_i$) based on the estimate in the $(k-1)th$ iteration. Specifically, the expectation in Eq. (IV.13) is replaced by $Q(\boldsymbol{\theta})$ which has an iterative formula

$$Q^{(k)}(\boldsymbol{\theta}) = Q^{(k-1)}(\boldsymbol{\theta}) + \gamma_k \left[l\left(\boldsymbol{\theta} \middle| \mathbf{y}_i, \mathbf{b}_i^{(k)}\right) - Q^{(k-1)}(\boldsymbol{\theta}) \right], \qquad \text{(IV.14)}$$

where $\mathbf{b}_i^{(k)}$ is sampled from an ergodic Markov chain with limiting distribution $p(\mathbf{b}_i|\widehat{\boldsymbol{\theta}}^{(k-1)}, \mathbf{y}_i)$ and transition probability $\pi_{\widehat{\boldsymbol{\theta}}^{(k-1)}}$, $l\left(\boldsymbol{\theta} \middle| \mathbf{y}_i, \mathbf{b}_i^{(k)}\right)$ is the likelihood function when $\mathbf{b}_i^{(k)}$ is used, and γ_k is a decreasing sequence satisfying $\sum \gamma_k = +\infty, \ \sum \gamma_k^2 < +\infty$.

M-step: Find the updated estimate of $\boldsymbol{\theta}$ that maximizes $\widehat{\boldsymbol{\theta}}^{(k+1)}$ in Eq. (IV.14)

$$\widehat{\boldsymbol{\theta}}^{(k+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \, Q^{(k)}(\boldsymbol{\theta}). \qquad \text{(IV.15)}$$

According to our preliminary study, the SAEM is much faster than the EM in general. For example, EM takes hours to estimate a cubic PRC model, while SAEM only takes minutes to obtain similar results. We also find that one explanation for EM's poor performance is the correlation of random effects; it works rather fast in the independent case. Since correlated random effects are assumed in this study, SAEM is a better choice.

IV.2.4. Change detection

Since random effects indicate the intrinsic process variation, the REML estimates of $\mathbf{b}_i$ will be monitored using a T^2 control chart for change detection. In Phase I analysis, we calculate the T^2 statistic based on the successive-difference estimator of the variance-covariance matrix (Holmes and Mergen, 1993)

$$T_i^2 = \hat{\mathbf{b}}_i^T \hat{\mathbf{G}}^{-1} \hat{\mathbf{b}}_i \quad \text{for } i = 1, 2, \ldots, m, \qquad \text{(IV.16)}$$

where $\hat{\mathbf{G}} = \frac{\mathbf{D}^T\mathbf{D}}{2(m-1)}$, $\mathbf{D} = [(\hat{\mathbf{b}}_2 - \hat{\mathbf{b}}_1), \ldots, (\hat{\mathbf{b}}_m - \hat{\mathbf{b}}_{m-1})]^T$. This version of T^2 statistic performs better in detecting sustained step shifts in the mean than the version based on the sample variance-covariance matrix of random effects estimates (Sullivan and Woodall, 1996).

When $m > (p+1)^2 + 3(p+1)$, the upper control limit is $\chi^2_{p+1,\alpha}$, where α is the significance level. Assume there are L control charts, one for each piece, we can set the overall significance level $\alpha_{overall}$ to be a certain value, and then the significance level of each chart is $\alpha_{chart} = 1 - (1 - \alpha_{overall})^{\frac{1}{L}}$. Since each chart monitors m statistics, this requires $\alpha = 1 - (1 - \alpha_{chart})^{\frac{1}{m}}$. If any of these control charts signals, the whole profile that contains the outlying segment will be removed. Then models will be re-estimated using the remaining data and the T^2 statistics will be re-calculated based on the updated model estimates. This repeats until no signal is generated from the control charts. Note that the specification of the overall significance level depends on the number of pieces. In general, when L is larger, $\alpha_{overall}$ needs to be higher to enable a reasonable significance level for each chart. When $2 \leq L \leq 10$ (which holds in most cases of optical profile monitoring), $\alpha_{overall} = 0.1$ is suggested, which leads to $0.01 \leq \alpha_{chart} \leq 0.05$.

In online (i.e., Phase II) monitoring, the statistic for a newly observed profile $\mathbf{y}_0$ is

$$T_0^2 = \hat{\mathbf{b}}_0^T \hat{\mathbf{G}}^{-1} \hat{\mathbf{b}}_0, \qquad \text{(IV.17)}$$

where $\hat{\mathbf{b}}_0 = \hat{\boldsymbol{\beta}}_0 - \hat{\boldsymbol{\beta}}$. $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{G}}$ are the final estimates of fixed effects and variance-covariance matrix of random effects from Phase I analysis, and $\hat{\boldsymbol{\beta}}_0$ is the least squares estimate of coefficients in the $(p+1)$-degree polynomial model of $\mathbf{y}_0$.

IV.3. The highly-correlated-random-effects problem and regressor transformation

IV.3.1. The HCRE problem and its impacts

The variance-covariance matrix of the random effects in Eq. (IV.5) can be decomposed as

$$
\mathbf{G} = \boldsymbol{\tau} \times \mathbf{R} \times \boldsymbol{\tau} =
\begin{bmatrix}
\tau_0 & & & \\
& \tau_1 & & \\
& & \ddots & \\
& & & \tau_p
\end{bmatrix}
\cdot
\begin{bmatrix}
1 & r_{01} & \cdots & r_{0p} \\
r_{01} & 1 & \cdots & r_{1p} \\
\vdots & \vdots & \vdots & \vdots \\
r_{0p} & r_{1p} & \cdots & 1
\end{bmatrix}
\cdot
\begin{bmatrix}
\tau_0 & & & \\
& \tau_1 & & \\
& & \ddots & \\
& & & \tau_p
\end{bmatrix}
\qquad \text{(IV.18)}
$$

where $\mathbf{R}$ is the correlation matrix of the random effects. When some random effects have high correlations, which is very likely to happen for the smooth optical profiles, the HCRE problem arises. For example, if b_{i0} and b_{i1} are highly correlated, $r_{01} \approx 1$, $r_{02} \approx r_{12}, \dots, r_{0p} \approx r_{1p}$, giving a correlation matrix

$$
\mathbf{R} =
\begin{bmatrix}
1 & \approx 1 & r_{02} & \cdots & r_{0p} \\
\approx 1 & 1 & \approx r_{02} & \cdots & \approx r_{0p} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
r_{p0} & r_{p1} & r_{p2} & \cdots & 1
\end{bmatrix},
$$

which is near singular as the first two rows are similar (or, nearly linearly dependent). As a result, $\mathbf{G}$ is near singular and $\mathbf{G}^{-1}$ is ill-conditioned.

The HCRE problem will impact the monitoring of optical profiles in three aspects: (i) *Model estimation*. The EM algorithm may not converge as the expectation of log-likelihood in Eq. (IV.13) depends on $\mathbf{G}^{-1}$. When the SAEM algorithm is used, this problem still exists because it involves computing the log-likelihood function which depends on $\mathbf{G}^{-1}$ too (see Eqs. (IV.10), (IV.11) and (IV.14)). In general, the higher the correlation level among the random effects, the more serious the non-convergence problem in model estimation. (ii) *Change detection*. Note that the T^2 statistics in Eqs.

(IV.16) and (IV.17) contain $\mathbf{G}^{-1}$. Thus, the HCRE problem will also make it difficult to apply the control chart. (iii) *Interpretability*. Highly correlated random effects represent similar information on the process variation, and thus monitoring them simultaneously makes little sense and cannot lead to easily interpretable results for root cause identification.

IV.3.2. Regressor transformation

A similar problem, which is due to correlated coefficient estimates, exists in the literature of polynomial regression and profile monitoring based on polynomial models. Popular remedies for the problem are centering and other transformations of the regressors. For example, Kazemzadeh et al. (2009) use the orthogonal polynomial model (OPM) to replace the regular polynomial model in profile monitoring. The regressors in the OPM are transformed polynomials that are orthogonal to each other; the corresponding coefficient estimates are independent so that they can be monitored separately using univariate control charts.

There are also a few studies on correlated random effects in linear mixed-effect (LME) models. Morrell et al. (1997) show how linear transformations of regressors affect the random effects. Zhang and Chen (2013) propose an adaptive algorithm for fitting LME models, which applies location transformation to the regressor to make resulting random effects independent. However, both of the two studies focus on simple LME models with a single regressor, so their results cannot be easily generalized to the PRC model in Eq. (IV.1) which has multiple regressors. Compared to general LME models, our model has another special aspect: the regressors are polynomials which themselves might be highly

correlated; as a result, the transformation of each regressor will not only affect its associated random effect, but also affect the random effects of other regressors.

In this study, we propose a regressor transformation method to solve the HCRE problem. We first show two simulation examples on the effect of centering as the simplest form of location transformation to provide some insights. Consider a quadratic PRC model with random effects b_{i0}, b_{i1}, b_{i2}, and assume the original design points are evenly distributed at $x = [0, 0.2, 0.4, ..., 6.0]$. In each example, $m = 100$ profiles are generated from the model, and then estimates of the random effects and their pairwise correlations are obtained from the data. Then the values of x are centered, i.e., $x' = [-3.1, -2.9, -2.7, ..., 2.9]$, with the response unchanged, and model estimation is performed based on the transformed data in a similar way. In Figure IV-4, the upper row displays the profiles before and after the centering, and the lower row shows the corresponding pairwise scatter plots of the estimated random effects (histograms are given in the diagonal).
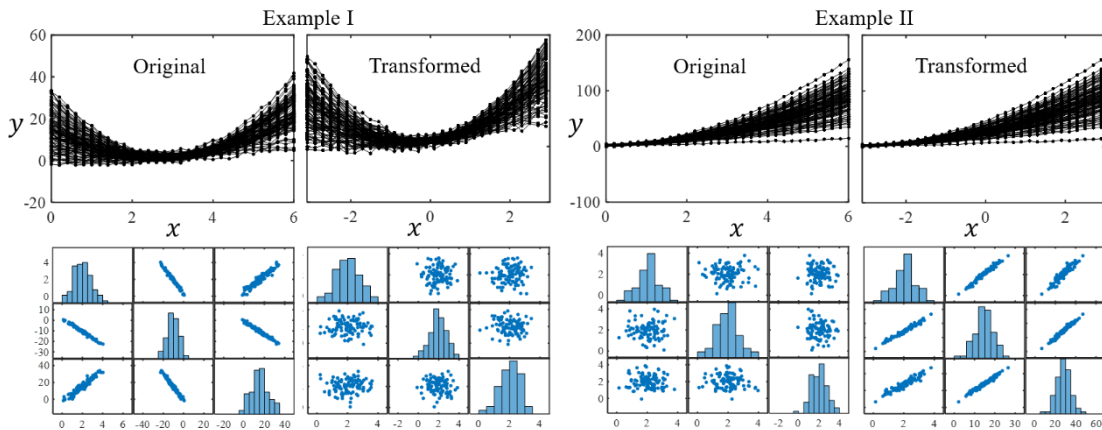
Figure IV-4. The effect of centering on correlations of random effects

In Example I, the original random effects are highly correlated, as indicated by the line patterns in the scatter plots, while after centering, they become independent, as indicated by the random clouds in the scatter plots. Example II shows the opposite: the independent random effects become highly correlated after the centering. These results suggest that the location transformation of regressors can significantly change the correlation structure of random effects. On the other hand, a simple transformation like centering may not necessarily lead to reduction of such correlations; more flexible transformations are needed to solve the HCRE problem.

This problem can be formulated as follows. Using the pointwise expression of the PRC model in Eq. (IV.3), the original model is

$$y_{ij} = (\beta_p + b_{ip})x_j^p + \cdots + (\beta_1 + b_{i1})x_j + (\beta_0 + b_{i0}) + \varepsilon_{ij}. \qquad \text{(IV.19)}$$

We can apply a location shift to each polynomial regressor, giving the transformed model

$$y'_{ij} = (\beta'_p + b'_{ip})(x_j - u_p)^p + \cdots + (\beta'_1 + b'_{i1})(x_j - u_1) + (\beta'_0 + b'_{i0}) + \varepsilon'_{ij}, \quad \text{(IV.20)}$$

where $\mathbf{u} = [u_1, \dots, u_p]$ is the vector of location shifts. Note that here we assume that appropriate centering and/or scaling has been performed in data pre-processing, so transformation on the intercept (i.e., u_0) is not considered in Eq. (IV.20). It is easy to find the relationship between the original and transformed coefficients

$$\boldsymbol{\beta}' + \mathbf{b}'_i = \Psi(\mathbf{u})(\boldsymbol{\beta} + \mathbf{b}_i), \qquad \text{(IV.21)}$$

where $\Psi(\mathbf{u})$ is a $(p+1) \times (p+1)$ transformation matrix which is a function of $\mathbf{u}$. The derivation of Eq. (IV.21) and the specific transformation matrices for $p = 1, 2, 3$ are given in Appendix H. Consequently, the transformed variance-covariance matrix of random effects is

$$G' = \Psi(\mathbf{u})G\Psi^T(\mathbf{u}). \tag{IV.22}$$

This indicates that the variances/covariances of the transformed random effects are functions of the original ones and the location shifts. So we can write the transformed pairwise correlations as $r'_{01}(\boldsymbol{\theta}, \mathbf{u}), r'_{02}(\boldsymbol{\theta}, \mathbf{u}), \dots, r'_{p-1,p}(\boldsymbol{\theta}, \mathbf{u})$ since G depends on $\boldsymbol{\theta}$ by Eqs. (IV.5) and (IV.18).

A critical question is how to find the optimal location shifts $\mathbf{u}^*$ that reduce the correlations of random effects as much as possible. According to Eq. (IV.22), if we can obtain an estimate of $\boldsymbol{\theta}$ and thus of G from the original data, then the transformed correlations will only depend on $\mathbf{u}$. Thus, we can formulate this as the following optimization problem

$$\mathbf{u}^* = \underset{\mathbf{u}=[u_1,\dots,u_p]}{\operatorname{argmin}} \left|r'_{01}(\mathbf{u}|\widehat{\boldsymbol{\theta}})\right| + \left|r'_{02}(\mathbf{u}|\widehat{\boldsymbol{\theta}})\right| + \cdots + \left|r'_{p-1,p}(\mathbf{u}|\widehat{\boldsymbol{\theta}})\right|$$

$$s.t. \ \left|r'_{01}(\mathbf{u}|\widehat{\boldsymbol{\theta}})\right| \leq c,$$

$$\left|r'_{02}(\mathbf{u}|\widehat{\boldsymbol{\theta}})\right| \leq c,$$

$$\dots$$

$$\left|r'_{p-1,p}(\mathbf{u}|\widehat{\boldsymbol{\theta}})\right| \leq c, \tag{IV.23}$$

where the constant c (e.g., 0.7) is a threshold for each correlation. The above formulation aims to minimize the sum of the (absolute) correlations, with each of them constrained by the threshold. The objective function takes the form of l_1-norm as in LASSO (Hastie, et al., 2009), which is convex and easy to solve.

Note that the estimation of $\boldsymbol{\theta}$ may not converge due to the HCRE problem. In this case, we can first apply some simple transformations, e.g., centering, to obtain the estimate and then apply the formulation in Eq. (IV.23). In cases where the estimate of $\boldsymbol{\theta}$ does not

have good precision, we can find the optimal transformation in an iterative way: given a crude estimate of $\boldsymbol{\theta}$, solution of Eq. (IV.23) will be obtained; then given the optimally transformed regressors, an updated estimate of $\boldsymbol{\theta}$ will be found, which will produce the updated optimal shifts $\mathbf{u}^*$. Another potential problem is that the specified value of c may not produce a solution in some extreme cases. A higher value can be used to relax the constraints.

One special case that deserves to mention is $p = 1$, i.e., a linear PRC model, which is equivalent to the LME model with a single regressor considered in the literature (Morrell et al., 1997; Zhang and Chen, 2013). In this case, there are only two random effects b_{i0} and b_{i1}, and accordingly one pairwise correlation r_{01} and one location shift u_1. Thus, it is possible to find an optimal u_1^* such that $r_{01}' = 0$, i.e., the two random effects become independent. The solution is

$$u_1^* = -\frac{\tau_0}{\tau_1} r_{01}.$$
(IV.24)

The derivation is given in Appendix I. This is a desired case as independent random effects have good interpretability. When $p > 1$, the number of location shifts (i.e., p) is smaller than the number of pairwise correlations (i.e., $p(p+1)/2$), so the desired case is not possible in general. Considering the complex shape of optical profiles, linear PRC models are not adequate for profile segments. Thus the generic formulation of optimal location shifts in Eq. (IV.23) makes more sense.

IV.4. Numerical study

We provide two numerical examples to demonstrate the regressor transformation method in different scenarios. The first example shows a case with low correlations of

random effects and how different types of transformations work in this case, and the second example shows a case with high correlations and the corresponding optimal location shifts. In both examples, a quadratic PRC model ($p = 2$) without the random error term is used to generate the profile data; the SAEM algorithm is used for model estimation through the MATLAB function *nlmefitsa*; and the interior-point algorithm is used for the optimization problem in Eq. (IV.23) through the MATLAB function *fmincon*. For notational simplicity, the subscript "i" and/or "j" will be dropped in related quantities (e.g., Eq. (IV.19)) in the following.

IV.4.1.Example 1

The data generation model used in this example is $y = (2 + b_2)x^2 + (2 + b_1)x + (2 + b_0)$, with the variance-covariance matrix of random effects

$$\mathbf{G} = \begin{bmatrix} 0.3 & & \\ & 0.5 & \\ & & 0.3 \end{bmatrix} \begin{bmatrix} 1 & 0.3 & 0.25 \\ 0.3 & 1 & 0.2 \\ 0.25 & 0.2 & 1 \end{bmatrix} \begin{bmatrix} 0.3 & & \\ & 0.5 & \\ & & 0.3 \end{bmatrix}.$$

$m = 500$ profiles are generated from this model, which are shown in the left panel of Figure IV-5. We first estimate parameters of the PRC model. Then, assuming location shifts u_1, u_2 for the regressors x, x^2 respectively, three types of location transformation are applied and transformed correlations of random effects are obtained. The scenarios of transformation are summarized in Table IV-1, and results of pairwise correlations $(r'_{01}, r'_{02}, r'_{12})$ vs. location shift are given in Figure IV-6.
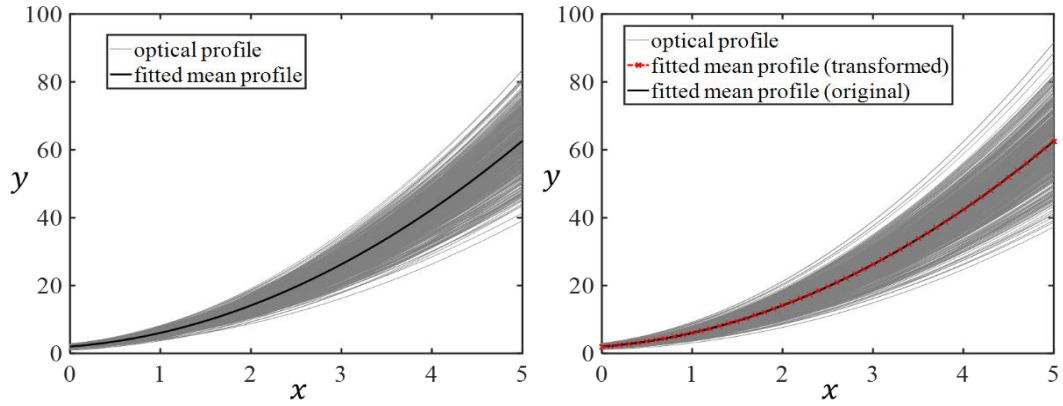
Figure IV-5. Simulated profiles and fitted mean profile in Example 1 (left) and Example 2 (right)

Table IV-1. Scenarios of location transformation in Example 1

Scenario	Setting of shifts	Setting of magnitude
I	$u_1 = u_2 \neq 0$	$-5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5$
II	$u_1 = 0, u_2 \neq 0$	$-2.5, -2, -1.5, -1, -0.5, 0, 0.5, 1, 1.5, 2, 2.5$
III	$u_1 \neq 0, u_2 = 0$	$-2.5, -2, -1.5, -1, -0.5, 0, 0.5, 1, 1.5, 2, 2.5$
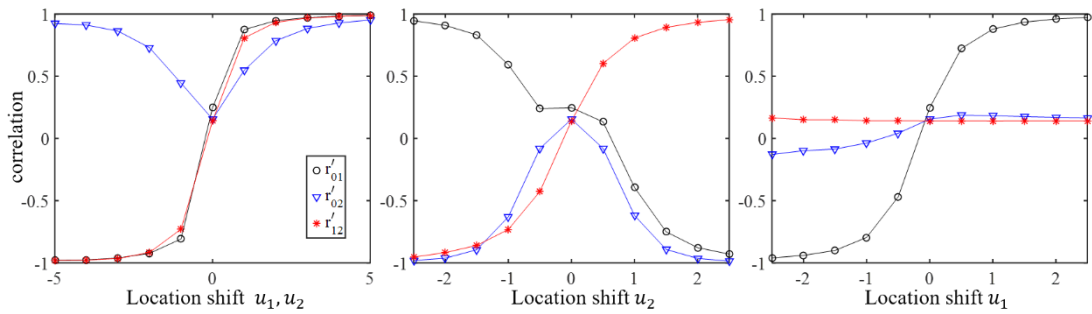
Figure IV-6. Transformed correlations of random effects in Scenario I (left), II (center), and Scenario III (right)

91

The REML estimates of model parameters are

$$\widehat{\boldsymbol{\beta}} = [1.994\ 1.976\ 2.015]',$$

$$\widehat{\mathbf{G}} = \begin{bmatrix} 0.301 & & \\ & 0.488 & \\ & & 0.304 \end{bmatrix} \begin{bmatrix} 1 & 0.251 & 0.258 \\ 0.251 & 1 & 0.226 \\ 0.258 & 0.226 & 1 \end{bmatrix} \begin{bmatrix} 0.301 & & \\ & 0.488 & \\ & & 0.304 \end{bmatrix},$$

which are close to the true values. The fitted mean profile is shown in the left panel of

Figure IV-5, which matches the trend of the data well.

In Figure IV-6, the pairwise correlations of random effects change with the

magnitude of location shift dramatically and exhibit different patterns in the three

scenarios. In Scenario I with two equal-size location shifts, r'_{01} and r'_{12} vary from -1 to

1, while r'_{02} varies in the range of $(0.2, 0.9)$. At $u_1 = u_2 = -0.15$, $r'_{01} \approx 0, r'_{12} \approx 0$,

meaning that b_1 is independent of both b_0 and b_2, which is a desired result. In Scenario II

with only a location shift u_2, r'_{01} and r'_{12} experience large changes as in Scenario I, except

that r'_{01} has an opposite pattern from r'_{12}; r'_{02} shows a symmetric pattern in the range of

$(-1, 0.2)$. By appropriately choosing the magnitude of the location shift, we can make

some of the correlations to zero. In Scenario III with only a location shift u_1, r'_{01} shows

dramatic changes as in Scenario I, while r'_{02} and r'_{12} change mildly as b_2 is not affected

by the shift in this case; under certain value of the location shift, r'_{01} and r'_{02} can be very

close to zero.

In summary, the above results validate that the location shifts have significant

impacts on the correlations of random effects in the PRC model. Moreover, the setting of

the location shifts, such as which regressors they are applied to and magnitudes, is critical.

When the original correlations of random effects are not high, a well-designed regressor

transformation, like the case given in the left panel of Figure IV-6, can lead to desired independent random effects.

IV.4.2.Example 2

In this example, we generate $m = 500$ profiles from the same quadratic PRC model as in Example 1, except that the correlation matrix now contains large elements

$$\mathbf{G} = \begin{bmatrix} 0.3 & & \\ & 0.5 & \\ & & 0.3 \end{bmatrix} \begin{bmatrix} 1 & 0.9 & 0.85 \\ 0.9 & 1 & 0.7 \\ 0.85 & 0.7 & 1 \end{bmatrix} \begin{bmatrix} 0.3 & & \\ & 0.5 & \\ & & 0.3 \end{bmatrix}.$$

The data are shown in the right panel of Figure IV-5. We first obtain the REML estimates which are very close to the true values. The fitted mean profile is also given in Figure IV-5.

$$\hat{\boldsymbol{\beta}} = [1.994 \; 1.989 \; 1.999]',$$

$$\hat{\mathbf{G}} = \begin{bmatrix} 0.290 & & \\ & 0.484 & \\ & & 0.299 \end{bmatrix} \begin{bmatrix} 1 & 0.900 & 0.855 \\ 0.900 & 1 & 0.716 \\ 0.855 & 0.716 & 1 \end{bmatrix} \begin{bmatrix} 0.290 & & \\ & 0.484 & \\ & & 0.299 \end{bmatrix},$$

Then the regressor transformation method is applied to reduce the correlations of random effects. In this specific case, the optimization problem in Eq. (IV.23) is

$$\mathbf{u}^* = \underset{\mathbf{u}=[u_1,u_2]}{\operatorname{argmin}} \left| r'_{01}(\mathbf{u}|\hat{\boldsymbol{\theta}}) \right| + \left| r'_{02}(\mathbf{u}|\hat{\boldsymbol{\theta}}) \right| + \left| r'_{12}(\mathbf{u}|\hat{\boldsymbol{\theta}}) \right|$$

$$s.t. \; \left| r'_{01}(\mathbf{u}|\hat{\boldsymbol{\theta}}) \right| \leq 0.7,$$

$$\left| r'_{02}(\mathbf{u}|\hat{\boldsymbol{\theta}}) \right| \leq 0.7, ..., \left| r'_{12}(\mathbf{u}|\hat{\boldsymbol{\theta}}) \right| \leq 0.7,$$

where the transformed correlations can be obtained based on the transformed matrix $\boldsymbol{\Psi}(\mathbf{u})$ for quadratic model given in Appendix H, and the threshold is set to be 0.7. The solution is $u_1^* = -0.252, u_2^* = -0.783$, and correspondingly, by Eqs. (IV.21)-(IV.22),

93

$$\widehat{\boldsymbol{\beta}}' = [1.056 - 1.140 \ 1.999]',$$

$$\widehat{\mathbf{G}}' = \begin{bmatrix} 0.136 & & \\ & 0.359 & \\ & & 0.299 \end{bmatrix} \begin{bmatrix} 1 & -0.000 & 0.700 \\ -0.000 & 1 & -0.339 \\ 0.700 & -0.339 & 1 \end{bmatrix} \begin{bmatrix} 0.136 & & \\ & 0.359 & \\ & & 0.299 \end{bmatrix}.$$

We also find the REML estimates of parameters under the optimal shifts, i.e.,

using regressors $(x - u_1^*), (x - u_2^*)^2$,

$$\widehat{\boldsymbol{\beta}}'_{REML} = [1.040 - 1.1214 \ 1.996]',$$

$$\widehat{\mathbf{G}}'_{REML} = \begin{bmatrix} 0.136 & & \\ & 0.351 & \\ & & 0.298 \end{bmatrix} \begin{bmatrix} 1 & 0.051 & 0.671 \\ 0.051 & 1 & -0.329 \\ 0.671 & -0.329 & 1 \end{bmatrix} \begin{bmatrix} 0.136 & & \\ & 0.351 & \\ & & 0.298 \end{bmatrix}.$$

The REML estimates are very similar to those from Eqs. (IV.21)-(IV.22). The

fitted mean profile of the transformed model, given in the right panel of Figure IV-5, is

almost identical to the original mean.

This example demonstrates the effectiveness of the proposed regressor

transformation method. It will reduce correlations of random effects, while maintaining

the good fitting performance of the original model. Also, note that in this example the

optimal location shifts lead to not only reduced correlations, but two independent random

effects $(r'_{02} \approx 0)$.

IV.5. Case study

This section applies the proposed methodology for Phase I analysis to real data

from low-E glass manufacturing. The dataset, shown in Figure IV-2, was from a glass

producer in the US. For confidentiality reasons, information on the company and their

products is not disclosed in this text. It contains reflectance profiles collected from $m = 191$ glass ribbons, with $n = 125$ design points evenly distributed on the spectrum

λ=[380nm, 1005nm]. In the data pre-processing, the x values are centered (original values are shown in the following figures for better interpretability).

First, the optical profiles are segmented using the proposed procedure. Then a PRC model is fit to profile segments at each piece of the spectrum, and performances of models with correlated and independent random effects are compared. For the pieces where the HCRE problem occurs, the regressor transformation method is applied, and computational advantages of the transformed models are discussed. Finally, the T^2 chart in Eq. (IV.16) is used to detect outlying profiles in the dataset. Details and results in each step are given as follows.
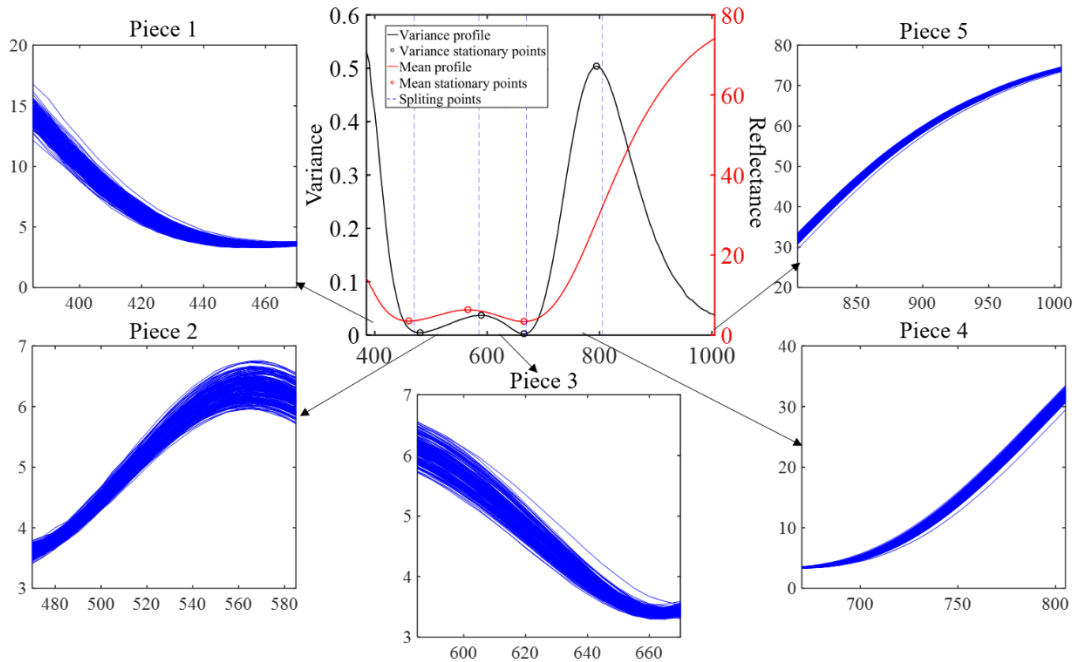
Figure IV-7. Segmentation of profiles using the proposed method: the mean and variance profiles (center) and the five pieces (around)

95

IV.5.1.Step 1: Profile segmentation

To apply the proposed method, we first obtain the sample mean profile and variance profile, which are shown in the center plot of Figure IV-7. There are four empirical stationary points on the variance profile (black circles) and three on the mean profile (red circles). Considering identical stationary points on these two profiles, the set of initial breaking points contains six points. Next, segments between some adjacent breaking points are very short, so we combine those short segments by removing their boundaries. This reduces the number of breaking points to four. As a result, the RSS in fitting the variance profile increases slightly (about 1%), while that of the mean profile remains the same. Furthermore, we adjust locations of those four points within a small neighborhood to improve the fit. Finally, the breaking points (vertical dashed lines) are determined at λ_1=470nm, λ_2=585nm, λ_3=670nm, and λ_4=805nm, leading to five pieces, which is reasonable for monitoring and root cause diagnosis. As shown in the surrounding plots of Figure IV-7, the five pieces have different shapes and variation patterns, but follow similar characteristics as data in Figure IV-5 which are simulated from PRC models.

IV.5.2.Step 2: Model estimation

A cubic PRC model is fit to each of the five pieces (denoted as P1~P5) in Figure IV-7. Table IV-2 lists the estimates of fixed effects, standard deviations/correlation matrix of random effects, and standard deviations of random errors for each piece. The pairwise correlations of random effects are very high in all the pieces, especially P2, P4 and P5. P1 has relatively weaker correlations than others, with only one high correlation. Also, as

96

pointed out previously, the variation due to random errors is much smaller than that due to random effects; for example, the ratios of $\hat{\tau}_0/\hat{\sigma}_\varepsilon$ are 2.0, 7.3, 33.4, 43.9 and 39.4 in P1~P5. Note that though the values of $\hat{\tau}_3$ are small, they may have large impact on the fit because of the cubic regressor (i.e., x^3).

The fitted mean profiles for each piece are shown in Figure IV-8. We can see that the fitted mean matches the shape of the profiles very well in each piece. For comparison, we also fit a cubic PRC model with independent random effects to each piece. The fitted mean profile, also shown in Figure IV-8, exhibits a clear deviation from its counterpart of the model with correlated effects and does not match the data well. These results validate that correlated random effects must be used for adequate fitting of optical profiles.

Table IV-2. Parameter estimates of the original cubic PRC models with correlated random effects

	$[\beta_0, \beta_1, \beta_2, \beta_3]'$	$[\tau_0, \tau_1, \tau_2, \tau_3]'$	R				σ_ε
P1	$\begin{bmatrix} 14.3972 \\ -32.4287 \\ 27.2772 \\ -4.6176 \end{bmatrix}$	$\begin{bmatrix} 0.7078 \\ 1.5440 \\ 3.8242 \\ 2.5952 \end{bmatrix}$	1	−0.2889	−0.4396	0.5888	0.3468
			−0.2889	1	−0.7317	0.5998	
			−0.4396	−0.7317	1	−0.9840	
			0.5888	0.5998	−0.9840	1	
P2	$\begin{bmatrix} 8.0788 \\ -15.4786 \\ 15.4725 \\ -4.1172 \end{bmatrix}$	$\begin{bmatrix} 1.0243 \\ 2.1354 \\ 1.4548 \\ 0.3084 \end{bmatrix}$	1	−0.9955	0.9674	−0.9196	0.1396
			−0.9955	1	−0.9864	0.9513	
			0.9674	−0.9864	1	−0.9889	
			−0.9196	0.9513	−0.9889	1	
P3	$\begin{bmatrix} -83.5117 \\ 121.5186 \\ -53.4717 \\ 7.5564 \end{bmatrix}$	$\begin{bmatrix} 4.1140 \\ 5.4000 \\ 2.3490 \\ 0.3431 \end{bmatrix}$	1	−0.9905	0.9566	−0.8945	0.1233
			−0.9905	1	−0.9865	0.9433	
			0.9566	−0.9865	1	−0.9846	
			−0.8945	0.9433	−0.9846	1	
P4	$\begin{bmatrix} 312.9730 \\ -253.7014 \\ 64.5434 \\ -4.7761 \end{bmatrix}$	$\begin{bmatrix} 12.7913 \\ 12.4688 \\ 3.9477 \\ 0.3976 \end{bmatrix}$	1	−0.9934	0.9792	−0.9654	0.2915
			−0.9934	1	−0.9960	0.9889	
			0.9792	−0.9960	1	−0.9982	
			−0.9654	0.9889	−0.9982	1	
P5	$\begin{bmatrix} -290.6179 \\ 123.3512 \\ -12.5272 \\ 0.3414 \end{bmatrix}$	$\begin{bmatrix} 11.3402 \\ 7.3960 \\ 1.5008 \\ 0.0977 \end{bmatrix}$	1	−0.9947	0.9884	−0.9843	0.2879
			−0.9947	1	−0.9988	0.9971	
			0.9884	−0.9988	1	−0.9996	
			−0.9843	0.9971	−0.9996	1	
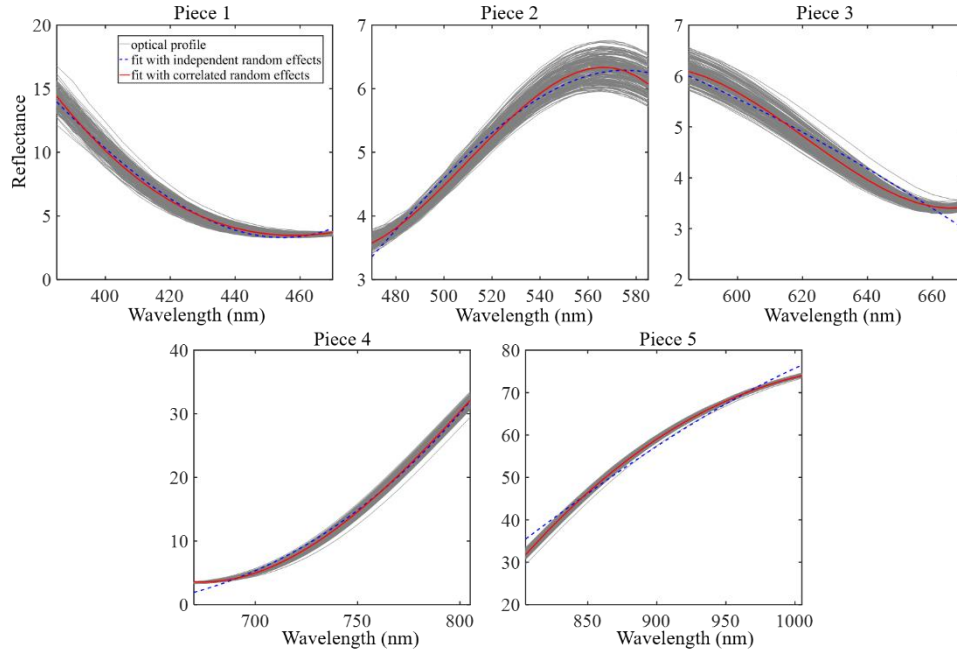
Figure IV-8. Fitted mean profiles of PRC models with correlated vs. independent random effects

IV.5.3.Step 3: Regressor transformation

As the HCRE problem occurs in all the five pieces, regressor transformation is applied to each piece. In this case, the optimization problem in Eq. (IV.23) is

$$\mathbf{u}^* = \underset{\mathbf{u}=[u_1,u_2,u_3]}{\operatorname{argmin}} \left|r'_{01}(\mathbf{u}|\widehat{\boldsymbol{\theta}})\right| + \left|r'_{02}(\mathbf{u}|\widehat{\boldsymbol{\theta}})\right| + \left|r'_{03}(\mathbf{u}|\widehat{\boldsymbol{\theta}})\right| + \left|r'_{12}(\mathbf{u}|\widehat{\boldsymbol{\theta}})\right| + \left|r'_{13}(\mathbf{u}|\widehat{\boldsymbol{\theta}})\right| + \left|r'_{23}(\mathbf{u}|\widehat{\boldsymbol{\theta}})\right|$$

$$s.t. \max(\left|r'_{01}(\mathbf{u}|\widehat{\boldsymbol{\theta}})\right|, \left|r'_{02}(\mathbf{u}|\widehat{\boldsymbol{\theta}})\right|, \left|r'_{03}(\mathbf{u}|\widehat{\boldsymbol{\theta}})\right|, \left|r'_{12}(\mathbf{u}|\widehat{\boldsymbol{\theta}})\right|, \left|r'_{13}(\mathbf{u}|\widehat{\boldsymbol{\theta}})\right|, \left|r'_{23}(\mathbf{u}|\widehat{\boldsymbol{\theta}})\right|) \le 0.7.$$

The transformation matrix $\boldsymbol{\Psi}(\mathbf{u})$ for cubic PRC models (Appendix H) is used in solving the above problem. The resulting optimal shifts and correlations of random effects are listed in Table IV-3.

Table IV-3. Optimal location shifts and correlations of random effects after transformation

	$[u_1, u_2, u_3]'$		R		
P1	$\begin{bmatrix}0.6809\\0.2275\\0.4815\end{bmatrix}$	$\begin{bmatrix}1 & \mathbf{-0.0126} & -0.7000 & 0.6895\\-0.0126 & 1 & -0.7000 & -0.6902\\-0.7000 & -0.7000 & 1 & -0.0208\\0.6895 & -0.6902 & -0.0208 & 1\end{bmatrix}$			
P2	$\begin{bmatrix}0.6985\\1.1582\\1.3264\end{bmatrix}$	$\begin{bmatrix}1 & \mathbf{0.0000} & \mathbf{0.0000} & \mathbf{0.0320}\\0.0000 & 1 & -0.2762 & -0.4674\\0.0000 & -0.2762 & 1 & -0.7000\\0.0320 & -0.4674 & -0.7000 & 1\end{bmatrix}$			
P3	$\begin{bmatrix}2.0180\\2.0261\\2.0700\end{bmatrix}$	$\begin{bmatrix}1 & \mathbf{-0.0000} & -0.7000 & \mathbf{0.0000}\\-0.0000 & 1 & -0.4396 & -0.2652\\-0.7000 & -0.4396 & 1 & -0.4059\\0.0000 & -0.2652 & -0.4059 & 1\end{bmatrix}$			
P4	$\begin{bmatrix}2.8402\\2.8265\\3.1098\end{bmatrix}$	$\begin{bmatrix}1 & 0.6130 & -0.6602 & \mathbf{0.0000}\\0.6130 & 1 & \mathbf{-0.0144} & -0.7000\\-0.6602 & -0.0144 & 1 & -0.7000\\0.0000 & -0.7000 & -0.7000 & 1\end{bmatrix}$			
P5	$\begin{bmatrix}7.0411\\4.0715\\4.9792\end{bmatrix}$	$\begin{bmatrix}1 & 0.6807 & -0.7000 & \mathbf{-0.0011}\\0.6807 & 1 & \mathbf{-0.0000} & -0.6699\\-0.7000 & 0.0000 & 1 & -0.6999\\-0.0011 & -0.6699 & -0.699 & 1\end{bmatrix}$			

From Table IV-3, the optimal location shifts for P1 are the smallest, while those for P5 are the largest, which is reasonable as P1 has the weakest HCRE problem and P5 has the most serious. The correlations of random effects after transformation are much smaller than their original values in all the pieces, with many of them far below 0.7. Especially, there are one or more near-zero correlations (marked in bold in Table IV-3) in each piece, meaning that some transformed random effects are near independent. It is also found that the fitted mean profiles and estimated standard deviations of random errors in the transformed models (results not shown here for space limit) are almost identical to those in the original models given in Table IV-2.
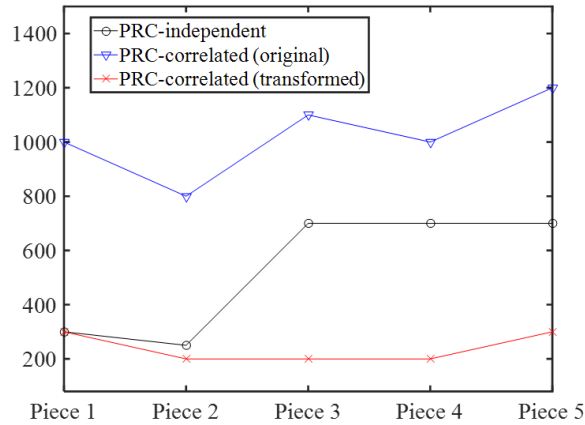
Figure IV-9. Convergence rate in estimating τ_1

It is interesting to compare the convergence behaviors in estimating the original model, the transformed model and the model with independent random effects. Figure IV-9 shows the number of iterations to achieve convergence in estimating τ_1 for each piece. The transformed model converges the fastest in all the five pieces, whereas the original model is the slowest. The estimates of other parameters have similar trends. For more details on the difference between the original and transformed models, Figure IV-10 shows their convergence process in parameter estimation for P4. The transformed model converges smoothly, while the original model exhibits jumps and jaggedness in the process. These results suggest that the proposed location transformation can enhance computation speed and efficiency in model estimation.
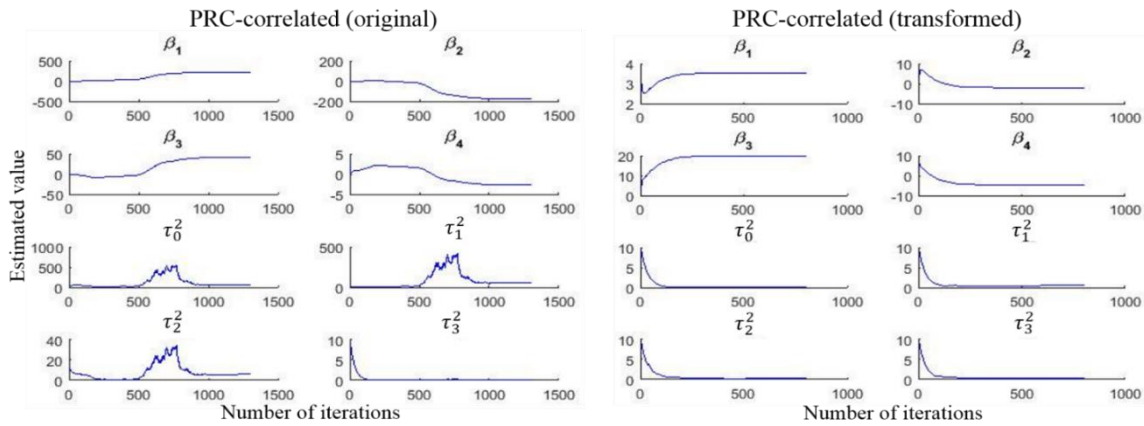
Figure IV-10. The convergence process in parameter estimation for Piece 4

IV.5.4. Step 4: Change detection

With $L = 5$ and $\alpha_{overall} = 0.1$, each control chart has a significance level $\alpha_{chart} \approx 0.021$. For an individual statistic in each chart that monitors $m = 191$ profiles, $\alpha \approx 0.00011$, hence the upper control limit is $\chi^2_{p+1,\alpha} = 23.30$. In the first run, we construct the T^2 control chart for each piece based on estimates of the transformed models. The resulting charts are displayed in the upper panel of Figure IV-11. Most data points are within the control limit, with only six, two, four, seven and five points out of control in P1~P5. Detailed information (sample index and value of T^2 statistic) of those points are listed in the first row of Table IV-4. Sample #40 is a common out-of-control point in most pieces, which looks like an obvious outlier. The lower panel of Figure IV-11 shows the original optical profiles associated with those points. The outlying profiles look clearly abnormal from the majority of profiles (shadow in each plot) in location and/or shape. Sample #40 is a salient case, which is outside of the data zone.

Table IV-4. Out-of-control points in the T^2 control charts

	Segment 1		Segment 2		Segment 3		Segment 4		Segment 5	
	#	Statistic	#	Statistic	#	Statistic	#	Statistic	#	Statistic
1st run	4	25.73	11	26.76	7	26.86	7	28.69	19	23.40
	23	24.87	150	29.76	19	36.65	8	26.02	40	23.46
	40	33.98			39	24.49	19	25.51	125	24.99
	49	44.62			40	46.73	20	30.60	172	25.60
	117	26.36					35	30.85	173	24.93
	118	32.68					40	80.82		
							166	25.06		
2nd run	22	23.57	41	37.03	5	26.08	3	25.23		
	109	26.56	167	30.78	14	26.31	5	31.71		
					109	26.08	6	34.31		
							27	23.37		
							68	79.07		

In the second run, the outlying profiles detected in the first run are removed from the dataset. Then the transformed model is fit to the remaining data and T^2 chart is constructed based on the model estimates. Figure IV-12 (upper) shows the updated control charts for each piece. P5 has no out-of-control points now, whereas other pieces still contain several weak ones. Detailed information of those points is given in the second row of Table IV-4. The original profiles associated with them are shown in the lower panel of Figure IV-12. Those outlying profiles have slightly different shapes from the majority profiles. In the third run, we remove them, refit the models, and construct the new control charts, as shown in Figure IV-13. Now all profiles are in control in each piece, which ends the Phase I analysis. The parameter estimates, including fixed effects $\widehat{\boldsymbol{\beta}}$ and variance-covariance matrix of random effects $\widehat{\mathbf{G}}$, obtained in the last run will be saved for online monitoring.
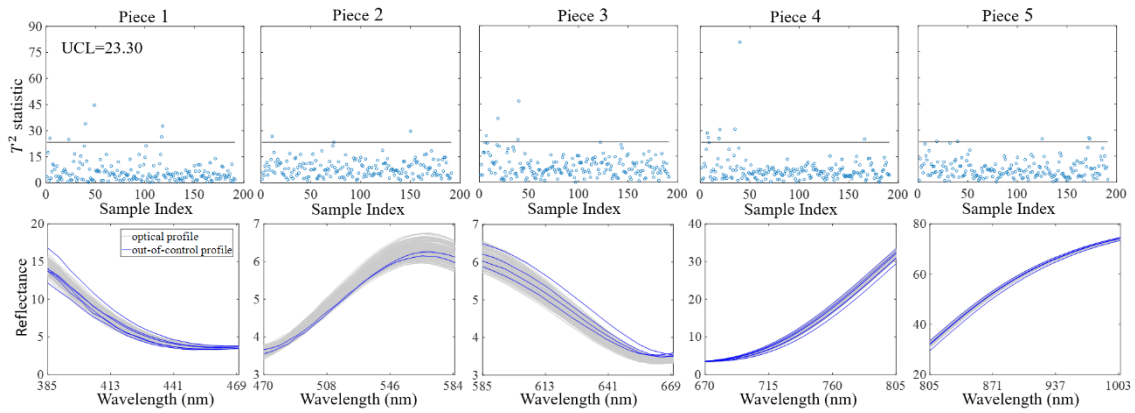
**Figure IV-11.** T^2 **charts and detected outlying profiles in the first run**
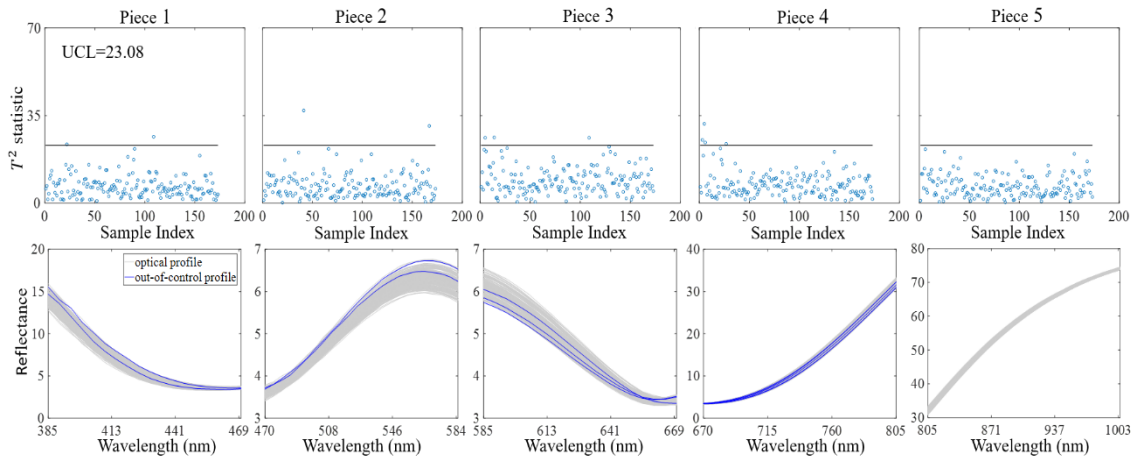
Figure IV-12. T^2 **charts and detected outlying profiles in the second run**
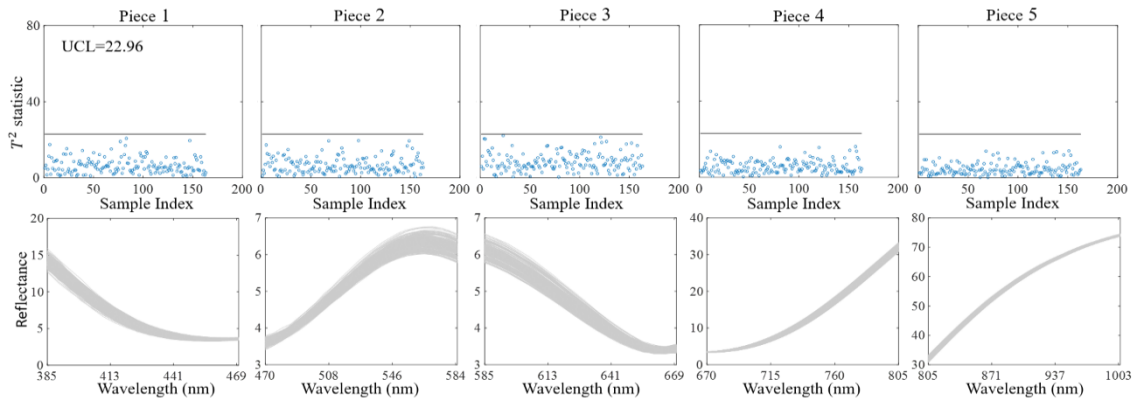
Figure IV-13. T^2 **charts and in-control profiles in the third run**

IV.6. Conclusions

This study investigates the monitoring of optical profile data from low-E glass manufacturing. The proposed methodology is based on a piecewise modeling of profiles using PRC model and T^2 charting of the estimated random effects for change detection. We also examine the HCRE problem and develop the regressor transformation method as a remedy. The case study shows that the piecewise PRC model can well capture the shape and intrinsic variations of optical profiles, and the proposed Phase I monitoring methodology can successfully screen out outlying profiles in the given dataset.

This study may be useful in applications of mixed-effect models with correlated random effects. As illustrated in Figure IV-9, taking correlations of random effects into account can significantly improve the model fitting. Moreover, the proposed regressor transformation method can enhance model estimation efficiency and interpretability. Another way to solve the HCRE problem is to simply cut each piece into shorter pieces to reduce the polynomial orders. Since more pieces is not a good strategy for monitoring purposes, we did not pursue this line.

One limitation of this study is that the model selection problem, i.e., determining which terms in the PRC model have random effects, was not considered. One reason is that there is no easy way for random effects selection (Pauler el al., 1999). In general, frequentist methods are inaccurate, while Bayesian methods are complex. For practitioners, one simple method is to fit a regular polynomial model to each profile segment, and then compare the variance of estimates of each coefficient among all segments. But be cautious that this method may fail because a coefficient with small

variance may have large impact on the fitting if it is associated with a high-order polynomial term (e.g., x^3). This is another special aspect of polynomial mixed-effect models, which will be explored in our future research.

Another direction of our future study is whether it is possible and how to transform the PRC model to enable independent random effects. As pointed out at the end of the section on location transformation, theoretically the proposed method is not able to make all random effects independent; but, it can produce some independent random effects, as shown in the numerical study and case study. We can formulate the optimization problem in Eq. (IV.23) in terms of independence between random effects, e.g., maximizing the number of zero correlations, to further improve the interpretability of this method.

IV.7. References

Abdel-Salam, A. G., Birch, J. B., and Jensen, W. A. (2013). "A Semiparametric Mixed Model Approach to Phase I Profile Monitoring". *Quality and Reliability Engineering International*, **29,** 555-569.

Arasteh, D., Carmody, J., Lee, E.S., and Selkowitz, S. (2004). *Window Systems for High-performance Buildings*. W. W. Norton & Company, New York, NY.

Carmody, J., Selkowitz, S., and Heschhong, L. (1996). *Residential Windows: A Guide to New Technologies and Energy Performance*. W. W. Norton & Company, New York, NY.

Carmody, J., and Haglund, K. (2012). "*Measure Guideline: Energy-efficient Window Performance and Selection*, Department of Energy report. Golden, CO.

Chang, S. I., Tsai, T.-R., Lin, D. K. J., Chou, S.-H., and Lin, Y.-S. (2012). "Statistical Process Control for Monitoring Nonlinear Profiles: A Six Sigma Project on Curing Process". *Quality Engineering*, **24**, 251-263.

Chang, S. I., and Yadama, S. (2010). "Statistical Process Control for Monitoring Non-linear Profiles Using Wavelet Filtering and B-Spline Approximation". *International Journal of Production Research*, **48**(4), 1049-1068.

Chou, S.-H., Chang, S. I., and Tsai, T.-R. (2014). "On Monitoring of Multiple Nonlinear Profiles". *International Journal of Production Research*, **52**(11), 3209-3224.

Dempster, A.P.; Laird, N.M.; Rubin, D.B. (1977). "Maximum Likelihood from Incomplete Data via the EM Algorithm". *Journal of the Royal Statistical Society, Series B*. **39** (1): 1-38.

Ding, Y., Zeng, L., and Zhou, S. (2006). "Phase I Analysis for Monitoring Nonlinear Profiles in Manufacturing Processes". *Journal of Quality Technology*, **38**(3), 199-216.

Fernandez-Oliveras, A., Rubiño, and Pérez, M. M. (2013). "Scattering and Absorption Properties of Biomaterials for Dental Restorative Applications". *Journal of the European Optical Society Rapid Publications*, **8**: 13056.

Gaponenko, S. V. (1998). *Optical Properties of Semiconductor Nanocrystals*. Cambridge University Press, New York, NY.

Gumedze, F. N., and Dunne, T. T. (2011). "Parameter estimation and inference in the linear mixed model". Linear Algebra and Its Applications, 435(8), 1920-1944.

Hartley, H. O., and Rao, J. N. K. (1967). "Maximum-Likelihood Estimation for the Mixed Analysis of Variance Model". *Biometrika*, 54, 93-108.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning, 2nd Edition*. Springer, New York, NY.

Holmes, D. S and Mergen, A. E. (1993). "Improving the Performance of the T^2 Control Chart". *Quality Engineering*, **5**, 619-625.

Hunter Lab (2008). CIE L*a*b* Color Scale. *Insight on Color*, **8**(7), 1-4.

Jensen, W. A., and Birch J. B. (2009). "Profile Monitoring via Nonlinear Mixed Models". *Journal of Quality Technologies*, **41**(1), 18-34.

Jensen, W. A., Grimshaw, S. D., and Espen, B. (2016). "Nonlinear Profile Monitoring for Oven-Temperature Data". *Journal of Quality Technologies*, **48**(1), 84-97.

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

Jin, J., and Shi, J. (1999). "Feature-Preserving Data Compression of Stamping Tonnage Information Using Wavelets". *Technometrics*, **41**(4), 327-339.

Kang, L., and Albin, S. (2000). "On-line Monitoring When the Process Yields A Linear Profile". *Journal of Quality Technology*, **32**(4), 418-426.

Kazemzadeh, R. B., Noorossana, R., and Amiri, A. (2008). "Phase I Monitoring of Polynomial Profiles". *Communications in Statistics—Theory and Methods*, **37**, 1671-1686.

Kazemzadeh, R. B., Noorossana, R., and Amiri, A. (2009). "Monitoring Polynomial Profiles in Quality Control Applications". *International Journal of Advanced Manufacturing Technology*, **42**, 703-712.

Kutner, M. H., Nachtsheim, C. J., and Neter, J. (2004). *Applied Linear Regression Models*. McGraw-Hill/Irwin, New York, NY.

Mahmoud M. A. (2008). "Phase I Analysis of Multiple Linear Regression Profiles. *Communications in Statistics—Simulation and Computation*, **37**, 2106-2130.

Morrell, C. H., Pearson, J. D., and Brant, L. J. (1997). "Linear Transformations of Linear Mixed-effects Models". *The American Statistician*, **51**(4), 338-343.

Mosesova, S. A., Chipman, H. A., MacKay, R. J. and Steiner, S. H. (2007). "Profile Monitoring Using Mixed-Effects Models". *Technical Report*.

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

Patterson, H. D., and Tompson, R. (1971). "Recovery of Inter-Block Information when Block Sizes are Unequal". *Biometrika*, **58,** 545-554.

107

Pauler, D. K., Wakefield, J. C., and Kass, R. E. (1999). "Bayes Factors and Approximations for Variance Component Models". *Journal of the American Statistical Association*, **94**(448), 1242-1253.

Paynabar, K., Jin, J., and Pacella, M. (2013). "Monitoring and Diagnosis of Multichannel Nonlinear Profile Variations Using Uncorrelated Multilinear Principal Component Analysis". *IIE Transactions*, **45**, 1235-1247.

Qiu, P., Zou, C. and Wang, Z. (2010). "Nonparametric Profile Monitoring by Mixed Effects Modeling". *Technometrics*, **52,** 265-277.

Seber, G. A. F., and Lee, A. J. (2003). *Linear Regression Analysis*. John Wiley & Sons, Hoboken, New Jersey.

Sullivan, J. H. and Woodall, W. H. (1996). "A Comparison of Multivariate Control Charts for Individual Observations". *Journal of Quality Technology*, **28**(4), 398-408.

West, B. T., Welch, K. B., and Gatecki, A. T. (2015). *Linear Mixed Models: A Practical Guide Using Statistical Software, 2nd Edition*. CRC Press, Boca Raton, FL.

Williams, J. D., Woodall, W. H., Birch, J. B., and Sullivan J. H. (2006). "Distribution of Hotelling's T^2 Statistic Based on the Successive Differences Estimator". *Journal of Quality Technology*, **38**(3), 217-229. T2

Williams, J. D., Birch, J. B., Woodall, W. H., and Ferry, N. M. (2007a). "Statistical Monitoring of Heteroscedastic Dose-Response Profiles from High-Throughput Screening". *Journal of Agricultural, Biological, and Environmental Statistics*, **12**(2), 216-235.

Williams, J. D., Woodall, W. H., and Birch, J. B. (2007b). "Statistical Monitoring of Nonlinear Product and Process Quality Profiles". *Quality and Reliability Engineering International*, **23**, 925-941.

Woodall, W. H., Spitzner, D. J., Montgomery, D. C., and Gupta, S. (2004). "Using Control Charts to Monitor Process and Product Quality Profiles". *Journal of Quality Technologies*, **36**(3), 309-320.

Zeng, L., and Chen, N. (2015). "Bayesian Hierarchical Modeling for Monitoring Optical Profiles in Low-E Glass Manufacturing". *IIE Transactions*, **47**(2), 109-124.

Zhang, H., and Albin, S. (2009). "Detecting Outliers in Complex Profiles Using a χ^2 Control Chart". *IIE Transactions*, **41**, 335-345.

Zhang, G., and Chen, J. J. (2013). "Adaptive Fitting of Linear Mixed-effects Models with Correlated Random Effects". *Journal of Statistical Computation and Simulation*, **83**(12), 2291-2314.

Zou, C., Tsung, F., and Wang, Z. (2008). "Monitoring Profiles based on Nonparametric Regression Methods". *Technometrics*, **50**(4), 512-526.

Zou, C., Tseng, S.-T., and Wang, Z. (2014). "Outlier Detection in General Profiles Using Penalized Regression Method". *IIE Transactions*, **46**, 106-117.

Zude, M (edited) (2009). *Optical Monitoring of Fresh and Processes Agricultural Crops.* CRC Press, Boca Raton, FL.

CHAPTER V

CONCLUSIONS AND DISCUSSION

V.1. Conclusions

For the purpose of optimizing a system delivering products or service, two tasks arise: (1) quantifying the performance of the system and then (2) tuning the process parameters to improve the system performance. Since a system often contains multiple components whose effects interact with each other, analysis from the physiochemical perspective only is difficult. Therefore, statistical methods and data driven approaches are considered. One type of data that is getting more and more popular in characterizing performance of systems is the output quality profiles. In this dissertation, we focus on using quality profile data with panel structure to solve two types of problems corresponding to the tasks mentioned above. First, modeling the relationship between process parameter of the system and the resulting quality of the output. Second, quantifying the variation of the output quality measures among individual product or service. The first problem aims at connecting the process parameter to the system performance from the perspective of output quality using statistical models. The desired system performance can then be achieved by selecting the optimal process parameter. The second problem targets on measuring the variation among the products or services from the system. This variation can be used as an approximation to reflect the stability of the system. These two types of problems are discussed and addressed via three studies in manufacturing and biomedical engineering domains in this dissertation.

For the first type of problem, we consider modeling the relationship between process parameter of the system and the resulting quality of the output in different scenarios depending on the availability of related expert knowledge. Specifically, we focus on the scenario where only partial information about the underlying knowledge is known, which is often the case when the system is complex or involves novel technologies. We develop a constrained semiparametric varying coefficient model where the domain knowledge is accommodated as a constraint on the effect of process parameter of interest and a parametric additive term for other factors in the system. A case study in biomedical fabrication is provided to demonstrate the performance of the model. We first mathematically formulate the dynamic stability constraint, which is a common observed behavior in engineering and then provide the corresponding estimation algorithms to implement the dynamic stability constraint. The results of the case study indicate both good interpretability and fitting accuracy of the proposed model.

For the second type of problem, we give the analysis of individual variation in two case studies. In Chapter III, the individual variation is modeled as functional random effects in a biomedical engineering application. The results show that the estimated random effects fit the quality profile data well and capture the variation pattern as well. In Chapter IV, a nonlinear mixed-effect model is used for anomaly detection in a manufacturing application. The results show that the proposed model fits the data well and is able to identify abnormal products based on the estimated random effects.

V.2. Future work

There are several topics which can be explored in my future research.

111

First, the dynamic stability constraint in Chapter II is applied to one process parameter in the system. In other cases, it is possible that multiple process parameters have a similar converging behavior over time. In such a situation, the dynamic stability constraint needs to be extended to all the process parameters.

Second, our study considers modeling for a single process parameter. It is reasonable in the practice of experimental studies as considering multiple process parameters would be complex and involve interaction effects of process parameters. However, for datasets retrieved from systems other than time-course experiments, it is possible that multiple process parameters are present and hence need to be modeled simultaneously. Then the model in Chapter II will need to involve multiple smoothing components and their interactions.

Third, there are many other types of constraints than the dynamic stability constraint one may encounter in reality, such as shape constraints and sign constraints. Therefore, another potential research direction is to consider multiple types of constraints in the modeling.

Fourth, when using mixed-effect models to quantify the variation of individual products and services, one of the problems is the model selection issue. In the case studies of Chapter III and IV, we simply assume that randomness occurs at all terms for all covariates for the purpose of covering all sources of variation. In other cases, it is not always the best way to assume all terms in the model contain randomness and a model selection step is necessary.

Finally, another possible direction to extend the work is considering multivariate quality profiles. For each individual, there may be multiple quality measurements taken and affected by the process parameter of interest simultaneously. The models developed in the dissertation can be extended to multivariate cases. By involving the covariance among the multiple quality measures, the performance of the model may be improved in both fitting accuracy and interpretability.

SMOOTHING SPLINE ESTIMATION OF VCM IN THE STUDY OF EUBANK ET

AL. (2004)

A p-dimensional varying coefficient model is defined as

$$y_{ij} = \boldsymbol{x}_{ij}\boldsymbol{\beta}_\lambda(t_j) + \epsilon_{ij}, \qquad \boldsymbol{x}_{ij} = [x_{ij1}, \ldots, x_{ijp}],$$

where y_{ij} represents the observed value on the ith curve at time t_j and x_{ijr} is the rth

component of $\boldsymbol{x}_{ij}$, a p-dimensional time-dependent covariate of the ith curve at time t_j,

$i = 1, \ldots, m, j = 1, \ldots, n, r = 1, \ldots, p$. The original paper provides an analytical smoothing

spline estimator of the varying coefficient $\boldsymbol{\beta}_\lambda(t) = [\beta_1(t), \ldots, \beta_p(t)]$ by solving the

generalized objective function

$$J_{\lambda,W} = \sum_{i=1}^n \sum_{j=1}^m \left[\left[y_{ij} - \sum_{r=1}^p x_{ijr}\, \beta_r(t_{ij}) \right]^2 \right] + \sum_{r=1}^p \lambda_r \int \beta_r^{(z)}(t)^2 dt, \qquad \text{(A1)}$$

where z is the order of derivatives required for smoothness, for example $z = 2$ for second-

order derivative. The proof is summarized as below.

For the data regarding objective function (A1), we rewrite them as

$$Y = [y_{11}, \ldots, y_{m1}, \ldots, y_{1n}, \ldots, y_{mn}]^T$$

$$X_{\cdot j} = \left[\boldsymbol{x}_{1j}^T, \ldots, \boldsymbol{x}_{mj}^T \right]^T, j = 1, \ldots, n$$

and then define $H = \begin{bmatrix} X_{\cdot 1} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & X_{\cdot n} \end{bmatrix}$, $\tilde{Q} = \left\{ \int_0^{\min(t_{j_1}, t_{j_2})} \frac{(t_{j_1} - s)^{z-1}(t_{j_2} - s)^{z-1}}{[(z-1)!]^2} ds \right\}_{j_1, j_2 = 1, \ldots, n}$

$$\tilde{t}_j = \left[1, t_j, \ldots, t_j^{z-1} \right]^T$$

$$K = \tilde{Q} \otimes I_p$$

$$T = \begin{bmatrix} \tilde{t}_1^T \otimes I_p \\ \vdots \\ \tilde{t}_n^T \otimes I_p \end{bmatrix}$$

$$\Lambda = I_n \otimes diag(\lambda_1, \dots, \lambda_p)$$

$$V = HT$$

where $\otimes$ represents the Kronecker product of two matrices and I_p is the identity matrix of size p.

The minimization of objective function (A1) can be reduced to solving for unknown coefficient c and d to minimize the quadratic function below

$$(Y - H(Kc + Td))^T (Y - H(Kc + Td)) + c^T \Lambda Kc, \tag{A2}$$

where $c = [c_{11}, \dots c_{1p}, \dots, c_{n1}, \dots, c_{np}]^T$, $d = [d_{11}, \dots, d_{1p}, \dots, d_{z1}, \dots, d_{zp}]^T$. To minimize (A2), we take the derivative of (A2) with respect to c and let it be zero and obtain

$$(KH^T HK + K\Lambda)c = KH^T (Y - Vd). \tag{A3}$$

Similarly, let the partial derivative with respect to d be zero and then

$$-V^T Y + V^T HKc + V^T Vd = 0. \tag{A4}$$

By applying Sherman-Morrison –Woodbury formula to Eq. (A3), we have

$$Kc = (H^T H + \Lambda K^{-1})^{-1} H^T (Y - Vd). \tag{A5}$$

Again apply Sherman-Morrison –Woodbury formula to $(H^T H + \Lambda K^{-1})$, we have

$$(H^T H + \Lambda K^{-1})^{-1} = K\Lambda^{-1} - (K\Lambda^{-1})(H^T)[I_{nm} +$$

$$H(K\Lambda^{-1})H^T]^{-1} H(K\Lambda^{-1}). \tag{A6}$$

Eq. (A6) can be further simplified as

$$(H^T H + \Lambda K^{-1})^{-1} = Q - QH^T M^{-1} HQ \tag{A7}$$

by letting $Q = K\Lambda^{-1}$ and $M = I_{nm} + HQH^T$. Also the algebra shows that

$$M^{-1} = I_{nm} - H(H^T H + K^{-1}\Lambda)^{-1}H^T. \tag{A8}$$

We then replace Kc in Eq. (A4) by the right hand side of Eq. (A5) along with Eq. (A8) and obtain

$$d = (V^T M^{-1} V)^{-1} V^T M^{-1} Y. \tag{A9}$$

Now with Eqs. (A5), (A7) and (A9), we have

$$Kc + Td = C(\lambda)Y, \tag{A10}$$

where

$$C(\lambda) = QH^T M^{-1} - (QH^T M^{-1} V - T)(V^T M^{-1} V)^{-1} V^T M^{-1}. \tag{A11}$$

As a result, the minimizer of the objective function (A1) is $\boldsymbol{\beta}_\lambda(t) = [\beta_1(t), \dots, \beta_p(t)]$, where each component $\beta_r(t)$ is the natural cubic spline through knots $[b_r, b_{m+r}, \dots, b_{(m-1)p+r}]$ and b_k is the kth element of $C(\lambda)Y$.

APPENDIX B

DERIVATION OF CLOSE-FORM SOLUTION TO EQ. (II.7)

Based on the smoothing spline estimate in Appendix A, we give an analytical minimizer $\boldsymbol{\beta}_\lambda(t) = [\beta_1(t), \dots, \beta_p(t)]$ to the generalized weighted smoothing spline with objective function

$$J_{\lambda,W} = \sum_{i=1}^n \sum_{j=1}^m w_{ij} \left[\left[y_{ij} - \sum_{r=1}^p x_{ijr} \beta_r(t_{ij}) \right]^2 \right] + \sum_{r=1}^p \lambda_r \int \beta_r^{(z)}(t)^2 dt, \quad \text{(B1)}$$

where the weight w_{ij} is assigned to each observed sample point y_{ij} and $\lambda = [\lambda_1, \dots, \lambda_p]$ is given. Using the same notations as in Appendix A, we provide the proof as follows.

We first write the weight matrix as $W = diag(w_{11}, \dots, w_{m1}, \dots, w_{1n}, \dots, w_{mn})$. Following a similar fashion shown in (Eubank et al., 2004), the objective function (B1) can be reduced to

$$\left(Y - H(Kc + Td)\right)^T W \left(Y - H(Kc + Td)\right) + c^T \Lambda Kc, \quad \text{(B2)}$$

where $c = \left[c_{11}, \dots c_{1p}, \dots, c_{n1}, \dots, c_{np} \right]^T$, and $d = \left[d_{11}, \dots, d_{1p}, \dots, d_{z1}, \dots, d_{zp} \right]^T$. We then solve for the unknown coefficients c and d to minimize the objective function by taking derivatives with respect to c and d. Consequently, we have

$$(KH^T WHK + K\Lambda)c = KH^T W(Y - Vd) \quad \text{(B3)}$$

and

$$-V^T WY + V^T WHKc + V^T WVd = 0. \quad \text{(B4)}$$

We apply Sherman-Morrison –Woodbury formula to obtain

$$(KH^T WHK + K\Lambda)^{-1} = (KH^T WHK)^{-1} - \frac{(KH^T WHK)^{-1} K\Lambda (KH^T WHK)^{-1}}{I + \Lambda (KH^T WHK)^{-1} K}. \quad \text{(B5)}$$

Eqs. (B3) and (B5) allow us to derive

$$Kc = (H^T W H + \Lambda K^{-1})^{-1} H^T W (Y - V d). \tag{B6}$$

Again Sherman-Morrison –Woodbury formula leads to

$$(H^T W H + \Lambda K^{-1})^{-1} = K\Lambda^{-1} - (K\Lambda^{-1})(H^T W)[I_{nm} + H(K\Lambda^{-1})H^T W]^{-1} H(K\Lambda^{-1}), \tag{B7}$$

which can be further simplified to

$$(H^T W H + \Lambda K^{-1})^{-1} = Q - Q H^T W M_0^{-1} H Q \tag{B8}$$

by letting $Q = \tilde{Q} \otimes diag(\lambda_1^{-1}, \dots, \lambda_p^{-1})$ and $M_0 = I_{nm} + HQH^T W$.

Noticing that

$$M_0^{-1} = I_{nm} - H(H^T W H + K^{-1}\Lambda)^{-1} H^T, \tag{B9}$$

we can replace Kc in Eq. (B4) by the right side of Eq. (B6) along with Eq. (B9) and obtain

$$d = (V^T W M_0^{-1} V)^{-1} V^T W M_0^{-1} Y. \tag{B10}$$

Now with Eqs. (B6), (B8) and (B10), we have

$$Kc + Td = C(\lambda)Y, \tag{B11}$$

where

$$C(\lambda) = Q H^T W M_0^{-1} - (Q H^T W M_0^{-1} V - T)(V^T W M_0^{-1} V)^{-1} V^T W M_0^{-1} \tag{B12}$$

As a result, each component $\beta_r(t)$, $r = 1, \dots, p$, in the minimizer $\boldsymbol{\beta}_\lambda(t)$ to objective function (B1) is computed as the natural cubic spline through knots $[b_r, b_{m+r}, \dots, b_{(m-1)p+r}]$, where b_k is the kth element of $C(\lambda)Y$.

When the smoothing parameter λ is not given, the generalized cross-validation smoothing parameter λ_{GCV} can be selected that

$$\lambda_{GCV} = \underset{\lambda}{\mathrm{argmin}} \frac{(Y - A_\lambda Y)^T (Y - A_\lambda Y)}{\left\{1 - \frac{tr(A_\lambda)}{nm}\right\}^2} \tag{B13}$$

where $A_\lambda = I_{mn} - (M_0^{-1} - M_0^{-1}V(V^TWM_0^{-1}V)^{-1}V^TWM_0^{-1})$ is the hat matrix of Y, i.e

$\hat{Y}(\lambda) = A_\lambda Y$. As Q, M_0 hence A_λ are matrices of unknown λ, the objective function is a

function of λ and the minimizer is the selected to be λ_{GCV} using constrained optimization

such that all elements in λ_{GCV} are positive.

APPENDIX C

ALGORITHM FOR MODEL ESTIMATION IN SCENARIO 1

Algorithm 4.1 Estimate parameter θ of $\beta_0(t; \theta)$ and $\beta_1(t)$

1. **Initialize** $k \leftarrow 0$; $\theta^{(k)} = \theta_0$.
2. **Repeat**
3.   Set $k \leftarrow k + 1$
        Estimate $\widehat{\boldsymbol{\beta}}_1(\boldsymbol{t})$
4.   Compute $\widetilde{\boldsymbol{Y}}^{(k-1)}\left(t_j\right) \leftarrow \boldsymbol{Y}\left(t_j\right) - \beta_0\left(t_j; \theta^{(k-1)}\right)\boldsymbol{I}_{m \times 1}, \qquad j = 1, \dots, n$
5.   Compute the optimal smoothness parameter λ_{GCV}
6.   Estimate $\beta_1^{(k)}\left(t_j\right)$ using λ_{GCV} from $\widetilde{\boldsymbol{Y}}^{(k-1)}\left(t_j\right) = \beta_1^{(k)}\left(t_j\right)\boldsymbol{X} + \boldsymbol{\epsilon}$, $j = 1, \dots, n$
        Estimate $\widehat{\boldsymbol{\theta}}$
7.   Compute $\boldsymbol{R}^{(k)}\left(t_j\right) \leftarrow \boldsymbol{Y}\left(t_j\right) - \beta_1^{(k)}\left(t_j\right)\boldsymbol{X}, \qquad j = 1, \dots, n$
8.   Estimate $\theta^{(k)}$ from $\boldsymbol{R}^{(k)}\left(t_j\right) = \beta_0\left(t_j; \theta^{(k)}\right)\boldsymbol{I}_{m \times 1} + \boldsymbol{\epsilon}, \qquad j = 1, \dots, n$
9. **Compute** $\left\|\theta^{(k)} - \theta^{(k-1)}\right\|_1$
10. **Until** $\left\|\theta^{(k)} - \theta^{(k-1)}\right\|_1 \leq \omega$ for some small ω.
11. $\widehat{\theta} = \theta^{(k)}$, $\hat{\beta}_1^*(t) = \hat{\beta}_1^{(k)}(t)$.

Implementing the above algorithm requires specifying the value of ω in the stopping rule. A larger value of ω allows larger difference between current and updated estimates, or, in other words, earlier termination of iteration. To guarantee stability of the estimation, small ω is preferred. The value can be determined according to the required precision of estimation. For example, if it needs to be precise to l decimal places, then ω will be set to $10^{-(l+1)}$. $\omega = 10^{-3}$ is used in the case study.

ALGORITHM FOR CONSTRAINED ESTIMATION OF $\beta_1(t)$ IN SCENARIO 2

Algorithm 4.2 Incorporate convergence constraint in the estimation of $\beta_1(t)$

Initialize: $k \leftarrow 0$, $\hat{\beta}_1^{(k)}(t) \leftarrow \hat{\beta}_1^*(t)$ from Algorithm 4.1,

$\quad\quad w_{ij}^{(k)} = 1, \ j = 1, \dots, n, \ a \in \{1, \dots, n | t_a = t^*\}$

$\tilde{y}_i(t_j) \leftarrow y_i(t_j) - \beta_0(t_j; \hat{\theta}), \quad\quad i = 1, \dots, m \quad j = 1, \dots, n$

$A_\alpha\left(\tilde{y}_i(t_j)\right) \leftarrow \tilde{y}_i(t_j) \pm t_{1-\frac{\alpha}{2}, n-1} \times \frac{\hat{\sigma}}{\sqrt{n}}, \quad i = 1, \dots, m \quad j = 1, \dots, n$

$\psi^{(k)} \leftarrow$ the total number of points in $x_i \hat{\beta}_1^{(k)}(t_j)$ that falls outside A_α, $i = 1, \dots, m, j = 1, \dots, n$

Repeat:

$\quad k \leftarrow k + 1$

$\quad$ Compute residuals $r_{ij}^{(k-1)} = y_i(t_j) - \beta_0(\hat{\theta}; t_j) - x_i \hat{\beta}_1^{(k-1)}(t_j), \ i = 1, \dots, m, j = 1, \dots, n$

$\quad$ Update the weights $w_{ij}^{(k)} = \min_{i,j}(|r_{ij}|)/|r_{ij}|, i = 1, \dots, m \quad j = a, \dots, n$

$\quad$ Update $\hat{\beta}_1^{(k)}(t) \leftarrow argmin J_{\lambda_{GCV}, w^{(k)}}$

$\quad$ Update $\psi^{(k)}$

Until $\psi^{(k)} > \psi^{(k-1)}$, then $K \leftarrow k$, where K is the total number of iterations.

$\hat{\beta}_1(t) = arg \min_k C_{\hat{\beta}_1^{(k)}}, \ k = 1, 2, \dots, K - 1$

APPENDIX E

ESTIMATES OF PARAMETERS IN PARAMETRIC MIXED-EFFECT MODELS

Model I						
Parameter	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	σ_{u_1}	σ_{u_2}	σ_{u_3}	σ_e
Estimate	−0.0183	0.4052	−0.41810	0.0103	0.1857	0.1340	0.1904
Model II						
Parameter	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	σ_{u_1}	σ_{u_2}	σ_{u_3}	σ_e
Estimate	−0.0913	2.230	0.2078	0.0443	0.6847	0.0849	0.1990

APPENDIX F

NATURAL CUBIC SPLINE

In general, a natural cubic spline of time t with K knots can be represented by K basis functions $\{C_j(t)\}_{j=1}^{K}$ as below

$$Z(t) = \sum_{j=1}^{K} \zeta_j C_j(t),$$

where ζ_j are coefficients of the basis functions. Let the knots be $\{t_1, t_2, \ldots, t_K\}$ in ascending order, the basis functions can be written as

$$\begin{cases} C_1(t) = 1 \\ C_2(t) = t \\ C_{k+2}(t) = R_k(t) - R_{k-1}(t) \end{cases}$$

where $1 \leq k \leq K - 2$, $R_k(t) = \dfrac{(t-t_k)_+^3 - (t-t_K)_+^3}{t_K - t_k}$, and $t_1 \leq t \leq t_K$.

APPENDIX G

ESTIMATION OF FUNCTIONAL MIXED-EFFECT MODEL

In the functional mixed-effect model, the minimizer of the objective function in Eq. (III.6) is a natural cubic spline (Fitzmaurice et al,2018). So $\beta(t)$ and $\alpha_i(t)$ can be written as

$$\beta(t) = \sum_{j=1}^{n} b_j C_j(t), \ \alpha_i(t) = \sum_{j=1}^{n} a_{ij} C_j(t),$$

where $C_j(t), j = 1, \ldots, n$, are basis functions of natural cubic spline at knots $t_1, \ldots t_n$ as given above. The objective function then becomes

$$(\boldsymbol{y_i} - \boldsymbol{Qb} - \boldsymbol{Qa_i})^T (\boldsymbol{y_i} - \boldsymbol{Qb} - \boldsymbol{Qa_i}) + \lambda_\beta \boldsymbol{b}^T \boldsymbol{\Omega} \boldsymbol{b} + \lambda_\alpha \boldsymbol{a_i}^T \boldsymbol{\Omega} \boldsymbol{a_i} \,,$$

where $\boldsymbol{y_i} = [y_{i1}, \ldots, y_{in}]^T, \boldsymbol{b} = [b_1, \ldots, b_n]^T, \boldsymbol{a_i} = [a_{i1}, \ldots, a_{in}]^T, \boldsymbol{Q}$ and $\boldsymbol{\Omega}$ are $n \times n$ matrices with elements $Q_{kj} = C_j(t_k), \ \Omega_{kj} = \int C_k''(t) C_j''(t) dt$. The unknown parameters $\boldsymbol{b}, \Sigma_\alpha, \sigma_e$ can be estimated using EM algorithm (Kuhn and Lavielle, 2005) and computed using the *sme* algorithm in R package (Berk, 2018).

APPENDIX H

DERIVATION OF EQ. (IV.22) AND TRANSFORMATION MATRIX FOR $p = 1,2,3$

Since random errors in the optical profiles are small and the transformation will not change the fitting in general, we have $y_{ij} \approx y'_{ij}$. By Eqs. (IV.19) and (IV.20),

$$(\beta_p + b_{ip})x_j^p + \cdots + (\beta_0 + b_{i0}) \approx (\beta'_p + b'_{ip})(x_j - u_p)^p + \cdots + (\beta'_0 + b'_{i0})$$

Comparing the coefficients associated with each polynomial term of x_j, i.e., $x_j, x_j^2, \ldots, x_j^p$, at the two sides and using the binomial expansion, we will obtain a set of equations

$$\begin{cases} \beta_p + b_{ip} = \beta'_p + b'_{ip} \\ \beta_{p-1} + b_{i,p-1} = p(-u_p)(\beta'_p + b'_{ip}) + (\beta'_{p-1} + b'_{i,p-1}) \\ \vdots \\ \beta_0 + b_{i0} = (-u_p)^p(\beta'_p + b'_{ip}) \\ \qquad + (-u_{p-1})^{p-1}(\beta'_{p-1} + b'_{i,p-1}) + \cdots + (\beta'_0 + b'_{i0}) \end{cases},$$

which will lead to Eq. (IV.21).

When $p = 1$, i.e., a linear PRC model is used,

$$\Psi_1(\mathbf{u}) = \begin{bmatrix} 1 & u_1 \\ 0 & 1 \end{bmatrix}.$$

When $p = 2$, i.e., a quadratic PRC model is used,

$$\Psi_2(\mathbf{u}) = \begin{bmatrix} 1 & \mu_1 & 2\mu_1\mu_2 - \mu_2^2 \\ 0 & 1 & 2\mu_2 \\ 0 & 0 & 1 \end{bmatrix}.$$

When $p = 3$, i.e., a cubic PRC model is used,

$$\Psi_3(\mathbf{u}) = \begin{bmatrix} 1 & \mu_1 & 2\mu_1\mu_2 - \mu_2^2 & \mu_3^2 - 3\mu_2^2\mu_3 + 6\mu_1\mu_2\mu_3 - 3\mu_1\mu_3^2 \\ 0 & 1 & 2\mu_2 & 6\mu_2\mu_3 - 3\mu_3^2 \\ 0 & 0 & 1 & 3\mu_3 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

APPENDIX I

OPTIMAL LOCATION SHIFT FOR $p = 1$

From $\boldsymbol{\Psi}_1(\mathbf{u})$ given in Appendix H, we can obtain

$$\beta_0' + b_{i0}' = (\beta_1 + b_{i1})u_1 + (\beta_0 + b_{i0}),$$

$$\beta_1' + b_{i1}' = \beta_1 + b_{i1}.$$

Thus

$$\text{cov}(\beta_0' + b_{i0}', \beta_1' + b_{i1}') = \text{cov}(b_{i0}', b_{i1}') = \text{var}(b_{i1}) \cdot u_1 + \text{cov}(b_{i0}, b_{i1}).$$

Letting $\text{cov}(b_{i0}', b_{i1}') = 0$, we obtain

$$u_1^* = -\frac{\text{cov}(b_{i0}, b_{i1})}{\text{var}(b_{i1})}.$$

By the definition of variance and correlation of random effects in Eq. (IV.5),

$$\text{var}(b_{i1}) = \tau_1^2,$$

$$cov(b_{i0}, b_{i1}) = \sqrt{var(b_{i0})} \cdot \sqrt{var(b_{i1})} \cdot r_{01} = \tau_0 \tau_1 r_{01},$$

which gives Eq. (IV.22).