Profile Monitoring Using Mixed-Effects Models

Sofia A. Mosesova
Department of Statistics and Actuarial Science
University of Waterloo
Waterloo, ON, Canada N2L 3G1

Hugh A. Chipman
Department of Mathematics and Statistics
Acadia University
Wolfville, NS, Canada B4P 2R6

R. Jock MacKay
Department of Statistics and Actuarial Science
University of Waterloo
Waterloo, ON, Canada N2L 3G1

Stefan H. Steiner
Department of Statistics and Actuarial Science
University of Waterloo
Waterloo, ON, Canada N2L 3G1

Abstract

In this paper we generalize some of the traditional process monitoring techniques to functional data in which each observation is a curve. We consider two types of models for functional data: a b-spline fit via linear regression and a nonlinear model. After outlining how these can be used to characterize a single curve, the two approaches are incorporated into mixed effects models for summarizing collections of profiles. Multivariate process monitoring techniques are then applied to the predicted random effects in order to identify in-control and anomalous curves. These techniques are illustrated with a data set taken from the automotive industry.

Key Words: functional data; random effects; multivariate process monitoring; statistical process control.

1 INTRODUCTION

Functional data analysis is a branch of statistics dealing with the analysis of data observed in the form of curves. Examples of functional data include any type of measurement recorded as a function of time, such as a growth curve or a profile of force applied over a time period. Functional data are becoming increasingly prevalent in process monitoring applications because of advances in sensor technology.

A particular application that motivated this work comes from automotive manufacturing. Force applied by two rams inserting four valve seats at a time (eight in total) into an aluminum cylinder head is automatically recorded every 400^{th} of a second for about five seconds. Each observation includes about 2000 force values, but only about 100 of these reveal any significant activity. This part of a typical curve lasts only about a quarter of a second, as shown in Figure 1. A total of 6000 cylinder heads, each with eight valves were examined over 41 days spanning two months (January and February) at the start of production.

In this paper we generalize process monitoring methodologies to functional data. In the context of the above example, we develop statistical models and associated control charts for detecting anomalous force curves in real time.

The wide range of names for functional data in the process-monitoring community include "signals", "waveform signals", "signatures", "profiles" and "curves". We will use the last two terms interchangeably. Previous work in this area is well summarized by Woodall et. al. (2004), who observed that simple linear regression is the most common tool for summarizing curves. Kang and Albin (2000), Kim et al. (2003), and Mahmoud and Woodall (2003) propose numerous ways to use multivariate T^2-charts to monitor curves that are well summarized by a slope and an intercept.

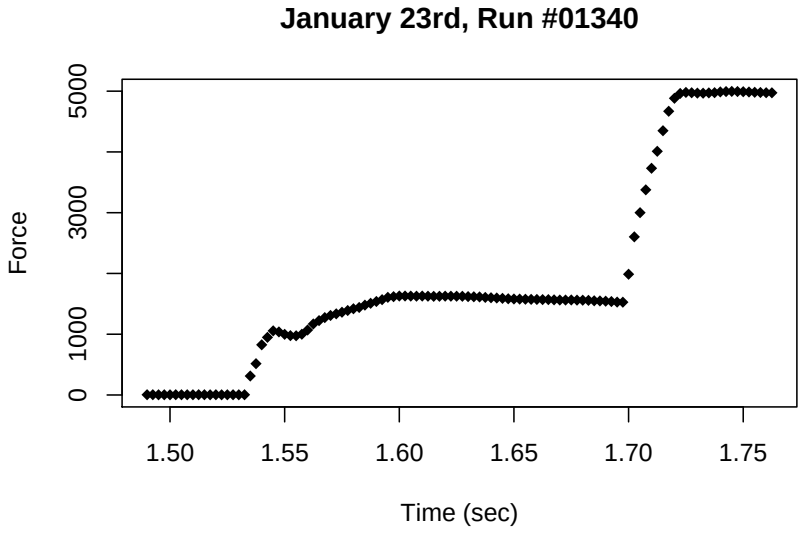

Figure 1: Observed force against time recordings for a single valve-seat insertion. Each second about 400 recordings (solid points) are made. The dashed lines are times corresponding to abrupt changes in the process.

With respect to process monitoring nonlinear profiles, Jin and Shi (2001) describe an adaptive feature-extracting technique for data with limited or no prior "in-control" information. They use an iterative procedure to compare the Hotelling T^2 criteria for each new observation to the base set of "normal" observations. At each step, an observation that is deemed unusual is either classified into a known cluster of outliers or becomes a part of a new outlying cluster.

Walker and Wright (2002) use generalized additive models to compare collections of curves. They

model the functional component of the data with smoothers, and the remaining variability with standard regression parameters. Unusual curves can be flagged by conducting an appropriate F-test. Williams et al. (2003) extend this idea by using a multivariate T^2 control chart statistics to monitor the curves.

The novelty of our approach is in monitoring random coefficients obtained from mixed effects modeling of data. This allows us to monitor small departures from the basic shape of the profiles. By using considerably fewer random effects than time points on a curve, we are also able to monitor in a lower dimension than the original data which may increase the power of our monitoring procedure.

We consider two mixed effects models: linear b-splines and a nonlinear model, but the basic ideas can be generalized to any other linear or nonlinear case. Abramovich & Angelini (2006), Guo (2002), Antoniadis & Sapatinas (2004), Morris & Carroll (2006) and Morris, Arroyo et al. (2006) are just a few of the papers that discuss fitting random effects models to functional data; however this work is constrained to linear models only. The last two papers by Morris in particular are closely related to our b-spline model, with the exception that wavelet basis functions are used in place of the b-spines. One disadvantage of wavelets for the purposes of process-monitoring is the fact that they require fitting many more coefficients than b-splines, consequently leading to a large number of random effects to be predicted and monitored.

The remainder of the paper is organized as follows. The next section discusses a key preprocessing step in the analysis of functional data. In Section 3 we discuss two possible regression models for describing a single curve and in Section 4 we generalize this to the analysis of collections of curves using mixed effects models. Control charts for monitoring these data are developed in Section 5, followed by final results for the insertion data presented in Section 6. We conclude with a discussion in Section 7.

2 CURVE REGISTRATION

A distinct characteristic of functional observations is the need to align the curves prior to analysis. This is referred to as the registration process as it involves registering or identifying landmark features in the curves and aligning them across all curves. Once aligned (e.g. start and end times are more or less the same across all curves), it is easier to isolate anomalous curves during monitoring.

Ramsay and Silverman (2005) provide an excellent reference for a variety of ways to register functional data. In this section we outline a registration algorithm developed for the example data, which combines ideas from some of these methods. It is summarized below, and illustrated using five observed curves in Figure 2. The top panel of the plot shows the raw data, with two landmark features corresponding to the beginning and end of the process chosen as registration points and marked along the curves. The middle

panel shows the cropped curves before they have been aligned, and the bottom panel illustrates the curves after the data has been registered, cropped and interpolated. Further details about the algorithm can be found in Mosesova (2002).

1) Identify landmark features in the data.

For each curve, pick two points corresponding to the beginning and end of the insertion process (Figure 2(a)). These were identified as time points where first differences of force values crossed a threshold value.

2) Pick the registration points.

Treating the two landmarks from step 1 as registration points, the time scale of the curve can be shifted (registered) so that the start point represents time point 0 and the end point represents time point 100. A local search in the two-dimensional space of registration points is then conducted, attempting to minimize the sum of squared differences between the registered curve and a "target curve". For this example, the target curve is taken as the mean of the first twenty curves registered using the registration points obtained in the previous step.

3) Crop and interpolate the data.

For each curve, discard time points before the left registration point and after the right registration point (Figure 2(b)). The remaining points are mapped to $(0, 100)$, and linear interpolation is used to generate new force values y_t at time indexes $t = 1, 2, ..., 100$ (Figure 2(c)). We chose 100 values because the cropped data has roughly this many observations before interpolation.

Although we use registration strictly as a pre-processing tool, registration information could be used for process-monitoring. That is, it might be interesting to consider monitoring registration points (e.g. start or end of the process) as well as the predicted random effect parameters or the curves themselves.

3 MODELING A SINGLE CURVE

In this section we describe two ways to characterize functional behavior of a single curve - smoothing and nonlinear modeling. A fundamental goal of both approaches is to provide a low-dimensional representation of the data. The first is an off-the-shelf technique which can be applied to any curve, whereas the latter utilizes functional information about the dynamics of the system and is tailored specifically to the force profiles presented here. In both cases, dimension reduction is achieved by using considerably fewer

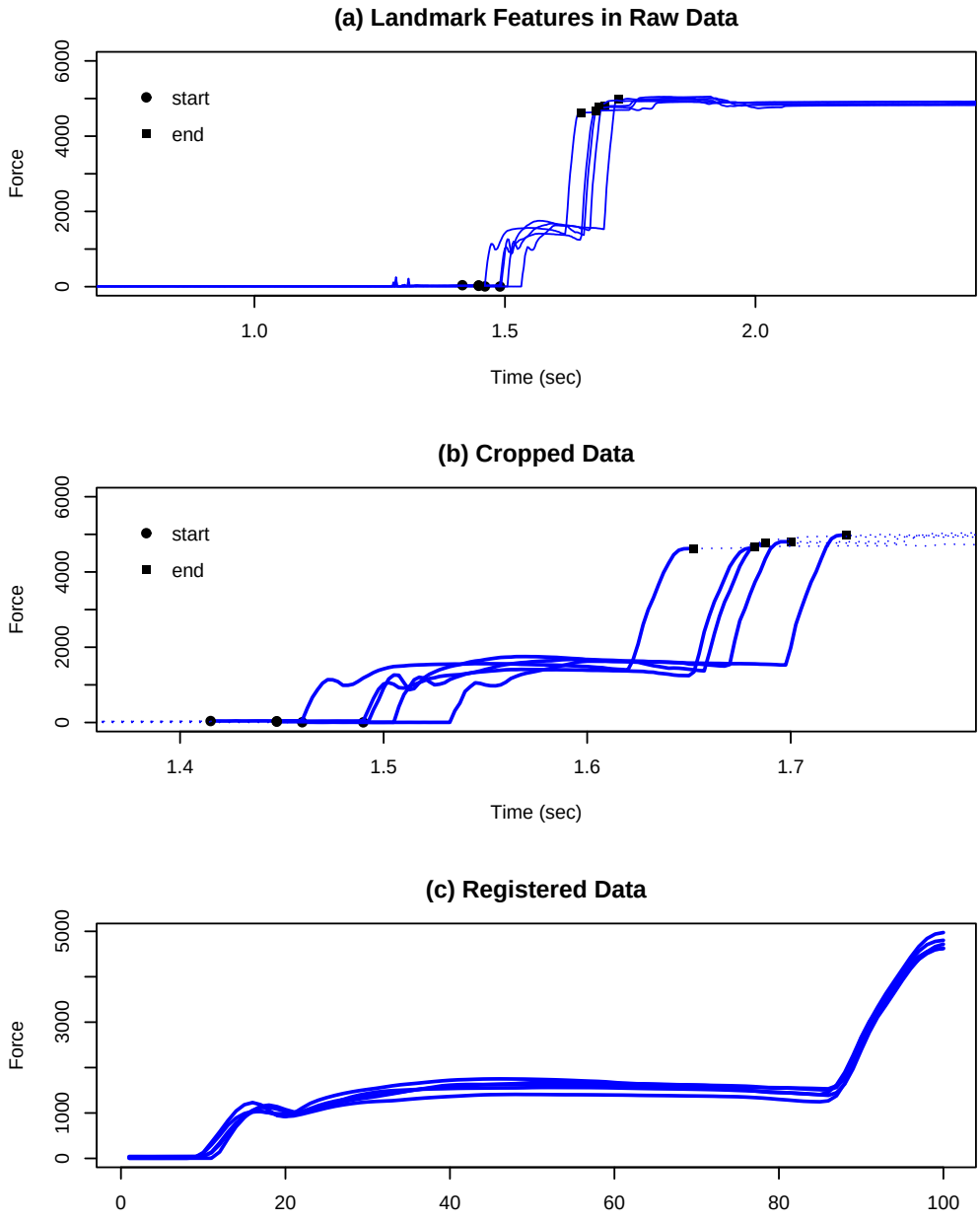

Figure 2: Registration process for five runs of the force-exertion data.

coefficients than time points on a curve to model each profile. A way of combining both methods called profiling is presented in Ramsay, Hooker et al. (2006).

3.1 Smoothing Splines

The idea behind smoothing is simple. It assumes that a vector of functional response values $\mathbf{y} = [y_1, y_2, ..., y_n]^T$ and discrete times $\mathbf{t} = [t_1, t_2, ..., t_n]^T$ at which they are observed have a functional relationship of form $f(\mathbf{t})$. That is, let

$$\mathbf{y} = f(\mathbf{t}) + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N_n\left(\mathbf{0}, \sigma^2 \mathbf{I}_n\right). \tag{1}$$

Although the exact form of $f(t)$ is unknown, one possibility is to assume that it is a linear combination of p known basis functions, such that

$$f(t) = \sum_{k=1}^{p} \theta_k b_k(t) = \boldsymbol{\theta}^T \mathbf{b}(t), \tag{2}$$

where $\mathbf{b}(t) = [b_1(t), ..., b_p(t)]^T$ are called basis functions because (2) is a basis-expansion of $f(t)$. We estimate the set of unknown coefficients $\boldsymbol{\theta} = [\theta_1, \theta_2, ..., \theta_p]^T$ by minimizing a least-squares criterion

$$SSE = \sum_{j=1}^{n} \{y_j - \boldsymbol{\theta}^T \mathbf{b}(t_j)\}^2$$

with respect to $\boldsymbol{\theta}$.

Smoothness of the fitted curve is controlled by the number, the type, and the placement of basis functions being used. Following James and Sugar (2001), b-splines basis functions were used to smooth the force-exertion data. By examining residual sums of squares, $p = 20$ was chosen as the optimal number of basis functions required to provide adequate fit. A general rule of thumb is that the smoother the curve, the fewer basis functions are needed.

Figure 3 provides an example of a b-spline fit to one of the curves from the force-exertion process. The 20 basis functions being used are plotted in heavier lines below the original data (circled points) and the fitted smooth (solid black line). The fitted line was obtained by regressing the observed data onto the matrix of basis functions generated using the `bs` function of the `splines` library in `R` (R Development Core Team, 2006). For simplicity, we choose basis functions that are equal in width and lie uniformly across the domain of the curve - a default choice in `bs`.

3.2 Nonlinear Model

While linear models simplify computation, a broader range of nonlinear functions is often appropriate. One such model for the force-exertion example is generated by observing the dynamics of the process,

which are reflected in the general shape and prominent features of the curve. For example, based on Figure 1, at least three important events appear to occur during the insertion process. These are marked by the two significant spikes in the force values at 1.53, 1.56 and 1.7 seconds. Each of the three parts of the curve can be modeled using a nonlinear exponential function. Combining this information, we get

$$f(t) = \sum_{k=1}^{3} \left\{ \alpha_k \left(1 - e^{-\beta_k (t-\delta_k)} \cos(\gamma_k (t - \delta_k)) \right) \cdot I(t > \delta_k) \right\} \tag{3}$$

where

$$I(t > a) = \begin{cases} 1 \text{ if } t > a \\ 0 \text{ otherwise.} \end{cases}$$

with the function $I(t > \delta_k)$ constraining each of the three piecewise solutions to specific regions of the curve for which $t > \delta_k$. A brief motivation for our particular choice of nonlinear functions is provided in Section 7.3.

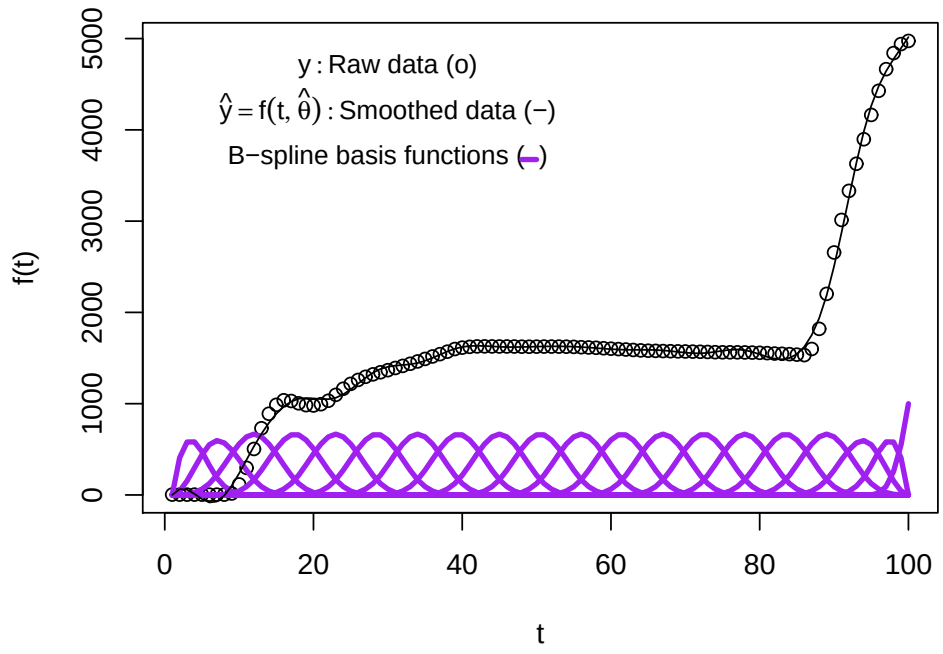

Figure 3: A b-spline fit $\hat{\mathbf{y}} = f(\mathbf{t}; \hat{\boldsymbol{\theta}})$ for one realization of the force-exertion data (points). Heavier curves at the bottom are the basis functions $b_1(t), ..., b_{20}(t)$ used to fit the data.

Following Ramsay (2000), unknown parameters in (3) can be found using nonlinear least squares. That

is $\boldsymbol{\theta} = (\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3, \gamma_1, \gamma_2, \gamma_3, \delta_1, \delta_2, \delta_3)$ can be estimated by minimizing

$$SSE = \sum_{j=1}^{n} \{y_j - f(t_j; \boldsymbol{\theta})\}^2 = [\mathbf{y} - f(\mathbf{t}; \boldsymbol{\theta})]^T [\mathbf{y} - f(\mathbf{t}; \boldsymbol{\theta})]$$

with respect to $\boldsymbol{\theta}$. As before $\mathbf{y} = [y_1, y_2, ..., y_n]^T$ represents a vector of force values observed at times $\mathbf{t} = [t_1, t_2, ..., t_n]^T$. We used the `nls` function in `R/S-PLUS` (R Development Core Team, 2006) to accomplish the minimization; another good option is the `lsqcurvefit` function in `MATLAB`'s Optimization Toolbox.

An estimate of (3) for a single curve taken from January 23^{rd} is displayed as a solid black line in the top panel of Figure 4. The three distinct regions of the curve are modeled using three different nonlinear functions, which are plotted in the bottom panel of Figure 4. These likely correspond to different events in the valve-seat insertion process.

It is clear that $\hat{\delta}_1 = 9, \hat{\delta}_2 = 12$, and $\hat{\delta}_3 = 87$, marked by dashed vertical lines in Figure 4, correspond to the times at which each of the three events was initiated. An engineering term for these parameters is reaction times, because they measure the amount of time it takes for the system to react to the initiated process. The fact that the $\hat{\delta}_1 \approx \hat{\delta}_2$ may imply that the process is dominated by two rather than three major events.

Gain parameters ($\hat{\alpha}_1 = 662, \hat{\alpha}_2 = 874, \hat{\alpha}_3 = 3079$) quantify a jump in force-exertion going from one major event to the next. Cumulative gain parameters ($\hat{\alpha}_1$, $\hat{\alpha}_1 + \hat{\alpha}_2$, and $\hat{\alpha}_1 + \hat{\alpha}_2 + \hat{\alpha}_3$) are marked by horizontal dashed lines in Figure 4.

Exponential decay or response speed parameters ($\hat{\beta}_1 = 0.187, \hat{\beta}_2 = 0.051, \hat{\beta}_3 = 0.098$) indicate how quickly changes occur. More precisely, $4/\beta_k$ approximates how long it takes for the force to plateau to a constant value during the k^{th} part of the process. This is because at $t = \frac{4}{\beta}$, $\alpha(1 - e^{-\beta t}) \approx \alpha$ in (3). The higher value of $\hat{\beta}_3$ than say $\hat{\beta}_2$ indicates that force is increasing more rapidly at the end of the process.

The sinusoidal parameters ($\hat{\gamma}_1 = 0.479, \hat{\gamma}_2 = 0.064, \hat{\gamma}_3 = -0.158$) measure volatility in the force values. Higher absolute values of these parameters correspond to highly-oscillating parts of the force curve, implying a measure of instability. One caution is that by definition $\cos(x) = \cos(-x)$, and therefore γ_k and $-\gamma_k$ generate the same curve estimates. To alleviate any confusion, we recommend either constraining these estimates to be positive during the nonlinear estimation procedure or taking absolute values after they are estimated.

In summary, we have the following interpretation for the parameters:

- α_k equals the gain, or the amount of force accumulated during event k. The overall gain of the system is given by $\sum_k \alpha_k$;

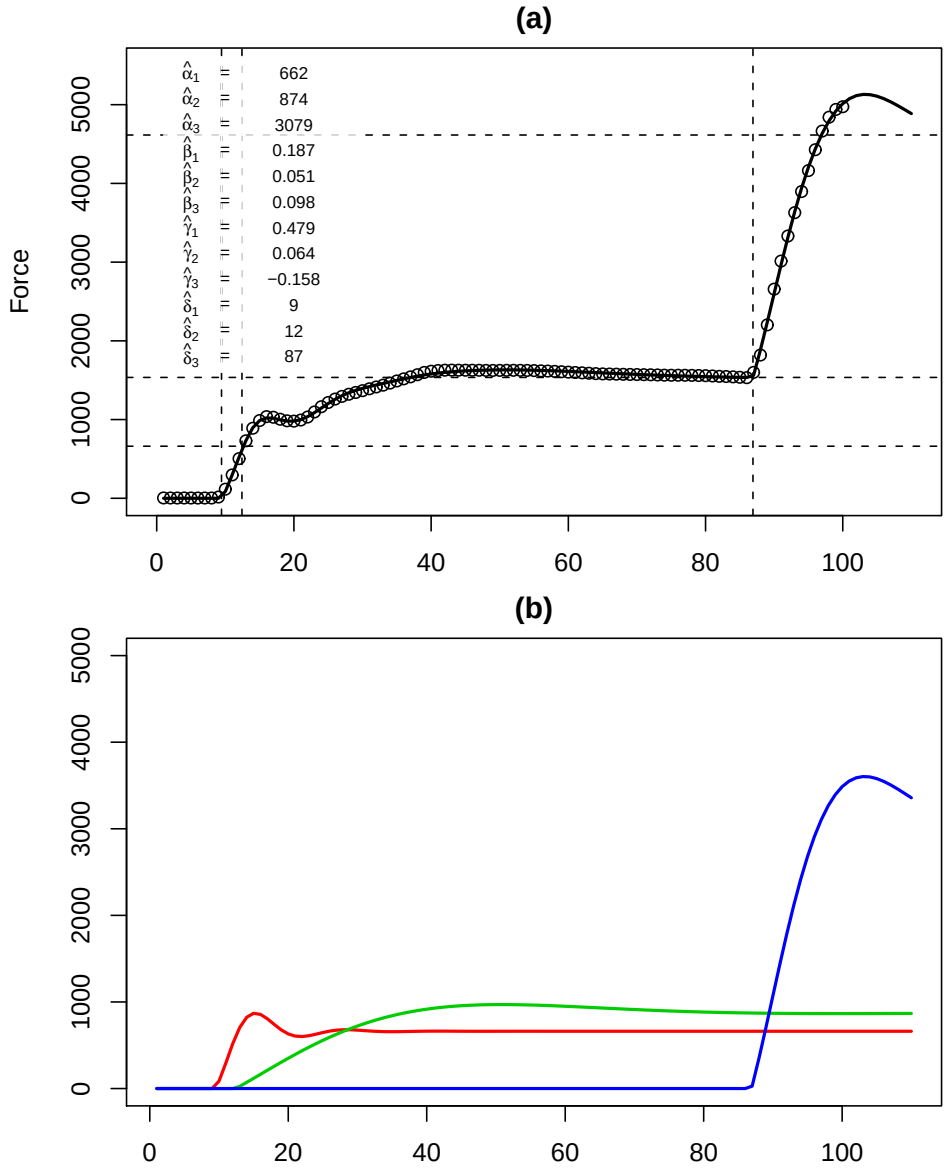

Figure 4: (a) Nonlinear least-squares fit to the second-order nonlinear model for one realization of the force-exertion data (points). (b) Three piecewise functions summed to obtained the fit in the top panel.

- β_k measures the decay, or the speed with which force stabilizes toward the gain;
- $|\gamma_k|$ represents volatility in the force values;
- Difference in reaction times, $\delta_{k+1} - \delta_k$, indicates the duration of the k^{th} part of the process.

We can think of decay and volatility as shape parameters, and gain and reaction time as level parameters. That is decay and volatility carry information about the twists and turns in the curve, whereas gain and reaction time simply shift parts of the plot toward higher or lower values.

All of this demonstrates that the key advantage of the nonlinear model over the b-spline smoothing model is interpretability of the parameters. Using relatively few parameters (12 as opposed to at least 20 in smoothing), we obtain a model that makes sense in the context of our problem. A main disadvantage of this approach is the fact that this is a nonlinear model, which has to be fit numerically with hefty computational costs.

4 MODELING COLLECTIONS OF CURVES

Now consider a collection of m independent profiles: $\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_m$. One possible extension of (1) to describing the curve $\mathbf{y}_i$ is to allow the structural parameters to vary from curve to curve, leading to the model:

$$\mathbf{y}_i = f(\mathbf{t}; \boldsymbol{\theta}_i) + \boldsymbol{\epsilon}_i; \qquad \boldsymbol{\epsilon}_i \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n), \tag{4}$$

where $\boldsymbol{\theta}_i = [\theta_{i1}, ..., \theta_{ip}]^T$ is a vector of p unknown model parameters specific to curve i, and $\boldsymbol{\epsilon}_i$ is a random component measuring pointwise roughness in the curve. To emphasize our interest in the structural parameters, we write $f(\boldsymbol{\theta}_i)$ instead of $f(\mathbf{t}; \boldsymbol{\theta}_i)$. A time-dependence suppressed by this notation is implied. For b-splines $f(\boldsymbol{\theta}_i)$ is the linear function $f(\boldsymbol{\theta}_i) = B\boldsymbol{\theta}_i$, and for the nonlinear model, it is the nonlinear function of form (3).

This model has far too many parameters to be useful in a monitoring context. An extension of (4) is the mixed effects model, which employs a set of fixed parameters ($\boldsymbol{\mu}$) that is the same across all curves and characterizes their common shape, and random parameters ($\boldsymbol{\eta}_i$ for $i = 1, ..., m$) that describe curve-specific variation from the overall shape. That is, we let $\boldsymbol{\theta}_i = (\boldsymbol{\mu}, \boldsymbol{\eta}_i)$, where $\boldsymbol{\mu}$ is a fixed vector of length p, and $\boldsymbol{\eta}_i$ is vector of length q following a $N_q(\delta, \Sigma)$ distribution.

By including both fixed ($\boldsymbol{\mu}$) and random ($\boldsymbol{\eta}_i$) parameters, each curve is modeled individually, at the same time borrowing strength from similarities amongst all profiles. Making some of the parameters random allows for two sources of variability associated with each curve - within curve variability (σ^2) and

variability in the structural parameters (Σ) contributing to differences between the individual curves. That is, by treating some of the structural parameters as random effects, we allow the profiles to differ in both roughness ($\boldsymbol{\epsilon}_i$) and shape ($\boldsymbol{\mu}, \boldsymbol{\eta}_i$).

The number of random effects used to model each curve can either be the same or fewer than the number of fixed effects ($q \leq p$). For parsimony and to ease computation, we suggest setting q to be small.

In the remainder of this section, we consider two possible applications of the mixed effects formulation to b-spline and nonlinear models.

4.1 Linear Mixed Effects for Curve Data

A straightforward application of mixed effects to b-splines is the Laird & Ware (1982) model:

$$\underset{n\times 1}{\mathbf{y}_i} = \underset{n\times p}{B_1}\ \underset{p\times 1}{\boldsymbol{\mu}} + \underset{n\times q}{B_2}\ \underset{q\times 1}{\boldsymbol{\eta}_i} + \underset{n\times 1}{\boldsymbol{\epsilon}_i}, \tag{5}$$

$$\boldsymbol{\eta}_i \sim N_q(\mathbf{0}, \Sigma) \quad \text{and} \quad \boldsymbol{\epsilon}_i \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n),$$

Here B_1 is an $n \times p$ matrix whose columns contain p known basis functions evaluated at times $\mathbf{t}$, and $\boldsymbol{\mu}$ is a fixed vector of corresponding b-spline coefficients. B_2 is a matrix of q basis functions and $\boldsymbol{\eta}_i$s are the random coefficients for each curve.

An example of a single curve fit using (5) for $p = 20$ and $q = 4$ is displayed as a solid line in Figure 5(a) alongside the raw data. The 24 basis functions used to model the data are plotted below, with the 20 columns of B_1 and four columns of B_2 shown in Figures 5(b) and 5(c) respectively.

According to model (5), the shape of each profile is dictated by an overall mean curve $B_1\boldsymbol{\mu}$ plus a curve-specific departure $B_2\boldsymbol{\eta}_i$. For monitoring, we restrict our attention to the latter. The hope is that predictions of the random effects $\boldsymbol{\eta}_i$ will capture enough key information about unique attributes of each profile to allow monitoring the process by observing changes in $\hat{\boldsymbol{\eta}}_i$ rather than the original data $\mathbf{y}_i$. This is a form of dimension reduction in that the $\hat{\boldsymbol{\eta}}_i$s serve as compact low-dimensional summaries of the differences among the curves. In an effort to summarize curve-specific variation in as few parameters as possible, we used a large number of b-splines ($p = 20$) to model the mean profile, but only a few ($q = 4$) to describe curve-specific departures from the overall mean.

A conceptual equivalent of the b-spline mixed-effects model involves subtracting the functional mean of all the profiles from each curve, and modeling the residual curves using

$$\mathbf{y}_i - \bar{\mathbf{y}} = B_2\boldsymbol{\eta}_i + \boldsymbol{\epsilon}_i,$$

$$\boldsymbol{\eta}_i \sim N_q(\mathbf{0}, \Sigma) \quad \text{and} \quad \boldsymbol{\epsilon}_i \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n).$$

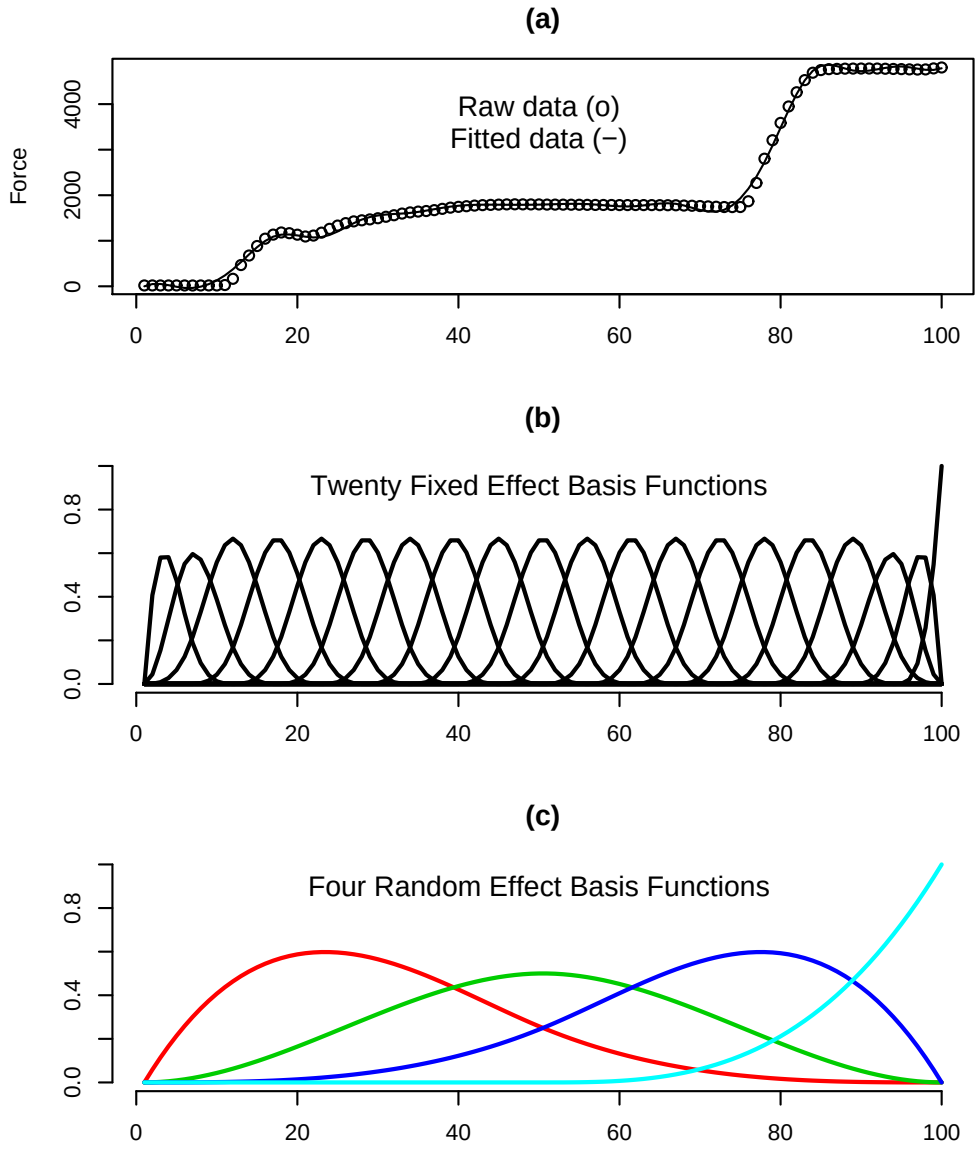

Figure 5: Linear mixed-effect model fitted to a single observation in the force-exertion data (a), and 24 basis function used to obtain the fit ((b) and (c)).

While this leads to similar results, an advantage of (5) is that it generalizes easily to more complicated scenarios. For example, rather than fitting a single fixed effect, we may also wish to control for the valve effect by including a separate fixed parameter for each valve (recall there are eight in total).

4.2 Nonlinear Mixed Effects for Curve Data

An extension of the Laird & Ware (1982) mixed model to the nonlinear case is discussed in Lindstrom & Bates (1990). We adopt a special case by letting

$$\underset{n\times 1}{\mathbf{y}_i} = f(\underset{p\times 1}{\boldsymbol{\mu}} + \underset{p\times q}{Z}\ \underset{q\times 1}{\boldsymbol{\eta}_i}) + \underset{n\times 1}{\boldsymbol{\epsilon}_i}, \tag{6}$$

$$\boldsymbol{\eta}_i \sim N_q(\mathbf{0}, \Sigma) \quad \text{and} \quad \boldsymbol{\epsilon}_i \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$$

for $i = 1, ..., m$ and $q \leq p$. Here $\boldsymbol{\theta}_i = \boldsymbol{\mu} + Z\boldsymbol{\eta}_i$, where Z is an indicator matrix dictating which of the p parameters have a random component.

In the application, the nonlinear form of $f(\boldsymbol{\theta}_i)$ is given by

$$f(\boldsymbol{\theta}_i) = \sum_{k=1}^{3} \left\{ \alpha_{ik} \left(1 - e^{-\beta_{ik}(t-\delta_{ik})} \cos(\gamma_{ik}(t - \delta_{ik})) \right) \cdot I(t > \delta_{ik}) \right\}.$$

The structural parameters are $\boldsymbol{\theta}_i = (\alpha_{i1}, \alpha_{i2}, \alpha_{i3}, \beta_{i1}, \beta_{i2}, \beta_{i3}, \gamma_{i1}, \gamma_{i2}, \gamma_{i3}, \delta_{i1}, \delta_{i2}, \delta_{i3})$, and their interpretation is discussed in Section 3.2. The formulation of our model in (6) allows all 12 of these parameters to differ for each curve (e.g. $Z = I_p$ or equivalently $\boldsymbol{\theta}_i = \boldsymbol{\mu} + \boldsymbol{\eta}_i$). However it seems likely that only a handful of the parameters capture significant changes in the curves. It is only these parameters that we wish to have random components, with the remaining fixed across all of the curves.

A key distinction between (6) and the linear model (5) is that there is an additive relationship between the fixed and random effects (i.e. $\boldsymbol{\theta}_i = \boldsymbol{\mu} + Z\boldsymbol{\eta}_i$), which implies that some (or all) of the fixed effects have associated random components. This is different from the linear case, where we use fixed effects to fit the mean curve ($B_1\boldsymbol{\mu}$) and random effects ($B_2\boldsymbol{\eta}_i$) to fit the residual curves after subtracting off the mean profile. In the linear case, all we have to decide is how many of the parameters we wish to make random; for the nonlinear model, we must decide which of the fixed effects will have associated random components. An example of how to make this choice for the force-exertion data is presented in Section 6.

5 PROFILE MONITORING

Recall that our ultimate goal is to set forth statistical machinery for detecting curves that are "unusual" during the course of the valve-seat insertion process. Process monitoring allows us to do this. The usual

course of action in this procedure consists of two phases (Ryan, 1989). In Phase I, a base set of observations is selected as representative of normal operation, and scrutinized for unusual behavior. Detected outliers (if any) are investigated and removed if and only if there are attributable causes of their abnormality. The base set is then re-examined, and control limits indicating bounds for "normal" or "in-control" activity are calculated. Phase II of the analysis involves assessing new curves from ongoing production using the control limits obtained from the base set. The process is flagged as out-of-control whenever new observations fall outside the bounds of the control limits. Unusual observations are then investigated for potential sources of the problem, and corresponding adjustments to the process are made.

The fact that we are dealing with functional observations complicates the task. Monitoring observed profiles as multivariate observations is known not to work well in practice due to such issues as correlation between neighboring values within a single curve, and the curse of dimensionality for curves that are observed at many time points. A natural extension of the process-monitoring procedure to functional data is to monitor estimated structural parameters instead of the observed data (Williams et al, 2003). In this section we describe a procedure for monitoring the predicted random effects of a functional observation under the mixed-effects model.

Key theoretical results employed throughout this section are provided in the Appendix. To be consistent with the language used in this field, we distinguish between the two different ways in which the fixed and random effects are estimated by calling the former estimates and the latter predictions.

5.1 Chart for Individual Curves

Starting with m curves in Phase I, let $\hat{\boldsymbol{\eta}}_i$ denote predicted random effect vectors for each curve approximately following a $N_q(\mathbf{0}, W)$ distribution. Then a Hotelling T^2 measure of atypical behavior for each curve is given by:

$$T_i^2 = \hat{\boldsymbol{\eta}}_i' W^{-1} \hat{\boldsymbol{\eta}}_i.$$

Large values of this test statistic imply that $\hat{\boldsymbol{\eta}}_i$ is far from its expected value of zero, and corresponding curves would be flagged as "unusual". To define large, we must determine the distribution of T_i^2.

If $\hat{\boldsymbol{\eta}}_i \sim N_q(\mathbf{0}, W)$ exactly, this is trivial. Let $W^{-1/2}$ denote a $q \times q$ symmetric standard error matrix of $\hat{\boldsymbol{\eta}}_i$, obtained using eigen-decomposition of W. Then $\boldsymbol{\zeta}_i = W^{-1/2}\hat{\boldsymbol{\eta}}_i \sim N_q(\mathbf{0}, I_q)$, so that $T_i^2 = \boldsymbol{\zeta}_i'\boldsymbol{\zeta}_i$ is a sum of q standard normals squared. It follows that $T_i^2 \sim \chi^2(q)$.

Of course W is unknown and the exact distribution of $\hat{\boldsymbol{\eta}}_i$ cannot be determined. A natural solution is

to approximate W using a sample covariance matrix of the random predictions:

$$\hat{W} = \frac{1}{m-1}\sum_{i=1}^{m}(\hat{\boldsymbol{\eta}}_i - \bar{\hat{\boldsymbol{\eta}}})^T(\hat{\boldsymbol{\eta}}_i - \bar{\hat{\boldsymbol{\eta}}}),$$

where $\bar{\hat{\boldsymbol{\eta}}}$ is the average of the predicted random effects. Another possibility is to estimate W by $\tilde{W}$, where $\tilde{W}$ is a model-based estimate of the covariance matrix for the predicted random effect, found by fitting a mixed-effects model using all of the Phase I data. Based on the results in the Appendix, $\tilde{W} = \hat{\Sigma}\hat{J}\hat{V}^{-1}\hat{J}^T\hat{\Sigma}$ in the linear case and $\tilde{W} = \hat{\Sigma}Z'\hat{J}\hat{V}^{-1}\hat{J}^TZ\hat{\Sigma}$ in the nonlinear scenario.

Based on simulation results and for the insertion data, both $\tilde{W}$ and $\hat{W}$ produce nearly identical estimates of W. We chose to use $\hat{W}$ because it is consistent with the usual way process-monitoring is conducted on multivariate data and most easily understood. Other possible covariance estimators discussed in Williams et al. (2003) but not explored in this paper include the successive differences estimator (Hawkins & Merriam, 1974) and the minimum volume ellipsoid estimator (Rousseeuw, 1984).

Using either $\tilde{W}$ or $\hat{W}$ in place of W, the chi-squared distribution of T_i^2 can be approximated by an F distribution with q and $m-q$ degrees of freedom. This is formalized in the following extension of a well-known result from multivariate statistics (Johnson & Wichern, 1992).

Result 5.1 *Suppose that* $\hat{\boldsymbol{\eta}}_i \sim N_q(\mathbf{0}, W)$ *is an unbiased estimator of a* $\boldsymbol{\eta}_i$ *($i = 1, ..., m$ where m is the total number of profiles) and let* $T_i^2 = \hat{\boldsymbol{\eta}}_i'\hat{W}^{-1}\hat{\boldsymbol{\eta}}_i$ *such that* $\hat{W} \approx W$ *as* $m \to \infty$. *If curve i is in-control, then*

$$F_i = \frac{m-q}{q(m-1)}T_i^2 \sim F(q, m-q).$$

It follows that curve i is unusual if $F_i \geq F_\alpha(q, m-q)$ for some significance level α. As a general rule throughout the paper we let $\alpha = 0.0027$. Combining all of the ideas presented thus far, our approach to monitoring individual curves in the force-exertion data can be summarized as follows:

PHASE I

1. Select a base set of m curves. Fit a mixed model to these data. Use it to obtain predictions of the random effects $\hat{\boldsymbol{\eta}}_i$ for each curve and an estimate of the associated covariance W.

2. Calculate Hotelling T^2 statistics for each curve, $T_i^2 = \hat{\boldsymbol{\eta}}_i'\hat{W}^{-1}\hat{\boldsymbol{\eta}}_i$. Use a QQ plot and/or a Kolmogorov-Smirnov test to verify that $F_i = \frac{m-q}{q(m-1)}T_i^2 \sim F(q, m-q)$ approximately.

3. Confirm that there are no unusual curves in the base set by examining a control chart of the test statistics. Investigate observations corresponding to $F_i \geq F_\alpha(q, m-q)$ for unusual behavior. Remove these if and only if the production process can be adjusted to avoid such observations in the future. $F_\alpha(q, m-q)$ is called an upper control limit (UCL) because it provides an upper bound for plausible values of F_i at the α level of significance.

PHASE II

1. For each new profile, predict its random effects $\hat{\boldsymbol{\eta}}_i$ under the mixed-model from Phase I.
2. Calculate Hotelling T^2 statistics for each new profile using $\hat{W}$ from Phase I.
3. Verify that the process is in-control by examining control charts of the test statistics. Flag observations that fall outside of the UCL established in Phase I as "unusual".

5.2 Chart for Subgrouped Curves

An important process-monitoring technique for situations when there are small persistent changes in the process is subgrouping. The basic idea is to monitor the means of subgrouped data rather than the individual observations themselves (Ryan, 1989), which allows the monitoring procedure to be more sensitive to small departures.

An extension of profile-monitoring to subgrouped data is straightforward. Let m denote the total number of curves in Phase I. Fit a mixed-effects model to these data and obtain predicted random effects for each curve. Group these into k sets of equal size $m_g = m/k$, and calculate corresponding subgroup means of the predicted random effects. Proceed to monitor the subgrouped averages as if they correspond to individual curves.

The following result formalizes profile-monitoring for subgrouped data (Ryan, 1989).

Result 5.2 *Suppose that* $\hat{\boldsymbol{\eta}}_i \sim N_q(\mathbf{0}, W)$ *is an unbiased predictor of* $\boldsymbol{\eta}_i$ $(i = 1, ..., m)$. *Group these into* k *sets* $S_1, S_2, ..., S_k$ *of equal size* m_g, *and let*

$$\bar{\hat{\boldsymbol{\eta}}}_g = \frac{1}{m_g} \sum_{i \in S_g} \hat{\boldsymbol{\eta}}_i \quad \texttt{for} \quad g = 1, ..., k.$$

denote the means of predicted coefficients for each subgroup. Then a Hotelling T^2 *criterion for testing that the functional mean of the curves in subgroup* g *is "unusual" is given as*

$$T_g^2 = \frac{m}{k} \left[\bar{\hat{\boldsymbol{\eta}}}_g^T \hat{W}^{-1} \bar{\hat{\boldsymbol{\eta}}}_g \right]$$

where $\hat{W} \approx W$ *as* $m \to \infty$. *When profiles in subgroup* g *are in-control,*

$$F_g = \frac{m-k-q+1}{mq-kq-mq/k+q} T_g^2 \sim F(q, m-k-q+1).$$

Phase I and II steps of the analysis for subgrouped data extend directly from Section 5.1 by replacing $\hat{\boldsymbol{\eta}}_i$, F_i and T_i^2 with $\hat{\boldsymbol{\eta}}_g$, F_g and T_g^2 as defined in Result 5.2, and setting the UCL equal to $F_\alpha(q, m-k-q+1)$. For the subgrouped data, we let $\hat{W}$ equal the mean of the within-subgroup sample covariance estimates. This is consistent with what is done in multivariate process-monitoring (Ryan, 1989). Another possibility is to let $\hat{W} = \sum_{i \in S_g} \tilde{W}_i / nk$ or $\hat{W} = \tilde{W}/nk$, which does not seem to change the results for our data, but might be more appropriate depending on the problem at hand.

An alternative to the subgrouped chart formulated here is to subgroup and average the curves first, then fit a mixed effects model to the subgroup averages, and monitor predicted random effects obtained from the model as you would individual observations. This is not explored here.

6 EXAMPLE

The force-exertion data that motivated our work in this area is taken from the process of inserting valve seats onto cylinder heads that is part of the assembly of automobile engines. A subset of 3776 force profiles is analyzed. These include all of the insertions made in January and 1008 curves from the last 6 days in February. February data are used in Phase I and January data in Phase II of the analysis. Note that the process was started on January 1. Although monitoring the past is not representative of what would happen in reality, this was done for illustrative purposes only, as we expected that the production process stabilize as time progressed. This rationale is discussed in section 7.2.

Our goal is to identify unusual insertions made in January (if any) using the profile-monitoring techniques developed in this paper. This is accomplished by following the steps outlined in Section 5. Results from the two phases of the analysis are summarized below. To save space and avoid repetition only results for the first valve are presented here.

R 2.0.0 statistical software (R Development Core Team, 2006) was used for computation, utilizing the NLME package to fit mixed effects models (Pinheiro & Bates, 2001).

6.1 Principal Component Analysis

Before modeling the data, all 3776 curves are registered, and principal component analysis (PCA) is used to understand the structure in the processed curves. PCA seeks a low-dimensional representation of p dimensional data (here $p = 100$) that captures most of the data variability. Based on 1008 force curves from the last six days in February, the first two principal components explain 89% of the overall variation in the curves.

Combined with process knowledge, information contained in the first two PCs can be used to decide which parameters should have random components in the nonlinear model. Interpretability of the parameters in the nonlinear model presented here makes incorporating known information relatively straightforward.

We use a technique suggested by Jones & Rice (1996), which involves projecting the data onto the first few principal components and plotting the curves corresponding to the smallest and largest projected values of the component. Figure 6 demonstrates this idea for the first two PCs of the force-exertion profiles. Curves with extreme values in the first principal component (Figure 6(a)) differ the most in the long flat section spanning time points 25 through 75. Variation is also evident in both the time points at which sharp vertical increases occur and in the extent of the sinusoidal loop in the curves between time points 10 and 20. This is further emphasized by looking at the bottom panel describing variability contained in the second PC. Variation in the second PC also stems from the flat part of the curve between time points 85 through 100, and the amount of decay in the long flat bit at times 25 through 75.

Based on a combination of process-knowledge and the results of the PCA analysis, α_{i2}, α_{i3}, δ_{i1}, δ_{i2}, and δ_{i3} are chosen to have random components in the nonlinear model. In an earlier discussion, we referred to these as the level parameters, because they determine the amount of vertical and horizontal stretch and skew of the force curves. In the context of model (6), this particular choice of random effects implies setting

$$\boldsymbol{\theta}_i = (\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3, \gamma_1, \gamma_2, \gamma_3, \delta_1, \delta_2, \delta_3) + (0, \eta_{i1}, \eta_{i2}, 0, 0, 0, 0, 0, 0, \eta_{i3}, \eta_{i4}, \eta_{i5})$$

or equivalently $\boldsymbol{\theta}_i = \boldsymbol{\mu} + Z\boldsymbol{\eta}_i$, where $\boldsymbol{\mu} = (\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3, \gamma_1, \gamma_2, \gamma_3, \delta_1, \delta_2, \delta_3)$, Z is a 12×5 sparse matrix of zeros everywhere except for entries (2,1), (3,2), (10,3), (11,4) and (12,5), which are equal to 1, and $\boldsymbol{\eta}_i = (\eta_{i1}, \eta_{i2}, \eta_{i3}, \eta_{i4}, \eta_{i5})$.

6.2 Phase I Analysis

The first part of monitoring involves looking at the February data. B-spline and nonlinear mixed models are fit, producing predicted random coefficients and their estimated covariances. For b-splines, 20 fixed

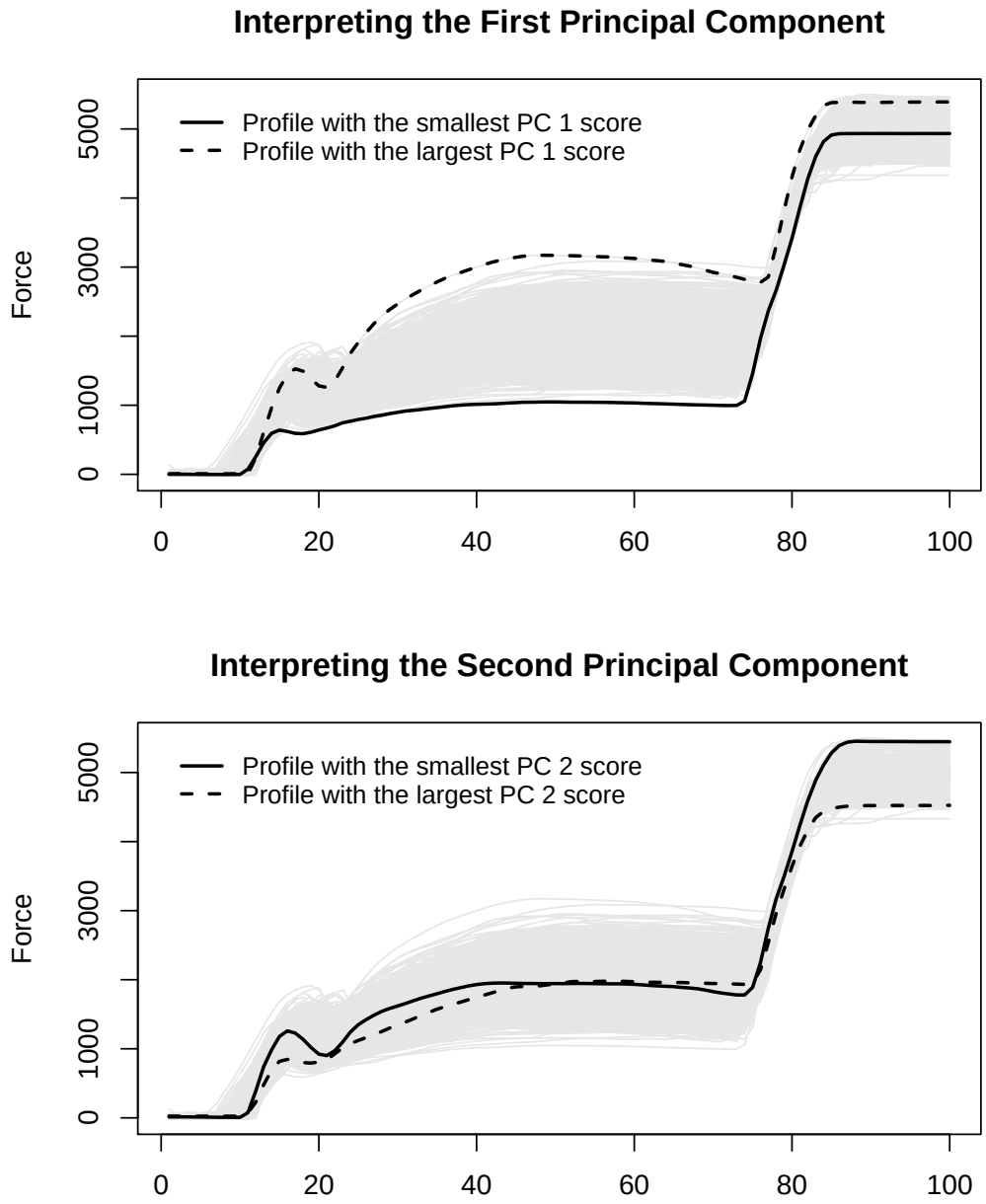

Figure 6: Exploring variability in the February force-exertion data as summarized by the first two principal components (PCs). The data are plotted in light gray, and curves corresponding to the smallest and largest PC scores are highlighted in black.

and four random parameters are fit. For the nonlinear model, we use 12 fixed parameters, of which five (α_{i2}, α_{i3}, δ_{i1}, δ_{i2}, and δ_{i3}) are chosen to have random components.

For purpose of illustration, we fit each model to the base set and get the predicted random effects for each of the 1008 curves, then form subgroups of five consecutive predictions. Since there is 100% inspection, it is more appropriate to chart individual insertions. From the February base set, we have 202 subgroups.

For each model, 1008 predicted random coefficient vectors are subgrouped into 202 sets of five parameter vectors per group. Such subgrouping implies 100% inspection, meaning that all of the Phase I data are used. This is done mostly for illustration purposes as well as because we had access to large amounts of data. An alternative not explored in this paper would be to group five predicted parameters from say every day.

We fit each model to the base set and get the predicted random effects for each of the 1008 curves. For each subgroup ($g = 1, ..., 202$), corresponding F_g statistics are calculated as defined in Result 5.2. Our first goal is to verify that these test statistics follow an appropriate F-distribution. For b-splines, the null hypothesis is that $\frac{797}{3184}T^2 \dot{\sim} F(4, 797)$. The corresponding Kolmogorov-Smirnov test statistic and p-value are 0.0578 and 0.5095, meaning that the null hypothesis should not be rejected. That is, assuming the $F(4, 797)$-distribution for the test statistics is not unreasonable. For the nonlinear model, the test statistic and p-value are 0.0585 and 0.4931 respectively, leading to the same conclusion.

Next, control charts are built for February data using distributional properties of the T^2 statistic based on subgrouped data. This did not reveal any outliers (graphics not shown).

6.3 Phase II Analysis

The second stage is to examine January data for potential faults in the online production process. Covariance estimates from Phase I are used to calculate predicted values of the random coefficients for both linear b-spline and nonlinear models. In each case, these are split into subgroups of five and corresponding subgrouped T^2 statistics obtained as stated in Result 5.2.

Subgrouped control charts for the first valve are provided in Figures 7(a) and 7(c). The first plot is for the b-spline model and the second is for the nonlinear model. Dashed horizontal lines mark the upper control limits in each case. For b-splines this is $F_{0.0027}(4, 2218) = 4.08$ at the 0.0027 significance level. For the nonlinear model $F_{0.0027}(5, 2217) = 3.65$. Any observations that fall above these values for each respective model are regarded as "out-of-control".

One striking feature of these results is a number of the flagged insertions occur during the first few days of January. This makes perfect sense. Problems are likely to occur in the initial stages of production

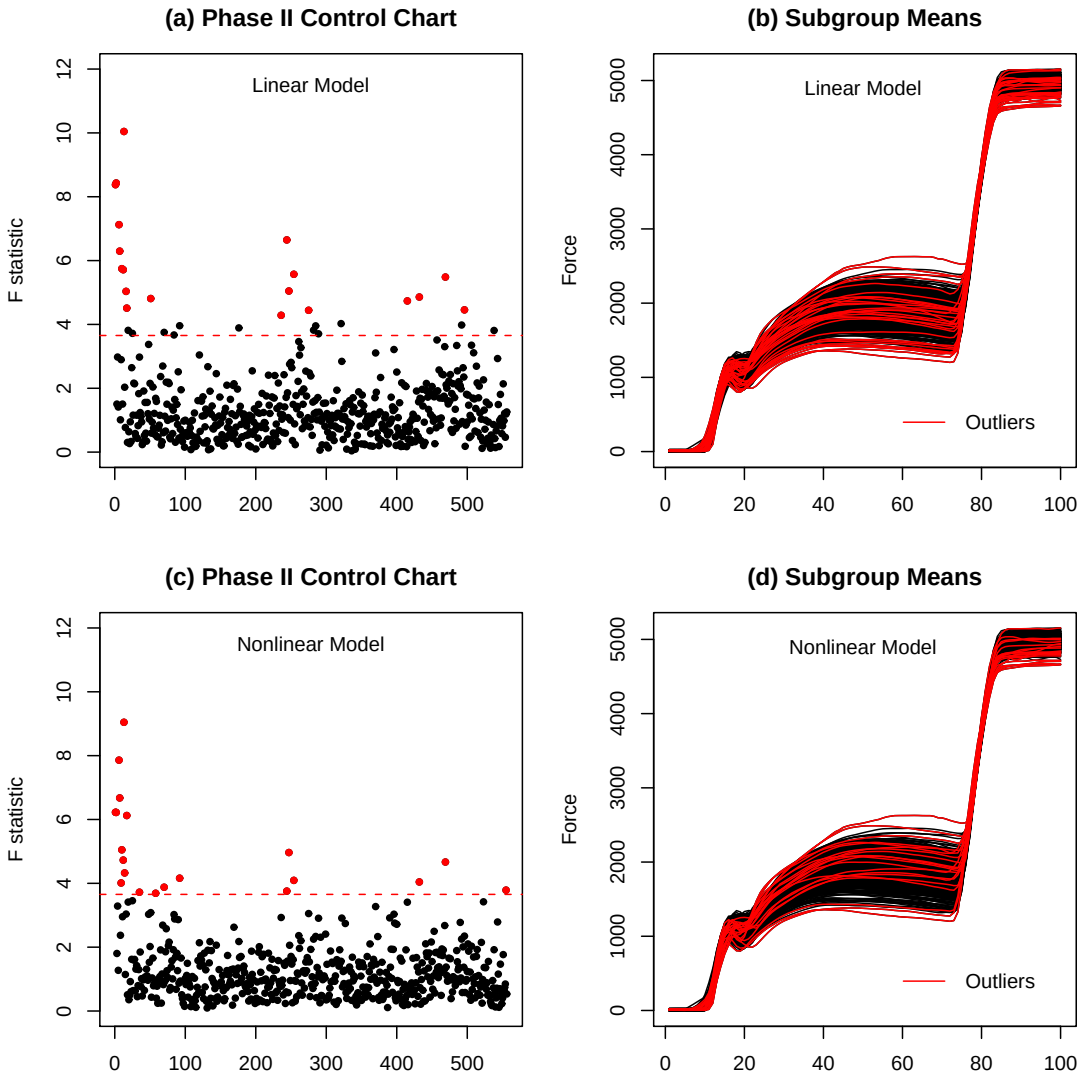

Figure 7: Phase II control charts for valve one under the linear b-spline (a) and nonlinear (c) models. Functional means of the outlying subgroups of profiles flagged by each chart are displayed in red alongside all of the subgroup means (black lines) in (b) and (d).

as the engine-assembly process is set up. Over time these problems tend to be resolved in an effort to improve performance - an intuitive notion that seems to be supported by the control charts. Two other less prominent sets of outlying curves occur in the mid-January, and during the last week of that month.

Averages of the outlying subgrouped curves for the linear and nonlinear models are displayed to the right of the control charts in Figures 7(b) and 7(d) respectively. Consistency between the fundamentally different b-spline and nonlinear models is reassuring. The seven most outlying subgroups of curves flagged by at least one or both of the linear and nonlinear procedures include two on January 7^{th}, four on the 8^{th}, and one on the 9^{th}.

Looking at the shapes of the average of the most outlying subgroup of curves, they appear to differ from the rest in the dip of the middle flat part of the profile and have a relatively small amount of force gain at the end. Process experts might be able to attribute these features to specific faults in the manufacturing procedure. Visualization techniques may yield insight into why curves are flagged as outliers. For example, a plot comparing the outlying curve to the mean curve might highlight discrepancies. Also, treating the predicted random effects $\hat{\eta}$ for each curve as "data" and applying exploratory graphical analysis such as parallel coordinate plots and scatterplot matrices might indicate which random effect made the curve appear unusual.

Results for the remaining valves are very similar. Across all valves, the largest outliers tend to correspond to insertions made during the first few days of January.

7 DISCUSSION

The goal of this paper was to address key issues involved in monitoring functional process data such as the force-exertion example, and to develop tools for accomplishing the task. We began our discussion by presenting an algorithm for registering our data, followed by an introduction to b-spline and nonlinear modeling of a single functional observation. Linear and nonlinear mixed-effect models were proposed for the analysis of multiple curves, and Hotelling T^2 used to differentiate between "good" and "bad" insertions. Results based on the force-exertion data indicate reasonable performance. Some of the underlying issues raised by the proposed methodology and future research are considered next.

7.1 Choosing Random Effects

When formulating the mixed effects model, it is not clear how many random effects should be chosen. For both b-spline and nonlinear models, we recommend restricting the number of random effects to as

few as possible. This reduces the computational cost of the procedure, and allows us to monitor compact summaries of the curves in a low dimension.

For nonlinear models, it is also difficult to know which of the parameters should have random components. In Section 4.2, we advised using both process knowledge and conclusions drawn from PCA in selecting the fixed parameters that should have associated random components. Another possibility is to fit 12 different models, where each model has a single random effect. We could then use model-selection criteria such as AIC or BIC to identify the best-fitting few. Random effects associated with these fits are then selected to be used in the final model. Alternatively, we could adapt a step-wise model-selection approach or use a straightforward fixed-effect regression fit to identify which of the parameters are the most significant, then make these random in a re-fitted mixed-effects model. In both cases, it is important to standardize the parameters as they are not measured on the same scale.

In general, deciding how many random effects should be used and which parameters to make random is a challenging problem. The approach of shrinking a large number of random effects taken by Morris and Carrol (2006) does not seem to be in the spirit of a parsimonious representation of the data. Rice and Wu (2001) suggest the use of information criteria (AIC, BIC, cross-validation) for selection of the number of equally-spaced splines used in a random effects model. Such a technique might be appropriate for the approach considered here.

7.2 Choice of Phase I Data

One problem common to all control chart applications is the choice of a base set in Phase I. For the force-exertion data, we monitored the January data and used all 1008 insertions made in the last six days of February as a baseline set indicative of a well-behaved process. This cannot represent what would be done in practice (historical not future data is used as a baseline), but was done for demonstration purposes only due to the suspicion that the insertion-process stabilized as time went on. Even so, with the limited amount of information that we have about our data, it is unclear whether or not February observations are truly representative of "normal operation".

The definition of a subgroup and the number of subgroups in the base set is also subjective. One helpful tactic is to re-fit the data after a while and adjust control limits using appropriate sampling and/or subgrouping schemes.

7.3 Linear versus Nonlinear Models

Several factors are worth noting when deciding whether the b-spline or the nonlinear model is best suited for the problem. Advantages of b-splines are that they are linear in the parameters and generic enough to be applied to any data without the need for understanding the dynamics of the process. This also means that they are easier and faster to fit computationally. In contrast, the nonlinear model involves parameters that are easy to interpret in the context of the problem. As with any situation where there is a choice between two different modeling techniques, there is no "best" model, but there might be one that is more appropriate than the other in terms of the trade-offs between computation and interpretability.

For the force exertion data, we also believe that the nonlinear model is better at explaining key sources of variability because it is application-specific. The model uses derivatives to characterize observable changes in the process. Specifically, each of the three spikes in the force curves is the solution to a second-order differential equation (DE). We chose to use second-order DEs as they allowed a lot more flexibility than the simplest possible first-order DEs in modeling curvature at the endpoints. For more information on the use of DEs to model functional data, see Ramsay & Silverman (2005).

7.4 Sequential Charts

The basic idea of monitoring the predicted random effects of a mixed-effects model can easily be extended to include a wide range of other multivariate charts. These include sequential charts and charts to monitor the covariance matrix, with the type of approach used depending on the context of the problem.

We considered one such possible extension - the use of exponentially weighted moving average (EWMA) monitoring procedure for profile data. Following a multivariate version of the EWMA chart developed by Lowry (1992), our EWMA procedure monitors weighted averages of the estimated random effects in a way that gives less weight to observations further in the past. More formally, using Lowry's approach, we monitor

$$EWMA_i = Z_i' W_Z^{-1} Z_i \geq UCL.$$

where

$$Z_i = \lambda \boldsymbol{\eta}_i + (1-\lambda) Z_{i-1}; Z_0 = 0$$

and

$$W_Z = Cov(Z_i) = \frac{\lambda(1-(1-\lambda)^{2i})}{2-\lambda} * W.$$

As before $\hat{\boldsymbol{\eta}}_i$ denotes the i^{th} predicted random effect $(i = 1, ..., m)$ and W is replaced by an appropriate estimate $\hat{W}$. A corresponding extension to subgrouped data involves using $\bar{\hat{\boldsymbol{\eta}}}_g$ and $W = \hat{W}$, where $\hat{W}$ is

the average of the within-subgroup covariance matrices.

A univariate alternative is a EWMA chart based on the Hotelling T^2s. That is, we can let

$$Z_i = \lambda T_i^2 + (1-\lambda) Z_{i-1}$$

and

$$Var(Z_i) = \frac{\lambda}{2-\lambda} * Var(T_i^2).$$

Multivariate EWMA chart for the insertion data provided results similar to those of the Hotelling T^2 (not shown here).

7.5 Charts Using Both Fixed and Random Effects

The underlying theme throughout this paper has been the monitoring of random effects parameters. Another possibility is to monitor a measure of overall variation explained by the fitted mixed effects model, such as the residual sum of squares, in addition to the random effects.

7.6 Charts for Multiple Valves

With respect to the force-exertion example, we focused our analysis on data from only one of the eight valve-seat insertions per head. One possibility would be to extend some of the ideas developed in this paper to models that consider all eight valves simultaneously.

The easiest way to include information on multiple valves is to include a deterministic valve effect in the model. For b-splines this is equivalent to fitting

$$\mathbf{y}_{ij} = B_1\boldsymbol{\mu} + B_2\boldsymbol{\gamma}_j + B_3\boldsymbol{\eta}_i + \boldsymbol{\epsilon}_{ij},$$

$$\boldsymbol{\eta}_i \sim N_q(\mathbf{0}, \Sigma) \quad \text{and} \quad \boldsymbol{\epsilon}_{ij} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I}_n),$$

where $i = 1, ..., m$ indexes the insertion, and $j = 1, ..., 8$ indexes the valve number.

Both the μ and the valve parameter $\boldsymbol{\gamma}_j$ in this model are fixed, which allows us to control for both the overall trend and any systematic valve effects in the data. The same basic idea can be extended to controlling for the ram effect (we have two rams, each making four valve insertions) by including a fixed effect for the ram rather than the valve.

ACKNOWLEDGMENTS

We would like to thank Professor Jim Ramsay of McGill University for sharing his valuable time and experience with us. Collaborative work with Professor Ramsay lead to the formulation of the nonlinear model presented here.

APPENDICIES

In this section we summarize some of the theoretical results for the b-spline and nonlinear mixed-effect models. Resources used in this section include results and discussions by Laird & Ware (1982), Lindstrom & Bates (1988 and 1990), Demidenko (2004) and Pinheiro & Bates (2001). The last two books in particular provide an excellent overview of linear and nonlinear mixed-effect models.

A Linear Model

For linear mixed-effects model, the following hold true (Pinheiro & Bates 2001, §2.5).

Result A.1 *Let* $\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_m$ *denote a collection of* m *independent profiles modeled by the linear mixed-effects model (5), such that* $V = B_2 \Sigma B_2' + \sigma^2 \mathbf{I}_n$ *denotes overall variability in the curves. Then for* σ^2 *and* Σ *known,*

- *The best linear unbiased estimator (BLUE) of* $\boldsymbol{\mu}$ *is given by* $\hat{\boldsymbol{\mu}} = (B_1' V^{-1} B_1)^{-1} B_1' V^{-1} \bar{\mathbf{y}}$*;*
- *The best linear unbiased predictor (BLUP) of* $\boldsymbol{\eta}_i$ *is given by* $\hat{\boldsymbol{\eta}}_i = \Sigma B_2' V^{-1} (\mathbf{y}_i - B_1 \hat{\boldsymbol{\mu}})$.

It follows that $\hat{\boldsymbol{\mu}} \sim N_p(\boldsymbol{\mu}, (m B_1' V^{-1} B_1)^{-1})$ *and* $\hat{\boldsymbol{\eta}}_i \sim N_p(\mathbf{0}, \Sigma B_2' V^{-1} B_2 \Sigma)$.

When variance components are not known, these results are no longer applicable. The best we can do is estimate the variances from the data, and assume that for large sample size m, the distributional properties hold true approximately. Since none of the estimators have closed-form expressions, it is not clear how to prove this mathematically. However, empirical results based on the analysis of both simulated and real data seem to support this claim.

B Nonlinear Model

Estimation and prediction for the nonlinear model is complicated by the fact that $f(\boldsymbol{\theta}_i)$ where $\boldsymbol{\theta}_i = (\boldsymbol{\mu}, \boldsymbol{\eta}_i)$ is nonlinear in terms of the structural parameters. Even if the variance components σ^2 and Σ were known, there are no closed-form maximum likelihood (ML) or generalized least squares (GLS) solutions for $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\eta}}_i$. Instead these must be calculated via numerical optimization. The usual approach is an adaptation of the Gauss-Newton technique, which works iteratively by linearizing $f(\boldsymbol{\theta}_i)$ using a Taylor series expansion about some initial value and GLS or ML estimates of the parameters under the approximation. These replace the initial values and the process is repeated in this fashion until convergence is attained.

Detailed descriptions of the Gauss-Newton approach to nonlinear regression can be found in Gallant (1975), Bates & Watts (1988) and Greenwood (2004), with discussions in the context of mixed models developed in Lindstrom & Bates (1990), Wolfinger (1993) and Pinhero & Bates (2001).

Determining distributional properties of $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\eta}}_i$ for the nonlinear model is challenging as well. Some progress can be made by linearizing $f(\boldsymbol{\theta}_i)$ and then applying Result A.1, which should hold approximately true depending on the accuracy of the linearization. Using first-order Taylor series expansion of a nonlinear vector function $f(\boldsymbol{\theta}_i)$ about some true value of the structural parameters denoted by $\boldsymbol{\theta}_i^*$, we get

$$f(\boldsymbol{\theta}_i) \approx f(\boldsymbol{\theta}_i^*) + J(\boldsymbol{\theta}_i^*)^T(\boldsymbol{\theta}_i - \boldsymbol{\theta}_i^*)$$

where $J(\boldsymbol{\theta}_i^*) = \frac{df(\boldsymbol{\theta})}{d\theta}|_{\boldsymbol{\theta}_i=\boldsymbol{\theta}_i^*}$ is a $p \times n$ matrix whose rows correspond to partial derivatives of $f(\boldsymbol{\theta}_i)$ with respect to $\boldsymbol{\theta}_i$ evaluated at $\boldsymbol{\theta}_i = \boldsymbol{\theta}_i^*$.

It follows from (6) that

$$\mathbf{y}_i \approx f(\boldsymbol{\theta}_i^*) - J(\boldsymbol{\theta}_i^*)^T\boldsymbol{\theta}_i^* + J(\boldsymbol{\theta}_i^*)^T\boldsymbol{\theta}_i + \boldsymbol{\epsilon}_i,$$

or equivalently

$$\mathbf{y}_i - f(\boldsymbol{\theta}_i^*) + J(\boldsymbol{\theta}_i^*)^T\boldsymbol{\theta}_i^* \approx J(\boldsymbol{\theta}_i^*)^T\boldsymbol{\theta}_i + \boldsymbol{\epsilon}_i.$$

Setting $\mathbf{y}_i^* = \mathbf{y}_i - f(\boldsymbol{\theta}_i^*) + J(\boldsymbol{\theta}_i^*)^T\boldsymbol{\theta}_i^*$ and substituting $\boldsymbol{\theta}_i = \boldsymbol{\mu} + \boldsymbol{\eta}_i$, we get the linearized mixed effects model similar to (5):

$$\mathbf{y}_i^* \approx J(\boldsymbol{\theta}_i^*)^T\boldsymbol{\mu} + J(\boldsymbol{\theta}_i^*)'Z\boldsymbol{\eta}_i + \boldsymbol{\epsilon}_i, \quad \text{for} \quad i = 1, ..., m$$

$$\boldsymbol{\eta}_i \sim N_q(\mathbf{0}, \Sigma) \quad \text{and} \quad \boldsymbol{\epsilon}_i \sim N_n(\mathbf{0}, \sigma^2\mathbf{I}_n).$$

All of the usual results from the linear case should roughly extend here. For example, when variances are known and $\boldsymbol{\theta}_i^*$ is some fixed estimate of the structural parameters at the present stage of the iteration, approximate estimates of the fixed effects and predictions of random effects are given by

$$\hat{\boldsymbol{\mu}} = [J(\boldsymbol{\theta}_i^*)V^{-1}J(\boldsymbol{\theta}_i^*)^T]^{-1}J(\boldsymbol{\theta}_i^*)V^{-1}\bar{\mathbf{y}}^*$$

and

$$\hat{\boldsymbol{\eta}}_i = \Sigma Z' J(\boldsymbol{\theta}_i^*)V^{-1}\left[\mathbf{y}_i^* - J(\boldsymbol{\theta}_i^*)^T\hat{\boldsymbol{\mu}}\right]$$

respectively. This leads to the following result.

Result B.1 *Let $\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_m$ denote a collection of m independent profiles modeled by a nonlinear mixed-effects model (6). If σ^2 and Σ are known and $J(\boldsymbol{\theta}_i^*) = \frac{df(\boldsymbol{\theta})}{d\theta}|_{\boldsymbol{\theta}_i=\boldsymbol{\theta}_i^*}$ for some known true value $\boldsymbol{\theta}_i^* = \boldsymbol{\mu}^* + \boldsymbol{\eta}_i^*$, sampling distributions of the structural parameters are approximated by*

- $\hat{\boldsymbol{\mu}} \dot{\sim} N_p\left(\boldsymbol{\mu}, (mJ(\boldsymbol{\theta}_i^*)V^{-1}J(\boldsymbol{\theta}_i^*)')^{-1}\right)$, *and*
- $\hat{\boldsymbol{\eta}}_i \dot{\sim} N_q\left(\mathbf{0}, \Sigma Z' J(\boldsymbol{\theta}_i^*)V^{-1}J(\boldsymbol{\theta}_i^*)'Z\Sigma\right)$ *independently for* $i = 1, ..., m$.

Since $J(\boldsymbol{\theta}_i^*)$ depends on the true (unknown) value of $\boldsymbol{\theta}_i$, an approximation $\hat{J} = \frac{1}{m}\sum_i J(\hat{\boldsymbol{\theta}}_i)$ can be used. Unknown variance components would also be replaced by their sample ML or REML estimates. While there is no hope of determining the exact distribution of $\hat{\boldsymbol{\eta}}_i$ after so many approximations, in Section 6 we present empirical evidence suggesting that (similarly to the linear case) result B.1 generalizes approximately in the presence of a large amount of data.

References

Abramovich, F. and Angelini, C. (2006), "Testing in Mixed-Effects FANOVA Models", *Journal of Statistical Planning and Inference*, 136, 4326-4348.

Antoniadis, A. and Sapatinas, T. (2004), "Estimation and Inference in Functional Mixed-Effects Models", Technical Report, TR/15/2004, Department of Mathematics and Statistics, University of Cyprus, Cyprus.

Demidenko, E. (2004), *Mixed Models: Theory and Applications*. New York: Wiley.

Fox, J., (2002), "Linear Mixed Models," *Appendix to An R and S-PLUS Companion to Applied Regression*.

Gallant, A. R. (1975), "Nonlinear Regression," *The American Statistician*, 29 (2), 73-81.

Gardner, M. M., Lu, J., Gyurcsik, R. S., Wortman, J. J., Hornung, B. E., Heinisch, H. H., Rying, E. A., Rao, S., Davis, J. C. and Mozumder, P. K. (1997), "Equipment Fault Detection Using Spatial Signatures," *IEEE Transactions on Components, Packaging, and Manufacturing Technology - Part C*, 20, 295-304.

Greenwood, M. C. (2004), "Functional Data Analysis for Glaciated Valley Profile Analysis", Ph.D. Thesis, Department of Statistics, University of Wyoming.

Guo, W. (2002), "Functional mixed effects models", *Biometrics*, 58, 121128

Harville, D. A. (1990), "BLUP (Best Linear Unbiased Prediction) and Beyond," *Advances in Statistical Methods for Genetic Improvement of Livestock*, 239-276, New York: Springer.

Hawkins, D. M. and Merriam, D. F. (1974), "Zonation of Multivariate Sequences of Digitized Geologic Data," *Mathematical Geology*, 6, 263-269.

James, G. and Sugar, C. (2003), "Clustering for Sparsely Sampled Functional Data," *Journal of the American Statistical Association (JASA)*, 98, 397-408.

Jeong, M. K., Lu, J. C., Huo, X., Vidakovic, B., and Chen, D. (2006), "Wavelet-based Data Reduction Techniques for Process Fault Detection," *Technometrics*, 48(1), 26-40.

Jin, J. and Shi, J. (2001), "Automatic feature extraction of waveform signals for in-process diagnostic performance improvement," *Journal of Intelligent Manufacturing*, 12, 257-268.

Johnson, R.A. and Wichern, D.W. (1992). *Applied Multivariate Statistical Analysis*. Prentice-Hall, Third Edition.

Jones, M. and Rice, J. (1996), "Displaying the Important Features of Large Collections of Similar Curves," *The American Statistician*, 46, 140-145.

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

Kim, K., Mohamoud, M. A., and Woodall, W. H. (2003), "On the Monitoring of Linear Profiles," *JQT*, 35, 317-328.

Koulis, T., Ramsay, J. O. and Levitin, D. (2006), "Input-Output Systems in Psychoacoustics," Submitted to *Psychometrika*.

Lada, E. K., Lu, J. and Wilson, J. R. (2002), "A Wavelet-Based Procedure for Process Fault Detection," *IEEE Transactions on Semiconductor Manufacturing* 15, 79-90.

Laird, N. M. and Ware, J. H. (1982), "Random Effects Model for Longitudinal Data," *Biometrics* 38 (4), 963-974.

Lange, N. and Ryan, L. (1989), "Assessing Normality in Random Effects Models," *The Annals of Statistics* 17 (2), 624-642.

Lindstrom, M. J. and Bates, D. M. (1988), "Newton-Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measures Data," *JASA*, 83 (404), 1014-1022.

Lindstrom, M. J. and Bates, D. M. (1990), "Nonlinear Mixed Effects Models for Repeated Measures Data," *Biometrics* 46 (3), 673-687.

Lowry, C.A., W.H. Woodall, C.W. Champ and S.E. Rigdon (1992), "A multivariate exponentially weighted moving average control chart", *Technometrics*, 34, 46-53.

Mahmoud, M. A., and Woodall, W. H. (2004), "Phase I Monitoring of Linear Profiles with Calibration Applications," Submitted to *Technometrics*, 46 (4), 380391.

Morris, J. S., Arroyo, C., Coull, B. A., Ryan, L.M., Herrick, R., Gortmaker, S. L. (2006), "Using wavelet-based functional mixed models to characterize population heterogeneity in accelerometer profiles: a case study", To appear in *JASA*.

Morris, J. S., Carroll, R. (2006) "Wavelet-based functional mixed models", *Journal of the Royal Statistical Society: Series B*, 68(2), 179-199.

Mosesova, S. A., *Mining Functional Data in High Dimensions*. Master's research paper, University of Waterloo, May 2002.

Ogunnaike, B. A. and Ray, W. H. (1994). *Process Dynamics, Modeling, and Control*. New York: Oxford University Press.

Pinheiro, J. and Bates, D.M. (2001) *Mixed Effects Models in S and S-PLUS*. New York: Springer-Verlag.

R Development Core Team (2006). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Ramsay, J. O. , Hooker, G, Cao, J., and Campbell D. (2006), "Estimating Differential Equations", Submitted to *Journal of the Royal Statistical Society, Series B*.

Ramsay, J. O. and Cao, J. (2006), "Smoothing, Nuisance Parameters and Profiling in Functional Data Analysis," Submitted to *Computational Statistics and Data Analysis*.

Ramsay, J. O., Silverman, B. W. (2005), *Functional Data Analysis*. New York: Springer.

Ramsay, J. O. (2000), "Differential equation models for statistical functions," *Canadian Journal of Statistics*, 28, 225-240.

Rice, J. A. and Wu, C. O., (March 2001), "Nonparametric Mixed Effects Models for Unequally Sampled Noisy Curves," *Biometrics*, 57, 253-259.

Rousseeuw, P. J. (1984), "Least Median of Squares Regression," *JASA*, 79, 871-880

Ryan, T. P. (1989), *Statistical Methods For Quality Improvement*. New York: Wiley.

Walker, E. and Wright, W. P., (2002), "Comparing Curves with Additive Models," *JQT*, 34 (1), 118-129.

Williams, J.D., Woodall, W. H., and Birch, J.B. (2003), "Phase I Analysis of Nonlinear Product and Process Profiles," Technical Report No. 03-5, Department of Statistics, Virginia Polytechnic Institute and State University.

Wolfinger, R., (1993), "Laplace's Approximations to Nonlinear Mixed Models," *Biometrika*, 80 (4), 791-795.