Profile Monitoring via Nonlinear Mixed Models

WILLIS A. JENSEN and JEFFREY B. BIRCH

Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0439

Profile monitoring is a relatively new technique in quality control best used where the process data follows a profile (or curve) at each time period. Little work has been done on the monitoring on nonlinear profiles. Previous work has assumed that the measurements within a profile are uncorrelated. To relax this restriction we propose the use of nonlinear mixed models to monitor the nonlinear profiles in order to account for the correlation structure. We evaluate the effectiveness of fitting separate nonlinear regression models to each profile in Phase I control chart applications for data with uncorrelated errors and no random effects. For data with random effects, we compare the effectiveness of charts based on a separate nonlinear regression approach versus those based on a nonlinear mixed model approach. Our proposed approach uses the separate nonlinear regression model fits to obtain a nonlinear mixed model fit. The nonlinear mixed model approach results in charts with good abilities to detect changes in Phase I data and has a simple to calculate control limit.

KEY WORDS: Blups, Multivariate Statistical Process Control, Nonlinear Model, Phase I, T^2 statistic

Introduction

It is crucial in Phase I of a control chart scheme to determine which of the data points are similar to each other and which ones are outlying in some way. This ensures that the Phase II application will be adequate for real time monitoring. Due to advances in technology, it is becoming much more common to obtain profiles (a series of data points forming a curve) at each time period that represents the quality state of a process. As such, recent research has focused on how to determine which profiles are outlying in Phase I applications. A good introduction to the concept of profile monitoring, and examples of its application can be found in Woodall et al. (2004).

The majority of work in profile monitoring has focused on situations where the profiles are linear. For example, see the work of Kang and Albin (2000), Kim, Mahmoud, and Woodall (2003), Mahmoud and Woodall (2004), Wang and Tsung (2005), and Jensen, Birch, and Woodall (2006b). These methods often fit separate linear regression models and monitor the coefficients of the fitted regression model to determine outlying profiles. Thus the profiles have been reduced to a smaller set of values that simplifies the monitoring scheme. However, it is often the case that the profiles are better described by a nonlinear function than by a linear function. Nonetheless, the majority of existing profile monitoring research deals with linear profiles. Williams, Woodall, and Birch (2003) gave a broad treatment of nonlinear profile monitoring where separate nonlinear (NL) regression models are fit to each profile. Williams et al. (2006a) gave an application of NL profile monitoring to dose-response data.

The previous work on nonlinear profile monitoring has assumed that the measurements within a profile are independent of each other. This is often an unrealistic assumption in practice for many types of data. Therefore, we propose the use of a nonlinear mixed (NLM) model to monitor the profiles in order to account for the correlation structure within profiles.

After discussing the formulations of the NL and NLM models, we show via simulation situations where the NLM model approach is superior to the NL approach in detecting changes in the Phase I dataset. We propose a method that supplements the approach of Williams, Woodall, and Birch (2003) with a NLM model to improve the control chart procedure. We demonstrate our proposed method by applying it to the particle board data of Walker and Wright (2002).

NL Model Formulation

We assume that we have m profiles of data, each of which has n_i measurements where i refers to the ith profile. We can then fit a separate NL model to each profile. Let y_{ij} refer to the jth measurement for the ith profile. The model for each of the separate nonlinear regressions is given by

$$y_{ij} = f(x_{ij}, \boldsymbol{\theta_i}) + \epsilon_{ij} \text{ for } i = 1, 2, \ldots, m, j = 1, 2, \ldots, n_i, \tag{1}$$

where $f(.)$ is some nonlinear function, x_{ij} is a regressor variable for the measurement, $\boldsymbol{\theta_i}$ is a $px1$ vector of parameters for each profile, and ϵ_{ij} is the error associated with the jth measurement of the ith profile. The errors are often assumed to be independent and normally distributed, that is $\epsilon_{ij} \sim N(0, \sigma_i^2)$, implying that the measurements within a profile are uncorrelated. We will assume throughout that the profiles have the same variability in the error term, thus $\sigma_i^2 = \sigma^2$ for $i = 1, 2, \ldots, m$.

If the responses for the ith profile are stacked to form a vector, $\mathbf{y}_i$, then we have the alternative form of the model from (1) given by

$$\mathbf{y}_i = f(\mathbf{x}_i, \boldsymbol{\theta_i}) + \boldsymbol{\epsilon}_i \text{ for } i = 1, 2, \ldots, m, \tag{2}$$

where $\mathbf{x}_i$ is a vector of the values of the predictor variable and $\boldsymbol{\epsilon}_i$ is a vector of errors that has a multivariate normal distribution, $\boldsymbol{\epsilon}_i \sim MN(\mathbf{0}, \mathbf{R}_i)$ with $\mathbf{R}_i$ being a n by n positive definite variance-covariance matrix. If the errors are correlated, $\mathbf{R}_i$ is often assumed to be a simple form such as compound symmetry (CS) or autoregressive (AR) in order to reduce the number of covariance parameters that need to be estimated. For more details on the various types of correlated error structures that can be assumed for $\mathbf{R}_i$, see Littell et al. (1996) or Schabenberger and Pierce (2002). If the measurements within a profile are uncorrelated then $\boldsymbol{\epsilon}_i \sim MN(0, \sigma_i^2 \mathbf{I})$.
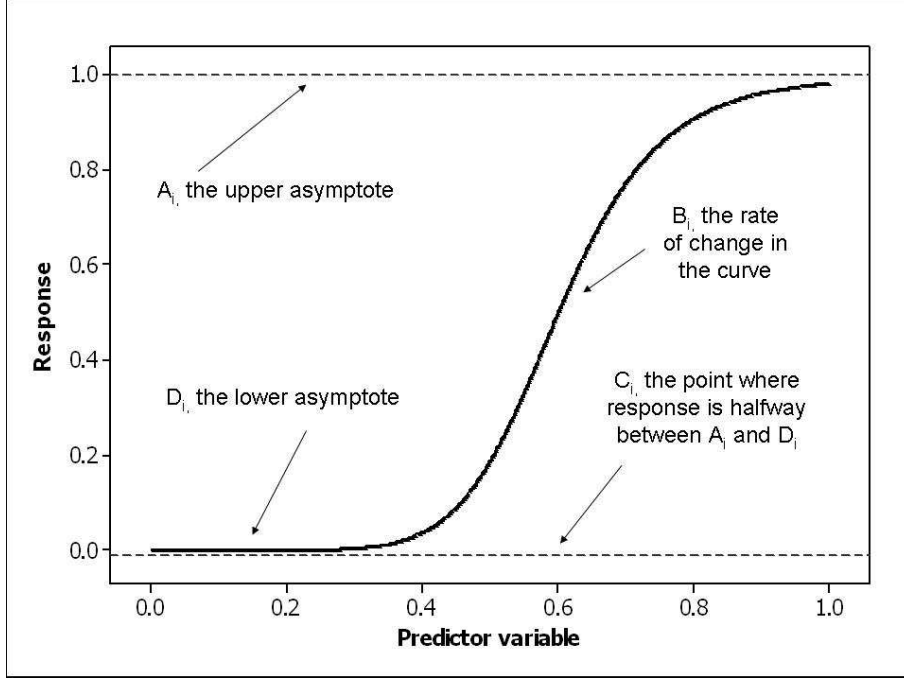
There is a wide variety of nonlinear functions that have been used for applications. A book covering nonlinear models (Gallant, 1987; Ratkowsky, 1990; Schabenberger and Pierce, 2002; Seber and Wild, 2003) will discuss many of the different types of functions that have been utilized. For example, consider the 4-parameter logistic model, which has been used frequently for dose-response studies. This model is given by

$$y_{ij} = A_i + \frac{D_i - A_i}{1 + \left(\frac{x_{ij}}{C_i}\right)^{B_i}} + \epsilon_{ij} \text{ for } i = 1, 2, \ldots, m, j = 1, 2, \ldots, n_i, \tag{3}$$

where A_i is the upper asymptote, D_i is the lower asymptote, C_i is the point where the curve reaches halfway between A_i and D_i, and B_i is a parameter representing the rate of increase or decrease from D_i to A_i. The larger the value of B_i, the steeper will be the slope of the curve. See Figure 1 for an example of the 4-parameter logistic curve where $A_i = 1$, $B_i = 8$, $C_i = .6$, and $D_i = 0$.

When the parameters in (2) are all fixed, then the only variability in the responses will be due to the $\boldsymbol{\epsilon}_i$ term. To illustrate the impact of the variability of the errors on profiles, consider Figure 2, which shows randomly generated 4-parameter logistic curves with uncorrelated errors with different values of σ^2 for all generated profiles. Here $m = 30$, $n = 10$, and the values of A_i, B_i, C_i, and D_i are the same as those in Figure 1. We see that the larger the

3

Figure 1: Example of a 4-parameter logistic curve where $A_i = 1$, $B_i = 8$, $C_i = .6$, and $D_i = 0$.

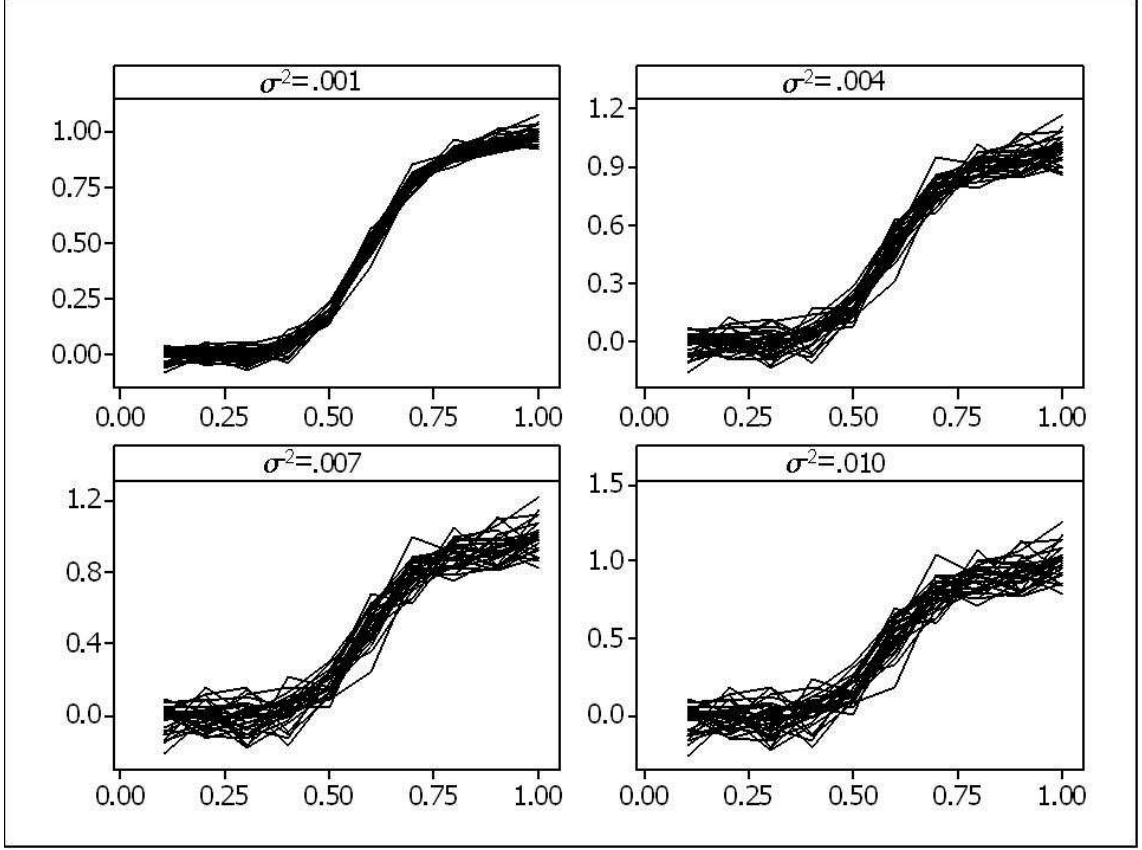variability of the errors the greater the differences between the profiles.

NL Model Estimation

For the NL model in (1) and (2), the maximum likelihood estimator (MLE) of $\boldsymbol{\theta_i}$, $\widehat{\boldsymbol{\theta}}_{\boldsymbol{i}}$, is that estimator that will minimize the residual sum of squares given by

$$
\begin{aligned}
SS(\boldsymbol{\theta_i}) &= \sum_{j=1}^{n_i} [y_{ij} - f(x_{ij}, \boldsymbol{\theta_i})]^2 \text{ for } i = 1, 2, \ldots, m \\
&= [\mathbf{y}_i - f(\mathbf{x}_i, \boldsymbol{\theta_i})]' [\mathbf{y}_i - f(\mathbf{x}_i, \boldsymbol{\theta_i})] \text{ for } i = 1, 2, \ldots, m.
\end{aligned}
\tag{4}
$$

Because of the nonlinearity introduced in (4) by $f(.)$ there is not a closed form expression for the estimator that will minimize $SS(\boldsymbol{\theta_i})$. There are two major iterative algorithms used for obtaining the parameter estimates for this NL model.

4

Figure 2: Illustration of the difference in the generated profiles due to differences in the variability of the errors. For these profiles, $m = 30$, $n = 10$, and the values of A_i, B_i, C_i, and D_i are the same as those in Figure 1.

The first, the Gauss-Newton (GN) algorithm, replaces $f(\mathbf{x}_i, \boldsymbol{\theta_i})$ in (4) with a Taylor series approximation about $\boldsymbol{\theta_i}$ so that the minimization can then proceed in an iterative fashion. We denote $\widehat{\boldsymbol{\theta_i}}^0$ as the vector of initial starting values for the iterative algorithm, and denoting the matrix of derivatives of the NL function evaluated at the initial starting values as

$$\widehat{\mathbf{F}}_i = \frac{\partial f(\mathbf{x}_i, \boldsymbol{\theta_i})}{\partial \boldsymbol{\theta_i}}\big|_{\boldsymbol{\theta_i} = \widehat{\boldsymbol{\theta_i}}^0}. \tag{5}$$

The GN algorithm computes

$$\widehat{\boldsymbol{\theta_i}}^1 = \widehat{\boldsymbol{\theta_i}}^0 + \left(\widehat{\mathbf{F}}_i'\widehat{\mathbf{F}}_i^{-1}\right)\widehat{\mathbf{F}}_i'\left[\mathbf{y}_i - f(\mathbf{x}_i, \widehat{\boldsymbol{\theta_i}}^0)\right]. \tag{6}$$

After each iteration $\widehat{\boldsymbol{\theta}_i}^0$ is replaced by $\widehat{\boldsymbol{\theta}_i}^1$ until $\widehat{\boldsymbol{\theta}_i}^0 \approx \widehat{\boldsymbol{\theta}_i}^1$. Thus the algorithm stops when the change in estimated parameters from one iterate to the next is sufficiently small. In this case the algorithm has converged and $\widehat{\boldsymbol{\theta}_i} = \widehat{\boldsymbol{\theta}_i}^1$. Adjustments to this basic algorithm are often implemented in practice to avoid numerical issues and ensure convergence (Schabenberger and Pierce, 2002, Chapter 5; Seber and Wild, 2003, Chapter 14).

The second, the Newton-Raphson (NR) algorithm, replaces the entire function in (4) by a Taylor series approximation and similar to the GN method, iteratively solves for the parameter estimates. For simple NL models, there will often not be much difference in the estimates obtained by either of the two algorithms. Thus the default method in $SAS^{\circledR}$, the GN algorithm, will be our choice for what follows.

With the normality assumption in the NL model, we can state the distributional properties of the vector of obtained parameter estimators, $\widehat{\boldsymbol{\theta}_i}$. In contrast to the standard linear model, normality of the parameter estimator will only hold asymptotically, that is,

$$\widehat{\boldsymbol{\theta}_i} \overset{A}{\sim} MN \left[\boldsymbol{\theta}_i, \sigma_i^2(\mathbf{F}_i'\mathbf{F}_i)^{-1}\right] \text{ for } i = 1, 2, \ldots, m. \tag{7}$$

The result in (7) holds whether or no the error are uncorrelated (Seber and Wild, 2003, Chapter 12.2).

NLM Model Formulation

In the NLM model we extend the NL model in (2) to allow for any random effects. In vector form it is given by

$$\mathbf{y}_i = f(\mathbf{x}_i, \boldsymbol{\theta}, \mathbf{b}_i) + \boldsymbol{\epsilon}_i \text{ for } i = 1, 2, \ldots, m, \tag{8}$$

where $\boldsymbol{\theta}$ is a vector of fixed effects common to all profiles, $\mathbf{b}_i$ is the r by 1 vector of random effects with $\mathbf{b}_i \sim MN(\mathbf{0}, \mathbf{D})$. $\mathbf{D}$ is assumed to be a diagonal matrix with the diagonal

6

elements referred to as the variance components. A good introduction and review of the literature on the NLM model can be found in Davidian and Giltinan (2003) or books by Pinheiro and Bates (2000), Schabenberger and Pierce (2002), and Demidenko (2004).

Demidenko (2004, Chapter 6.1) proposed a restriction of (8) by forcing the random effects to enter the model linearly so that we can rewrite (8) as

$$\mathbf{y}_i = f(\mathbf{x}_i, \boldsymbol{\theta}) + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i \text{ for } i = 1, 2, \ldots, m. \tag{9}$$

where $\mathbf{Z}_i$ is a n_i by r matrix of values corresponding to the random effects. While this may simplify the computational effort required to obtain estimates, we will not pursue it here because it is less flexible.

The model in (8) allows for two levels of correlation for the measurements within a profile. The first results from the random effects which cause all the measurements within a profile to be correlated to each other. The second results from the within-profile variance-covariance matrix of the errors, $\mathbf{R}_i$. A NLM model that uses neither of the two levels of correlation is simply the NL model in (2) with uncorrelated errors because $\mathbf{Z}_i = \mathbf{0}$ and $\boldsymbol{\epsilon}_i \sim MN(\mathbf{0}, \sigma_i^2\mathbf{I})$. See Davidian and Giltinan (2003, pp. 395-400) for more discussion of the within profile correlation and its interpretation in the NLM model.

NLM Model Estimation

As noted by Schabenberger and Pierce (2002), given the random effects in the NLM model, one can write the marginal density of $\mathbf{y}_i$ as

$$g(\mathbf{y}_i) = \int g(\mathbf{y}_i|\mathbf{b}_i) \, g(\mathbf{b}_i)d\mathbf{b}_i. \tag{10}$$

where $g(.)$ is the assumed probability density function. The evaluation of this integral is required for inference but the distribution $g(\mathbf{y}_i)$ is not known even when the errors and

7

random effects have a multivariate normal distribution. As a result, numerical methods are needed. The most common methods are the linearization and the integral approximation methods.

Earlier work on estimation of NLM models focused on the linearization approach which is an approximation of the nonlinear function in (8) by some linear function. This replacement results in a multivariate normal density function for which an estimator can be obtained. On the other hand, the integral approximation approach has increased in popularity in recent years due to advances in computing power and existence of Monte Carlo (MC) based methods for directly evaluating the integral in (10). The nlmixed procedure of $SAS^{\circledR}$ uses the integral approximation approach with adaptive MC procedures that are more computationally efficient than standard MC procedures. Thus our analysis of the NLM model will be based on the integral approximation method.

Diagnostics in NL and NLM models

Davidian and Giltinan (1995, p. 328) noted that diagnostic methods for NL and NLM models are underdeveloped and nearly non-existent but would be very useful. While there are computational difficulties in dealing with the nonlinearity of the models, they recommended more research in the area of diagnostics and noted that new research will increase the utilization of NL and NLM models. We review here some of the methods that have appeared but believe the area to still be underdeveloped.

Pinheiro and Bates (2000) proposed to check assumptions of the NLM model by using the classical regression diagnostic plots such as the normal probability plot of errors and a scatter plot of residuals versus the fitted values. They do not give justification for why such plots would work other than that they have been used for classical regression problems.

Demidenko (2004, Chapter 9.5) discussed some methods to determine the influence of unusual data points on NL regression models. Lee and Xu (2004) considered diagnostic methods for the NLM model based on the case deletion and local influence approach. However, they had to use MC methods in order to compute the diagnostics and this would seem to limit their applicability.

Our approach, which combines the separate NL regressions with a NLM model, differs from these previous approaches appearing in the literature. Prior to explaining the approach, we first wish to compare approaches based on the NL and NLM models, which we do in the next section.

T^2 Statistic for NL and NLM Models

For most control chart applications, where the profiles occur at regular time periods, the data collection is well controlled as if from a designed experiment. Thus the number of measurements per profile will often be the same and at the same locations along the profile. Once the profiles are obtained, they can be fit with separate NL models or with a NLM model. The NLM model has the advantage of pooling information from the profiles together and allows us to model the random effects. When utilizing the NL approach, that of fitting separate NL regression models to each profile, we have reduced the profiles to a series of time-ordered vectors, $\widehat{\boldsymbol{\theta_i}}$. For the NLM model we have reduced the profiles to the estimated fixed effects vector, $\widehat{\boldsymbol{\theta}}$, and the vectors containing the estimated random deviations from the fixed effects vector, $\widehat{\mathbf{b}}_i$.

Once we have obtained the estimates we use the T^2 statistic to determine outlying profiles. The T^2 statistic is widely used in multivariate quality control and a comprehensive review of its properties and alternative forms can be found in Mason and Young (2002). A T^2 statistic

based on the sample mean vector and sample variance-covariance matrix is widely used but is not very effective in detecting anything more than a single moderately sized outlying profile (Vargas, 2003). Its distribution is proportional to a beta distribution (Mason and Young, 2002). The alternative is to base the T^2 statistic on the sample mean and the variance-covariance matrix of the successive differences between vectors (Holmes and Mergen, 1993). Sullivan and Woodall (1996) showed that using successive differences is effective in detecting sustained step changes in the process that occur in Phase I data. While the distribution of the T^2 statistic based on successive difference does not have a simple closed form, its asymptotic distribution is χ_p^2. A discussion of the various approximate distributions and the preferred χ_p^2 approximation for large samples is given in Williams et al. (2006b). The sample sizes that we use here are large enough to justify use of the χ_p^2 approximation to obtain the control limit.

For the NL approach we use the $\widehat{\boldsymbol{\theta_i}}$ vectors to calculate the T^2 statistics. The T^2 statistic for the separate NL regression models based on sample mean and variance-covariance is denoted by $T_{1,i,NL}^2$ and the T^2 statistic based on successive differences by $T_{2,i,NL}^2$. They are given by

$$
\begin{aligned}
T_{1,i,NL}^2 &= (\widehat{\boldsymbol{\theta_i}} - \overline{\boldsymbol{\theta_i}})' \mathbf{S}_{1,NL}^{-1} (\widehat{\boldsymbol{\theta_i}} - \overline{\boldsymbol{\theta_i}}) \\
&= (\widehat{\boldsymbol{\theta_i}} - \overline{\boldsymbol{\theta_i}})' \left[\frac{\sum_{i=1}^m (\widehat{\boldsymbol{\theta_i}} - \overline{\boldsymbol{\theta_i}})(\widehat{\boldsymbol{\theta_i}} - \overline{\boldsymbol{\theta_i}})'}{m-1} \right]^{-1} (\widehat{\boldsymbol{\theta_i}} - \overline{\boldsymbol{\theta_i}}) \text{ for } i = 1, 2, \ldots, m, \quad (11)
\end{aligned}
$$

where

$$
\overline{\boldsymbol{\theta_i}} = \frac{\sum_{i=1}^m \widehat{\boldsymbol{\theta_i}}}{m}, \tag{12}
$$

and by

$$
\begin{aligned}
T_{2,i,NL}^2 &= (\widehat{\boldsymbol{\theta_i}} - \overline{\boldsymbol{\theta_i}})' \mathbf{S}_{2,NL}^{-1} (\widehat{\boldsymbol{\theta_i}} - \overline{\boldsymbol{\theta_i}}) \\
&= (\widehat{\boldsymbol{\theta_i}} - \overline{\boldsymbol{\theta_i}})' \left[\frac{\sum_{i=1}^{m-1} (\widehat{\boldsymbol{\theta_{i+1}}} - \widehat{\boldsymbol{\theta_i}})(\widehat{\boldsymbol{\theta_{i+1}}} - \widehat{\boldsymbol{\theta_i}})'}{m-1} \right]^{-1} (\widehat{\boldsymbol{\theta_i}} - \overline{\boldsymbol{\theta_i}}) \text{ for } i = 1, 2, \ldots, m. \quad (13)
\end{aligned}
$$

10

Because the fixed effects vector, $\widehat{\boldsymbol{\theta}}$, is the same for all profiles in the NLM model, the T^2 statistic will only depend on the predicted random effects. They will be denoted by $T_{1,i,NLMM}^2$ and $T_{2,i,NLMM}^2$ and are respectively,

$$T_{1,i,NLMM}^2 = (\widehat{\mathbf{b}}_i - \overline{\mathbf{b}})' \left[\frac{\sum_{i=1}^{m}(\widehat{\mathbf{b}}_i - \overline{\mathbf{b}})'(\widehat{\mathbf{b}}_i - \overline{\mathbf{b}})}{m-1} \right]^{-1} (\widehat{\mathbf{b}}_i - \overline{\mathbf{b}}) \text{ for } i = 1, 2, \ldots, m, \qquad (14)$$

and

$$T_{2,i,NLMM}^2 = (\widehat{\mathbf{b}}_i - \overline{\mathbf{b}})' \left[\frac{\sum_{i=1}^{m-1}(\widehat{\mathbf{b}}_{i+1} - \widehat{\mathbf{b}}_i)'(\widehat{\mathbf{b}}_{i+1} - \widehat{\mathbf{b}}_i)}{2(m-1)} \right]^{-1} (\widehat{\mathbf{b}}_i - \overline{\mathbf{b}}) \text{ for } i = 1, 2, \ldots, m, \quad (15)$$

where

$$\overline{\mathbf{b}} = \frac{\sum_{i=1}^{m} \widehat{\mathbf{b}}_i}{m}. \tag{16}$$

Simulation Study Setup

We now explain the general procedure for the simulation studies used to compare the NL and NLM methods. To generate the multivariate normal errors and random effects we first generate univariate normal data and use the Cholesky decomposition to transform the generated univariate data to multivariate data. The multivariate data is then added appropriately to the nonlinear function to get the generated values of the response variable. The data are fit with separate NL regression models or a NLM model using the nlin and nlmixed procedures of $SAS^{\circledR}$ with the correct model specification.
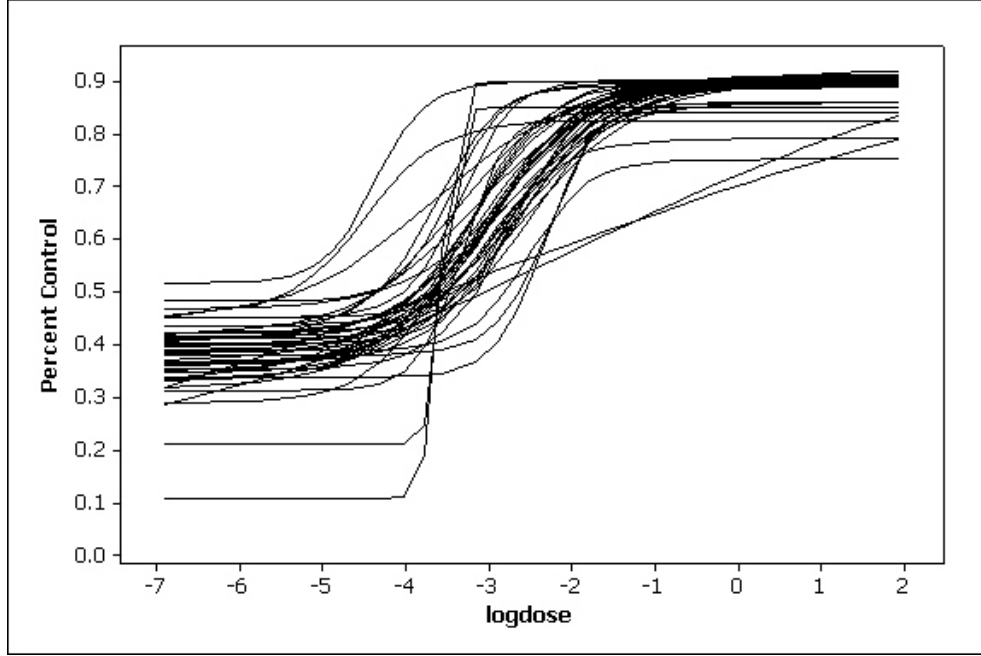
The control limit is established using the appropriate percentiles of the beta or χ^2 distributions so that the probability of signal for the in-control data is .05, the nominal value. The actual probability of signal is estimated by the proportion of datasets where there was a signal. That is, a signal occurs when at least one of the T^2 statistics exceeds the control limit.

When obtaining the estimates in a NL or NLM model, non-convergence of the iterative algorithm can occur. In some cases the frequency of non-convergence can be problematic. We found that in general, the more variability there is in the simulated data, either due to increased variability in the errors or larger variance components of the random effects distribution, the more frequent the non-convergence. To reduce the frequency of non-convergence, it is often recommended to use good starting values for the fixed parameters and components of the variance-covariance matrix. These starting values can be obtained via graphical methods (Schabenberger and Pierce, 2002). In some situations where the non-convergence was more likely to be present, we used in our simulations the known parameter values used to generate the data as starting values of the iterative algorithm as was done by Hartford and Davidian (2000). This reduces the frequency of non-convergence just as would occur if a knowledgeable researcher were to spend a sufficient amount of time exploring, cleaning, and appropriately analyzing a single dataset.

We note that the regression equivariance property discussed in Rousseuw and Leroy (1987) does not hold for the NL and NLM models. The unfortunate consequence is the difficulty in obtaining broad conclusions from a smaller set of simulation studies because the obtained results will depend on the type of nonlinear function, its particular form, and the set values of the parameters. In order to investigate via simulation the differences obtained by using the NL versus the NLM approach, we picked a nonlinear function related to a real data situation to ensure that our results will hold when analyzing the corresponding dataset. We believe that the conclusions obtained here will hold for other types of functions but it would be very difficult to make a general conclusion to all functions.

We considered the dose-response data described in Williams et al. (2006a) which can be modeled by the 4-parameter logistic curve mentioned in (3). The fitted data curves are

Figure 3: Fitted curves for dose response data of Williams et al. (2006a).

shown in Figure 3. Note that the values for the dose were not equally spaced but the log of the values of the dose were equally spaced. So the curves in Figure 3 show equally spaced data because the horizontal axis is the log of the dose. In considering these dose-response data curves, exploratory data analysis shows that after eliminating several of the profiles due to lack of model fit, the median values of the 4 parameters are very close to $A_i = .9$, $B_i = 2$, $C_i = .05$, and $D_i = .4$.

We use this function and its parameter values as typical for this particular application. We also considered the four parameter logistic curve with two other sets of parameter values representing functions more extreme than the median function which we denote the maximum curve and the minimum curve. The maximum curve has parameter values of $A_i = 1$, $B_i = 4$, $C_i = .05$, and $D_i = 0$ and has a steeper rate of change than the median curve with the asymptotes further apart from each other. The minimum curve has parameter values of

$A_i = .75$, $B_i = 1$, $C_i = .05$, and $D_i = .5$ and has a more gradual rate of change with the asymptotes closer together.
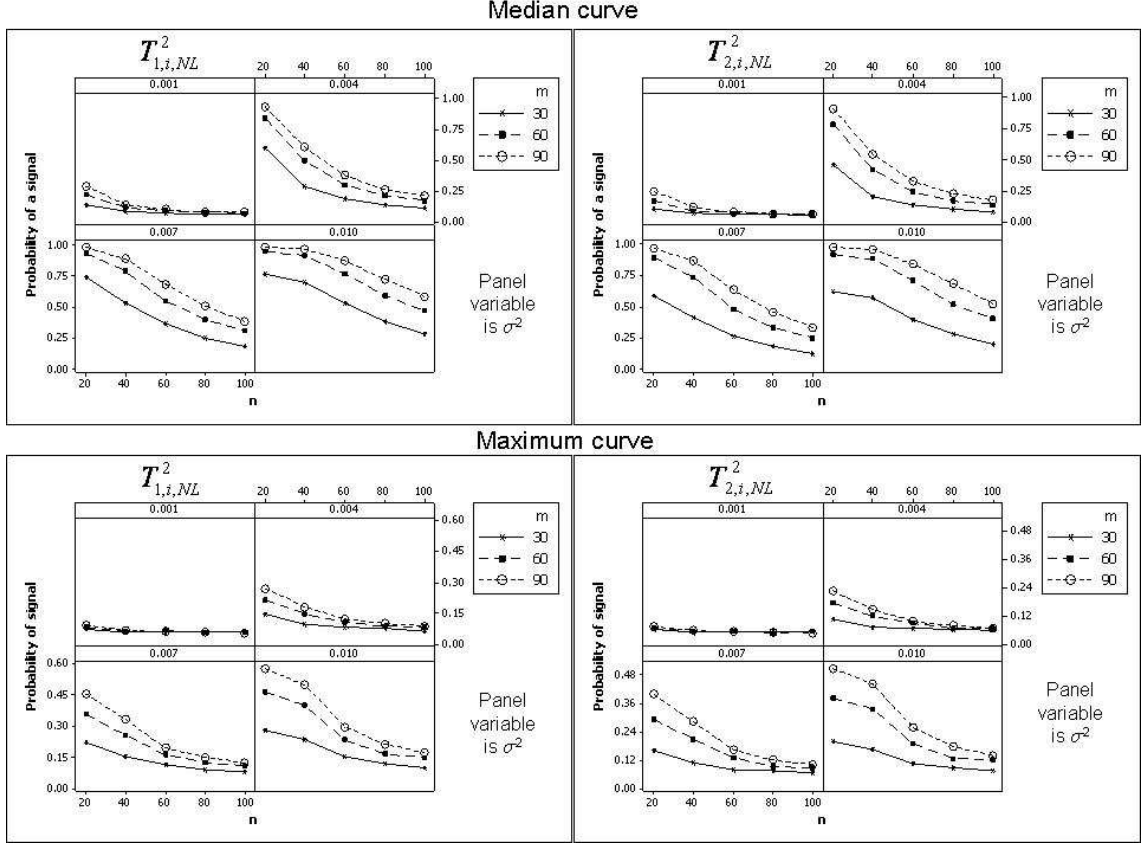
Uncorrelated Data with no Random Effects

We first investigated the probability of signal of the method of fitting separate NL regression models as proposed by Williams, Woodall, and Birch (2003). We considered the situation where there are no random effects and the errors are independent. For randomly generated in-control data, we want to determine if the control limit based on beta or χ_p^2 distributions is appropriate. Here the data are balanced and equally spaced. $10,000$ datasets were generated for each run of the simulation studies.

Figure 4 shows the probability of signal for in-control data generated from both the 4-parameter logistic median and maximum curves. The horizontal axis is the number of observations per profile, n, and the vertical axis is the probability of signal for various values of m and σ^2 for both $T_{1,i,NL}^2$ and $T_{2,i,NL}^2$. We see that while $T_{1,i,NL}^2$ and $T_{2,i,NL}^2$ have similar performance, the probability of signal can be much larger than the desired .05 level, particularly as σ^2 increases. When there is little variability in the errors the nominal probability of signal will be maintained.

The probability of a signal decreases as n increases. Thus, if there are more observations per profile, the parameter estimators are more stable and the control limit based on the beta or χ_p^2 distributions is appropriate. This is to be expected because the NL model parameter estimators are only asymptotically (in n) normal. Normality is required for the T^2 statistics in (11) and (13) to have a beta or χ_p^2 distribution, respectively. Thus when using a smaller number of observations per profile it is not appropriate to use a control limit that requires the assumption of normality.
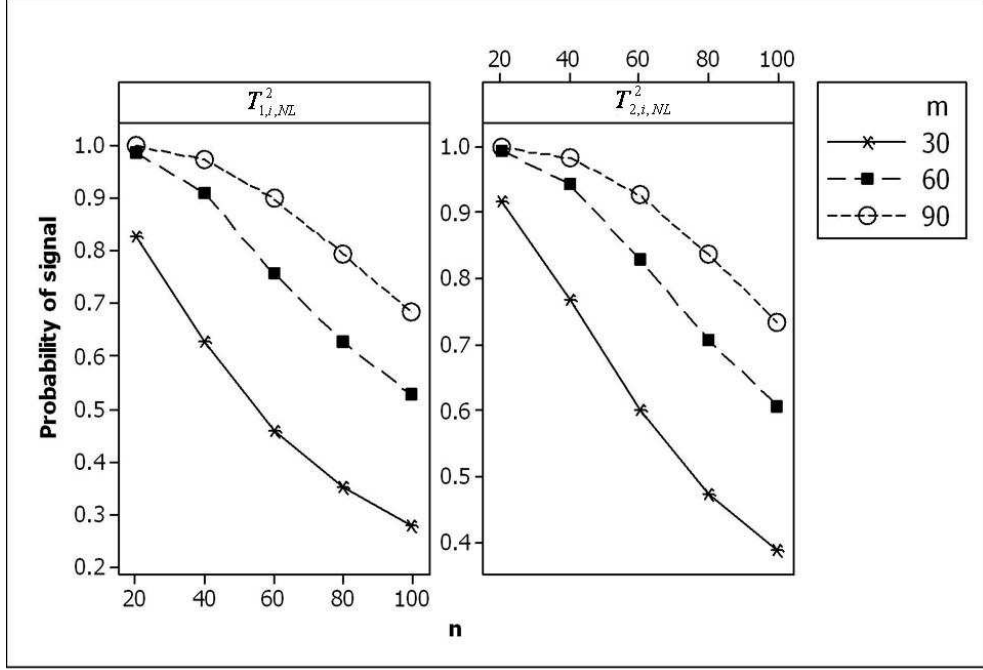
14

Figure 4: Probability of signal for two T^2 statistics for simulated in-control data following the median and maximum curve for various values of m, n, and σ^2.

On the other hand, the probability of signal increases when m increases suggesting that when there are more profiles present, it is more likely that at least one of them will be declared an outlying profile than when there are not many profiles present.

Figure 5 shows the probability of signal for in-control data generated from the minimum curve. The probability of signal was calculated only for a single, smaller value of $\sigma^2 = .001$, because larger amounts of variability of the errors made the probability of signal virtually 1, and the programming code had difficulty handling larger amounts of variability without convergence problems.

Figure 5: Probability of signal of two T^2 statistic for simulated in-control data following the minimum curve for various values of m and n, with $\sigma^2 = .001$.

In conclusion, for this particular 4-parameter logistic model, the control limit based on beta or χ^2_p distributions will not be appropriate for any of the three types considered unless n is sufficiently large and σ^2 is sufficiently small. Results not shown here for other values of A_i, B_i, C_i, and D_i concur with our conclusions here even though it should be noted that because regression equivariance does not hold, there are some situations where the control limit will be sufficiently accurate. Otherwise, the control limit will have to be simulated for practical applications.

It would be possible to consider other data scenarios such as balanced, unequally spaced data or unbalanced data but we do not pursue it here in light of the results of Jensen, Birch, and Woodall (2006b) for linear profiles. If the control limit is inadequate for balanced, equally

spaced data then they will be even more inadequate for data that are unequally spaced, or unbalanced. In addition, because the control limit is inadequate for complete datasets, then they will be even more inadequate for profile data that have missing observations within the profile.

Uncorrelated Data with 1 Random Effect

Because of the inherent difficulty in modeling multiple random effects as mentioned previously, we next considered data where a single random effect is present. We wish to compare the NL approach with a NLM approach when analyzing the 4-parameter logistic model. To determine which parameter to set at random we analyzed the dose-response data of Williams et al. (2006a) shown earlier in Figure 3. After initial data cleaning and test of lack of fit, there remained 32 profiles for analysis. We calculated the parameter estimates for separate NL regression models for each profile. The mean and variance of the parameter estimates for the 32 profiles are shown in Table 1.

Table 1: Mean and variance of the parameter estimates obtained from separate NL models for the dose-response data of Williams et al. (2006a).

Parameter	Mean	Variance
A_i	0.8984	0.0002
B_i	1.9353	0.2917
C_i	0.0543	0.0003
D_i	0.3919	0.0022

It is clear that B_i has the largest amount of variability among the profiles and thus it is the best candidate to be modeled as a random effect in a NLM model. We rewrite (3) as

$$y_{ij} = A_i + \frac{D_i - A_i}{1 + \left(\frac{x_{ij}}{C_i}\right)^{B+b_i}} + \epsilon_{ij} \text{ for } i = 1, 2, \ldots, m, j = 1, 2, \ldots, n_i, \tag{17}$$

17

where b_i is the random effect that represents how much the slope parameter of the ith profile differs from the overall slope parameter, B.

To determine the values of σ_B^2 that we will use when generating the simulated data, we chose to use values similar in magnitude to the estimated variance of the random effect from Table 1. The values for σ_B^2 that we considered were all between .1 and .5 We generated in-control data that followed the median 4-parameter logistic curve with a random effect in B_i and uncorrelated errors. In one study, we set $\sigma^2 = .001$ and $\sigma_B^2 = .5$ and performed $1,000$ simulation runs. Table 2 shows the probability of signal when using the approximate control limit for $m = 30$ and various values of n. A smaller number of runs was performed here when modeling the random effect than when there was no random effect because of the larger computational burden required to obtain estimates for the NLM approach.

Table 2: Probability of signal for the four T^2 statistics for simulated data with a random effect in B_i, n ranging from 10 to 500, $m = 30$, $\sigma^2 = .001$, and $\sigma_B^2 = .5$.

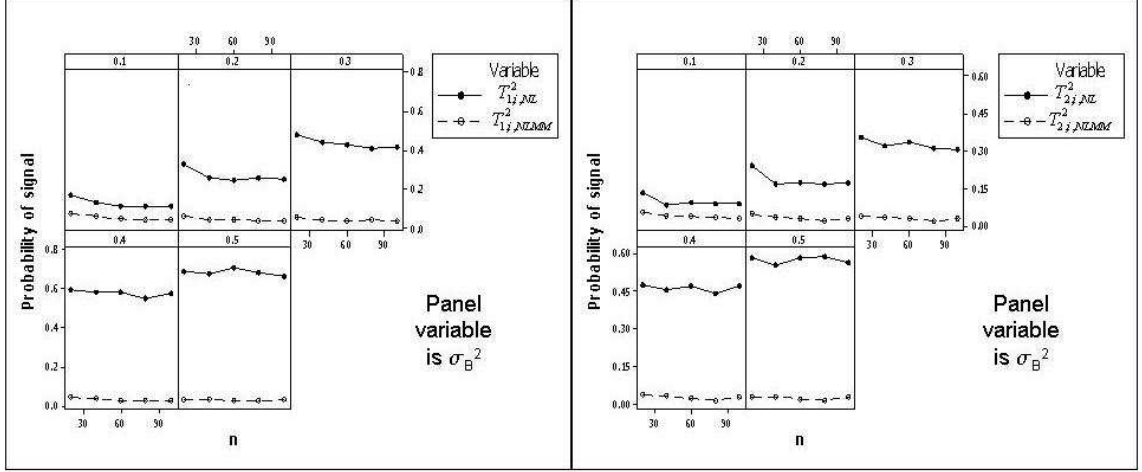| n	$T^2_{1,i,LS}$	$T^2_{2,i,LS}$	$T^2_{1,i,MIX}$	$T^2_{2,i,MIX}$	Non-convergence in NL	Non-convergence in NLM
10	0.740	0.606	0.070	0.043	0.0216	0.1240
20	0.692	0.586	0.040	0.031	0.0072	0.1245
40	0.681	0.557	0.037	0.033	0.0055	0.1361
50	0.674	0.587	0.038	0.035	0.0051	0.1613
60	0.708	0.585	0.030	0.023	0.0045	0.1670
80	0.685	0.592	0.030	0.017	0.0044	0.1763
100	0.667	0.566	0.036	0.031	0.0039	0.1808
200	0.679	0.581	0.033	0.028	0.0032	0.1737
500	0.676	0.570	0.043	0.028	0.0027	0.1750

The NL approach is the wrong approach here because it ignores the random effect, thus we see that the probability of signal for the NL method is quite large and that the NLM

method does a much better job of keeping the probability of signal close to the nominal .05 level. Notice that the probability of signal for the NL approach is higher in Table 2 than it was in Figure 4. Additional variability in the nonlinear data (due to the random effect) causes the performance of the NL approach to deteriorate. In contrast, the NLM approach does not worsen because it is correctly accounting for the increased variability. As a result the approximate control limit based on the asymptotic normality assumption will not be very accurate for the NL approach when random effects are present. As n increases, the NL method does not improve. This is because the asymptotic results of the NL estimators does not necessarily hold in the presence of random effects. One concern is that the frequency of non-convergence in the NLM method is higher and appears to increase slightly as n increases, nonetheless, the NLM approach is still the preferred approach. The probability of signal calculations for the $T^2_{1,i,MIX}$ and $T^2_{2,i,MIX}$ statistics shown in Table 2 involve only the simulated runs where there was no non-convergence.

Now consider how the variability of the random effect impacts the results shown in Table 2. We repeated the same simulation study used to generate Table 2 with different values of σ^2_B used to generate the data. Figure 6 shows the probability of signal for the T^2 statistics for different values of σ^2_B where $m = 30$ and $\sigma^2 = .001$. Figure 7 shows the proportion of non-convergence of the NLM approach as σ^2_B varies where $m = 30$ and $\sigma^2 = .001$.

From Figure 6 we see that as the variability of the random effect gets smaller, the difference between the NL and NLM methods nearly disappears. This is because as σ^2_B decreases, the profiles are more similar to each other and the more similar they are to profiles with no random effects. In addition, in Figure 7 we see that the proportion of non-convergence decreases as σ^2_B decreases. There are slight increases in the proportion of non-convergence as n increases, likely due to the increased computational difficulty for increased sample sizes.
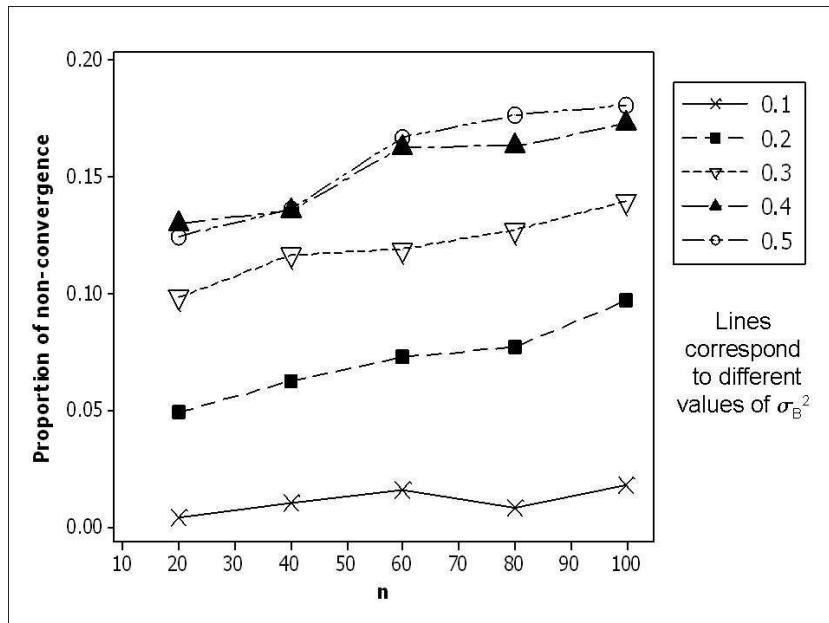
Figure 6: Probability of signal for $T_{1,i,LS}^2$, $T_{2,i,LS}^2$, $T_{1,i,MIX}^2$, and $T_{2,i,MIX}^2$ for in-control data for the median curve where σ_B^2 varies from .1 to .5. In this case $m = 30$ and $\sigma^2 = .001$.

We do not show specific results where σ^2 changes but note that its impact on data with random effects is the same as the impact on uncorrelated data with no random effects as was discussed earlier. That is, increasing σ^2 will increase the probability of signal for the NL approach with in-control data. We chose a smaller value of σ^2 for the results in this section, resulting in a smaller probability of signal for the NL method, in order to illustrate the impact of changing σ_B^2.

To illustrate the out-of-control performance of the NL and NLM methods we did a power study by introducing a change in the generated data and calculating the probability of signal under the change. With the 4-parameter logistic curve, a shift in the profiles can be introduced in any of the 4 parameters. We considered a step shift in B_i which represents a change in the steepness of the profiles. Figure 8 shows the out-of-control performance for the median curve with $T_{2,i,NL}^2$ and $T_{2,i,NLMM}^2$ when there is an increase in B_i after the fifth profile. Here $m = 30$, $\sigma^2 = .001$, n varies from 10 to 60, and the value of B_i increases by values ranging from 0 (the in-control case) to 5. We do not show the results for $T_{1,i,NL}^2$
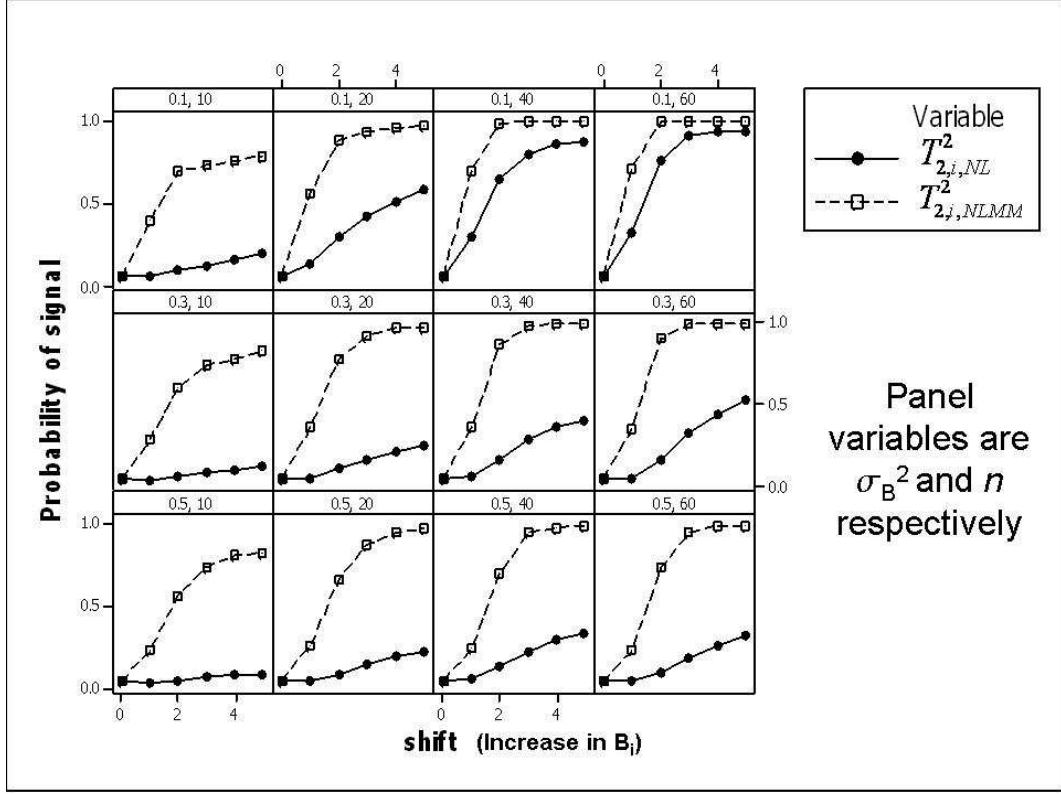
20

Figure 7: Proportion of non-convergence in NLM approach for the median curve when $m = 30$, $\sigma^2 = .001$ and n and σ_B^2 vary.

nor $T_{1,i,NLMM}^2$ because they have little ability in detecting step changes as demonstrated by Sullivan and Woodall (1996). To obtain these results in Figure 8, the control limit was simulated in order to ensure that the probability of signal for in-control data is equal to .05. $1,000$ datasets were simulated to generate the power curves.

We see that the NLM approach clearly has a much higher probability of detecting the change than the NL approach. The difference for the two approaches is larger when σ_B^2 is larger because the difference between the power curves is larger for the bottom row of Figure 8 than for the top row. This is because the NLM approach is taking into account the larger variability in the random effect that the NL approach ignores. As n increases, the closer the two approaches will be because they are both fitting the nonlinear curves equally well. When there is a smaller number of observations per profile, the NL approach will not fit the curves as well whereas the NLM approach pools information together across profiles to
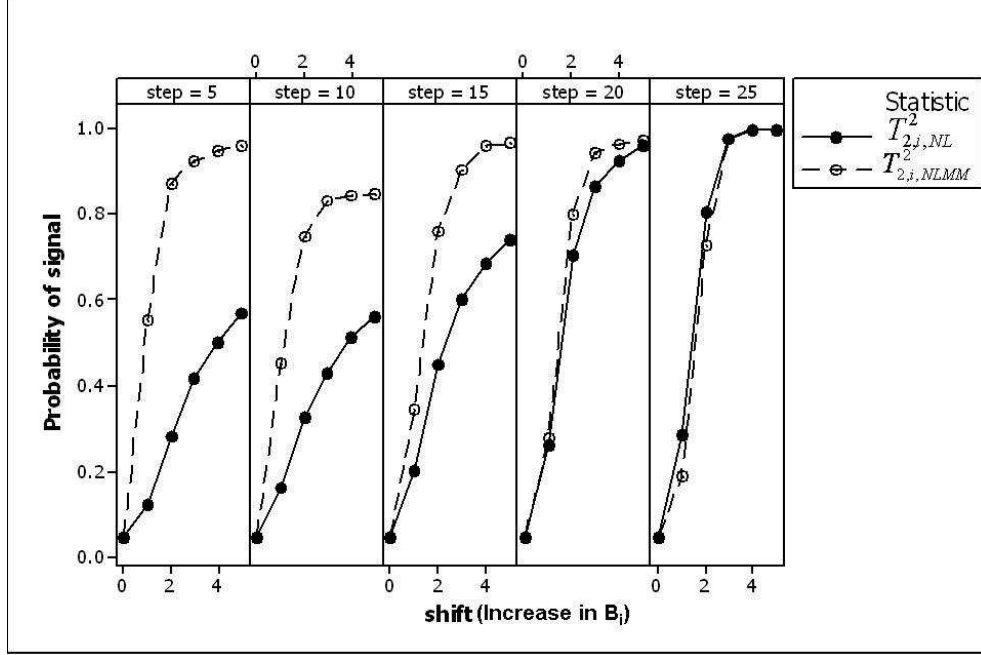
21

Figure 8: Probability of signal for $T^2_{2,i,NL}$ and $T^2_{2,i,NLMM}$ for out-of-control data for the median curve where σ^2_B and n vary. Here $m = 30$ and $\sigma^2 = .001$ and the step change in B_i occurred after the fifth profile.

achieve a better fit. As an added bonus and not shown here, we found that the frequency of non-convergence decreased for the NLM as the size of the shift increased. Similar results were obtained for the maximum and minimum curves, thus they are not presented here.

We do note that there is a dependence of the results on the location of the shift. Figure 9 shows the probability of signal for out-of-control data from the median curve for $m = 30$, $n = 20$ and when the step change occurred at different locations. We see that the NLM approach is always at least equivalent to the NL approach and often times far superior. The closer the shift is to the beginning of the data collection procedure the better the NLM approach will be.

22

Figure 9: Probability of signal for $T^2_{2,i,NL}$ and $T^2_{2,i,NLMM}$ for out-of-control data from median curve for various locations of the step change where there is an increase in B_i Here $m = 30$, $n = 20$, $\sigma^2_B = .1$ and $\sigma^2 = .001$.
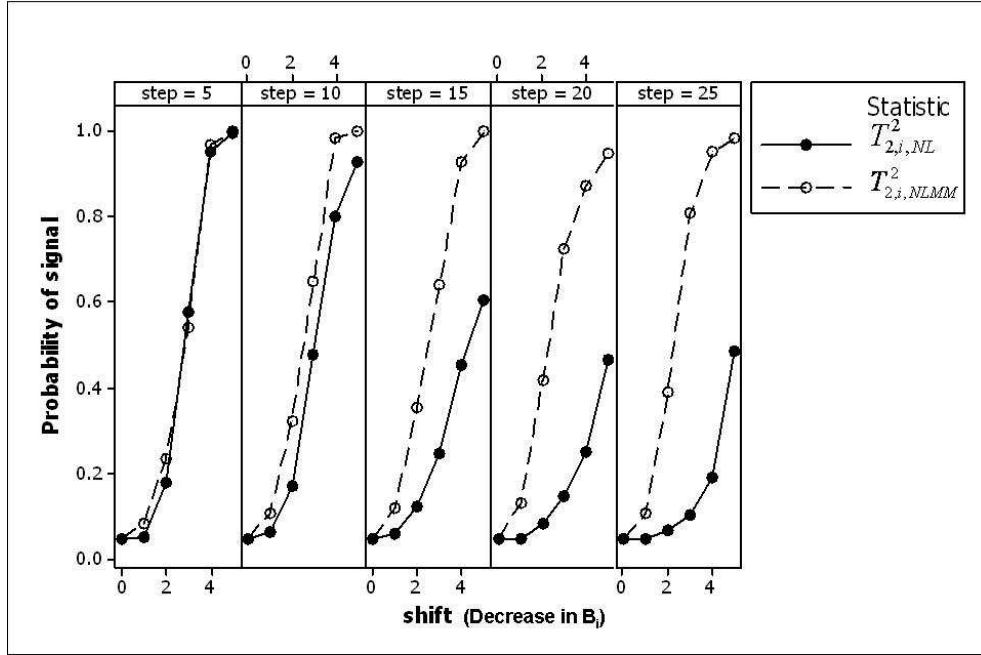

To explain the dependence on the locations of the shift, recall that because of regression equivariance, the appropriateness of the approximate control limit depends on the values of the parameters used in the function. For example, notice from Figure 4 that the probability of signal for the NL approach is different for the maximum curve than for the median curve. For each of the different locations where the step change occurs, there is a different mix of curves with different parameter values, thus the difference between the NL and NLM approaches is not the same across all values of the step change.

This lack of regression equivariance makes it difficult to generalize the conclusions regarding the increases to what occurs when there are decreases in B_i. To illustrate, compare Figure 9 with Figure 10 which shows the probability of signal for decreases in B_i. An issue

to be concerned with when studying decreases in B_i is that when B_i value is close to zero, the greater the chance that there will be instability in the numerical algorithms used to obtain the results. Thus the generated curves used in Figure 10 are the same as the median curve used to generate Figure 9, except that the in-control value of $B_i = 7$ instead of 2. We see again that the probability of signal depends on the location of the shift. In Figure 10, the difference between the NL and NLM approaches is largest for later shifts and negligible for earlier shifts. Note that the results for step=5 in Figure 9 are similar to the results for step=25 in Figure 10. This is because in the first case when step=5, the in-control curves are equivalent to the out-of-control curves from the second case and vice versa.

Figure 10: Probability of signal for $T^2_{2,i,NL}$ and $T^2_{2,i,NLMM}$ for out-of-control data for various locations of the step change where there is a decrease in B_i Here $m = 30$, $n = 20$, $\sigma^2_B = .1$ and $\sigma^2 = .001$.

Correlated Data With Random Effects

Now consider the situation where the profiles have correlated errors with random effects but still are balanced and equally spaced. In our review of the literature on the NLM model, we found that the $\mathbf{R}_i$ matrix is often assumed to be a diagonal or identity matrix, a fact also noted by Davidian and Giltinan (2003). This is because the variability between profiles represented by the random effects has a bigger impact on the variability in the response than does the correlation within profiles. This was noted for the linear mixed model by Verbeke and Molenbergs (2000) and for the NLM model by Davidian and Giltinan (1995, 2003). As a result software such as $SAS^{®}$ and S-Plus do not allow specification of a correlation structure of the errors in conjunction with integral approximation methods although they do allow for correlation with linearization methods (Schabenberger and Pierce, 2002, p. 538).

In addition, it was noted by Schabenberger and Pierce (2002) that modeling the correlation concurrently with a random effect makes little difference in the results obtained but can increase the likelihood of convergence problems. To model the correlation in the NLM approach would be difficult in practice. It would require one to program the numerical algorithms to obtain the estimates because they are not readily available in $SAS^{®}$ or S-Plus when using the integral approximation approach. Thus, although theoretically possible, it is not easy to obtain a NLM model fit with correlated errors. We present here empirical evidence showing that it may be sufficient to simply model the random effects and not model the correlation in our application.

For the median 4-parameter logistic curve we generated in-control data with uncorrelated and correlated errors where there was a single random effect in B_i. In Table 3 we show the probability of signal for the T^2 statistics based on the NL and NLM approaches for in-control data following the median curve with $m = 30$, $n = 20$, $\sigma_B^2 = 0.5$ and $\sigma^2 = .001$ and where

different amounts of correlation were introduced in the errors. There is a large difference in the probability of signal for the T^2 statistics based on the NL approach when compared to those based on the NLM approach, but there is little difference in the T^2 statistics as the amount of correlation in the errors increases. Thus the modeling of the random effect has a large impact on the appropriateness of the control limit.
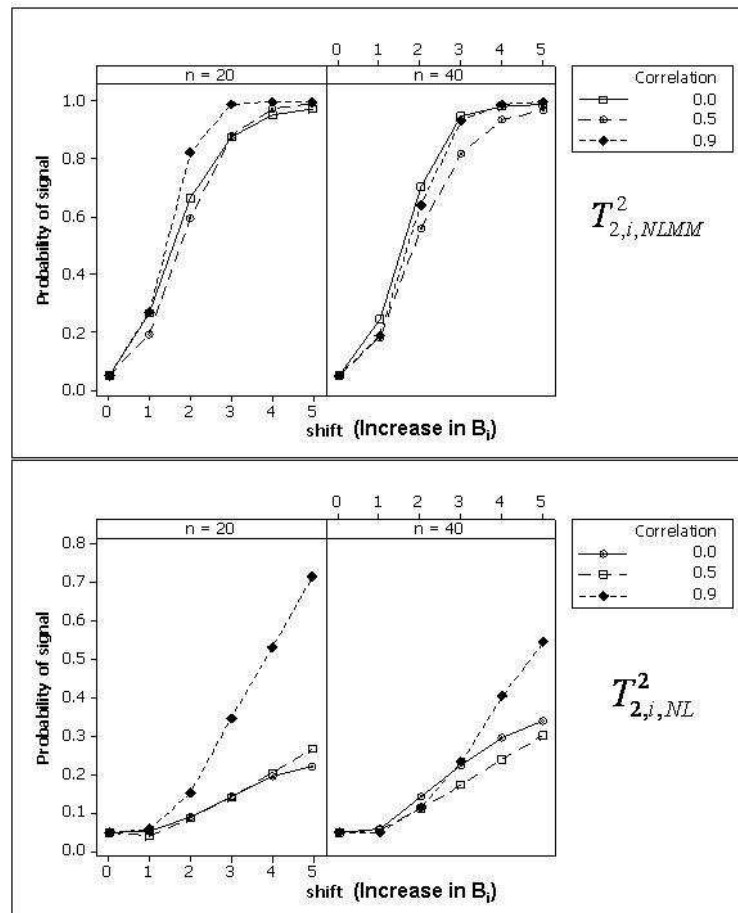
Table 3: T^2 statistics for the NL and NLM approaches for data with uncorrelated and correlated errors. The generated data follow the median curve with $m = 30$, $n = 20$, $\sigma_B^2 = 0.5$ and $\sigma^2 = .001$.

ρ	$T^2_{1,i,NL}$	$T^2_{2,i,NL}$	$T^2_{1,i,NLMM}$	$T^2_{2,i,NLMM}$
0.0	0.6784	0.5680	0.0502	0.0378
0.1	0.6766	0.5642	0.0470	0.0366
0.2	0.6810	0.5660	0.0428	0.0336
0.3	0.6836	0.5706	0.0404	0.0295
0.4	0.6786	0.5638	0.0360	0.0288
0.5	0.6742	0.5602	0.0310	0.0277
0.6	0.6716	0.5594	0.0292	0.0263
0.7	0.6542	0.5498	0.0275	0.0258
0.8	0.6358	0.5286	0.0302	0.0279
0.9	0.6226	0.5090	0.0433	0.0332

When considering the out-of-control case as done previously, we show in Figure 11 the probability of signal for $T^2_{2,i,NL}$ and $T^2_{2,i,NLMM}$ when a step change was introduced in the profiles when the data have uncorrelated or correlated errors. We see that the NLM approach gives us a higher probability of signal than the NL approach across the levels of correlation.

To summarize the results here, we see that the NLM approach that ignores the correlation in the errors is superior to the NL approach even when the errors are correlated. We note that this does not mean that we could not do even better with a NLM approach that does model the correlation in the errors. While we do not believe the NLM approach could be improved drastically by explicitly modeling the correlated errors, it remains to be seen how

Figure 11: Probability of signal for $T^2_{2,i,NL}$ and $T^2_{2,i,NLMM}$ for out-of-control data following the median curve where a step change occurred after the fifth profile. Here $m = 30$, $\sigma^2_B = 0.5$ and $\sigma^2 = .001$.

much improvement could be gained.

Our conclusions here agree with the results of Staudhammer et al. (2005) who modeled profiles in lumber manufacturing using mixed models. They modeled the autocorrelated errors with time series models and concluded that for SPC applications, ignoring the autocorrelation does not make much difference. This is true even though their profiles are rather complicated and there is an obvious autocorrelation in the data. The number of observations that they have per profile is much larger than the sample sizes considered here ($n > 2000$).

Proposed Method

Our proposed method of Phase I analysis when considering nonlinear profiles uses both the NL and NLM to determine outlying profiles. It builds on the approach of Williams et al. (2006a) who proposed the following steps:

1. If there is replication of the points taken at each location along the profile, then the homogeneity of variance can be checked. Profiles that do not have a homogeneous variance across the locations within the profile are eliminated from the dataset after inspecting the appropriate T^2 statistic, which is obtained by using a variance regression model.

2. Determine the appropriateness of the choice of the nonlinear function through a lack-of-fit (LOF) test. This test can be performed whether or not there is replication of the points although when there is no replication, the lack-of-fit test is model based.

3. Fit separate NL regression models to each of the profiles to obtain individual profile parameter estimates. Use the estimates for the basis of the T^2 statistics as in (13). This step corresponds to the NL approach studied earlier.

Our approach uses the first two steps of the approach outlined above and replaces step 3 with two new steps that utilize the NLM model approach. Thus the last two steps of our proposed approach are given by:

3. Fit separate NL regression models to each of the profiles to obtain individual profile parameter estimates. Compute the sample variances across the m profiles for each of the parameters. Upon inspecting the sample variances for large values relative to

the other values, determine which parameters, if any, should be modeled with random effects, trying to keep the number of random effects small. In many cases, there will be one or two of the nonlinear parameters that have much larger variability than the others; these should be prime candidates for inclusion of random effects.

4. Fit the NLM model using the determination of random effects from the previous step to obtain the predicted random effects. Use the predicted random effects as the basis for the T^2 statistics to determine outlying profiles as in (15). The control limit can be obtained by using the corresponding percentiles of the appropriate distribution. This step corresponds to the NLM approach studied earlier and compared with the NL approach.
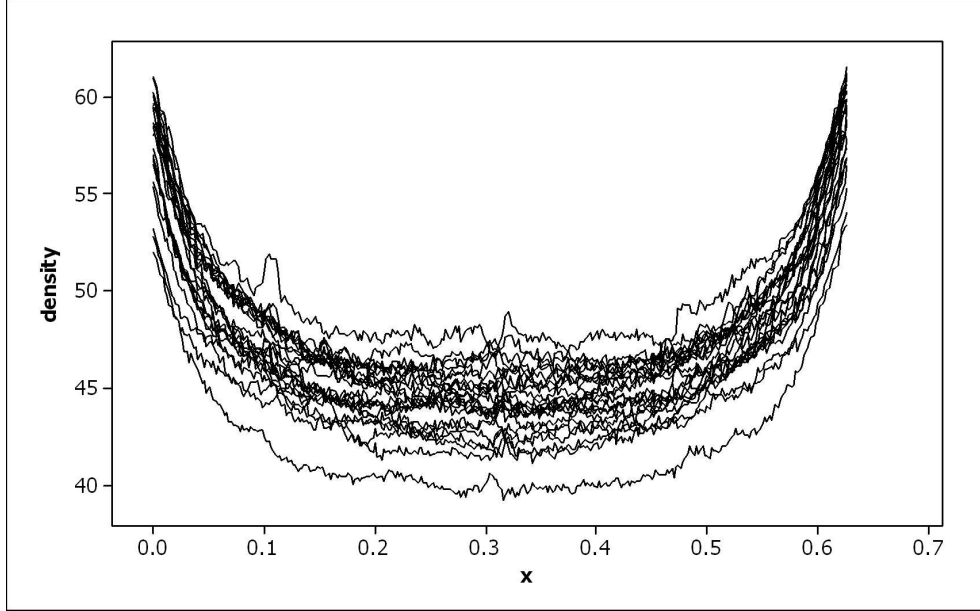
This proposed approach will be illustrated in the example that follows.

Example

To illustrate the differences between the proposed approach and the approach of Williams et al. (2006a) we consider the particle board data presented in Walker and Wright (2002) and studied by Williams, Woodall, and Birch (2003). In the initial analysis, there were 24 boards, each of which had 314 measurements along the profile. The raw data profiles are shown in Figure 12.

Williams, Woodall, and Birch (2003) fit a nonlinear function requiring six parameters to the raw data profiles. These parameters consisted of a lower asymptote, the center where the lower asymptote is achieved and four other parameters determining the flatness of the center of the curve and the rate at which the curve increases. They allowed the curve to be asymptotic about the center. However, we note that the profile are very nearly symmetric

Figure 12: Raw data profiles for the particle board data of Walker and Wright (2002).

so we reduce the number of parameters required by constraining the curve to be symmetric. Thus our model will only require 4 parameters to describe the function. The model is given by

$$
y_{ij} = \left\{
\begin{array}{l}
E_i \left(x_{ij} - H_i\right)^{F_i} + G_i \text{ if } x_{ij} > H_i \\[2mm]
E_i \left(-x_{ij} + H_i\right)^{F_i} + G_i \text{ if } x_{ij} \leq H_i
\end{array}
\right\} \text{ for } i = 1, 2, \ldots, m, j = 1, 2, \ldots, n_i, \quad (18)
$$

where E_i and F_i represent the rate of increase and flatness in the curve, G_i is the lower asymptote, and H_i is the center point where the curve attains the lower asymptote.

Because there is no replication of the points for the locations along the profile, we do not perform the first step of our proposed procedure, the test of homogeneity of variance within the profile.

The second step is to perform a LOF test. Because there are no replicates of the points at the different locations we cannot perform a classical lack of fit test discussed in many

30

textbooks on regression. An alternative is to divide the points along the profile into different groups and treat the points within a group as "pseudo-replicates". This idea is discussed in Su and Yang (2006) and its references. We do not pursue this idea here because the particle board data does not lend itself to natural groups within the profile.

In lieu of a LOF test, we test for a "lack of consistency" among the m profiles as discussed in Williams (2005). This test compares the fits within the profiles to see if any profiles have substantially worse fit than the other profiles. For each profile we calculate the sums of squares of the residual obtained by subtracting from the response its predicted value. This is given by

$$SSE_i = \sum_{i=1}^{m} (\mathbf{y}_{ij} - \widehat{y}_{ij})^2 \ \text{ for } i = 1, 2, \ldots, m, \tag{19}$$
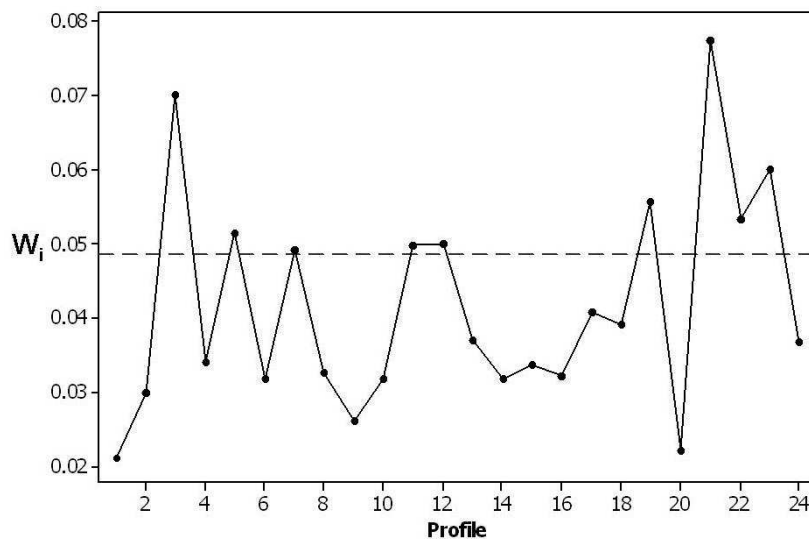
where $\widehat{y}_{ij}$ is the predicted value obtained from the nonlinear regression fit. The the statistic

$$W_i = \frac{SSE_i}{\sum_{i=1}^{m} SSE_i}. \tag{20}$$

will have an approximate beta distribution with n_i and $(\sum_{i=1}^{m} n_i) - n_i$ degrees of freedom respectively. Using (19) for the particle board data we obtain the SSE_i and W_i values and plot the W_i values in Figure 13. Also shown is the critical value (the dotted line) from the beta distribution.

We do not believe that the beta distribution is correct in this case because of the large number of profiles that would be declared as lacking fit. Further research is needed to determine the appropriate limit to use for this lack of consistency procedure where there are no replicates. Instead we inspect the chart for values of W_i that stand out from the others. Profiles 3 and 21 both have W_i values that are larger than the values for the other profiles and would be candidates for exclusion. We choose to keep them in the dataset in order to illustrate the control chart testing procedure.

31

Figure 13: Lack of consistency test values for the particle board data of Walker and Wright (2002).
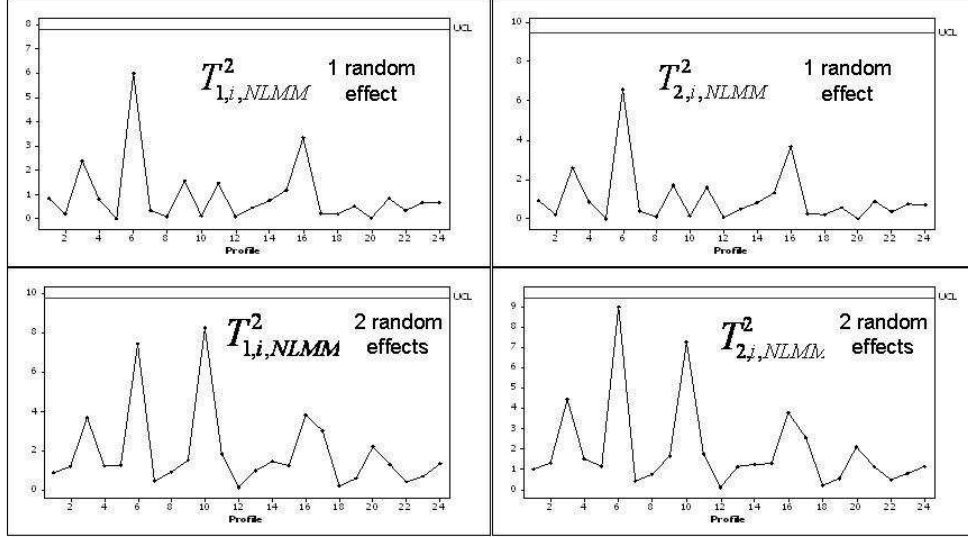

The third step is to fit the separate NL models and obtain the parameter estimates to determine which random effects to include. The mean and standard deviation of the parameter estimates of the separate NL fits for the particle board data is shown in Table 4. We see that the variability of E_i is much larger than any other parameter. Thus it is the most likely candidate for having a random effect. If a second random effect is included in the NLM model, the G_i would be included with a random effect.

Table 4: Mean and standard deviation of the parameter estimates for the particle board data of Walker and Wright (2002).

Parameter	Mean	Standard Deviation
E_i	9317.5	20965
F_i	4.7173	1.0273
G_i	44.428	1.7381
H_i	0.3104	0.0030

Figure 14: T^2 control charts for the NLM approach for the particle board data of Walker and Wright (2002). The top panel corresponds to the chart with a single random effect and the bottom panel corresponds to the chart with two random effects.
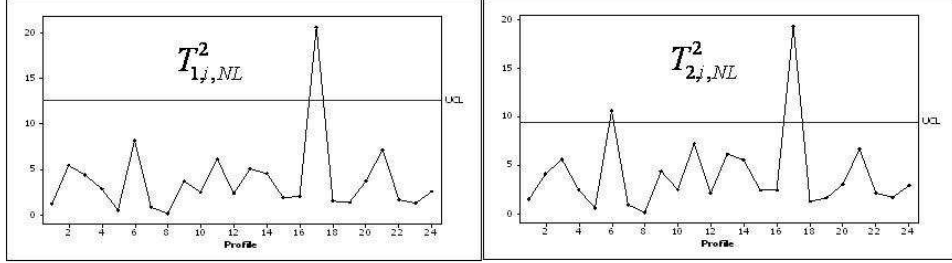

The final step is to fit the NLM model and obtain the T^2 statistics from (14) and (15). The resulting charts for the T^2 statistics based on the NLM model is shown in Figure 14. The top panel corresponds to the charts when we only modeled a single random effect, E_i and the bottom panel corresponds to the charts when we modeled two random effects, E_i and G_i. Compare Figure 14 with Figure 15 which shows the charts obtained for the separate NL regression models from (11) and (13) and which corresponds to the method of Williams et al. (2006a). We see that the use of $T^2_{2,i,NL}$ produces two signals on the control chart whereas the $T^2_{2,i,NLMM}$ chart doesn't produce any signals. Thus we conclude that we can use all 24 profiles to obtain the parameter estimates on which the Phase II control charts will be based.

Conclusions

For nonlinear profiles, we proposed to fit them with a NLM model to account for random

Figure 15: T^2 control charts for the NL approach for the particle board data of Walker and Wright (2002).

effects and correlated errors. We found that an approach that uses the separate NL regression models to determine needed random effects works well in setting up the analysis of the profiles with a NLM model. This proposed method uses an easy to calculate control limit and thus does not require extensive simulation to obtain the correct control limit as the approach of Williams et al. (2006a). We also found that we can ignore the correlation of the errors and concentrate our effort in modeling the random effects. Modeling the random effects allows us to use a Phase I control limit that does not need to be obtained via simulation as would be needed if we only obtained the estimates from separate NL regression models.

There are a number of questions not answered here. For example, we limited ourselves to the balanced, equally spaced data scenario but did not consider unbalanced data or data with missing observations. Because of the superiority of the NLM approach for balanced, equally spaced data, we believe that it will retain or even increase its advantage for unbalanced or missing data scenarios. Nor have we considered a T^2 statistic based on high breakdown estimators that are excellent at detecting multiple outliers (Vargas, 2003; Jensen, Birch, and Woodall, 2006a).

Profile monitoring is a very fruitful area of research. We believe that there is a tremendous reservoir of untapped data applications where this methodology can be utilized. Profile

monitoring is a tool of the future to match the technology of today that will realize more widespread acceptance as it becomes more readily accessible. We encourage further research and applications of profile monitoring.

REFERENCES

Davidian, M. and Giltinan, D.M. (1995). *Nonlinear Models for Repeated easurements Data.* Chapman and Hall, London, UK.

Davidian, M. and Giltinan, D.M. (2003). "Nonlinear Models for Repeated Measurements Data: An Overview and Update". *Journal of Agricultural, Biological, and Environmental Statistics* 8, pp. 387-419.

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

Gallant, A.R. (1987). *Nonlinear Statistical Models.* John Wiley and Sons, New York, NY.

Hartford, A. and Davidian, M. (2000). "Consequences of Misspecifying Assumptions in Nonlinear Mixed Effects Models". *Computational Statistics and Data Analysis* 34, pp. 139-164.

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

Jensen, W.A., Birch, J.B., and Woodall, W.H. (2006a). "High Breakdown Estimation Methods for Phase I Multivariate Control Charts". Submitted to *Journal of Quality Technology.*

Jensen, W.A., Birch, J.B., and Woodall, W.H. (2006b). "Profile Monitoring via Linear Mixed Models". Submitted to *Journal of Quality Technology.*

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

Kim, K., Mahmoud, M. A., and Woodall, W. H. (2003). "On the Monitoring of Linear Profiles". *Journal of Quality Technology* 35, pp. 317-328 .

Lee, S.-Y. and Xu, L. (2004). "Influence Analyses of Nonlinear Mixed-Effects Models". *Computational Statistics & Data Analysis* 45, pp. 321-341.

Littell, R.C., Milliken, G.A., Stroup, W.W., and Wolfinger, R.D. (1996). *SAS System for Mixed Models*, SAS Institute Inc. Cary, NC.

Mahmoud, M. A. and Woodall, W. H. (2004). "Phase I Analysis of Linear Profiles with Calibration Applications". *Technometrics* 46, pp. 380-391.

Mason, R. L. and Young, J. C. (2002). *Multivariate Statistical Process Control with Industrial Applications*, Philadelphia: SIAM.

Pinheiro, J.C. and Bates, D.M. (2000). *Mixed-Effects Models in S and S-PLUS*, Springer-Verlag, New York, NY.

Ratkowsky, D.A. (1990). *Handbook of Nonlinear Regression Models*. Marcel Dekker, New York, NY.

Rousseeuw, P.J. and Leroy, A.M. (1987). *Robust Regression and Outlier Detection*. John Wiley & Sons, New York, NY.

Schabenberger, O. and Pierce, F.J. (2002). *Contemporary Statistical Models for the Plant and Soil Sciences*. CRC Press, Boca Raton, Florida.

Seber, G.A.F. and Wild, C.J. (2003). *Nonlinear Regression*. Wiley-Interscience, Hoboken, NJ.

Staudhammer, C.L., Lemay, V.M., Kozak, R.A, and Maness, T.C. (2005). "Mixed-Model Development for Real-Time Statistical Process Control Data in Wood Products Manufacturing". *Forest Biometry, Modelling and Information Sciences* 1, pp. 19-35.

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

Su, Z. and Yang, S.-S. (2006). "A Note on Lack-of-Fit Tests for Linear Models without Replication". *Journal of the American Statistical Association* 101, pp. 205-210.

Vargas, J.A. (2003). "Robust Estimation in Multivariate Control Charts for Individual Observations". *Journal of Quality Technology* 35, pp. 367-376.

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

Walker, E. and Wright, S. (2002). "Comparing Curves Using Additive Models". *Journal of Quality Technology* 34, pp. 118-129.

Wang, K. and Tsung, F. (2005). "Using Profile Monitoring Techniques for a Data-Rich Environment with Huge Sample Sizes". *Quality and Reliability Engineering International* 21, pp. 677-688.

Williams, J.D. (2005). "Contributions to Profile Monitoring and Multivariate Statistical Process Control". Unpublished doctoral dissertation, Department of Statistics, Virginia Polytechnic Institute & State University.

Williams, J.D., Birch, J.B., Woodall, W.H., and Ferry, N.M. (2006a). "Statistical Monitoring of Heteroscedastic Dose-Response Profiles from High-Throughput Screening". Submitted to *JABES*.

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

Williams, J.D., Woodall, W.H., Birch, J.B., and Sullivan, J.H. (2006b). "On the Distribution of T^2 Statistics Based on Successive Differences". to appear in *Journal of Quality Technology*.

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