Statistical Methods for Non-Linear Profile Monitoring

Ana Valeria Quevedo Candela

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Statistics

Gordon G. Vining, Chair

Robert Gramacy

Peter A. Parker

William H. Woodall

December 6, 2019

Blacksburg, Virginia

Keywords: non-linear profile monitoring, non-linear mixed model, practical significance,

Gaussian process model, heteroscedasticity

Statistical Methods for Non-Linear Profile Monitoring

Ana Valeria Quevedo Candela

(ABSTRACT)

We have seen an increased interest and extensive research in the monitoring of a process over time whose characteristics are represented mathematically in functional forms such as profiles. Most of the current techniques require all of the data for each profile to determine the state of the process. Thus, quality engineers from industrial processes such as agricultural, aquacultural, and chemical cannot make process corrections to the current profile that are essential for correcting their processes at an early stage. In addition, the focus of most of the current techniques is on the statistical significance of the parameters or features of the model instead of the practical significance, which often relates to the actual quality characteristic. The goal of this research is to provide alternatives to address these two main concerns. First, we study the use of a Shewhart type control chart to monitor within profiles, where the central line is the predictive mean profile and the control limits are formed based on the prediction band. Second, we study a statistic based on a non-linear mixed model recognizing that the model leads to correlations among the estimated parameters.

Statistical Methods for Non-Linear Profile Monitoring

Ana Valeria Quevedo Candela

(GENERAL AUDIENCE ABSTRACT)

Checking the stability over time of the quality of a process which is best expressed by a relationship between a quality characteristic and other variables involved in the process has received increasing attention. The goal of this research is to provide alternative methods to determine the state of such a process. Both methods presented here are compared to the current methodologies. The first method will allow us to monitor a process while the data is still being collected. The second one is based on the quality characteristic of the process and takes full advantage of the model structure. Both methods seem to be more robust than the current most well-known method.

Acknowledgments

This dissertation would not have been possible without the guidance of my advisor, Dr. Geoffrey Vining. His remarkable insights and inspiring ideas have influenced me academically and professionally.

I thank the members of my committee, Dr. William H. Woodall, Dr. Robert Gramacy and Dr. Peter Parker, for their valuable comments and suggestions to this dissertation, and for their continued support during all the process.

I am grateful to the Universidad de Piura for the support and confidence placed in me. I specially thank Dr. Susana Vegas for her friendship and trust.

I also thank the friends we made in Blacksburg, especially Carmen. Thank you for making us feel always at home.

Finally, to my family. To Marcos, Mariana and Valentina, for your sacrifice, encouragement and unconditional love. To my mom, for your unconditional support every step since day one in your womb.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In recent years statistical process monitoring where process characteristics are represented mathematically in functional forms such as curves, surfaces, or images has received increasing attention. The research in this growing literature has made extensive use of both parametric and non-parametric methods in profile monitoring analysis [Woodall et al., 2004, Woodall, 2007, Colosimo and Pacella, 2010, Noorossana et al., 2011].

A typical approach to profile monitoring entails developing an appropriate model for the background structure and, upon observing a new profile, determining if the profile is different from "normal behavior" based on a given deviation metric. The literature shows a general common approach. First, model the profiles through a parametric or non-parametric technique. Then, estimate the parameters or relevant features that identify a profile. Finally, use a multivariate control chart on those parameters or features.

The current literature on parametric profile modeling tends to focus on the estimated parameters of the possibly non-linear model. The profile monitoring literature suggests the use of multivariate T^2 statistics to monitor simultaneously the stability of the parameters [Woodall et al., 2004, Williams et al., 2007, Colosimo and Pacella, 2010]. The use of non-parametric models as data reduction techniques in profile monitoring has also been studied [Zou et al., 2008, Chang and Yadama, 2010, Qiu et al., 2010].

While profile monitoring has been studied extensively, most of the current techniques are purely retrospective given that we cannot make process corrections within the current profile. These approaches require all of the data for each profile to determine the state of the process. Basically, these techniques assume that a profile is produced in such a short period of time that corrections to the observed process cannot be made.

In addition, most of the parametric and non-parametric techniques for profile monitoring focus on the parameters or features of the model instead of the actual quality characteristic observed. These methods fit separate non-linear regression models to monitor the stability of the parameter estimates simultaneously through a multivariate statistic. By doing so, their focus is on detecting process changes based on the statistical significance of the estimators, which often produce relatively unimportant changes in the true quality characteristic. The debate over statistical and practical significance is now more prevalent in the literature [Woodall and Faltin, 2019].

This dissertation attempts to address these issues. First, we study the use of a Shewhart-type control chart where the central line is the mean profile estimated with a parametric or non-parametric model which allows heteroscedasticity to account for the profile-to-profile variability and the prediction band becomes the control limits. The prediction band would define the normal behavior; a profile with an "abnormal number" of observations outside the prediction band represents a possible assignable cause. This control chart addresses both issues outlined above: it does not require all of the data for each profile to determine the state of the process and keeps the focus on the quality characteristic.

In the second part of the dissertation, we propose a control chart procedure that addresses the second issue directly. Instead of treating the estimated parameters of the model as separate, but correlated, quality characteristics of interest, we study a statistic based on a non-linear mixed model structure. We use the response as the actual quality characteristic and recognize that the model subsequently leads to correlations among the estimated parameters.

The remainder of this dissertation is organized as follows. Chapter 2 reviews current literature on parametric and non-parametric approaches for profile monitoring. Chapter 3 provides context based on a practical situation to illustrate statistical process control where the fundamental issue is practical significance given that the economic consequences that are attributable to assignable causes relate to the response. Chapter 4 demonstrates that when there is an unknown physical model behind the process, Gaussian process modeling is effective in predicting the mean profile that will converge to the true unknown model with replications. The advantage of using the GP model rather than other non-parametric techniques is that it naturally provides the prediction band around the mean profile. Chapter 5 proposes a new, more robust, procedure for monitoring profiles based on a non-linear mixed model that tests if a given profile is different from the rest of the profiles. This procedure considers the asymptotic variance of the fixed parameters obtained using the actual data, not the parameter estimates themselves. Finally, Chapter 6 ends with conclusions and possible extensions to this research.

Chapter 2

Review of Literature

Profile monitoring is concerned with the analysis of a product or process whose quality characteristics are best represented in functional relationships with one or more explanatory variables. The common approach is to develop an in-control (IC) model for the underlying structure and upon observing a new profile, determine if the profile is satisfactory based on a particular deviation metric. See Woodall et al. [2004], Woodall [2007], Colosimo and Pacella [2010], Noorossana et al. [2011], and Qiu [2013].

The development of an appropriate model for the profile can be separated into two categories, parametric and non-parametric modeling. Kang and Albin [2000], Kim et al. [2003], Ding et al. [2006], Zou et al. [2006], Williams et al. [2007], Zou et al. [2007], Noorossana et al. [2010], and Colosimo and Pacella [2010] study the use of linear and non-linear parametric profiles. The idea of monitoring parametric profiles is extended to include mixed-effects models by Jensen et al. [2008], and Jensen and Birch [2009]. Zou et al. [2008], Chang and Yadama [2010], and Qiu et al. [2010] study non-parametric techniques for profile monitoring. Recently, Gaussian process models also has been used for this purpose (see Zhang et al. [2014], Wang et al. [2014], Zhang et al. [2016], and Yongxiang et al. [2018]).

Any given non-linear profile can be expressed as:

$$y_{ij} = f(\mathbf{x}_{ij}^t, \boldsymbol{\theta}) + \epsilon_{ij} \tag{2.1}$$

where y_{ij} is the jth observation ($j = 1 \dots n_i$) of the ith profile ($i = 1 \dots m$), $\mathbf{x}_{ij}$ is a vector of predictors, $\boldsymbol{\theta}$ is the px1 vector of parameters or features of the model, and ϵ_{ij} is the error term whose structure depends on the model used. The modeling approach determines whether to use a parametric or non-parametric technique to estimate $f(\mathbf{x}_{ij}^t, \boldsymbol{\theta})$.

2.1 Parametric modeling approach

The profile monitoring literature suggests the use of multivariate T^2 statistics to monitor the stability of the parameters simultaneously (see Williams et al. [2007], Colosimo and Pacella [2010]). Generally, the analyst fits a non-linear parametric model to each profile from IC data.

Either for Phase I or II, the literature suggests the calculation of a measure of distance between the vector of parameter estimates from the ith profile, $\hat{\boldsymbol{\theta}}_i$, and a target or the in-control (IC) vector of estimates, $\boldsymbol{\theta}_0$. The metric most often used for this purpose is the T^2 statistic whose general form is the following:

$$T^2 = (\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_0)^t S^{-1} (\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_0),$$

where $\boldsymbol{\theta}_0$ is estimated by averaging the parameter estimates from the Phase I: $\bar{\hat{\boldsymbol{\theta}}} = \frac{\sum_{i=1}^{m} \hat{\boldsymbol{\theta}}_i}{m}$, and S is the estimated covariance matrix of the parameter estimators.

There are a few choices in the literature for estimating the covariance matrix [Williams et al., 2007]. The simplest one is the sample covariance matrix (S_1), where

$$S_1 = \frac{1}{m-1} \sum_{i=1}^{m} (\hat{\boldsymbol{\theta}}_i - \bar{\hat{\boldsymbol{\theta}}})(\hat{\boldsymbol{\theta}}_i - \bar{\hat{\boldsymbol{\theta}}})^t.$$

The distribution of $T^2(S_1)$ based on the sample covariance matrix is proportional to a beta distribution as proved by Gnanadesikan and Kettenring [1972], i.e.,

$$T_i^2(S_1) \frac{m}{(m-1)^2} \sim BETA(\frac{p}{2}, \frac{m-p-1}{2}).$$

Another measure proposed by Holmes and Mergen [1993] uses the successive differences of consecutive vectors of estimates (S_2), where

$$S_2 = \frac{1}{2(m-1)} \sum_{i=1}^{m-1} (\hat{\boldsymbol{\theta}}_{i+1} - \hat{\boldsymbol{\theta}}_i)(\hat{\boldsymbol{\theta}}_{i+1} - \hat{\boldsymbol{\theta}}_i)^t. \tag{2.2}$$

In a retrospective analysis, Sullivan and Woodall [1996] show that this is a more robust estimator. The exact distribution of a T^2 based on S_2 is unknown [Williams et al., 2007]. However, as Williams et al. [2006] show, when $m > p^2 + 3p$ for any value of p, the $T_i^2(S_2)$ can be approximated by a $\chi^2(p)$. On the contrary, if $m \leq p^2 + 3p$, the authors recommend the following distribution for $p < 10$:

$$T_i^2(S_2) \sim MV(m,i)BETA(\beta(m,p,i), \gamma(m,p,i)),$$

where $\beta(m,p,i)$ and $\gamma(m,p,i)$ are estimated functions shown in Williams et al. [2006]. Note that to use S_2 requires that the profiles are in time order which may not be always the case.

An alternative choice for estimating the covariance matrix S, studied by Vargas [2003], Williams et al. [2007] and Jensen et al. [2007], is the use of the minimum volume ellipsoid estimator. This leads to

$$S_3 = \frac{1}{m-1}\sum_{i=1}^{m}(\hat{\boldsymbol{\theta}}_i - \bar{\hat{\boldsymbol{\theta}}}_{MVE})(\hat{\boldsymbol{\theta}}_i - \bar{\hat{\boldsymbol{\theta}}}_{MVE})^t,$$

where $\bar{\hat{\boldsymbol{\theta}}}_{MVE}$ is the minimum value ellipsoid (MVE) estimator. This method finds the ellipsoid with the smallest value containing the least half of the $\boldsymbol{\theta}_i$ [Williams et al., 2007]. The distribution of $T^2(S_3)$ is unknown.

The methods outlined above monitor parametric profiles through their fixed parameter estimates. Jensen and Birch [2009] extend this idea to include mixed-effects. The authors propose a non-linear mixed model for Phase I analysis to account for the correlation structure of all of the profiles. For this case, the function shown in Equation 2.1 becomes $f(\mathbf{x}_{ij}^t, \boldsymbol{\theta}, \boldsymbol{\eta}_i)$ where $\boldsymbol{\theta}$ is the vector of fixed effects and $\boldsymbol{\eta}_i$ is the vector of the random effects of the ith profile. In this framework, the authors propose the use of the T^2 based on the predicted random effects parameters for each profile instead of the fixed effects:

$$T_i^2 = (\hat{\boldsymbol{\eta}}_i - \bar{\hat{\boldsymbol{\eta}}})^t S^{-1} (\hat{\boldsymbol{\eta}}_i - \bar{\hat{\boldsymbol{\eta}}}),$$

where $\hat{\boldsymbol{\eta}}_i$ is the predicted random effect of the ith profile, and $\bar{\hat{\boldsymbol{\eta}}} = \frac{\sum_{i=1}^m \hat{\boldsymbol{\eta}}_i}{m}$ is the mean predicted random effects estimate. As estimators for S, Jensen and Birch [2009] propose the use of either the sample covariance matrix or the successive differences of the predicted random effects.

2.2 Non-parametric modeling approach

The function $f(.)$ can be estimated through a non-parametric approach such as smoothing methods like spline-fitting techniques [Boeing Commercial Airplane Group, Walker and Wright, 2000], wavelets analysis [Jin and Shi, 2001, Ganesan et al., 2004, Jeong et al., 2006, Chang and Yadama, 2010], principal components analysis (PCA) [Jackson, 1991, Ding et al., 2006, Colosimo and Pacella, 2010], and Gaussian processes [Wang et al., 2014, Zhang et al., 2014, 2016, Yongxiang et al., 2018].

The literature outlines two approaches to calculate the metrics for determining how different a given profile is from the mean profile. In the first one, the metrics are based on the actual profile data that summarize the differences between a given profile and the mean profile. In the second approach, multivariate statistics based on the relevant features that identify a profile are used.

As part of the first approach to summarize the differences between the ith profile and the IC profile based on the profile data, a number of metrics can be used as shown below. We could evaluate the average difference or deviation M_{1i}, the average absolute deviation M_{2i}, the maximum absolute deviation M_{3i}, and the averaged squared deviation M_{4i} [Boeing Commercial Airplane Group, Gardner et al., 1997, Williams et al., 2007].

$M_{1i} = \Sigma_{j=1}^{n_i}(y_{ij} - \hat{f}(\mathbf{x}_{ij}^t))/n_i;$

$M_{2i} = \Sigma_{j=1}^{n_i}|y_{ij} - \hat{f}(\mathbf{x}_{ij}^t)|/n_i;$

$M_{3i} = \max_j|y_{ij} - \hat{f}(\mathbf{x}_{ij}^t)|;$

$M_{4i} = \Sigma_{j=1}^{n_i}(y_{ij} - \hat{f}(\mathbf{x}_{ij}^t))^2/n_i;$

where $i = \{1 \ldots m\}$ and m is the number of profiles.

If the profiles are produced in time order, Boeing Commercial Airplane Group suggests using individual and moving range charts on those metrics for monitoring the between profile variation.

Splines as part of a functional data analysis (FDA) methods has been largely used to estimate

$f(\cdot)$. Boeing Commercial Airplane Group suggests to fit a spline to each profile. It then suggests to average all the splines fits at each location or point in time to obtain the mean spline. Finally, it suggests comparing each fitted spline to the average by an appropriate measure of distance such as the ones presented above.

Gardner et al. [1997] use the same idea to monitor wafer surfaces. They use thin-plate spline fitting on wafer surfaces. To estimate an average spline surface, the authors suggest to average all the individual spline estimates or predictions evaluated on a dense grid. Subsequently, they calculate a metric that is the sum of all the deviations between each spline fit and the average spline at each location. They evaluate different types of deviation metrics as the ones shown above: squared metric, absolute value metric and a combination of both. They use bootstrapping to determine the distribution of the metrics.

The second and most used approach to calculate a metric for determining how different a given profile is from the IC profile is the use of multivariate statistics which depends on the non-parametric technique used.

PCA has been used to monitor non-linear profiles [Ding et al., 2006, Colosimo and Pacella, 2010]. These techniques estimate the IC profile through the PCA approach, and use a T^2 statistic proposed by Jackson [1991] that is based on the first k principal components (PC's),

$$T_i^2 = \frac{z_{i1}^2}{\lambda_1} + \cdots + \frac{z_{ik}^2}{\lambda_k},$$

where the λ's are the eigenvalues, and $z_{i1} \ldots z_{ik}$ are the scores for the k PC's. Also Jackson

[1991] suggests to use the relationship T^2 has with the F distribution to calculate the upper control limit, i.e.,

$$T_i^2 = \frac{k(n-1)}{n-k} F(n, n-k),$$

where n is the number of observation per profile.

Jackson [1991] also uses the Q statistics (Squared prediction error) based on the first k principal components,

$$Q_i = (\mathbf{y}_i - \hat{\mathbf{y}}_{i(k)})^t (\mathbf{y}_i - \hat{\mathbf{y}}_{i(k)}),$$

where $\mathbf{y}_i$ is the observed data of the ith profile, and $\hat{\mathbf{y}}_{i(k)}$ is the estimated profile based on the k PC's.

To find a threshold for this statistic, Jackson and Mudholkar [1979] and Jackson [1991] suggest the following metric based on the Q statistic:

$$c = \theta_1 \frac{\left[\left(\frac{Q}{\theta_1}\right)^h - \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} - 1\right]}{\sqrt{2\theta_2 h_0^2}},$$

where $\theta_q = \sum_{i=k+1}^{n} \lambda_i^q$ and $h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2}$. This metric is approximately standard normal.

Wavelet analysis is another technique used for modeling profiles. The monitoring scheme uses the wavelets coefficients obtained after applying the discrete wavelet transform (DWT). Since those coefficients can be still a large set, most of the proposed approaches use reduction techniques such as PCA [Jin and Shi, 2001, Jeong et al., 2006]. Subsequently, they apply the T^2 statistic to the reduced set of coefficients, yielding

$$T_i^2 = (\hat{\boldsymbol{\theta}}_i^* - \boldsymbol{\theta}_0^*)^t S^{-1} (\hat{\boldsymbol{\theta}}_i^* - \boldsymbol{\theta}_0^*),$$

where $\hat{\boldsymbol{\theta}}_i^*$ is the vector of the reduced wavelets coefficients for the ith profile, and $\boldsymbol{\theta}_0^*$ is the target reduced wavelet coefficients vector based on Phase I data. Depending on how the variance is estimated, the distribution of the T^2 changes accordingly. For more details, refer to Jin and Shi [2001], Ganesan et al. [2004], and Jeong et al. [2006].

Chang and Yadama [2010] argue that even after applying PCA on wavelet coefficients, high dimensionality may be an issue. They also highlight another major concern which is the loss of location information on the out-of-control signal when trying to reduce dimensionality. To remedy those problems, the authors suggest first to de-noise the profile through wavelet analysis. Once a de-noised profile is obtained, profiles are partitioned into a number of user-defined segments. They then apply B-spline smoothing to each segment in order to estimate its function. For the B-spline fitting, the control points in each segment are selected and used in the monitoring phase. At each segment for a given fitted profile, they calculate the Euclidean distance between the fitted profile and the standard profile (obtained by fitting B-spline on all profiles). As a result, a vector of Euclidean distances (one per segment) for each ith profile is obtained. Subsequently, the authors use a T^2 control chart to monitor those distances as follows,

$$T_i^2 = (\mathbf{z}_i - \bar{\mathbf{z}})^t S^{-1} (\mathbf{z}_i - \bar{\mathbf{z}}),$$

where $\mathbf{z}_i$ is the vector of Euclidean distances of the ith profile, $\bar{\mathbf{z}} = \frac{1}{m}\sum_{i=1}^{m}\mathbf{z}_i$ is the mean vector of Euclidean distances, and S is the covariance matrix of the $\mathbf{z}_i$'s which Chang and Yadama [2010] suggest to estimate through the successive differences estimator shown in Equation (2.2).

Gaussian process (GP) regression has also been applied to describe the within profile correlation. The GP regression model is popular in the machine learning community because of its flexibility. A GP is fully specified by its mean function and its covariance function. The parameters of its mean function are referred to as fixed parameters, and the parameters of its covariance function are commonly called hyper-parameters because they are tuned to the data [Rasmussen and Williams, 2006].

Wang et al. [2014], Zhang et al. [2014], Zhang et al. [2016], and Yongxiang et al. [2018] use GP for quality monitoring in Phase II. Using the data from Phase I, they fit a GP model to the profiles, obtaining the IC fixed and hyper-parameters estimates. Subsequently, they monitor the data obtained from a new profile through T^2-type control charts.

Zhang et al. [2014] propose GP regression for linear profiles where the mean function has a linear form. For estimating the IC fixed parameters and hyper-parameters, they fit a GP model to each profile in Phase I and calculate the average vector of estimates. They use a T^2 statistic chart to monitor the fixed parameters of the linear trend, and a univariate Shewhart moving range-type chart to monitor the shift in correlation among the profiles, i.e.,

$$T_i^2 = (\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_0)^t S_0^{-1} (\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_0)$$
$$M_i = \frac{1}{2}(\mathbf{y}_i - \mathbf{y}_{i-1})^t \Sigma_0^{-1} (\mathbf{y}_i - \mathbf{y}_{i-1}),$$

where $\hat{\boldsymbol{\beta}}_i$ is the vector of the fixed parameter estimates of the linear trend of the ith profile, $\boldsymbol{\beta}_0$ is the IC vector of fixed parameters that can be estimated by averaging over all the vectors of fixed parameter estimates from Phase I, S_0 is the covariance matrix of the IC fixed parameters $\boldsymbol{\beta}_0$, and Σ_0 is the kernel function that describes the correlation between any two points. Since $\hat{\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{\beta}_0, S_0)$, $T_i^2 \sim \chi^2(k)$ where k is the number of fixed parameters. Additionally, the authors argue that since a given profile $\mathbf{y}_i \sim \mathcal{N}(X\boldsymbol{\beta}_0, \Sigma_0)$, then M_i follows a $\chi^2(n)$ where n is the number of observations within the ith profile. One concern about this is that the authors are using the prior distribution of the observed data or data in Phase I D_N where N is the number of observations, not the distribution of a given profile conditioned on the observed data: $\mathbf{y}_i | D_N$.

Wang et al. [2014] assume that the IC parameters are known or estimated using historical data, although they do not specify how exactly it is done. Assuming the IC parameters are known, they propose two multivariate T^2 control charts to monitor the fixed and hyper-parameters respectively. These T^2 control charts are based on the assumption that the parameters follow asymptotically a multivariate normal distribution with the inverse of the

Fisher information matrix as the covariance function, i.e.,

$$\begin{aligned} T_{1i}^2 &= (\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_0)^t I(\boldsymbol{\beta}_0)(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_0) \\ T_{2i}^2 &= (\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_0)^t I(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_0). \end{aligned} \tag{2.3}$$

They use the inverse of the Fisher information matrix to estimate the covariance matrix of the parameters, $I(\boldsymbol{\beta}_0)$, and $I(\boldsymbol{\theta}_0)$. Assuming that $\hat{\Theta}\dot{\sim}\mathcal{N}\big(\Theta, I^{-1}(\Theta)\big)$ where $\Theta = (\boldsymbol{\beta}, \boldsymbol{\theta})^t$, T_i^2 is asymptotically distributed $\chi^2(k)$ where k is the total number of parameters to be estimated. They argue that the MLE of the two groups of parameters are approximately independent if there are enough observations. Subsequently, they suggest that $T_{1i}^2\dot{\sim}\chi^2(k_1)$, and $T_{2i}^2\dot{\sim}\chi^2(k_2)$ where k_1 and k_2 are the number of parameters in each set respectively.

Instead of fitting a GP model to each profile as the two previous methods suggest, Zhang et al. [2016] propose an Additive Gaussian process (AGP) model which uses two independent GP models with different covariance structures. The first one models the standard profile and the second one, which is fitted on the errors, quantifies the deviation from the standard profile. This model is based on the assumption of homoscedasticity with correlated noise. Taking the profiles from Phase I data as one big univariate vector, they estimate the IC parameters of the AGP model. For Phase II, when having a new profile, they test the measurements of the profile (not the parameters) using a T^2 statistic, i.e.,

$$T_i^2 = (\mathbf{y}_i - \tilde{\mu}(\mathbf{x}_i))^t \tilde{\sigma}^2(\mathbf{x}_i)^{-1}(\mathbf{y}_i - \tilde{\mu}(\mathbf{x}_i)), \tag{2.4}$$

where $\tilde{\mu}(\mathbf{x}_i)$ and $\tilde{\sigma}^2(\mathbf{x}_i)$ are the mean and covariance of ith profile given the Phase I data, D_N. Given that $\mathbf{y}_i|D_N \sim \mathcal{N}(\tilde{\mu}(\mathbf{x}_i), \tilde{\sigma}^2(\mathbf{x}_i))$, T_i^2 follows $\chi^2(n)$ where n is the number of the observations of the ith profile.

They also propose a Generalized Likelihood Ratio (GLR) test to specifically test for specific mean shifts, variance changes, and roughness changes. These changes are deviations from the IC parameters and are tested to see if they are significantly different from zero. They proposed the following test of those specific changes,

$$H_0 : \mathbf{y}_i|D_N \sim \mathcal{N}(\tilde{\mu}(\mathbf{x}_i), \tilde{\sigma}^2(\mathbf{x}_i)), \quad H_1 : \mathbf{y}_i|D_N \sim \mathcal{N}(\tilde{\mu}(\mathbf{x}_i) + \delta, \tilde{\sigma}^2(\mathbf{x}_i) + \Sigma_w).$$

Then, they form the GLR based on both likelihoods.

In Chapter 4, we study the use of GP model with heteroscedasticity proposed by Binois et al. [2018] to profile monitoring.

2.3 A note on statistical vs. practical significance

In control charting, the goal is to detect departures from the IC or underlying model which will indicate the presence of an assignable cause so the practitioners can take corrective actions if necessary. There has been a lot of research to study methods that can detect small departures from the IC model as quickly as possible. This high sensitivity to those small departures can become an issue in many industrial processes. This issue has been identified and discussed by the control charting community [Woodall and Faltin, 1996]. Small assignable causes may be inherent in the process and their removal can become either impossible, impractical or expensive. Thus, as the authors discuss, the control charts need to be "desensitized".

Woodall and Faltin [2019] point out that the current approaches of evaluating and comparing methods in control charting have high sensitivity in detecting small shifts that are not necessarily of practical significance. The authors encourage researchers to have methods whose signals are also associated with practical significance. They discuss the usefulness of a three region approach to deal with this sensitivity in classical control charts in which three zones of the parameter space are defined as in-control, "indifference", and out-of-control. The authors suggest to determine these zones based on engineering knowledge.

As seen in this chapter, most of the current parametric or non-parametric techniques focus on the estimated parameters or features of the model instead of the quality characteristic of the process. Through a multivariate statistic, their focus is on the signal detection of those parameters or features of the model being monitored. Often those can signal unimportant changes on the quality characteristic observed. Since those signals are based on the model parameters, we may often have signals for processes far from the specification limits which are set based on the tolerance given by the engineering requirements of the process or product, not by the model parameters. Additionally, it may become economically impractical if we try to remove these assignable causes. We recall Shewhart [1930] defines a product or process to be in-control within limits to which we "*expect to go economically in finding and removing* *causes of variability without changing a major portion of the manufacturing process*".

In this dissertation, we explore and study two procedures to deal with this issue. Chapters 3 and 4 study the use of a Shewhart-type control chart to monitor the quality characteristic directly instead of the parameters or features of the model. This procedure is based on a bandwidth which is defined to monitor the quality characteristic itself. In Chapter 5 we study a χ^2 statistic that is defined based on a model with the response as the actual quality characteristic observed, not the estimated parameters.

Chapter 3

Within Profile Monitoring Applied to Shrimp Farming

Each of these current profile monitoring approaches are based on the implicit assumption that the process produces the profile in such a short period of time that a person monitoring the process cannot make process corrections during the observance of this profile. In essence, these approaches are retrospective in terms of each profile observed. While this assumption is reasonable for a large number of applications, there are exceptions where the person monitoring the profile actually could intervene earlier and attempt to correct the process. We focus on the situation in which the production engineer must make these decisions as the current profile is being formed.

For example, many agricultural situations are prime candidates for profile monitoring where the basic profile represents the entire growing season. Those monitoring the process are actively engaged throughout the growing season and take corrective actions when they are needed. A proper profile monitoring scheme would allow these people to interact with their process in a more intelligent, efficient, and effective manner. This chapter illustrates how to approach this problem within the context of an actual agricultural process: shrimp farming in Peru.

We first provide some background on Peruvian shrimp farming, and outline a reasonable non-linear model that describes the basic shrimp growth and how to estimate a critical parameter that changes from growing season (a campaign) to growing season. In Section 3.2, we discuss the data cleansing required for the Phase I estimates of the control chart parameters. Section 3.3 outlines the current state of the Phase I analysis using a constant variance approach. In the Section 3.4, the results after fitting a heteroscedastic non-linear model are studied. Section 3.5 applies the Phase I results to the most recent campaign and discusses the updating of the control chart parameters as a result. The final section provides conclusions and potential next steps.

3.1 Background on Peruvian shrimp farming

The recent global increase in shrimp demand has driven the development of shrimp farming in northern Peru. Shrimp growth is very sensitive to changes in certain process variables. As a result, proper production requires good process monitoring. Currently, Peruvian shrimp farms use very simple methods to monitor these variables.

Typical shrimp farms have many growing ponds (more than 50) to be monitored in parallel. The data are collected during the day and at the end of the day. This information is interpreted by the production engineer, who then implements corrective actions if needed. The problem with this approach is that sometimes the corrective action occurs after the growth process has already been adversely affected. Management policies exacerbate this problem by requiring the production manager and engineer to monitor every pond closely. Proper statistical process monitoring that highlights the problem ponds would allow the farm management to concentrate its efforts where they are most needed.

Historically, there is very little need for active process monitoring for the first few weeks of the campaign. One of the company's protocols is to monitor the ponds closely once the average shrimp weight is 5 gr., which historically occurs around week five. Thus, the first four weeks provide a window that can allow the monitoring procedure to adapt itself to any unique conditions in the specific campaign. Since the shrimp farm wants to ensure the expected natural growth of the shrimp, our objective is to determine a baseline growth function that reflects this natural behavior in all the ponds. We use historical data to model the behavior of the log weight of the shrimp. For this study, we work with data from five harvesting campaigns between 2011 and 2016. The number of ponds varies in each campaign depending on production targets, but some campaigns involve up to 380 ponds.

3.1.1 Non-linear model describing shrimp growth

A reasonable basis for the profile of shrimp growth is a Gompertz model. The Gompertz model has been used to study the growth of animals, see Wright [1926], Medawar [1940], Seber and Wild [2003], and Bates and Watts [2007], and has been used to model the growth of shrimp in Tian et al. [1993], Yu et al. [2006], and Araneda et al. [2013]. Empirical analysis of the data from the specific farm used in this study confirms that such an approach is very reasonable. Let w_{ij} be the observed average weight of the shrimp for the i^{th} pond, and let t_{ij} be the time at week j from when the shrimp enter the pond. One form for this model is

$$w_{ij} = \phi_1 \,\mathbf{exp}\left[-\mathbf{exp}\left(\phi_2 - \phi_3 t_{ij}\right)\right].$$

An alternative form of the model focuses on $y_{ij} = \mathbf{ln}\; w_{ij}$, which produces the von Bertalanffy equation [Seber and Wild, 2003]. We thus may re-write the basic Gompertz model as

$$y_{ij} = \theta_1 - \theta_2 \,\mathbf{exp}\,(-\theta_3 t_{ij}). \tag{3.1}$$

where $\theta_1 = \mathbf{ln}\; \phi_1$, $\theta_2 = \mathbf{exp}\; \phi_2$, and $\theta_3 = \phi_3$. The primary benefit of the re-parameterized model in the log weight is the interpretation of the parameters: (i) θ_1 is the asymptotic log weight of the shrimp, (ii) $\theta_1 - \theta_2$ is the initial log weight of the shrimp, and (iii) θ_3 is the growth rate parameter.

At one level, one could construct a control chart to monitor if the Gompertz parameters change from campaign to campaign using T^2 statistics as described in Chapter 2. However, with it one only evaluates the process at the end of the campaign, when it is too late to correct assignable causes. Even worse, there essentially is only one campaign per year. As a result, a proper Phase I period would require decades of growing seasons.

The important management question is to identify within a campaign which ponds are under-performing in order to apply appropriate corrective actions. A weekly statistical monitoring procedure within a campaign is extremely important for the farm manager to prioritize his time with the ponds. Currently, the production engineer must check every pond each week. We propose a data driven technique that will help prioritizing which ponds need attention. In the longer run, the issue is to persuade the farm owners that a proper statistical monitoring procedure would allow the production engineer to check only the ponds that appear problematic rather than all ponds. The result would be more effective use of the production engineer's time.

For the remainder of this chapter, consider the log weight formulation of the model shown in Equation 3.1. The proper monitoring limits must take into account the uncertainty associated with the estimated parameters. We first consider a retrospective study for a single campaign or batch. Note that this approach does not allow us to monitor within the campaign; however, it does provide insight as to how this could be performed. Let PE_j be the proper prediction error for the predicted log weights for week j. In general, the resulting monitoring limits for the ith profile are

$$\hat{y}_{ij} \pm k \, PE_j, \tag{3.2}$$

where k depends on the sample size and desired coverage.

Note that these limits do depend upon week. Observed ponds below the lower limit require attention. Ponds above the limit indicate larger growth than expected, which may provide an opportunity to learn how to accelerate the shrimp growth, or could indicate that the shrimp are growing too quickly which may be due to higher expenses related to feed input ration.

Active monitoring within a future campaign must use the information from previous campaigns to create monitoring control limits. The process of analyzing historical data for this purpose is referred to as the Phase I portion of statistical process control. In general, a set of historical data is used to assess the process stability and determine if any outlying profiles are present. Once the set of profiles are deemed to be stable, these are then used to generate the limits for the Phase II, or prospective monitoring.

When analyzing the previous campaigns in the shrimp farm example, the data used for Phase I can be broken down into two separate components, the overall shrimp growth for each campaign and the individual pond growth of shrimp within each campaign. Typically campaigns involve more than seventeen weeks and usually more than 200 growing ponds. Clearly, there is more than the typical amount of information present in standard Phase

I studies as it pertains to the shrimp growth within each campaign. The number of past campaigns, however, is limited.

Using a small number of previous campaigns does pose some risk in the monitoring procedure. First, the assumption that the Gompertz parameters do not change over time can not be adequately assessed using the data. Fortunately, the interpretability of the Gompertz parameters can help guide our assumptions. Except under extraordinary circumstances, there should be very little, if any, change from campaign to campaign in θ_1, the asymptotic log weight of the shrimp and θ_3, the shrimp growth rate. These two parameters reflect the standard growth of shrimp which is not expected to vary substantially from campaign to campaign. On the other hand, there may be some differences in θ_2, which reflects the initial weight of shrimp as they enter the pond due to farm management practices.

The key point to a practical monitoring procedure is the ability to adapt information from previous campaigns as early as possible to the current campaign. The production engineer needs a useful monitoring process to prioritize decisions about ponds. Clearly, once the campaign is over, the analyst then needs to update the parameter estimates based on the new information for the next campaign.

3.2 The data cleansing prior to the Phase I analysis

The proposed monitoring process uses as many recent campaigns as possible to develop the baseline model. The analyst must estimate the Gompertz parameters within each campaign.

This analysis requires proper data cleansing techniques. Ponds that do not seem to follow the estimated model cannot be dropped from the analysis unless the assignable cause is known. The next step is to examine the stability of the estimated Gompertz parameters across the campaigns. In a similar manner, a campaign cannot be dropped from the analysis unless there is a proper, known assignable cause.

We next fit the model given by Equation (3.1). The starting values for the estimating procedure were $\theta_1 = 3.3$, $\theta_2 = 3.8$ and $\theta_3 = 0.1$. These values were chosen based on the data, taking into account the interpretation of each parameter. There were no problems with convergence in the estimation for any of the campaigns using these starting values.

The residuals obtained after fitting the model were analyzed. The ponds for which the model did not fit as expected were identified for further investigation to see if this particular behavior was due to assignable causes. If so, we removed them from the analysis since our objective at this stage was to find the natural growth function of the shrimp. This decision was performed by a team that included the production engineer. His experience, insights, and detailed data records were vital in this step. With this new set of ponds, we again estimated the non-linear model and studied the residuals. We repeated this process until no assignable causes were identified.

Table 3.1 presents the estimates of the model parameters by campaign (11, 13, 14, 15, and 16). As seen, campaign 13's parameter estimates are different from the rest of the campaigns indicating that there is a different growth behavior. Upon further investigation we learned that this campaign was anomalous because the shrimp were affected by white spot syndrome virus. They had to be placed in the growing ponds much younger than usual resulting in a different growth behavior.

Table 3.1: Parameter estimates from non-linear model assuming constant variance.

Campaign	11	13	14	15	16	$\bar{\bar{\theta}}$ excluding campaign 13	$S.D.(\bar{\bar{\theta}})$ excluding campaign 13
$\hat{\theta_1}$	3.4200	2.8352	3.4078	3.4247	3.3024	3.3887	0.0580
$\hat{\theta_2}$	3.3053	4.2377	3.4364	3.3781	3.1652	3.3212	0.1171
$\hat{\theta_3}$	0.1225	0.1609	0.1220	0.1127	0.1219	0.1198	0.0047

3.3 Non-linear model assuming constant variance

Table 3.1 indicates that, except for campaign 13, the parameter estimates for θ_1 and θ_3 are very close to each other and do not display as much variation as the estimates for θ_2. As noted previously, it is reasonable to expect that the asymptotic shrimp log weight, θ_1, and the growth rate, θ_3, should not vary much from campaign to campaign since these parameters are related to the standard growth of the shrimp under the farm's standard operating protocols. The averages excluding campaign 13 (shown in column 7) should be good estimates for the baseline growth function.

On the contrary, the estimates for θ_2 vary significantly across campaigns. Such variation seems plausible given that this parameter is directly related to the initial log weight of the shrimp. There are some factors inherent to each year that affect this initial weight such as the shrimp larvae suppliers, batch-to-batch variation, atmospheric temperature, and management practices which affect the growth process of younger shrimp before entering the growing ponds. Since we have just four campaigns, it is not appropriate to use a non-linear mixed model where the parameter θ_2 can be modeled as a random effect [Jensen and Birch, 2009].

For doing prospective monitoring within the campaign, the issue then becomes how to estimate θ_2 within the current campaign. It is important to recall that process monitoring is not critical until the fifth week, which suggests that we can use the log weights through week four to estimate this parameter, fixing θ_1 and θ_3 to their average values from the Phase I period. Table 3.2 compares the estimates of θ_2 from the non-linear regression model with constant variance and the estimates from the NL model based on first four weeks of each campaign. The estimates that are quite close suggesting that it is reasonable, for the purposes of monitoring within a campaign, to use the estimate based on the log weight through week four.

Table 3.2: Estimates for θ_2 after fitting the model using fixed $\bar{\hat{\theta}}_1$ and $\bar{\hat{\theta}}_3$. Top row shows the estimates when using the entire data of each campaign, and bottom row shows the estimates when using the first four weeks of data for each campaign.

	11	14	15	16
$\hat{\theta}_2$	3.3053	3.4364	3.3781	3.1652
$\hat{\theta}_2 \vert y_{1-4}$	3.2942	3.3907	3.3887	3.2609

3.3.1 Limits for prospective monitoring

From previous analysis we fit a model in which the estimates of θ_1 and θ_3 are the averages of the estimates for each campaign, and the estimate of θ_2 is based on the first four weeks of that campaign. Using this model as the central line of our control chart, we next need to construct an appropriate prediction error using equation (3.2) as the basis for the proposed monitoring procedure.

Assuming constant variance, a reasonable approach is to use an average measure of the residual variability across Phase I. Such an average involves a very large number of ponds, which implies that we can use an appropriate average RMSR. The correct average is a weighted average of the sum of mean squared residuals from each campaign, with the number of degrees of freedom for each sum of squared residuals as the weights. This estimate is 0.212618.

We use this baseline function and limits retrospectively to observe its performance. See Figure 3.1. These limits do not work properly during the first weeks nor the last weeks of the campaign periods. At the beginning of monitoring there are points falling outside the limits, and at the end of monitoring the limits appear to be too wide.

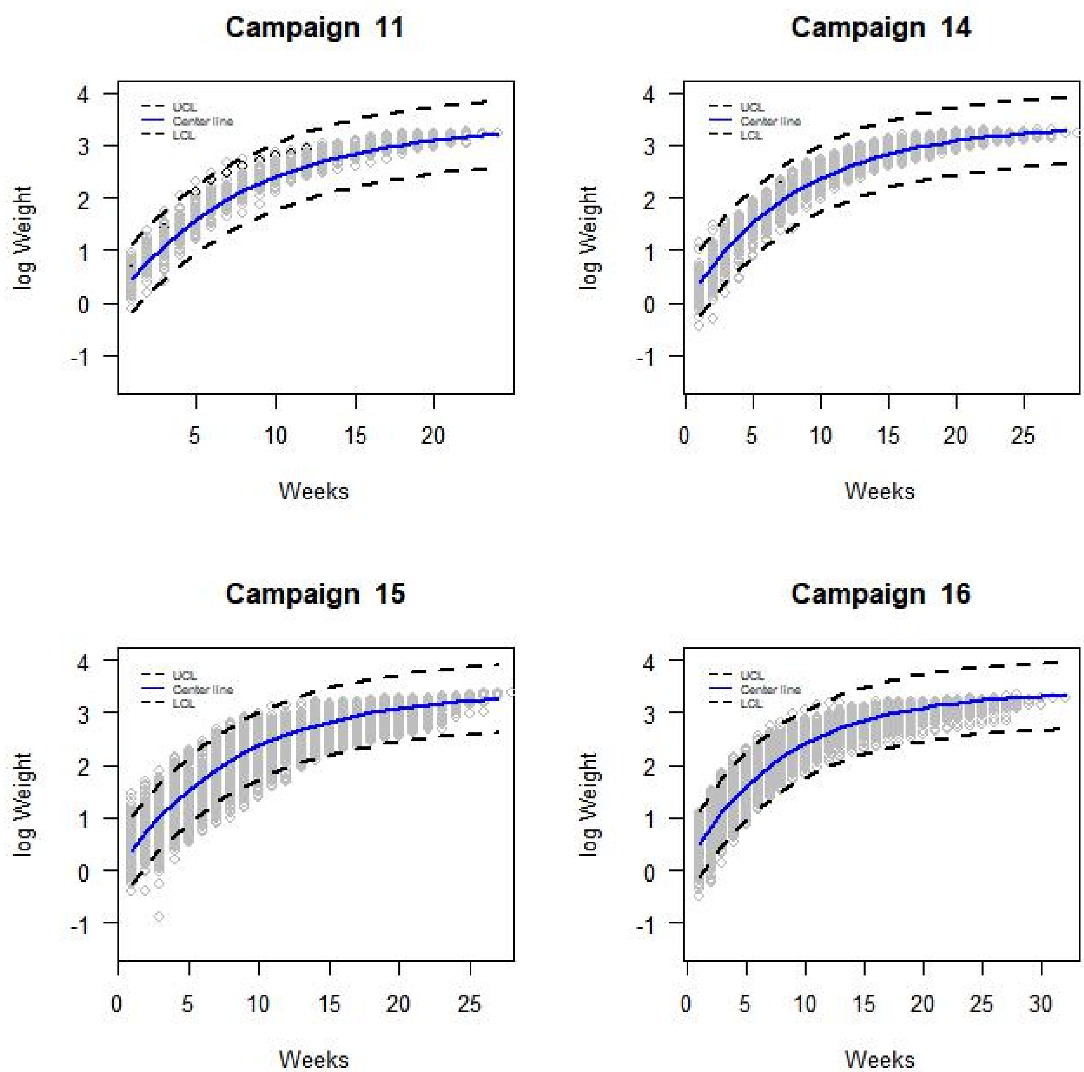

Figure 3.1: Control chart for campaigns 11, 14, 15 and 16 with constant limits when using non-linear model with constant variance.

3.4 Heteroscedastic non-linear model

From the previous analysis, we observe that the homoscedasticity assumption in the non-linear model proposed before is violated since the variance seems to depend on time. We can write our regression model with heteroscedasticity for a given campaign as follows,

$$E[y_{ij}] = f(t_{ij}, \boldsymbol{\theta}); \quad Var[y_{ij}] = \sigma_j^2, \tag{3.3}$$

where $f(\cdot)$ is the Gompertz model shown in Equation 3.1, and σ_j^2 is the variance at week j for a given campaign.

The literature suggests different parametric and non-parametric approaches that can be used to estimate a non-constant variance function. Davidian and Carroll [1987] suggested $\sigma_j^2 = \sigma^2 g^2(x_{ij}, \boldsymbol{\theta}, \beta)$, where σ^2 is a constant, and $g^2(\cdot)$ is a function that depends on the covariates x_{ij}, parameter vector $\boldsymbol{\theta}$ from the mean function, and another parameter β. Bellio et al. [2000] suggested the use of the "power of X" variance function, $\sigma_j^2 = \alpha t_{ij}^{\beta}$, which is independent of $\boldsymbol{\theta}$ parameters, and hence $\hat{\boldsymbol{\theta}}$ and $\hat{\beta}$ are asymptotically independent. The "power of X" variance function can also be transformed to $log(\sigma_j^2) = log(\alpha) + \beta log(t_{ij})$. In this case, the logarithm of the variance at week j is used as the dependent response for the variance linear model. This is consistent with what is suggested by the literature in the response surface methodology when modeling the variance [Vining and Myers, 1990]. Note that in this case, the errors are additive.

We studied two variance function models. One uses the log of the sample variance of the residuals from the Gompertz model fit as responses, and the second one uses the log of the sample standard deviation of those residuals. We show here the latter one since it satisfies the normality assumption best as shown later. The variance function we used is $g^2(\cdot)$, where

$$g(t_{ij}, \hat{\beta}) = log(\hat{\sigma}_j) = log(s_j) = \hat{\beta}_1 + \hat{\beta}_2 log(t_{ij}). \tag{3.4}$$

We used the generalized least squares (GLS) technique for estimating the parameters of the mean function $f(\cdot)$ and the variance function $g^2(\cdot)$. First, we estimated $\boldsymbol{\theta}$ using the unweighted nonlinear least squares to fit the mean function $f(\cdot)$. Then, based on the residuals, ordinary least squares was used to estimate the parameters $\boldsymbol{\beta}$ of the variance function, and we used those to update the estimates of $\boldsymbol{\theta}$ using weighted nonlinear least squares. This process was repeated until convergence.

To evaluate the model in Equation 3.3 with the variance function shown above, we performed Phase I analysis using the historical data. Table 3.3 shows the GLS parameters for the mean and variance functions. First, we observe that the parameters from $f(\cdot)$ are slightly different with respect to the estimated ones with the constant variance model. We also note that the β_2 estimates of the $g(\cdot)$ function are not that different from each other with respect to the campaign. However, the estimates for β_1 are quite different among the campaigns.

It is important to recall that the objective of this study is to actively monitor the shrimp growth within the campaign. Since θ_2 and β_1 estimates seem to be quite variable and depend on the data from the same campaign, we applied the same approach used in the constant variance model explained in section 5.1. Using the first four weeks of data from each campaign, we estimated θ_2 and β_1 by fixing θ_1, θ_3 and β_2 to their average values from the Phase I period. Table 3.4 shows the estimates for θ_2 and β_1 for both functions. The

θ_2 values are very close to the ones when using the entire data from each campaign shown in Table 3.3. This result suggests that it is reasonable to use them for monitoring the log weights within the campaign after week four. Tables 3.3 and 3.4 also show the RMSR after fitting the heteroscedastic non-linear model for each campaign when using all data and just the first four weeks of data, respectively.

Table 3.3: Generalized least squares estimates for parameters in mean function $f(\cdot)$ and variance function $g(\cdot)$ using all data from each campaign. Bottom row shows the weighted RMSR obtained after fitting the model.

		11	14	15	16
$f(t_i, \Theta)$	$\hat{\theta}_1$	3.3599	3.3981	3.3844	3.2869
	$\hat{\theta}_2$	3.3535	3.4458	3.4065	3.1813
	$\hat{\theta}_3$	0.1320	0.1235	0.1182	0.1245
$g(t_i, \beta)$	$\hat{\beta}_1$	-0.5018	-0.9233	-0.2116	-0.6301
	$\hat{\beta}_2$	-0.7751	-0.5611	-0.6632	-0.5523
RMSR		1.2752	1.1149	1.3168	1.0745

Table 3.4: Top rows show the GLS estimate for θ_2 from the mean model with fixed $\bar{\hat{\theta}}_1$ and $\bar{\hat{\theta}}_3$, and the GLS β_1 estimate from the variance model with fixed $\bar{\hat{\beta}}_2$, using data through week four. Bottom row shows the RMSR obtained after fitting this model for each campaign.

		11	14	15	16
$f(\cdot)$	$\hat{\theta}_2 \vert y_{1-4}$	3.2891	3.3825	3.4198	3.2706
$g(\cdot)$	$\hat{\beta}_1 \vert y_{1-4}$	-0.9270	-0.9912	-0.6239	-0.7015
RMSR		1.5199	1.4622	1.8832	1.5071

Figure 3.2 shows the unweighted residuals from the mean model $f(\cdot)$, using the first weeks of data to estimate θ_2 and β_1. Figures 3.3 shows the residuals from the variance model $g^2(\cdot)$. We can see that the variance function $g^2(\cdot)$ seem to capture the variability within each campaign.

An adaptative process for updating the GLS estimates for θ_2 and β_1 as soon as more data become available was also analyzed. This update can be done weekly from week 5 up to the the last one of the period within the campaign. We would have a $\hat{\theta}_2^{(i)}$ for week $i = \{5, ...T\}$ for campaign j. This adaptative procedure, not shown here, gave slightly better results.

3.4.1 Limits for prospective monitoring for the heteroscedastic model

For a new campaign at a given t_0, we can predict y_0, the log shrimp weight, for that particular week using $f(t_0, \hat{\boldsymbol{\theta}})$ where $\hat{\boldsymbol{\theta}}$ are the GLS estimates. Using the variance model from Equation 3.4, we estimate the variance $\hat{\sigma}_0^2$ for that prediction. Davidian and Carroll [1987] suggested the following approximate $(1-\alpha)100\%$ prediction interval, which Bellio et al. [2000] called a naive prediction interval:

$$f(t_0, \hat{\boldsymbol{\theta}}) \pm t_{(1-\alpha/2),N-p}\sqrt{g^2(t_0, \hat{\beta})}, \tag{3.5}$$

where $N = \sum_{i=1}^{m} n_i$ is the total number of observations for the new campaign with m profiles, and p is the number of parameters estimated in $f(\cdot)$. Davidian et al. [1988], and Carroll and Ruppert [1988] suggest the following $(1-\alpha)100\%$ prediction interval which takes into account the estimate of the mean function variance:

$$f(t_0, \hat{\boldsymbol{\theta}}) \pm t_{(1-\alpha/2),N-p}\sqrt{g^2(t_0, \hat{\beta}) + v\hat{a}r(f(t_0, \hat{\boldsymbol{\theta}}))}, \tag{3.6}$$

where, as shown in Carroll and Ruppert [1988],

$$v\hat{a}r(f(t_0, \hat{\boldsymbol{\theta}}) = N^{-1} f'_{\boldsymbol{\theta}}(t_0, \hat{\boldsymbol{\theta}})^t S_G^{-1} f'_{\boldsymbol{\theta}}(t_0, \hat{\boldsymbol{\theta}})$$

$$S_G = N^{-1} \sum_{i=1}^{m} \sum_{j=1}^{n_i} \frac{f'_{\boldsymbol{\theta}}(t_{ij}, \hat{\boldsymbol{\theta}}) f'_{\boldsymbol{\theta}}(t_{ij}, \hat{\boldsymbol{\theta}})^t}{g(t_{ij}, \hat{\beta})},$$

where $f'_{\boldsymbol{\theta}}(t, \hat{\boldsymbol{\theta}})$ is the derivative of $f(\cdot)$ with respect to $\boldsymbol{\theta}$ evaluated at $\hat{\boldsymbol{\theta}}$.

As the authors explained, when the size of the data is large, the error in the prediction is driven by the variance function $g^2(\cdot)$ since $v\hat{a}r\big(f(t_0, \hat{\boldsymbol{\theta}})\big)$ becomes very small. In our analysis, we used both prediction bands as shown in Equations 3.5 and 3.6, and they show almost no difference since the values of $v\hat{a}r\big(f(t_0, \hat{\boldsymbol{\theta}})\big)$ are very small.

Table 3.5 shows the number of out-of-control ponds per week for each campaign when applying retrospective monitoring (when using entire data) and prospective monitoring (when using first four weeks of data). Assuming that the heteroscedastic non-linear model using the entire data is adequate, we observe that our approach for monitoring within the campaign signals more out-of-control ponds especially at the boundaries.

Figures 3.4 and 3.5 show the mean profile and the 99% prediction band taking into account the heteroscedasticity of the non-linear model, using the entire data and first four weeks of data for the estimation respectively. We observe that the predictive band when using a variance function model captures the variability of all campaigns much better.

Table 3.5: A comparison between retrospective and prospective monitoring with respect to the number of out-of-control ponds for each week when using heteroscedastic non-linear model. For each case, the out-of-control ponds with values above the upper control limit (UCL) and below the lower control limit (LCL) are shown.

		Weeks															
		5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Campaign 11																	
Retrospective	$> UCL$	1	2	2	2	2	2	1	1	1	0	0	0	0	0	0	0
	$< LCL$	0	0	0	1	1	1	0	2	1	1	1	1	1	1	0	0
Prospective	$> UCL$	3	3	3	3	4	4	3	4	6	6	1	3	2	1	0	0
	$< LCL$	0	0	2	1	1	1	0	0	1	1	1	1	1	0	0	0
Campaign 14																	
Retrospective	$> UCL$	0	0	0	1	3	1	1	1	0	2	0	1	0	0	0	0
	$< LCL$	3	4	3	1	2	2	1	1	1	1	0	1	2	1	1	0
Prospective	$> UCL$	1	1	2	6	8	8	9	12	10	13	13	10	8	9	4	4
	$< LCL$	5	11	7	7	3	4	4	5	6	5	8	6	4	4	0	0
Campaign 15																	
Retrospective	$> UCL$	0	0	4	7	8	3	1	1	2	2	0	0	0	0	0	0
	$< LCL$	6	13	16	22	23	22	19	12	12	12	11	11	12	14	15	14
Prospective	$> UCL$	11	16	15	14	16	11	6	3	3	5	4	4	1	0	0	0
	$< LCL$	27	29	33	35	41	36	37	31	30	28	34	37	39	38	37	33
Campaign 16																	
Retrospective	$> UCL$	4	5	5	3	4	3	3	2	1	0	0	0	0	0	0	0
	$< LCL$	2	1	4	5	4	2	3	3	3	3	3	4	3	3	3	2
Prospective	$> UCL$	6	10	9	8	9	8	7	8	6	3	1	1	1	0	0	0
	$< LCL$	8	12	12	13	14	16	16	19	18	17	19	20	23	21	16	12

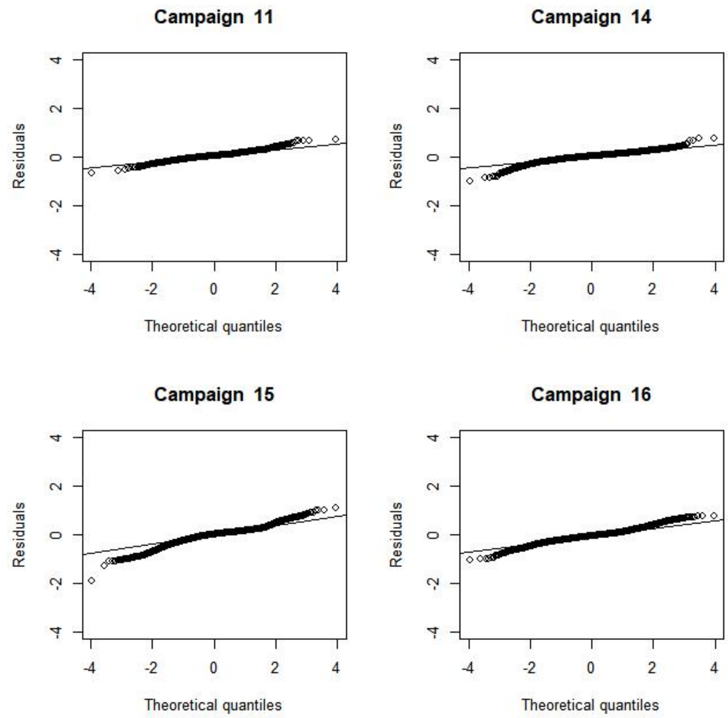

Figure 3.2: Q-Q plots of the residuals obtained after fitting the mean model, using the first four weeks of data from each campaign.

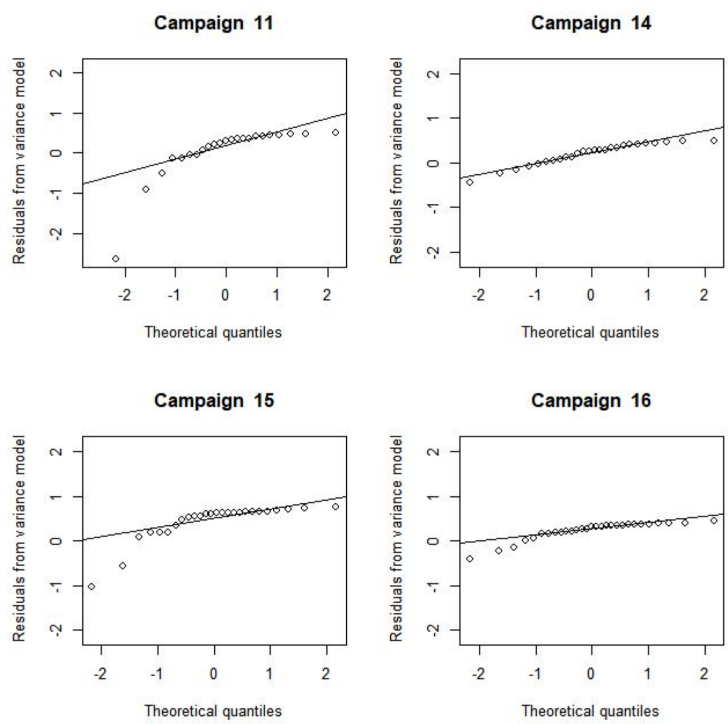

Figure 3.3: Q-Q plots of the residuals obtained after fitting the variance model $g^2(\cdot)$, using the first four weeks of data from each campaign.

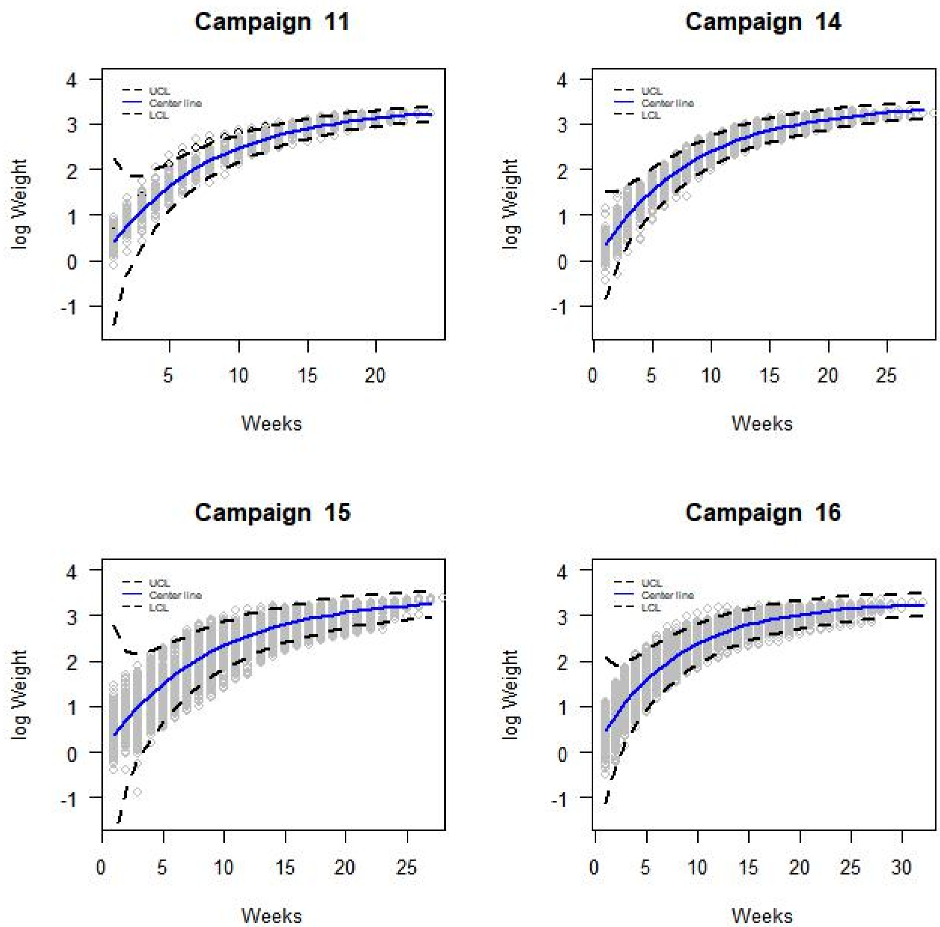

Figure 3.4: Control chart for the mean function and prediction band with heteroscedacity for campaigns 11, 14, 15 and 16, using the entire data for each campaign (retrospective monitoring).

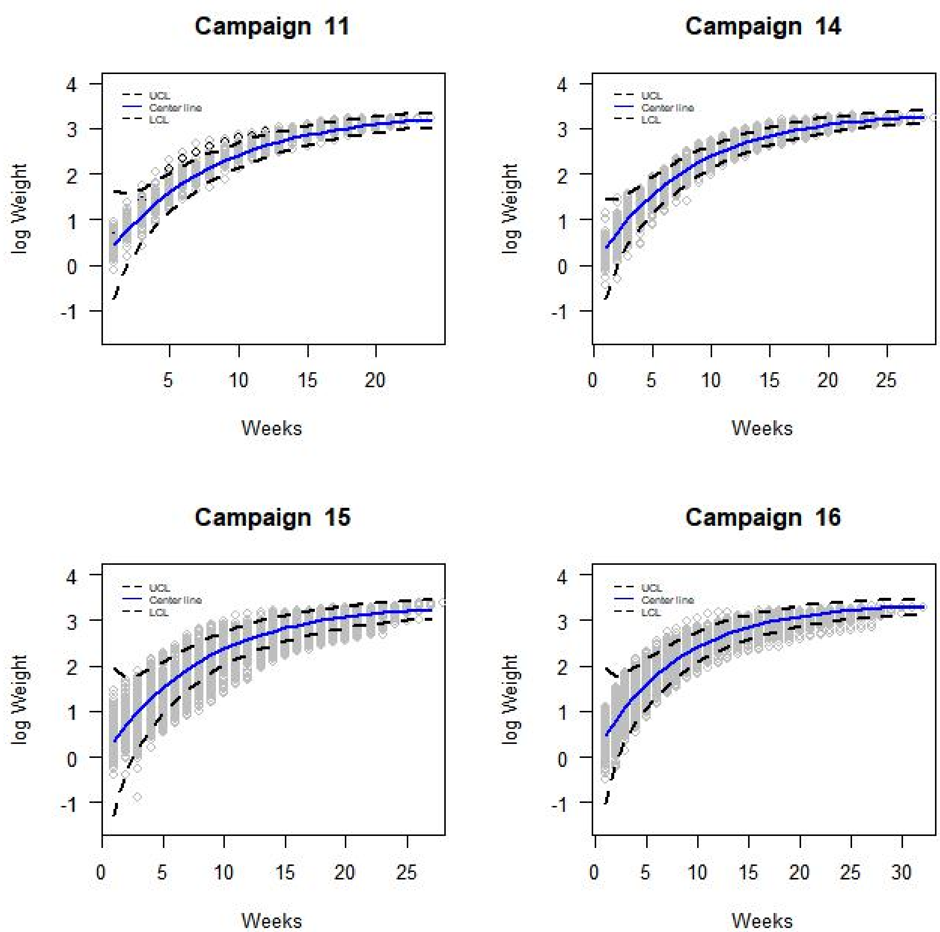

Figure 3.5: Control chart for the mean function and prediction band with heteroscedacity for campaigns 11, 14, 15 and 16, using the first four weeks of data for each campaign (prospective monitoring).

3.5 Phase II application to most recent campaign and future monitoring

To illustrate the proposed method with the baseline function and limits described above, we used the data from campaign 17. The data are for 372 ponds but the period is shorter than the rest of the campaigns. While the other campaigns lasted more than 25 weeks, most of the ponds in campaign 17 lasted between 18 and 20 weeks. This is because some of the shrimp got the white spot virus and to avoid losing the entire campaign, management decided to harvest the shrimp much younger than usual.

Figure 3.6 shows a control chart applying the proposed method. We notice that the baseline function describes the pattern of growth very well. As expected, we see that the shrimp did not reach the asymptote since they were harvested much younger. Clearly the heteroscedastic limits are better for capturing the variability within the campaign.

As seen, the number of batches used in this study to estimate the process parameters was limited so it is important to continue the Phase I process by updating such parameters as soon as more batches are available. As Jensen et al. [2006] point out after a thorough literature review on the effects of parameter estimation on control chart properties, "as more data become available and the process is still determined to be stable, the control limits should be updated". Quevedo et al. [2016] show the application of an iterative procedure on how to update the control limits when more data is obtained and the process is still unstable. Recent work in Goedhart et al. [2017] provides more insight as to how many

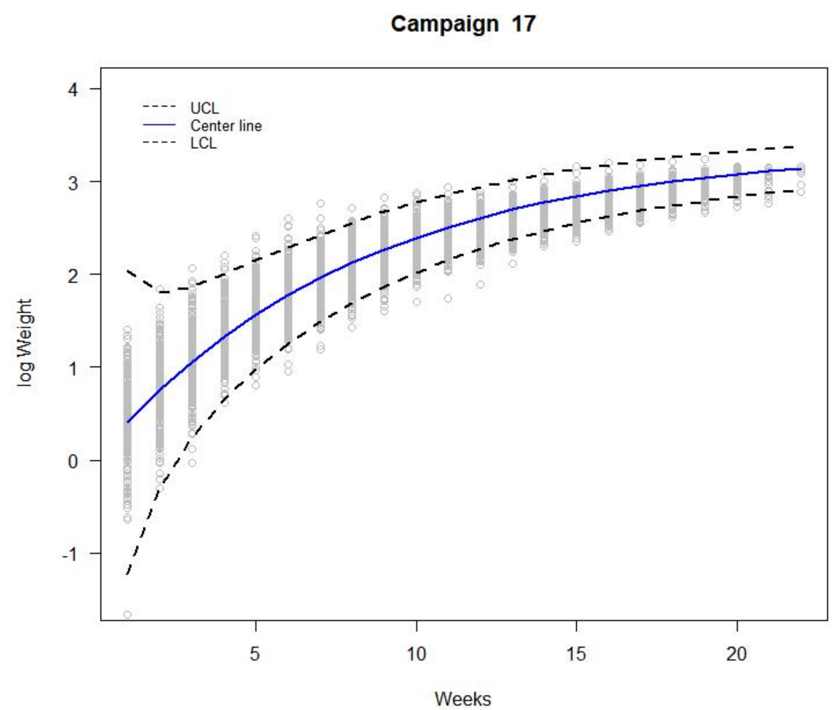

Figure 3.6: Control chart for prospective monitoring for campaign 17 with 95% prediction heteroscedastic limits.

rational subgroups are required to give good Phase II properties.

For this study, we recommend to keep updating the parameter estimates for prospective monitoring. Table 3.6 shows the updated estimates using the non-linear model with heteroscedasticity after campaign 17 ended. Following our proposed procedure, the averages of the GLS estimates within each of the five campaigns (including the recent one) are the updated estimates for θ_1 and θ_3 for the mean model and β_2 for the variance model. These will be used for the next campaign, while for estimating θ_2 for the mean model and β_1 for the variance model, data of the next campaign through week 4 is needed as described in previous sections.

Table 3.6: *GLS* estimates for parameters θ_1, θ_3 from mean function $f(\cdot)$, and β_2 from the $g(\cdot)$ function using historical data. Last column shows the averages of those estimates to be used as fixed parameters in campaign 18.

		11	14	15	16	17	Averages
$f(\cdot)$	$\hat{\theta}_1$	3.3599	3.3981	3.3844	3.2869	3.1779	3.3215
	$\hat{\theta}_3$	0.1320	0.1235	0.1182	0.1245	0.1328	0.1262
$g(\cdot)$	$\hat{\beta}_2$	-0.7751	-0.5611	-0.6632	-0.5523	-0.5044	-0.6112

It is especially encouraging that the shrimp farm's management has endorsed this approach for future campaigns. They see the value of being able to monitor the campaign and to take appropriate collective actions within the specific season.

3.6 Summary and conclusions

There are many cases where the processing time of the batch is sufficiently long that the production engineer can take corrective actions within the batch. In such situations the Shewhart chart on the residuals provides a powerful intuitive basis for controlling the process. The production engineer seeks out the assignable cause once the profile crosses the prediction bands. If she/he can determine the cause quickly enough, he/she can take appropriate actions, which in many cases prevents the production of an out-of-specification batch.

This chapter illustrates such a monitoring procedure for a shrimp farming process in Peru. Each year's campaign represents a batch, and the weekly average weights of the shrimp from each pond are the individual observations. A three-parameter Gompertz model provides the basis for describing the shrimp growth over the campaign. In this study, two of the three parameters are very stable over time; however, the parameter associated with the initial weight of the shrimp is not. This chapter demonstrates an approach that provides a reasonable estimate of this parameter soon enough in the campaign to be valuable.

It is important to note that in this situation, a rigid focus on maintaining specified average run lengths is less important than concentrating the production engineer's efforts to the critical few ponds that truly need attention. The proper goal is to produce a reasonable procedure that the engineer can use within the shortest time frame possible. The proposed procedure walks a very tight rope balancing the ability to monitor an important process as soon as possible with the ability to estimate the critical process parameters sufficiently well.

Finally, we note that the process variability changes over time, as with many processes, so we propose a linear model for the process variance to correct for this problem. This research is part of a project that is developing an application with an interface that allows the producers to collect data from the farms. An important proposal is to embed the proper code into this application to allow the generation of active, real-time control charts based on this research.

Chapter 4

Monitoring Profiles Using a Gaussian Process Model with Heteroscedasticity

The literature review in Chapter 2 shows that Gaussian process (GP) regression has also been applied to profile monitoring in Phase II [Wang et al., 2014, Zhang et al., 2014, 2016, Yongxiang et al., 2018]. Most of these approaches use either Hotelling's T^2 statistic or an analogue for defining the control charts to monitor the hyper-parameter estimates of the GP models. However, it can take a great deal of time to collect data for individual profiles for some chemical or agro-industry processes; hence, monitoring profiles only after all the data of the most recent profile is obtained is not necessarily useful.

The approach we propose here is similar to Zhang et al. [2016] given that we seek to model a mean or standard profile in addition to quantifying the deviation of a profile from the standard one. The main difference between their approach and ours is that they utilize a constant variance assumption. Additionally, as we will discuss later, their approach requires all observations from a new profile to determine the stability of the process, while ours does not and can monitor within the profile data collection period.

Qiu and Xiang [2014] propose the use of a CUSUM control chart of the normalized residuals to monitor the longitudinal behavior of patients' cholesterol levels. Even though their approach uses historical cholesterol level data from patients who have not suffered strokes to estimate the mean behavior, each patient is considered a separate process. Thus, the interest is in the monitoring of each individual behavior, not a process that produces different profiles.

To tackle the challenge of monitoring within and between the profiles, we propose using a Shewhart-type control chart to monitor profiles in Phase II where the central line is the mean profile estimated with the GP model with heteroscedasticity using the joint model proposed by Binois et al. [2018], which is based on a latent-variable idea borrowed from machine learning. The prediction band estimated from the GP model defines the normal behavior. For example, a profile with an observation outside the prediction band can be considered different. The data we use for estimating the GP model are assumed to be clean and contamination-free. With this chart, we will be able to monitor the profile as it evolves. We will not monitor the hyper-parameters of the model, which are tuned to the data and as such may generate misleading results. We demonstrate this point later in this chapter.

Multiple testing within the profile is an issue we need to deal with, particularly if the profile has a large number of observations. It is demonstrated in this chapter that when there are small number of observations per profile, Bonferroni correction works well. However, it is well known that for large number of observations the Bonferroni correction will fail. Herewith, we will refer to the profiles with large number of observations as "over-sampled" profiles, a term we borrow from signal processing in which observations are collected at a higher sampling rate than the actual information. We study a couple of approaches to deal with over-sampled profiles, and we show their results later in this chapter.

The remainder of this chapter is organized as follows. Section 4.1 introduces the GP regression model with heteroscedasticity and different approaches to estimate its parameters. In Section 4.2, we describe the statistical approach monitoring using the GP model with non-constant variance. Section 4.3 shows our control chart's performance from the simulation study. A case study is presented in Section 4.4. Finally, Section 4.5 concludes the chapter with a summary and discussion.

4.1 GP model with heteroscedasticity

Consider a dataset $D_N = \{X_N, \mathbf{y}_N\}$ where $X_N \in \mathbb{R}^{N\text{x}p}$ and $\mathbf{y}_N \in \mathbb{R}^N$, and the following regression model, Santner et al. [2003]:

$$y(\mathbf{x}_j) = \sum_{k=1}^{p} f_k(\mathbf{x}_j)\beta_k + \epsilon(\mathbf{x}_j) = f^t(\mathbf{x}_j)\boldsymbol{\beta} + \epsilon(\mathbf{x}_j),$$

where $j = 1 \dots N$, $f_k(\cdot)$ are known regression functions, and $\epsilon(\mathbf{x}_1) \dots \epsilon(\mathbf{x}_N) \sim \mathcal{N}(\mathbf{0}_N, \Sigma_N)$ with Σ_N as the NxN covariance matrix. Hence, $\mathbf{y}_N \sim \mathcal{N}(\boldsymbol{\mu}_N, \Sigma_N)$ where $\mu_j = f^t(\mathbf{x}_j)\boldsymbol{\beta}$. For example, in the linear regression case $\boldsymbol{\mu}_N = X_N\boldsymbol{\beta}$. Often in computer experiments the $\boldsymbol{\mu}_N = \mathbf{0}_N$ because the covariance term controls the actual GP modeling. This study uses this simplification, which implies estimating only the hyper-parameters.

The covariance Σ_N must be a positive definite matrix where each entry $\Sigma(\mathbf{x}_i^t, \mathbf{x}_j^t)$ is a correlation function between a pair of points $\mathbf{x}_i^t$ and $\mathbf{x}_j^t$. In the homoscedasticity case, the covariance is $\Sigma_N = \tau^2 K_N + \sigma_\epsilon^2 I_N$ where τ^2 is the process variance or scale and σ_ϵ^2 is the random noise. In the heteroscedasticity case, we have

$$\Sigma_N = \tau^2 K_N + \Sigma_\epsilon, \tag{4.1}$$

where K_N is the noise-free kernel function whose entries are defined as $k(\mathbf{x}_i^t, \mathbf{x}_j^t)$. A popular choice for this function is defined by the Euclidean space, whose simple form is $k(\mathbf{x}_i^t, \mathbf{x}_j^t) = exp(-||\mathbf{x}_i^t - \mathbf{x}_j^t||^2/\theta)$ where θ is called *isotropic* length-scale because the correlation decays radially. To generalize it we have a p-dimensional *separable Gaussian* $k(\mathbf{x}_i^t, \mathbf{x}_j^t) = exp(-\sum_{k=1}^{p}(x_{ki} - x_{kj})^2/\theta_k)$ where each variable is given a different parameter. The Σ_ϵ is the noise level matrix that can be expressed as $\Sigma_\epsilon = diag(r(\mathbf{x}_1), \dots r(\mathbf{x}_N))$. The conditional distribution of a new observation $y(\mathbf{x})$ given the IC data is $y(\mathbf{x})|D_N \sim \mathcal{N}(\mu(\mathbf{x}), \sigma^2(\mathbf{x}))$, where

$$\mu(\mathbf{x}) = k(\mathbf{x}, X_N)(K_N + \frac{1}{\tau^2}\Sigma_\epsilon)^{-1}\mathbf{y}_N \tag{4.2}$$
$$\sigma^2(\mathbf{x}) = \tau^2\big(k(\mathbf{x},\mathbf{x}) + \frac{1}{\tau^2}r(\mathbf{x}) - k(\mathbf{x}, X_N)(K_N + \frac{1}{\tau^2}\Sigma_\epsilon)^{-1}k^t(\mathbf{x}, X_N)\big),$$

and where $\mu(\mathbf{x})$ and $\sigma^2(\mathbf{x})$ are the predictive mean and variance respectively and can be used to predict a new observation.

4.1.1 In-control parameter estimation

We use data from the Phase I to estimate the scale, length-scale and varying noise level parameters of the covariance of the GP model shown previously, and we will refer to these as the *IC parameter estimates*. The estimation of the parameters from a GP model with heteroscedasticity is not trivial. A description of a few approaches to perform this estimation follows.

Stochastic kriging

Ankenman et al. [2010] presented the idea of using non-constant variance in the structure of a GP model. When there are replications, a simple solution is to estimate the noise level at each unique point empirically. With this noise level estimated at each unique point, the $\hat{\Sigma}_\epsilon$ is formed and substituted into the log-likelihood to estimate the hyper-parameters of the GP model through maximum likelihood estimation (MLE). A few approaches for estimating the noise level at each point are described below.

Empirical noise level

When there are replications at n unique points, we can utilize n-unique predictive equations (instead of the total number of observations N) and hence, an empirical nxn noise matrix, $\hat{\Sigma}_{\epsilon_n}$. To estimate this matrix, we have $\hat{\Sigma}_{\epsilon_n} = diag(\hat{r}(\mathbf{x}_i), \ldots \hat{r}(\mathbf{x}_n)) = diag(\frac{\hat{\sigma}^2_{\epsilon_1}}{m_1}, \ldots \frac{\hat{\sigma}^2_{\epsilon_n}}{m_n})$ where m_j is the number of observations at the jth unique point. Ankenman et al. [2010] propose estimating $\hat{\sigma}^2_{\epsilon_j}$ at each unique point empirically by $\hat{\sigma}^2_{\epsilon_j} = \frac{1}{m_j - 1}\Sigma_{i=1}^{m_j}(y_i(\mathbf{x}_j) - \bar{y}(\mathbf{x}_j))^2$.

The authors show that $\hat{\Sigma}_{\epsilon_n}$ does not imply a prediction bias as long as the number of replicates is not too small. They mention that it is difficult to make general recommendations for the number of replications required given that to minimize prediction bias, both the number of unique points and the number of replications must be considered. The authors mention that the number of replications at each unique point should no be less than 10. If the intrinsic variance is low, it is good to have more unique points to fill the space. On the contrary, if the intrinsic variance is high, the number of unique points should not be large.

Known variance function

If we have reasons to believe that the variability depends on the predictors, we can fit a parametric model to the noise level at each unique point. This is not uncommon in chemical, pharmacokinetic, or fisheries processes [Carroll and Ruppert, 1988]. As Davidian and Carroll [1987] suggest, we could set the variance as $\sigma^2_{\epsilon_j} = \sigma^2 g^2(\mathbf{x}_j, \boldsymbol{\theta})$, where σ^2 is a constant and $g^2(\cdot)$ is a function that depends on the covariates and the parameter vector $\boldsymbol{\theta}$. When modeling the variance, it is common to use the $log(\sigma^2_{\epsilon_j})$, or $log(\sigma_{\epsilon_j})$ [Vining and Myers, 1990]. Assuming a variance function $z(\mathbf{x}_j) = log(\hat{\sigma}_{\epsilon_j})$, then $\hat{r}(\mathbf{x}_j) = \{exp(\hat{z}(\mathbf{x}_j))\}^2$ can be used to form the empirical nxn noise matrix: $\hat{\Sigma}_{\epsilon_n} = diag(\hat{\sigma}^2_{\epsilon_1} \ldots \hat{\sigma}^2_{\epsilon_n}) = diag(\hat{r}(\mathbf{x}_1) \ldots \hat{r}(\mathbf{x}_n))$.

Assuming a GP on the noise level

Assuming that the noise follows a smooth function of the input, we could define a second GP model on the noise level as in Golberg et al. [1998]. However, the authors propose a model to estimate all the hyper-parameters and the noise parameters jointly. Here we can use a second GP on the noise level, $z(\mathbf{x}_j) = log(\hat{\sigma}_{\epsilon_j}) \sim \mathcal{N}(\mathbf{0}_n, C_n)$, to then estimate the nxn noise matrix $\hat{\Sigma}_{\epsilon_n} = diag(\hat{r}(\mathbf{x}_1) \ldots \hat{r}(\mathbf{x}_n))$ where $\hat{r}(\mathbf{x}_j) = \{exp(\hat{z}(\mathbf{x}_j))\}^2$. After substituting it into the log-likelihood function, the hyper-parameters can be estimated by MLE.

We can also apply a GP model on the noise level in an iterative procedure, similar to the generalized least squares (GLS) estimation. The first step is to assume that the data have constant variance and estimate a GP model with homoscedasticity. The next step is to calculate the residuals for the predictions. A GP model is then applied to the sample standard deviation of the residuals. This noise level matrix is substituted into the log-likelihood and the hyper-parameters of the model are estimated through weighted least squares. The process is repeated until convergence.

GP model with latent variance

Golberg et al. [1998] suggest the use of a joint model to estimate the parameters of the GP with heteroscedasticity. However, the method proposed requires the use of Markov Chain Monte Carlo (MCMC) to average over the latent variables, which can be computationally expensive. Other studies proposed alternatives to replace the MCMC through maximization [Kersting et al., 2007, Quadrianto et al., 2009, Lázaro-Gredilla and Titsias, 2011], but these still present the same level of complexity in the optimization. Binois et al. [2018] propose an alternative to estimate all parameters jointly without approximation as well. They propose to use the Woodbury formula to shorten the calculations and estimate the parameters, which reduces much of the computational burden.

The authors express the correlation function as $\tau^2 K_N + \Sigma_\epsilon = \tau^2(K_N + \Lambda_N)$, where $\Lambda_N = diag(\lambda_1, \ldots \lambda_N)$, with $\lambda_j = \sigma^2_{\epsilon_j}/\tau^2$. Taking the first derivative of the loglikelihood function for $\mathbf{y}_N$, we obtain: $\hat{\tau}^2 = \frac{1}{N}\mathbf{y}_N^t(K_N + \Lambda_N)^{-1}\mathbf{y}_N$, which can be substituted into the loglikelihood and obtain:

$$logL = -\frac{1}{N}log(2\pi) - \frac{N}{2}log(\hat{\tau}^2) - \frac{1}{2}log|K_N + \Lambda_N| - \frac{N}{2}. \tag{4.3}$$

Using the Woodbury formula, the authors establish in their Lemma 3.1:

$$\tau^2 k(\mathbf{x}, X_N)(\tau^2 K_N + \Sigma_\epsilon)^{-1}\mathbf{y}_N = k_n(\mathbf{x}, X_n)(K_n + \Lambda_n A_n^{-1})^{-1}\bar{\mathbf{y}} \tag{4.4}$$

$$\tau^2 k(\mathbf{x}, X_N)(\tau^2 K_N + \Sigma_\epsilon)^{-1}\tau^2 k^t(\mathbf{x}, X_N) = \tau^2 k_n(\mathbf{x}, X_n)(K_n + \Lambda_n A_n^{-1})^{-1}k_n^t(\mathbf{x}, X_n), \tag{4.5}$$

where n is the number of unique points and $A_n = diag(m_1, \dots m_n)$. These equations imply that the predictive mean and variance respectively at the n unique points will be the same as the one calculated with the usual GP model with all the observations. This will allow us to work with nxn instead of NxN matrices, which reduces calculation especially if N is large. These calculations are implemented in the R package *hetGP* [Binois and Gramacy, 2017].

4.2 Statistical monitoring based on GP model with heteroscedasticity

Instead of monitoring the hyper-parameters of a new profile, we propose to test if a jth observation of a new kth profile $y_{kj}(\mathbf{x}_{kj})$ follows the conditional multivariate normal distribution: $y_{kj}(\mathbf{x}_{kj})|D_N \sim \mathcal{N}(\tilde{\mu}(\mathbf{x}_{kj}), \tilde{\sigma}^2(\mathbf{x}_{kj}))$, where D_N is from Phase I, and

$$\tilde{\mu}(\mathbf{x}_{kj}) = k_n(\mathbf{x}_{kj}, X_n)(K_n + \Lambda_n A_n^{-1})^{-1}\bar{\mathbf{y}} \tag{4.6}$$

$$\tilde{\sigma}^2(\mathbf{x}_{kj}) = \tau^2\big(k_n(\mathbf{x}_{kj}, \mathbf{x}_{kj}) + \hat{\lambda}_j(\mathbf{x}_{kj}) - k_n(\mathbf{x}_{kj}, X_n)(K_n + \hat{\Lambda}_n A_n^{-1})^{-1}k_n^T(\mathbf{x}_{kj}, X_n)\big), \tag{4.7}$$

$\bar{\mathbf{y}}$, X_n are from IC data at the unique points from the IC data, and $j = \{1 \ldots, n_k\}$ where n_k is the number of observations of the kth profile.

Shewhart control charts are a natural choice for monitoring the mean. We use $\tilde{\mu}(\mathbf{x}_{kj})$ as the estimate of the jth observation of the predicted mean profile, and the prediction band as the control limits for that point. Since there are n_k observations for the kth profile, we need to deal with the multiple testing issue within the profile. Therefore, we study the following control limits:

$$\tilde{\mu}(\mathbf{x}_{kj}) \pm z_{\frac{\alpha}{2}} \tilde{\sigma}(\mathbf{x}_{kj}) \tag{4.8}$$

$$\tilde{\mu}(\mathbf{x}_{kj}) \pm z_{\frac{\alpha}{2}/n_k} \tilde{\sigma}(\mathbf{x}_{kj}) \tag{4.9}$$

$$\tilde{\mu}(\mathbf{x}_{kj}) \pm z_{\frac{\alpha}{2}/j} \tilde{\sigma}(\mathbf{x}_{kj}). \tag{4.10}$$

The first one has no correction while the last two have a Bonferroni correction. Equation 4.9 uses the Bonferroni correction which depends on the total number of data points the recent profile has or will have (n_k). Given that the monitoring procedure is applied at each location while the profile data are being collected, in Equation 4.10 the control limits depend on the number of observations up to that location (j). Herewith, we will refer to this correction as *Bonferroni varying correction.*

Using the proposed control chart, we can identify if the most recent profile is performing out of control at an early stage, and can investigate and take corrective actions if necessary. With this chart, we evaluate how each new profile evolves over time with respect to a mean profile function.

We recognize that when dealing with over-sampled profiles, the Bonferroni correction will not work. Some approaches on how to deal with this situation are also studied and discussed later in the next section.

4.3 Evaluating control chart performance

In this section, we study the performance of our proposed method in the monitoring phase and compare it to other methods from the literature. To generate the IC data for this analysis, we use the following function, that is used in Zhang et al. [2016]:

$$y = sin(x) + sin(10x/3) + ln(x) - 0.84x + 3, \quad \text{where } 2.5 \leq x \leq 7.5. \tag{4.11}$$

We generate a number of realizations from the previous function with an error $\epsilon(x)$ following GP model: $\epsilon(x_1) \ldots \epsilon(x_n) \sim \mathcal{N}(\eta_0 = \mathbf{0}_n, \Sigma_n)$ where $\Sigma_n = 0.05K_n(\theta_0 = 0.2) + \Sigma_\epsilon$ where K_n is the kernel function whose entries are $k(x_i, x_j) = exp(-||x_i - x_j||^2/\theta_0)$. Note that the authors do not consider the noise level matrix Σ_ϵ so in our baseline or IC simulations, we consider $\Sigma_\epsilon = 0$ to make it comparable. However, when we analyze the performance of our proposed method in the out-of-control (OOC) case later in the section, we will consider that Σ_ϵ linearly increases from 0 up to a maximum noise level. The generated profiles in the IC situation are shown in Figure 4.1.

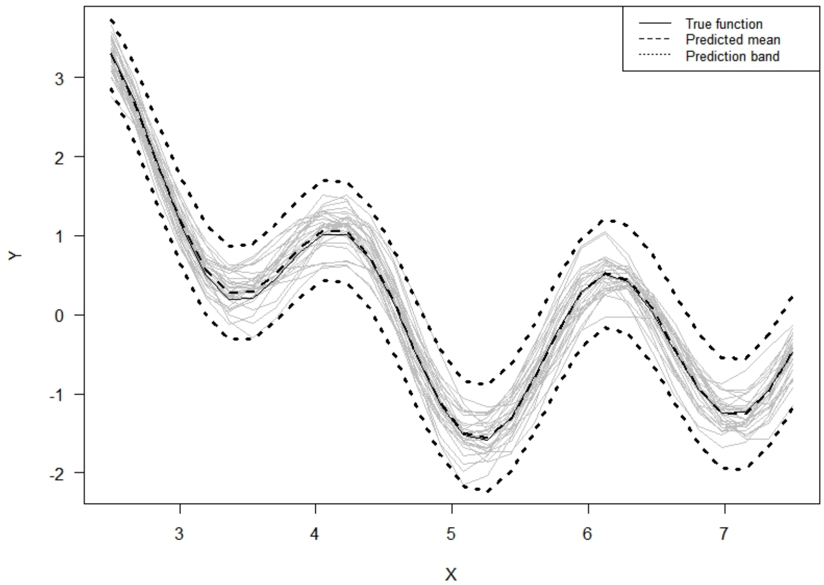

Figure 4.1: The grey lines are 50 realizations of the function shown in Equation 4.11 with an error under a GP model. The solid line shows the true function. The dashed lines are the predicted mean and the prediction band under the GP model with heteroscedasticity as described in Section 4.1.

We produce our Phase I data by generating a number of IC profiles based on Equation 4.11. Then, a GP model with heteroscedasticity is estimated; subsequently, this estimate is used to set the mean profile as described in Section 4.2. We study different IC scenarios based on the number of IC profiles generated in Phase I, $m = \{20, 30, 50\}$, and at the unique points used in each profile $n = \{20, 30, 40\}$.

Frequently, the performance of different monitoring approaches is evaluated through false alarm probabilities and detection rates. In our approach, we compare the three anomaly detection methods as defined in Equations 4.8-4.10. To estimate the false alarm rates, we generated 50 profiles without any anomaly based on the IC model shown in Equation 4.11

and calculate the percentage of the profiles that are considered as OOC. To estimate the detection rates, we create different "abnormal" or OOC scenarios, and for each one we measure the percentage of profiles that are detected as OOC.

4.3.1 Evaluating false alarm rates

Each scenario was run 10000 times, and the empirical false alarm rate is calculated for each one. Table 4.1 shows these rates with their standard errors for the cut-off values of $\alpha = \{0.05, 0.01\}$.

Table 4.1: Analyzing false alarm rate (and the standard error in parentheses) for the proposed method with three approaches to form the control limits.

		$\alpha = 0.05$			$\alpha = 0.01$		
Number of IC profiles in Phase I (m)	Number of unique points at each profile (n)	Prediction band ($z_{\frac{\alpha}{2}}$)	Prediction band with Bonferroni correction ($z_{\frac{\alpha}{2}/n}$)	Prediction band with Bonferroni varying correction ($z_{\frac{\alpha}{2}/j}$)	Prediction band ($z_{\frac{\alpha}{2}}$)	Prediction band with Bonferroni correction ($z_{\frac{\alpha}{2}/n}$)	Prediction band with Bonferroni varying correction ($z_{\frac{\alpha}{2}/j}$)
20	20	0.579(0.105)	0.067(0.046)	0.179(0.077)	0.196(0.082)	0.018(0.022)	0.049(0.039)
	30	0.622(0.103)	0.060(0.044)	0.169(0.075)	0.227(0.088)	0.018(0.022)	0.049(0.039)
	40	0.643(0.101)	0.056(0.043)	0.161(0.075)	0.245(0.091)	0.017(0.022)	0.04(0.039)
30	20	0.558(0.094)	0.056(0.039)	0.162(0.066)	0.176(0.070)	0.014(0.018)	0.041(0.033)
	30	0.596(0.093)	0.048(0.036)	0.149(0.064)	0.200(0.074)	0.013(0.017)	0.039(0.032)
	40	0.613(0.092)	0.043(0.034)	0.140(0.063)	0.214(0.078)	0.012(0.017)	0.038(0.032)
50	20	0.542(0.086)	0.049(0.034)	0.151(0.058)	0.163(0.062)	0.012(0.016)	0.036(0.028)
	30	0.576(0.086)	0.040(0.030)	0.137(0.056)	0.184(0.066)	0.010(0.015)	0.034(0.028)
	40	0.591(0.085)	0.034(0.028)	0.124(0.054)	0.193(0.067)	0.008(0.013)	0.031(0.027)

For the scenarios analyzed, the performance of the chart when using the Bonferroni correction ($z_{\frac{\alpha}{2}/n}$) shown in Equation 4.9 is better by far than the other two given that its false alarm rates are close to the nominal value. It can be seen that the false alarm rates slightly decrease as the number of IC profiles used to estimate the mean model increases; and they decrease, as expected, when the number of unique points per profile increases. When we do not use the Bonferroni correction, a high percentage of false alarms is present.

Assuming that the profiles are produced in time order, we can use an individual chart based on the metrics shown in Chapter 2, as suggested by Williams et al. [2007]. The control limits were calculated according to the standard univariate I-charts as suggested by Boeing Commercial Airplane Group. Table 4.2 shows the results of this performance. The average deviation (M_1), average absolute deviation (M_2) and maximum absolute deviation (M_3) show better performance given that the false alarm rates are closer to their nominal values. These appear stable for all the IC situations analyzed.

Table 4.2: Analyzing the false alarm rate (and the standard error in parentheses) for the univariate I-chart of the 4 deviation metrics shown in Chapter 2.

		$\alpha = 0.05$				$\alpha = 0.01$			
Number of IC profiles in Phase I(m)	Number of unique points at each profile (n)	M_1	M_2	M_3	M_4	M_1	M_2	M_3	M_4
20	20	0.048(0.029)	0.047(0.028)	0.048(0.027)	0.050(0.024)	0.010(0.014)	0.011(0.014)	0.013(0.015)	0.020(0.017)
	30	0.048(0.029)	0.047(0.028)	0.048(0.027)	0.050(0.024)	0.010(0.014)	0.011(0.014)	0.013(0.015)	0.021(0.017)
	40	0.048(0.029)	0.047(0.027)	0.047(0.028)	0.050(0.024)	0.010(0.014)	0.012(0.014)	0.013(0.015)	0.021(0.017)
30	20	0.048(0.029)	0.046(0.028)	0.048(0.027)	0.049(0.023)	0.009(0.014)	0.011(0.014)	0.013(0.015)	0.021(0.016)
	30	0.048(0.029)	0.047(0.027)	0.048(0.027)	0.050(0.024)	0.009(0.014)	0.012(0.014)	0.013(0.015)	0.021(0.017)
	40	0.048(0.028)	0.047(0.028)	0.048(0.027)	0.050(0.023)	0.010(0.014)	0.012(0.014)	0.013(0.015)	0.021(0.017)
50	20	0.048(0.029)	0.047(0.028)	0.047(0.027)	0.050(0.023)	0.010(0.014)	0.011(0.014)	0.013(0.015)	0.021(0.017)
	30	0.048(0.029)	0.047(0.028)	0.048(0.027)	0.050(0.024)	0.010(0.014)	0.012(0.014)	0.013(0.015)	0.021(0.017)
	40	0.048(0.029)	0.047(0.028)	0.048(0.027)	0.050(0.024)	0.010(0.014)	0.012(0.014)	0.013(0.015)	0.021(0.017)

Performance comparison with T^2 charts under GP models

As mentioned in the literature review, there are existing methods that propose the use of a T^2 control chart based on a GP model to determine if a profile is in or out of control. The T^2 statistic suggested by Wang et al. [2014] shown in Equation 2.3 monitors the hyper-parameters of the model. Their T^2 approximately follows $\chi^2(k)$ where k is the number of hyper-parameters to be estimated; for this simulation study, $k = 3$. To estimate the IC hyper-parameters, we used two different approaches. The first fits the model directly on the entire IC data. The second entails fitting a GP model on each IC profile and then averaging all the estimates as suggested by Zhang et al. [2014].

We also analyze the T^2 control chart based on the AGP model proposed by Zhang et al. [2016], which monitors the new profile data and compares it to the IC mean profile instead of monitoring the hyper-parameters. Their T^2 statistic follows $\chi^2(n)$, where n is the number of points of a given profile.

For each scenario, Table 4.3 shows the empirical average false alarm rate in 10000 runs for the T^2 charts under the GP model proposed by Wang et al. [2014] considering two IC parameter estimation procedures, and for the T^2 chart under the AGP by Zhang et al. [2016]. The control limits of the charts were established as suggested by the authors which is described in Chapter 2.

Our results suggest that the T^2 control chart proposed by Wang et al. [2014] which is based on the hyper-parameter estimates does not perform well. First, it is sensitive to the type of approach used to estimate the IC hyper-parameters given that we obtain very different results for each one. In addition, we observe that when using the first approach which fits the GP model on the entire data to estimate the IC hyper-parameters, we get a false alarm rate of 100%. This is because in this approach we fit the model on more data and the variability is being recognized as noise rather than signal, and hence we get a longer length-scale parameter (as defined Equation 4.1) which reflects on a smoother mean profile. Then, when the length-scale for a new single profile is estimated, we get a shorter length-scale value because the profile is less smooth given that the fit is done on only that profile data. In this sense, the approach that fits a GP model on the entire IC data is not appropriate to estimate the IC hyper-parameters. The second approach in which we fit a GP model on each profile and then average all the estimates works better but still presents a high percentage of false alarms.

On the contrary, the T^2 control chart based on the the AGP by Zhang et al. [2016] works quite well when the number of unique points at each profile is small (up to 30). The average false alarm rate in an IC situation seems to decrease when either the number of profiles used in Phase I or the number of unique points in a given profile increases.

Table 4.3: Analyzing the average false alarm rate (and the standard error in parentheses) for the T^2 chart by Wang et al. [2014] under two different IC parameter estimations and for the T^2 chart by Zhang et al. [2016].

		$\alpha = 0.05$			$\alpha = 0.01$		
Number of IC profiles in Phase I (m)	Number of unique points at each profile (n)	Wang et al. [2014]-fit on all IC data	Wang et al. [2014] with averaged estimates	Zhang et al. [2016]	Wang et al. [2014]-fit on all IC data	Wang et al. [2014] with averaged estimates	Zhang et al. [2016]
20	20	1.000(0.000)	0.429(0.295)	0.070(0.051)	1.000(0.000)	0.305(0.324)	0.018(0.023)
	30	1.000(0.000)	0.481(0.177)	0.063(0.047)	1.000(0.000)	0.330(0.133)	0.015(0.020)
	40	1.000(0.000)	0.563(0.229)	0.296(0.097)	1.000(0.000)	0.388(0.214)	0.235(0.100)
30	20	1.000(0.000)	0.460(0.294)	0.067(0.045)	1.000(0.000)	0.328(0.305)	0.016(0.020)
	30	1.000(0.000)	0.470(0.167)	0.061(0.042)	1.000(0.000)	0.321(0.122)	0.014(0.018)
	40	1.000(0.000)	0.558(0.210)	0.257(0.090)	1.000(0.000)	0.365(0.187)	0.182(0.089)
50	20	1.000(0.000)	0.480(0.264)	0.064(0.041)	1.000(0.000)	0.353(0.282)	0.015(0.019)
	30	1.000(0.000)	0.454(0.156)	0.059(0.038)	1.000(0.000)	0.307(0.113)	0.013(0.017)
	40	1.000(0.000)	0.581(0.182)	0.262(0.083)	1.000(0.000)	0.365(0.159)	0.159(0.078)

4.3.2 Evaluating detection rates

To generate anomalies in the profiles we consider four scenarios. The first is created by shifting the mean ($\eta_1 = \eta_0 + \delta_\eta$) where $\delta_\eta = \{0.2, 0.4, 0.6\}$. For the second, we analyze a shift in the process variance ($\tau_1^2 = \tau_0^2 + \delta_{\tau^2}$) where $\delta_{\tau^2} = \{0.05, 0.15, 0.25\}$. Shifts in the length-scale or correlation parameter ($\theta_1 = \theta_0 + \delta_\theta$) where $\delta_\theta = \{-0.167, 0, 0.8, 9.8\}$, and shifts in the maximum noise level ($\sigma^2_{\epsilon_n 1} = \sigma^2_{\epsilon_n 0} + \delta_{\sigma^2_\epsilon}$) where $\delta_{\sigma^2_\epsilon} = \{0.1, 0.2, 0.3\}$ are also studied. As an example, Figure 4.2 shows the IC profile and 5 generated profiles for these scenarios.

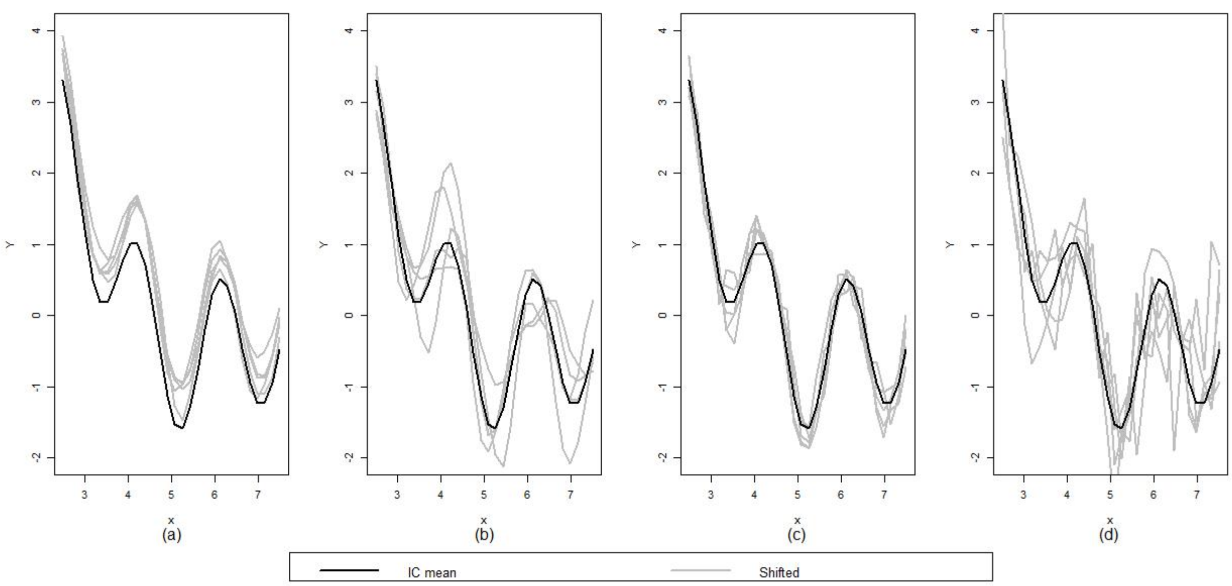

Figure 4.2: Generated profiles for four OOC scenarios: (a) shift of $\delta_\eta = 0.6$ in the mean, (b) shift of $\delta_{\tau^2} = 0.25$ in the process variance, (c) shift of $\delta_\theta = 9.8$ in the correlation, and (d) shift of $\delta_{\sigma^2_\epsilon} = 0.3$ in the maximum noise level.

Figure 4.3 shows the proposed method's performance with regard to detecting OOC profiles

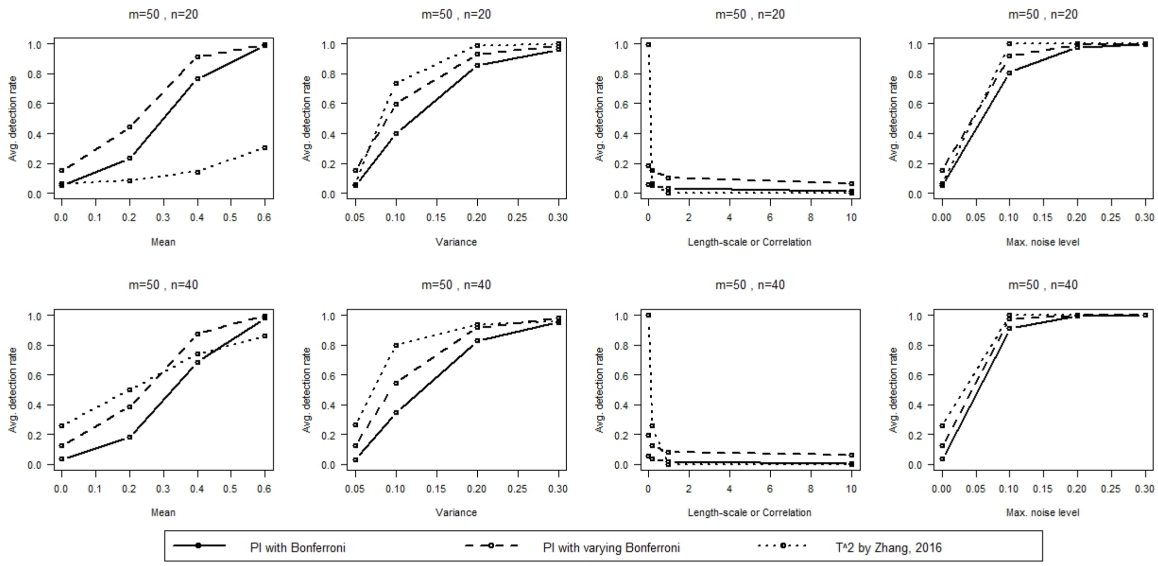

Figure 4.3: Detection rate performance of our method and T^2 chart by Zhang et al. [2016] in detecting different types of shifts.

for the three prediction band approaches when there is a shift in the mean, process variance, correlation and maximum noise level parameters. The T^2 charts by Wang et al. [2014] are not shown here because they did not present adequate IC behavior; hence, their use would be impractical. Our method without the Bonferroni correction is also impractical given that it generates high false alarm rates.

As noted in the the previous section, the T^2 chart by Zhang et al. [2016] presents a high average false alarm rate if number of observations per profile is higher than 30. However, if the number of observations per profile is 30 or less, the detection performance of this chart is good, but not as good as our method.

The use of Bonferroni correction shows good performance in detecting shifts in the mean, variance and maximum noise level. However, we do not see that our method detects the shifts in the correlation parameter. Given that this parameter models the rate of decay of spatial correlation among the points, as Figure 4.2 (c) shows, shifts in this parameter make the profile more or less smooth. Nonetheles, the shifts in this parameter are not enough to cause the profile to significantly depart from the mean IC profile and as such are not detected by our method. This method would not be appropriate if the objective were to detect profiles with greater or lesser correlation among the points within a profile, which is rarely the case in an industrial setting. This situation is a clear example of a statistical vs. practical significance case that we discussed in Chapter 2.

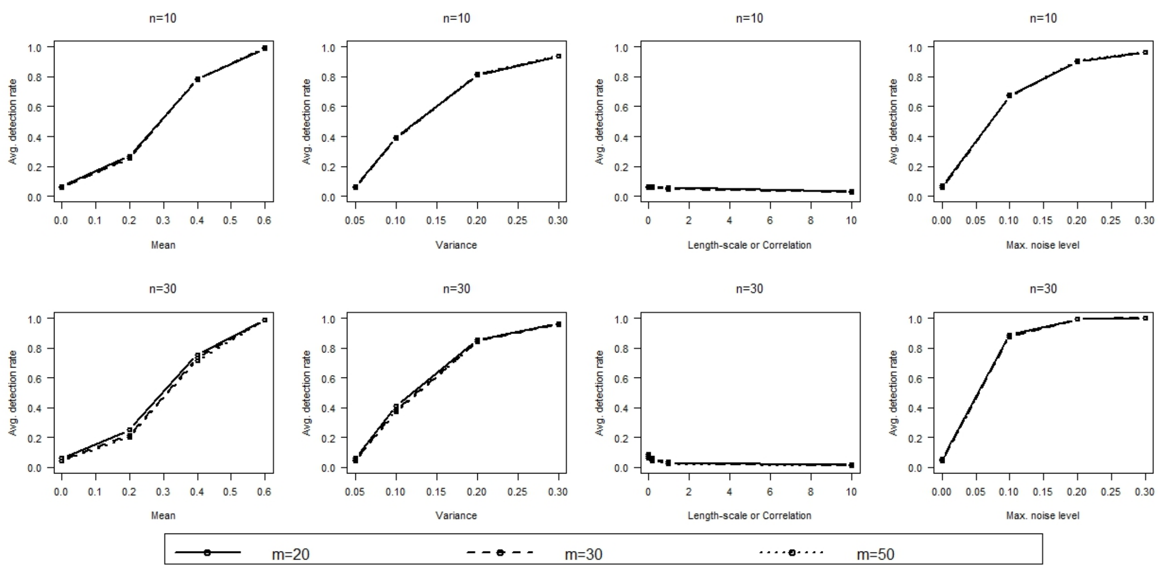

Figure 4.4: Detection average rate performance of our method with Bonferroni correction in detecting different types of shifts when using different numbers of profiles in Phase I.

Figure 4.4 shows the performance of our method with Bonferroni correction when using different numbers of profiles in Phase I. Its average performance does not depend on the

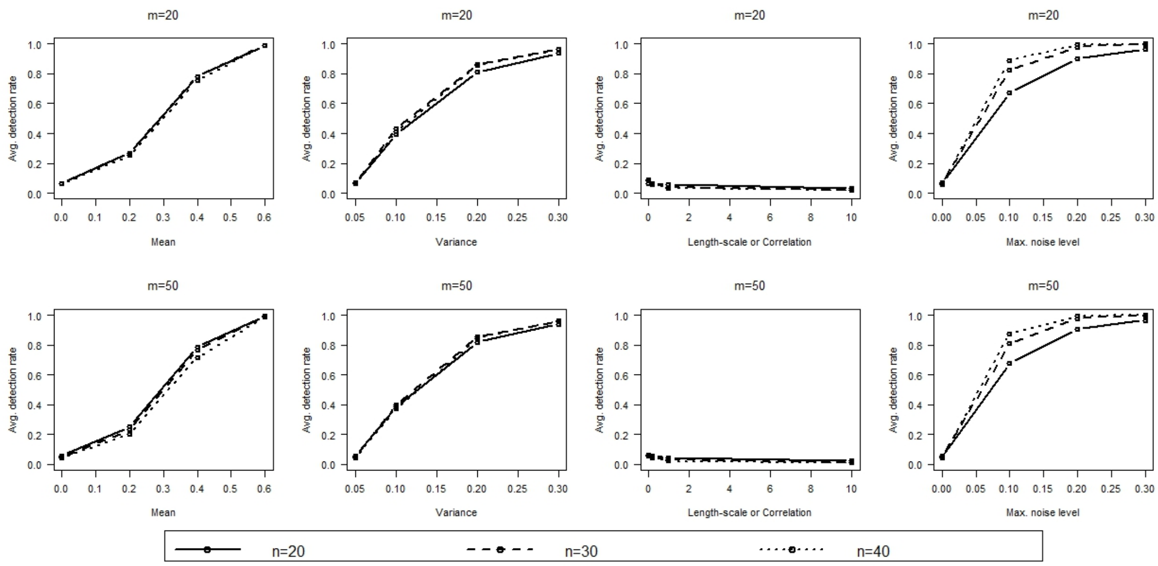

Figure 4.5: Detection average rate performance of our method with Bonferroni correction in detecting different types of shifts when using different numbers of observations per profile.

number of profiles used in Phase I. Figure 4.5 shows its performance when using different numbers of unique points or observations in each profile. When there is a shift in the maximum noise level, it appears that the detection average rate increases as the number of observations in each profile increases. This result is as expected; when we have more points in a very noisy profile, there is a greater chance that at least one will fall outside the prediction band.

4.3.3 Over-sampled profiles

For the cases analyzed above, the Bonferroni correction seems to be appropriate. However, this correction will not work if the profiles have a larger number of observations. Here, we analyze two approaches to deal with over-sampled profiles. The first is to select every kth point in each new profile and to execute the monitoring scheme on the selected points. Herewith, we will call this approach the *systematic selecting approach*. An alternative approach is to borrow Shewhart's concept of a rational subgroup. We group the observations of each new profile by subgroups of size k, calculate their averages and perform the monitoring scheme on those averages, using the location of the last observation from each subgroup. Herewith, we will refer to this approach as the *sub-grouping approach*.

We use the monitoring approach described in Section 4.2. Instead of considering the total number of unique points or total number of observations for the new profile, we consider only the selected points when using the *systematic selecting approach*, or the averages of each subgroup when using the *sub-grouping approach*. Tables 4.4 and 4.5 show the false alarm rates for these approaches.

There are a few comments to highlight here. In general, the *sub-grouping approach* works better than the *systematic selecting approach*. The *systematic selecting approach* seems to be dependent on either the number of profiles used in the IC phase (or D_N data) or the number of observations per profile. The number of false alarms increases as both the number of profiles used in Phase I and observations per profile increases. It seems that this approach also depends on the value of k used.

In the *sub-grouping approach*, when the number of subgroups per profile is small (less than 50), the Bonferroni correction as shown in Equation 4.9 ($z_{\frac{\alpha}{2}/n}$), works well. With a number of observations of up to 250 with a subgroup size of $k = 5$, our proposed method with

Table 4.4: Average false alarm rate when using *systematic selecting approach*.

Number of observations per each profile in IC data (m)	Number of unique points at each profile (n)	Number of sampled observations per profile ($n^* = n/k$)	$\alpha = 0.05$ Prediction band ($z_{\frac{\alpha}{2}}$)	Prediction band with Bonferroni correction ($z_{\frac{\alpha}{2}/n^*}$)	Prediction band with Bonferroni varying correction ($z_{\frac{\alpha}{2}/j}$)	$\alpha = 0.01$ Prediction band ($z_{\frac{\alpha}{2}}$)	Prediction band with Bonferroni correction ($z_{\frac{\alpha}{2}/n^*}$)	Prediction band with Bonferroni varying correction ($z_{\frac{\alpha}{2}/j}$)
30	30	30	0.596	0.048	0.150	0.201	0.013	0.040
	50	50	0.611	0.034	0.123	0.211	0.009	0.032
	1000	1000	0.620	0.003	0.062	0.219	0.001	0.014
	60	30	0.625	0.062	0.172	0.229	0.019	0.051
	100	50	0.640	0.046	0.145	0.242	0.014	0.043
	2000	1000	0.648	0.006	0.071	0.249	0.002	0.019
	150	30	0.761	0.074	0.224	0.359	0.024	0.071
	250	50	0.676	0.030	0.143	0.268	0.009	0.040
	5000	1000	0.620	0.002	0.062	0.219	0.000	0.014
100	30	30	0.781	0.093	0.255	0.392	0.034	0.090
	50	50	0.703	0.042	0.167	0.301	0.014	0.052
	1000	1000	0.649	0.004	0.071	0.250	0.001	0.019
	60	30	0.902	0.113	0.370	0.584	0.045	0.152
	100	50	0.758	0.027	0.177	0.355	0.009	0.054
	2000	1000	0.622	0.001	0.062	0.219	0.000	0.014
	150	30	0.911	0.136	0.405	0.612	0.060	0.181
	250	50	0.780	0.038	0.204	0.390	0.014	0.071
	5000	1000	0.648	0.002	0.070	0.249	0.001	0.018

Bonferroni correction works as expected. When using a larger number of subgroups per profile, meaning higher than 50 and up to 1000 (up to 5000 observations with a subgroup size of 5), our proposed method with Bonferroni varying correction ($z_{\frac{\alpha}{2}/j}$) works better. As described in Section 4.2, this correction shown in Equation 4.10 varies according to the number of points that has been collected up to the moment the monitoring procedure is applied. In addition, the performance of *subgrouping approach* does not depend on the number of observations per profile used in the IC phase or the subgroup size used in this simulation.

Figure 4.6 shows the detection rate performance of our method with Bonferroni correction ($z_{\frac{\alpha}{2}/n}$) using the *sub-grouping approach* in cases entailing observations of up to 150 per observed profile. Figure 4.7 shows our method's performance with Bonferroni varying cor-

Table 4.5: Average false alarm rate when using *sub-grouping approach*.

Number of observations per each profile in IC data	Number of unique points at each profile	Number of subgroups per profile	$\alpha = 0.05$ Prediction band ($z_{\frac{\alpha}{2}}$)	Prediction band with Bonferroni correction ($z_{\frac{\alpha}{2}/n}$)	Prediction band with Bonferroni varying correction ($z_{\frac{\alpha}{2}/j}$)	$\alpha = 0.01$ Prediction band ($z_{\frac{\alpha}{2}}$)	Prediction band with Bonferroni correction ($z_{\frac{\alpha}{2}/n}$)	Prediction band with Bonferroni varying correction ($z_{\frac{\alpha}{2}/j}$)
30	30	30	0.596	0.048	0.150	0.201	0.013	0.040
	50	50	0.611	0.034	0.123	0.211	0.009	0.032
	1000	1000	0.620	0.003	0.062	0.219	0.001	0.014
	60	30	0.625	0.062	0.172	0.229	0.019	0.051
	100	50	0.640	0.046	0.145	0.242	0.014	0.043
	2000	1000	0.648	0.006	0.071	0.249	0.002	0.019
	150	30	0.649	0.040	0.216	0.247	0.011	0.070
	250	50	0.629	0.022	0.145	0.227	0.006	0.041
	5000	1000	0.619	0.002	0.062	0.219	0.000	0.014
100	30	30	0.675	0.052	0.238	0.278	0.018	0.084
	50	50	0.658	0.032	0.166	0.259	0.010	0.052
	1000	1000	0.649	0.004	0.070	0.250	0.001	0.019
	60	30	0.736	0.038	0.324	0.337	0.013	0.129
	100	50	0.663	0.014	0.181	0.258	0.004	0.057
	2000	1000	0.622	0.001	0.063	0.218	0.000	0.014
	150	30	0.758	0.051	0.351	0.371	0.019	0.149
	250	50	0.688	0.022	0.202	0.291	0.008	0.071
	5000	1000	0.648	0.002	0.070	0.249	0.001	0.018

rection ($z_{\frac{\alpha}{2}/j}$) using the *sub-grouping approach* in cases entailing over-sampled profiles (up to 5000 observations per observed profile). Our method seems to work as expected in detecting shifts in the mean, variance, and maximum noise level. As we discussed in the previous section, the method fails to detect shifts in the correlation parameter given that these shifts affect the "wiggliness" of the profiles but the profiles tend to remain within the bandwidth.

4.3.4 Simulation study II

In this section we analyze the performance of our method using data generated based on a more realistic situation. With this example that is based on the case study seen in Chapter 3, we can produce OOC scenarios that have a practical interpretation. To produce the simulated data based on this case study data, we apply the double bootstrap. We fit a

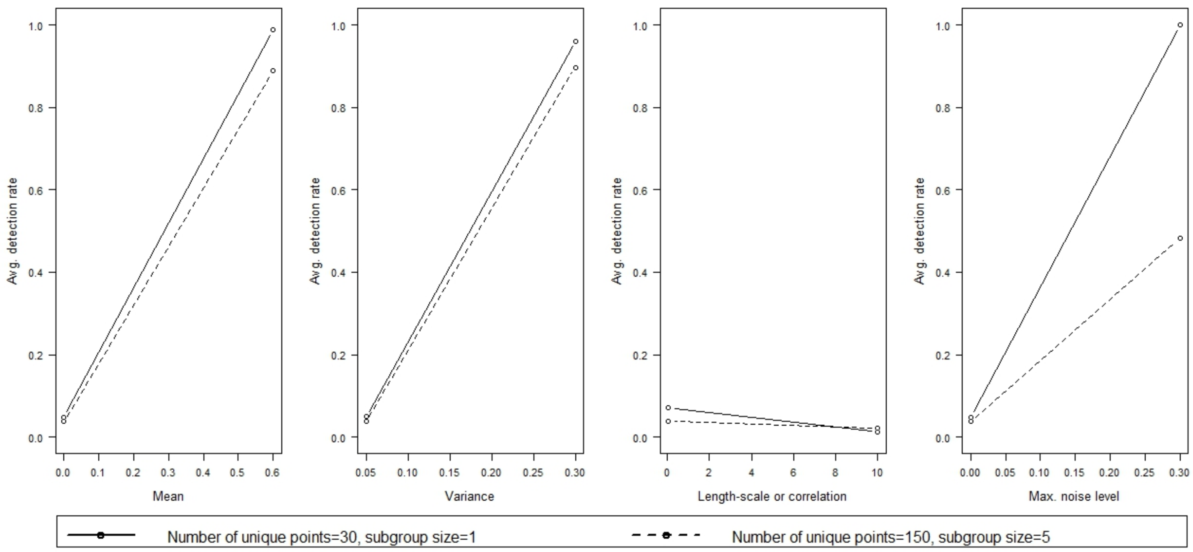

Figure 4.6: Detection rate performance of the proposed method with Bonferroni correction shown in in Equation 4.9 when using $m = 30$ profiles in the Phase I (or D_N).

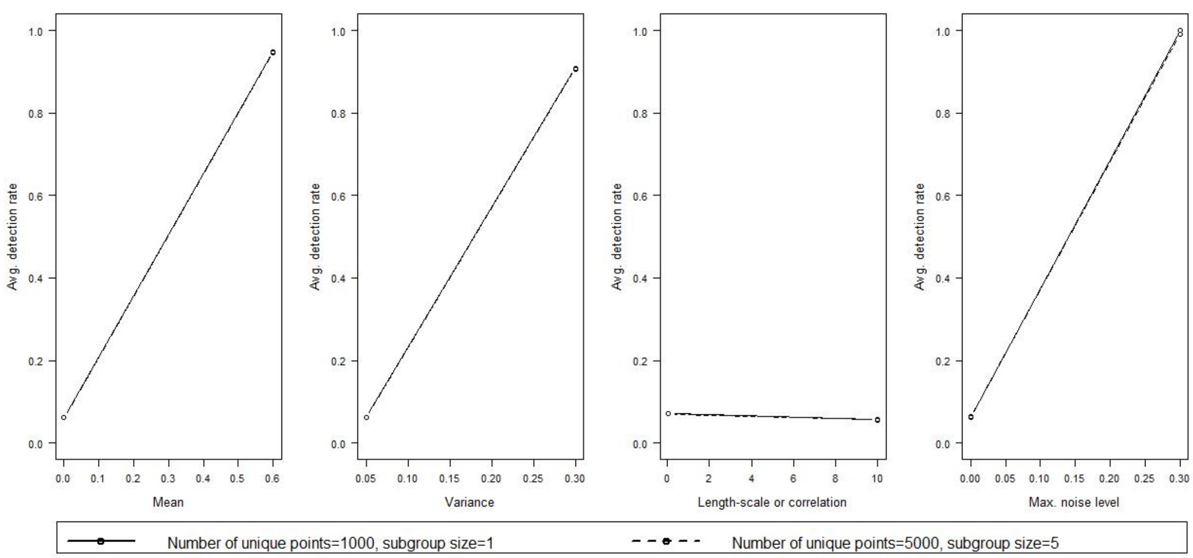

Figure 4.7: Detection rate performance of the proposed method with Bonferroni varying correction shown in Equation 4.10 when using $m = 30$ profiles in the Phase I (or D_N).

parametric Gompertz model to each profile and we store the residuals at each point in time. The first bootstrap is applied on the fits of these profiles and the second is applied on the residuals at each point in time. Figure 4.8 shows 52 generated profiles for one simulation run.

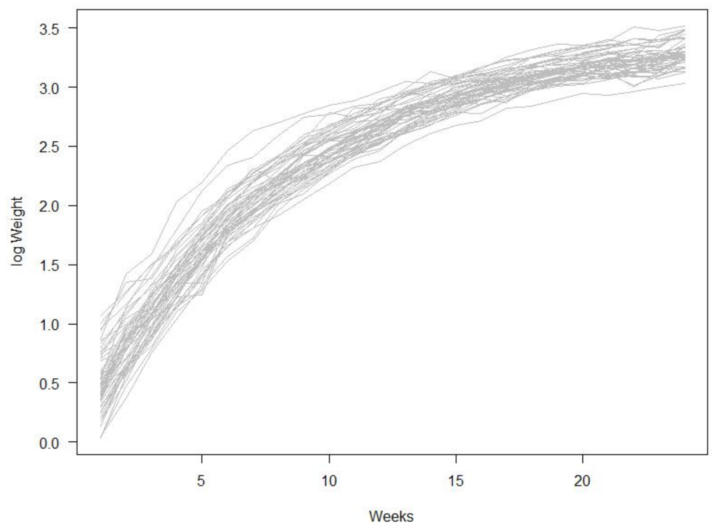

Figure 4.8: The grey lines are the profiles generated for one simulation run by applying double bootstrap on case study post cleansed data.

Evaluating false alarm rates

Table 4.6 shows the average false alarm rates (and the standard errors in parentheses) in 10000 simulation runs for the T^2 charts proposed by Wang et al. [2014] and Zhang et al. [2016], also evaluated in the previous section and described in Chapter 2, and for the proposed method with the three approaches to form the control limits. The Bonferroni correction works as expected since there are just 24 observations or unique points per profile.

Table 4.6: Analyzing false alarm rate (and the standard error in parentheses) for the T^2 chart by Wang et al. [2014], the T^2 chart by Zhang et al. [2016], and our proposed method using the three approaches to form the control limits.

	Wang et al. [2014] with averaged estimates	Zhang et al. [2016]	Prediction band ($z_{\frac{\alpha}{2}}$)	Prediction band with Bonferroni correction ($z_{\frac{\alpha}{2}/n}$)	Prediction band with Bonferroni varying correction ($z_{\frac{\alpha}{2}/j}$)
$\alpha = 0.01$	0.838(0.063)	0.157(0.032)	0.102(0.031)	0.007(0.015)	0.053(0.024)
$\alpha = 0.05$	0.874(0.055)	0.206(0.036)	0.300(0.044)	0.032(0.021)	0.136(0.034)

We obtain similar results as in Section 4.3.1. The T^2 control charts that monitor the hyper-parameters suggested by Wang et al. [2014] do not work well. The T^2 control chart proposed by Zhang et al. [2016] performs better, but not as good as our method with Bonferroni correction ($z_{\frac{\alpha}{2}/n}$). As seen previously, Bonferroni correction works well if the number of observations (or subgroups) per profile is less than 50. When we do not use the Bonferroni correction, a high percentage of false alarms is present.

Evaluating detection rates

Here we evaluate the performance of our proposed chart in detecting OOC profiles. We create OOC scenarios by shifting the IC parameters of the underlying parametric Gompertz model. We consider shifting: (i) the asymptotic log weight $\theta_1^* = \theta_{1_{IC}} + \delta$ where $\delta = \{-0.2, -0.1, 0, 0.1, 0.2\}$, (ii) the initial weight $\theta_2^* = \theta_{2_{IC}} + \delta$ where $\delta = \{-0.8, -0.4, 0, 0.4, 0.8\}$, and (iii) the growth rate $\theta_3^* = \theta_{3_{IC}} + \delta$ where $\delta = \{-0.04, -0.02, 0, 0.02, 0.06\}$. Figure 4.9 shows how the different shifts are reflected in the growth profiles.

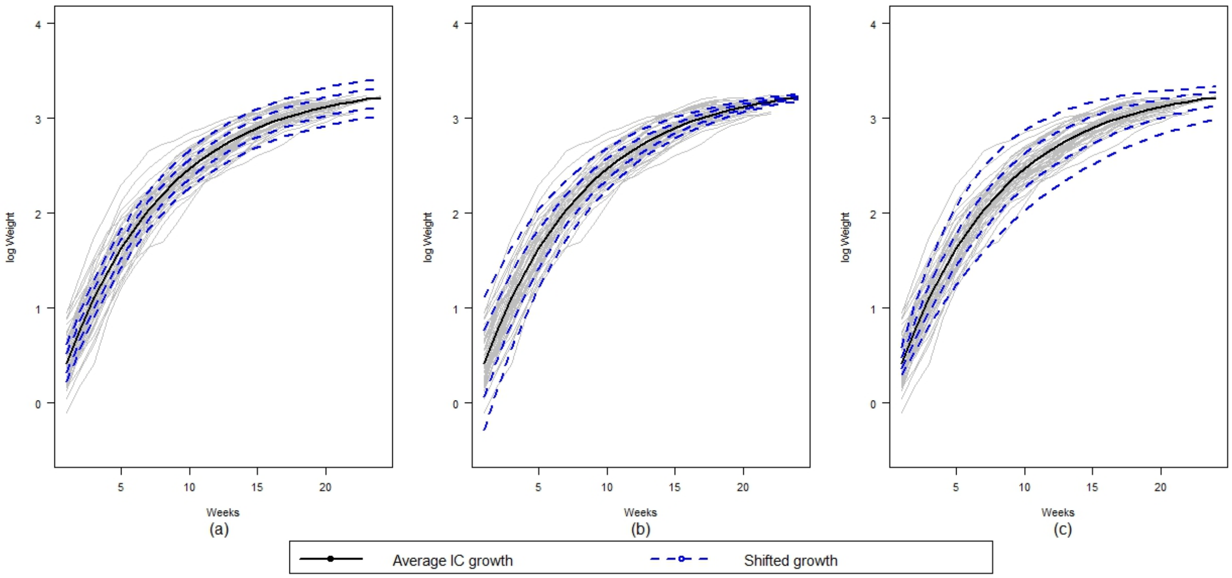

Figure 4.9: OOC scenarios: (a) shifts in the asymptotic log weight, (b) shifts in the initial log weight, and (c) shifts in the growth rate.

Figure 4.10 shows the performance of the proposed method with the three prediction band approaches, compared to the T^2 chart by Zhang et al. [2016]. The other T^2 charts are not shown here because they did not possess good IC behavior. The detection performance of our method in detecting OOC profiles when there is a shift in the parameters shows much better results than the method proposed by Zhang et al. [2016].

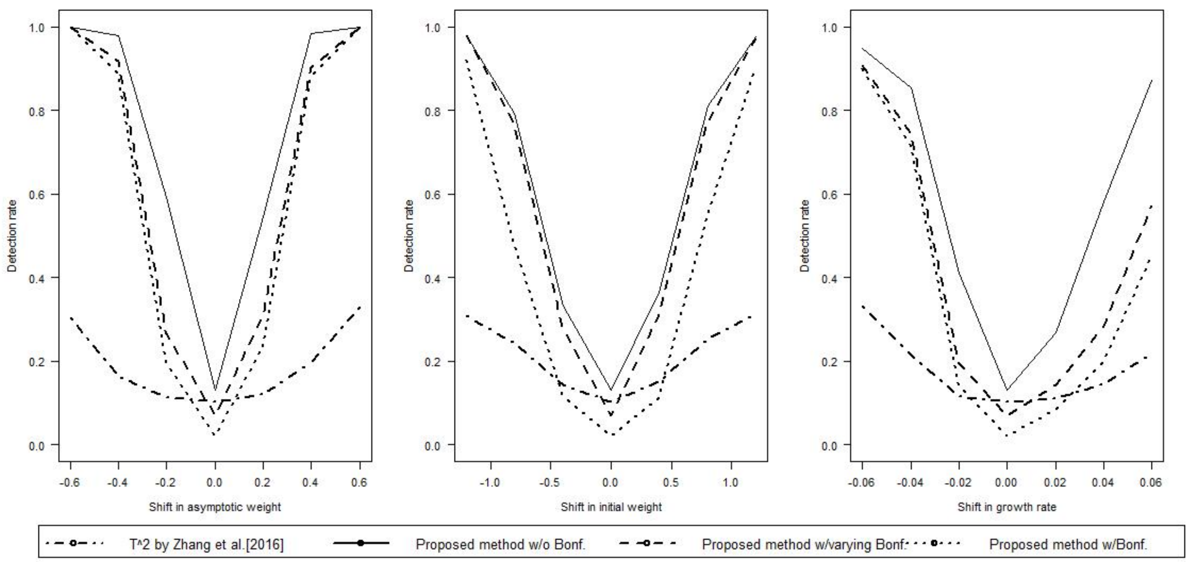

Figure 4.10: Comparison of detection rates performance between our method and the T^2 chart by Zhang et al. [2016], for different OOC scenarios.

4.4 Case study

This section illustrates how the GP model with heteroscedasticity can be applied to monitor the aquaculture process described in Chapter 3 through a Shewhart-type control chart, and compare its performance with the other methods that use GP modeling studied above. The number of observations for each week is not the same within the growing season or campaign. Also, the number of growth profiles for each campaign is different; there are 52, 223, 339 and 392 profiles for the four campaigns analyzed respectively. Given the high number of profiles, the computational savings of using the methodology proposed by Binois et al. [2018] are not negligible.

We use 75% of the data as training set or IC data and the rest as testing data. We use the training set to model the mean profile and its predictive band using the GP model with heteroscedasticity described in Section 4.1. Figure 4.11 shows the control charts for monitoring the process within each campaign. Each control chart shows the predicted mean profile for the shrimp growth and the 99% prediction interval obtained under the GP model with heteroscedasticity. We compare this predictive band to the ones obtained using a parametric Gompertz model suitable for this data described in Chapter 3. The grey lines represent the growth profiles from the testing data. We observe that both predictive bands are very close to each other except at the boundaries. Using *hetGP* R package from Binois and Gramacy [2017] we obtain the estimates for θ, τ^2 and all the $\sigma^2_{\epsilon_1}$ which are shown in Table 4.7.

Table 4.7: IC hyper-parameter estimates obtained when fitting a GP model with heteroscedasticity on the study case data, for each campaign. Due to space limitations, only the first and last noise level are shown for each campaign.

	Campaigns			
	11	14	15	16
θ	249.9642	249.9511	249.9764	249.9746
$\sigma^2_{\epsilon_i}$	$\sigma^2_{\epsilon_1} = 0.0540\ldots$ $\sigma^2_{\epsilon_{24}} = 0.0036$	$\sigma^2_{\epsilon_1} = 0.0632\ldots$ $\sigma^2_{\epsilon_{27}} = 0.0031$	$\sigma^2_{\epsilon_1} = 0.1056\ldots$ $\sigma^2_{\epsilon_{28}} = 0.0127$	$\sigma^2_{\epsilon_1} = 0.1194\ldots$ $\sigma^2_{\epsilon_{30}} = 0.0068$
τ^2	14.4299	18.7999	19.013	13.6891

4.4.1 Performance comparison with other GP methods

First we compare the mean function estimation using the different models. We use the testing data to calculate the Mean Squared Error (MSE) for each model analyzed. These

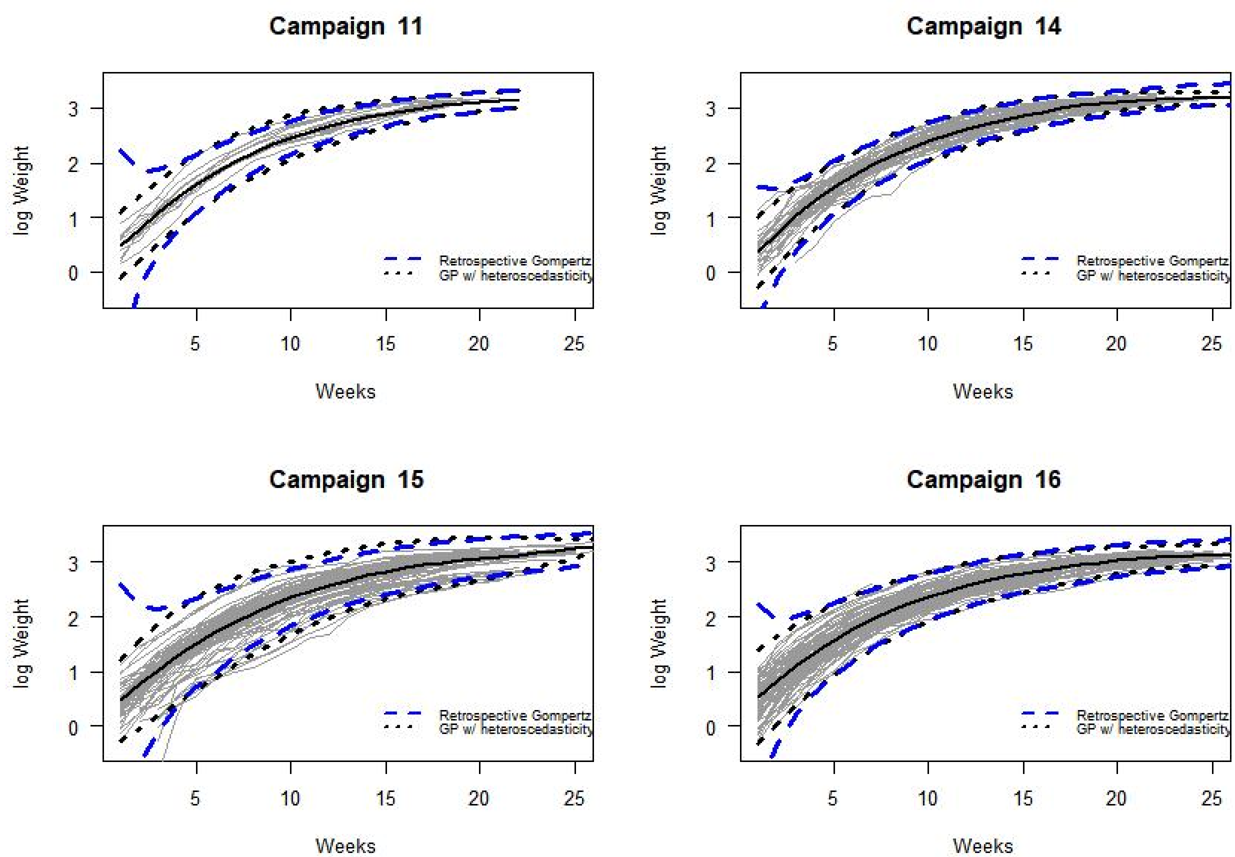

Figure 4.11: Control charts for monitoring the mean function for each campaign. The darker middle lines show the predicted mean function. The dashed lines are the 99% prediction limits. The grey profiles belong to the testing data. The blue lines are the bandwidth when using Gompertz model, and the black lines are the bandwidth when using GP model.

are shown in Table 4.8. This indicates that all the methods are effective in the overall mean prediction.

To compare our monitoring approach with the T^2 control charts, we used the three approaches for prediction band as control limits shown in Equations 4.8-4.10. Table 4.9 shows the number of OOC profiles that each method signals. Our control chart with the Bonferroni correction seems to have better performance given that this case study data can be consid-

Table 4.8: MSE comparison using testing data for estimating the IC mean profile. We used a parametric Gompertz model, GP model with heteroscedasticity, GP model by Zhang et al. [2014], GP model by Wang et al. [2014] and the AGP model by Zhang et al. [2016], for each campaign.

	Campaigns			
Models	11	14	15	16
Parametric Gompertz model	0.02745	0.01802	0.06774	0.03745
GP model with heteroscedasticity	0.02764	0.01799	0.06828	0.03748
GP model by [Wang et al., 2014] with averaged estimates	0.02764	0.01795	0.06830	0.03746
GP model by [Wang et al., 2014] on all data	0.02769	0.01800	0.06827	0.03746
AGP model by [Zhang et al., 2016]	0.02796	0.01813	0.08851	0.04434

ered clean with no anomalies because a thorough Phase I analysis has been applied to it. The T^2 control chart that monitors the hyper-parameters proposed by Wang et al. [2014], with either choice of IC parameter estimation procedure, signals a large number of profiles as seen in the simulation study. Given that this chart monitors the hyper-parameters and those are tuned to each profile, the values of T^2 obtained are large and have high variability; hence a high number of signals is present. On the contrary, we observe the T^2 control chart by Zhang et al. [2016] does not signal any profile.

Table 4.9: Number of out-of-control profiles signaled by the different methods.

	Campaigns			
Methods	11	14	15	16
Parametric Gompertz model	2	0	9	6
GP model with heteroscedasticity with no correction	0	3	6	7
GP model with heteroscedasticity with Bonferroni correction ($z_{\frac{\alpha}{2}/n}$)	0	0	0	1
GP model with heteroscedasticity with Bonferroni varying correction ($z_{\frac{\alpha}{2}/j}$)	0	2	2	1
GP model by Wang et al. [2014] with averaged estimates	10	49	78	87
GP model by Wang et al. [2014] on all data	13	55	84	98
AGP model by AGP model	0	0	0	0
Number of total profiles*	13	55	84	98
*From the testing set.				

4.5 Summary and discussion

We observe that all the methods using GP modeling are effective in the prediction of the mean profile which demonstrates that when having replications the GP model (either with constant variance or without) will converge to the true unknown model. This is based on the assumption that the data is IC. Further research has to be done to study the performance of this method in Phase I analysis where some anomalies are present.

In this chapter we suggest the use of the mean profile and the predictive band obtained from a GP model with heteroscedasticity as basis of a Shewhart-type control chart to monitor the within and between profile variation. Compared to any T^2 control chart, this chart has a huge advantage given that it can be applied to monitor within the profile and while the most recent profile data are being collected.

Our results indicate that the use of the T^2 chart to monitor the hyper-parameters is not a good choice given that these parameter estimates are tuned to the data, and hence sensitive to it. As a result, a high percentage of false alarms is present. On the other hand, the T^2 chart proposed by Zhang et al. [2016], which is based on the profile data, is a better choice. Although it appears that this chart would work well if the number of observations per profile is small (up to 30). However, it requires all the data for each profile to determine if the process is in or out of control.

If the profiles are produced in order and the purpose of the analysis is only monitoring between profiles, a much better choice is the use of an univariate individual chart based on a deviation metric as suggested by Williams et al. [2007]. We show that this chart has better performance and does not depend on either the number of unique points a profile has or the number of profiles used from Phase I.

In one of our simulation studies, in which the OOC scenarios are created by adding shifts to the hyper-parameters of the IC model, our method fails to detect shifts in the length-scale which models the spatial correlation among the points. Given that those shifts make the profile more or less smooth but the profiles tend to remain within the prediction band, the changes are not detected by our method. Generally, signals that result from using a chart based on the prediction band are associated with practical significance for which this method, as shown in this chapter, performs well. We recognize that there could be special cases in which the focus needs to be on the parameters such as the length-scale hyper-parameter; for this, certainly the approach we propose in this chapter will not be appropriate. In Chapter

5.2 we propose another procedure in which the focus is on the parameters.

We explore two simple approaches to deal with over-sampled profiles: i) The *systematic selecting approach* which systematically random selects observations per profile and performs the monitoring scheme on these, and ii) the *sub-grouping approach* which borrows Shewhart's concept of rational subgroup by applying our approach on the average of each subgroup. We show that the latter method performs well while the former yields higher false alarm rates for the scenarios analyzed here. It could be that the use of the *systematic selecting approach* entails loss of location information; more research needs to be done in this regard.

In regard to the *sub-grouping approach*, if the number of subgroups is small, the Bonferroni correction works well, but if the number of subgroups is higher up to 1000 (up to 5000 observations with a subgroup size of 5) our method with the varying correction works best. It appears that the *sub-grouping approach* does not depend on either the number of observations per profile used in the IC phase or the subgroup size used. Even though these results are pointing in the right direction, more research needs to be done in this regard such as testing more options for subgroup sizes, higher number of observations per profile, and different locations for the *systematic selecting approach* or *sub-grouping approach*.

Chapter 5

Proposed Method for Monitoring Non-Linear Profiles

The current literature on parametric profile monitoring tends to focus on the estimated parameters of the non-linear model. As a result, they often do not take full advantage of the model structure. An alternative approach in the literature uses non-parametric modeling in profile monitoring which in most cases does focus specifically on the quality characteristic of interest.

This chapter proposes an efficient control chart procedure based on the estimated parameters which are the result of a model that accounts for the within and between profile correlation. We focus our attention to those cases where a parametric model is appropriate.

5.1 Review of non-linear mixed model and estimation

The general form of a non-linear mixed model (NLMM) is given by

$$y_{ij} = f(x_{ij}, \boldsymbol{\theta}, \eta_i) + \epsilon_{ij}; \quad i = \{1 \dots m\}, j = \{1 \dots n_i\},$$

where $f(\cdot)$ is the non-linear function with a predictor x_{ij} for the jth observation for ith profile, $\boldsymbol{\theta}$ is the px1 vector of fixed parameters, and η_i is the vector of random effects for the ith profile. It is assumed that $\eta_i \sim N(0, \sigma_\eta^2 \psi_i)$, and $\epsilon_{ij} \sim N(0, \sigma^2)$.

For the entire data, this model can be expressed as

$$\mathbf{Y} = f(X, \boldsymbol{\theta}, \eta) + \boldsymbol{\epsilon}, \tag{5.1}$$

where $\eta \sim N(0, \sigma_\eta^2 \Psi)$ with $\Psi = diag(\psi_1, \dots \psi_m)$ and $\boldsymbol{\epsilon} \sim N(0, \sigma^2 I_N)$ with $N = \sum n_i$.

The usual approach for finding $\boldsymbol{\theta}$ linearizes the non-linear model followed by an iterative estimation method [Lindstrom and Bates, 1990]. Applying Taylor series approximation to $f(\cdot)$ near a current estimate $\hat{\boldsymbol{\theta}}$ and $\hat{\eta}$, we get

$$f(X, \boldsymbol{\theta}, \eta) \approx f(X, \hat{\boldsymbol{\theta}}, \hat{\eta}) + D(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) + Z(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}}) \tag{5.2}$$

where D and Z are the matrices of derivatives defined as follows:

$$D = \frac{\partial f}{\partial \boldsymbol{\theta}}\Big|_{\substack{\hat{\boldsymbol{\theta}} \\ \hat{\eta}_i}}$$

$$Z = diag(Z_1 \ldots Z_m); \quad Z_i = \frac{\partial f}{\partial \eta_i}\Big|_{\substack{\hat{\boldsymbol{\theta}} \\ \hat{\eta}_i}}.$$

From 5.1 and 5.2, we have

$$\mathbf{Y} - f(X, \hat{\boldsymbol{\theta}}, \hat{\eta}) + D\hat{\boldsymbol{\theta}} + Z\hat{\boldsymbol{\eta}} \approx D\boldsymbol{\theta} + Z\boldsymbol{\eta} + \boldsymbol{\epsilon}.$$

Consider $W = \mathbf{Y} - f(X, \hat{\boldsymbol{\theta}}, \hat{\eta}) + D\hat{\boldsymbol{\theta}} + Z\hat{\boldsymbol{\eta}}$, then,

$$W \approx D\boldsymbol{\theta} + Z\boldsymbol{\eta} + \boldsymbol{\epsilon}, \tag{5.3}$$

where $W \sim N(D\boldsymbol{\theta}, \Sigma); \quad \Sigma = Z\Psi Z^t + \sigma^2 I_N$

From this so-called pseudo model, we get the MLE estimates for the variance components and $\boldsymbol{\theta}$, then we calculate the estimates for $\hat{\eta}$, and the pseudo responses W are recalculated. This procedure is repeated until convergence (Lindstrom and Bates [1990] and Williams et al. [2015]).

Fitting a non-linear mixed model can be easily performed in SAS or R. There are several functions in R; however, *nlme*, which uses the two-step algorithm proposed by Lindstrom and Bates [1990], seems most popular. According to Stegmann et al. [2018], the *nlme* function is the fastest in terms of computing time compared to the other R functions that produce equivalent estimation results. We use this function for our evaluations shown later in this chapter.

5.2 Proposed procedure to monitor profiles

The profile monitoring literature on parametric modeling suggests the use of multivariate T^2 statistics to monitor the stability of the parameters simultaneously as we reviewed in Chapter 2. All these approaches focus on the parameters of the non-linear model and treat them as separate, but correlated, quality characteristics of the process. By doing so, the correlation structure dictated by the model is ignored. In this section, we propose a procedure based on a non-linear mixed model structure. This procedure recognizes and uses the correlations among the estimates of the fixed parameters created by the model.

Consider partitioning the data into two sets in which the first one contains the data from one profile (say the kth profile) and the second one contains the data from the rest of the profiles, as follows:

$$y_{ij} = f(x_i, \boldsymbol{\theta_k} * (1 - \delta_k) + \boldsymbol{\theta_{-k}} * \delta_k, \eta_k * (1 - \delta_k) + \eta_{-k} * \delta_k) + \epsilon_{ij} \tag{5.4}$$

where $\delta_k = 0$ for those points belonging to the kth profile and $\delta_k = 1$ if they do not belong to it. Also, $\boldsymbol{\theta}_k$ is the px1 vector of parameters of the kth profile, and $\boldsymbol{\theta}_{-k}$ is that one of the rest of the profiles.

Based on the model shown in Equation 5.4, our interest is to test whether the fixed parameters of the kth profile, $\boldsymbol{\theta}_k$, are different from those of all the other profiles, $\boldsymbol{\theta}_{-k}$, as follows:

$$H_0 : \theta_{k1} - \theta_{-k1} = \theta_{k2} - \theta_{-k2} = \ldots \theta_{kp} - \theta_{-kp} = 0.$$

We can re-express this hypothesis as the following:

$$H_0 : C\boldsymbol{\theta} = 0$$

$$H_1 : C\boldsymbol{\theta} \neq 0,$$

where

$$C\boldsymbol{\theta} = \begin{bmatrix} I_p & -I_p \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta}_k \\ \boldsymbol{\theta}_{-k} \end{bmatrix}.$$

Given the model shown in Equation 5.4, it is easy to see that Equation 5.3 can be partitioned as follows

$$\begin{bmatrix} W_k \\ W_{-k} \end{bmatrix} \sim N\left(\begin{bmatrix} D_k\boldsymbol{\theta}_k \\ D_{-k}\boldsymbol{\theta}_{-k} \end{bmatrix}, \begin{bmatrix} \Sigma_k & 0 \\ 0 & \Sigma_{-k} \end{bmatrix} \right),$$

where

$$\begin{bmatrix} \Sigma_k & 0 \\ 0 & \Sigma_{-k} \end{bmatrix} = \begin{bmatrix} Z_k\psi_k Z_k^t + \sigma^2 I_{n_k} & 0 \\ 0 & Z_{-k}\Psi_{-k}Z_{-k}^t + \sigma^2 I_{-k} \end{bmatrix}.$$

When this algorithm reaches convergence, we obtain the estimates for the $2p$ vector of parameters $\boldsymbol{\theta} = (\boldsymbol{\theta_k}, \boldsymbol{\theta_{-k}})^t$. We note that $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\theta}}_k, \hat{\boldsymbol{\theta}}_{-k})^t$ and that $\hat{\boldsymbol{\theta}}_k$ and $\hat{\boldsymbol{\theta}}_{-k}$ are asymptotically independent. Since the asymptotic variance of $\hat{\boldsymbol{\theta}}$ is

$$\begin{bmatrix} (D_k^t \Sigma_k D_k) & 0 \\ 0 & (D_{-k}^t \Sigma_{-k} D_{-k}) \end{bmatrix}^{-1},$$

the variance of $C\hat{\boldsymbol{\theta}}$ is

$$\begin{bmatrix} I_p & -I_p \end{bmatrix} \begin{bmatrix} (D_k^t \Sigma_k D_k) & 0 \\ 0 & (D_{-k}^t \Sigma_{-k} D_{-k}) \end{bmatrix}^{-1} \begin{bmatrix} I_p \\ -I_p \end{bmatrix},$$

which results in

$$(D_k^t \Sigma_k D_k)^{-1} + (D_{-k}^t \Sigma_{-k} D_{-k})^{-1}.$$

Then the resulting statistic is given by

$$\chi^2 = (C\hat{\boldsymbol{\theta}})' \left[(D_k^t \hat{\Sigma}_k D_k)^{-1} + (D_{-k}^t \hat{\Sigma}_{-k} D_{-k})^{-1} \right]^{-1} (C\hat{\boldsymbol{\theta}}) \tag{5.5}$$

which asymptotically follows a $\chi^2(1-\alpha, p)$. We note that this statistic can be used in either Phase I or Phase II.

In this chapter we evaluate the properties and performance of the proposed statistic through a simulation study assuming first an IC process and then an OOC process. For evaluating any method's performance in Phase II, most of the literature on profile monitoring simulate

OOC situations by adding a shift in the values of the parameters from the IC model. Too often, these shifts focus on the statistical significance, not practical significance.

In this study, we propose first to evaluate how the OOC changes in the response (the actual quality characteristic) bias the parameter estimates, and second to use that result to create practical meaningful shifts for our Phase-II performance evaluation.

5.2.1 Estimation of the asymptotic bias

The bias is the result of not fitting the additional regressors, X_2, that actually are important to the model. In the statistical process control setting, the bias is the result of the "assignable causes". Assume that the difference between the true and the presumed models, which typically is non-linear, is small enough, then we can use a first-order Taylor series in the regressors not fit to explain the difference. In this case, we fit the model shown in Equation 5.1, when the true model is

$$\mathbf{Y} = f(X, \boldsymbol{\theta}, \eta) + X_2\boldsymbol{\beta} + \boldsymbol{\epsilon}. \tag{5.6}$$

From 5.3, it is easy to see that $\hat{\theta} = \left(D^t\Sigma^{-1}D\right)^{-1}D^t\Sigma^{-1}W$, where W are the pseudo responses as seen in section 5.1, in an iterative procedure.

Since the true model has the extra term $X_2\boldsymbol{\beta}$, then the asymptotic bias is

$$\textbf{asym bias}(\hat{\boldsymbol{\theta}}) = (D^t\Sigma^{-1}D)^{-1}D^t\Sigma^{-1}X_2\boldsymbol{\beta}. \tag{5.7}$$

We use this result later in this chapter to create meaningful OOC situations in our simulation study.

5.3 Simulation study

The present study focuses on a situation where there is profile-to-profile variability that can be modeled by the following general form:

$$y_{ij} = f(x_{ij}, \boldsymbol{\theta}) + \eta_i + \epsilon_{ij}, \tag{5.8}$$

where $f(x_{ij}, \boldsymbol{\theta})$ is the non-linear function with $i = \{1 \dots m\}$, $j = \{1 \dots n_i\}$, and $\eta_i \sim N(0, \sigma_\eta^2)$ and $\epsilon_{ij} \sim N(0, \sigma^2)$.

In matrix form, we have

$$\mathbf{Y} = f(X, \boldsymbol{\theta}) + M\eta + \boldsymbol{\epsilon},$$

where

$$M = \begin{bmatrix} 1_{n_1} & 0 & \dots & 0 \\ 0 & 1_{n_2} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1_{n_m} \end{bmatrix}.$$

As seen in the previous section the general form of the covariance matrix is $\Sigma = Z\Psi Z^t + \sigma^2 I_N$. For this simplifying model version, that covariance matrix results in: $\Sigma = \sigma_b^2 MM^t + \sigma^2 I_N$.

There is a wide number of non-linear functions for different applications [Bates and Watts, 2007]. We consider the Gompertz function which has been used to study the growth of shrimp in Chapter 3 to provide an appropriate context to illustrate the proposed methodology:

$$y_{ij} = \theta_{01} - \theta_{02} * exp(\theta_{03} * t_{ij}) + \eta_i + \epsilon_{ij} \tag{5.9}$$

where y_{ij} is the average log shrimp weight for the ith pond where $i = \{1, \ldots m\}$ for the jth week where $j = \{1, \ldots n\}$, $\eta_i \sim N(0, \sigma_\eta^2)$, and $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$.

Consider the following IC parameters for generating our IC process: $\hat{\boldsymbol{\theta}}_{\mathbf{0}} = \{3.4612, 3.4624, -0.1173\}$, $\hat{\sigma}_\eta^2 = 0.01$, and $\hat{\sigma}_\epsilon^2 = 0.01$. Figure 5.1 shows the generated profiles based on this function.

We produce the Phase I data by generating profiles based on this function, analyze different scenarios based on the number of profiles $m = \{7, 15, 20, 50\}$ generated in Phase I, and vary number of observations per profile $n = \{20, 30\}$. We assume that the process is IC for the entire IC period. We evaluate the performance our proposed statistic by calculating the false alarm rates and detection rates. To calculate the former, we generate a number of profiles without any anomaly based on the IC model shown in Equation 5.9, and calculate the percentage of profiles that were considered OOC for each method. For this situation, the type-I error of the test is expected to be close to the nominal value of α. For the detection rates calculations, we create different "abnormal" situations that are described in Section 5.3.2, and for each one we measure how well the method detects them.

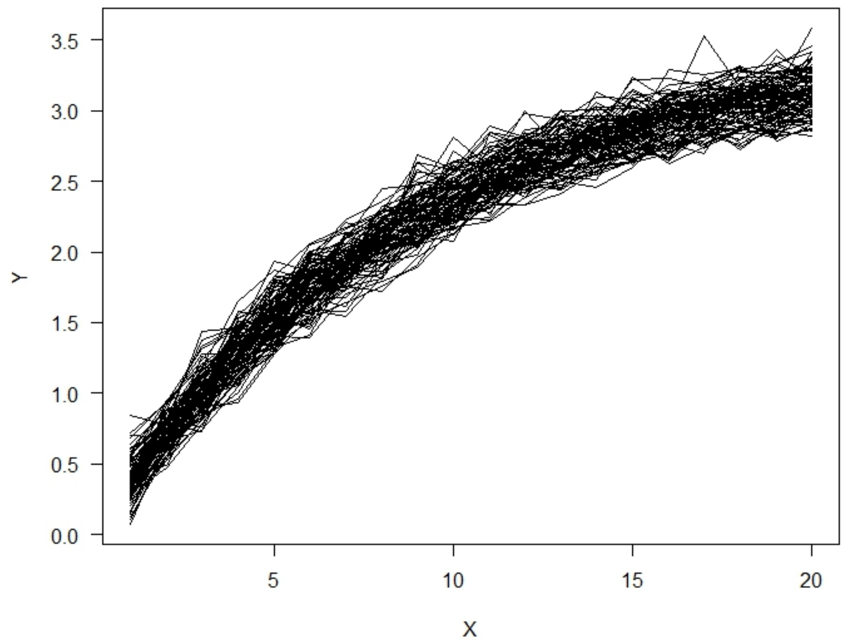

Figure 5.1: The lines represent 100 generated profiles.

5.3.1 False alarm rate performance

We evaluate our proposed statistic and compare it to the classical T^2 statistic with the sample covariance (S_1) and successive difference (S_2) as estimators for the covariance matrix of the fixed parameters as described in Chapter 2. The average false alarm rate for each method and for each scenario is calculated based on 10000 simulation runs. Table 5.1 shows these average rates with their standard deviation in parentheses at two cut-off values of $\alpha = \{0.01, 0.05\}$. Figures 5.2 and 5.3 show the distribution for each statistic for a given simulation when the number of profiles in Phase I (m) is 15 and 50 respectively.

We observe that the performance of our statistic is more robust than the other two in the

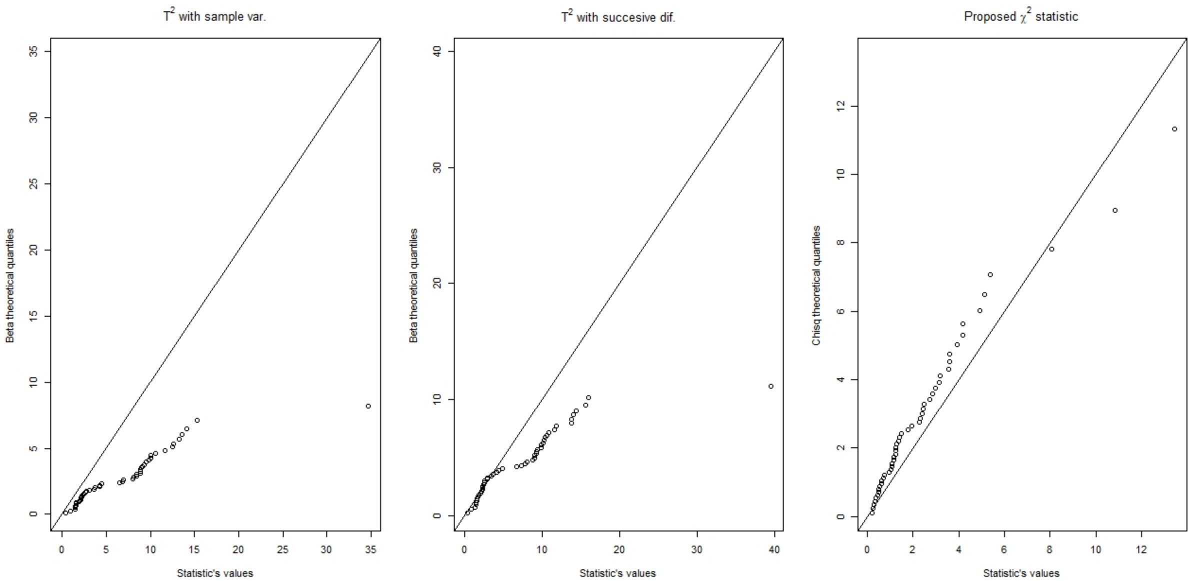

Figure 5.2: QQ plots of statistic's values for the different methods for a given simulation when $m = 15$.

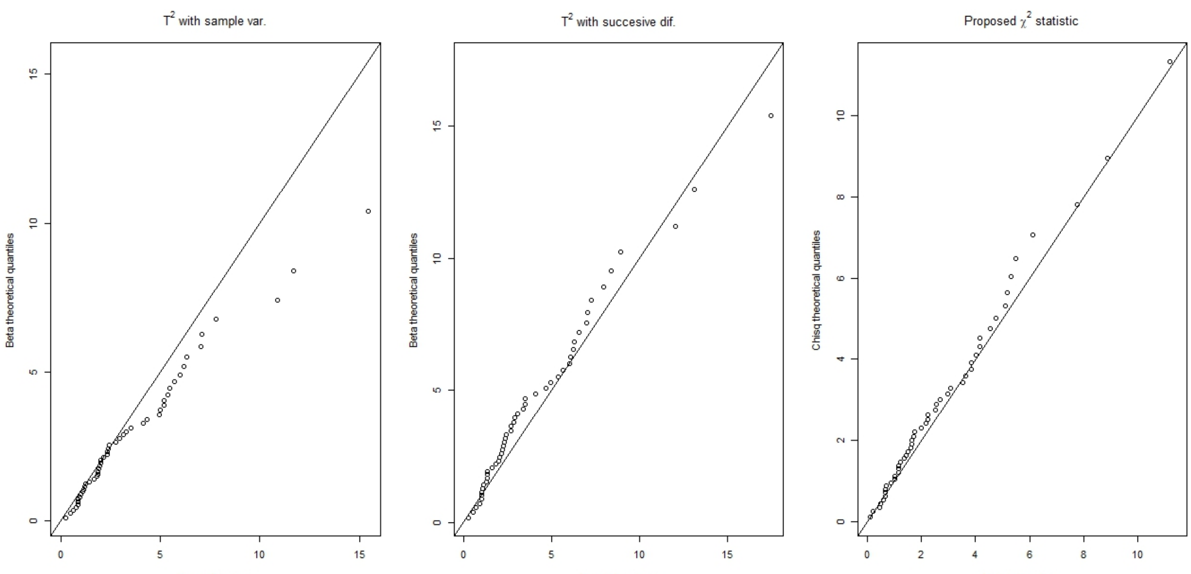

Figure 5.3: QQ plots of statistic's values for the different methods for a given simulation when $m = 50$.

sense that the percentage of false alarm is closer to the nominal value even for small sample of profiles used in Phase I. The T^2 control charts work well when a reasonable number of profiles in Phase I data is used; in our scenarios analyzed this number is 50. On the contrary, the T^2 charts do not show good performance when a small number of profiles in Phase I is used.

Table 5.1: Average false alarm rates (and their standard deviations in parentheses) for different scenarios for each method.

		$\alpha = 0.01$			$\alpha = 0.05$		
Number of profiles used in Phase I (m)	Number of obs. per profile (n)	$T^2(S_1)$	$T^2(S_2)$	Proposed χ^2 under NLMM	$T^2(S_1)$	$T^2(S_2)$	Proposed χ^2 under NLMM
7	20	0.495(0.187)	0.561(0.190)	0.028(0.060)	0.518(0.184)	0.561(0.190)	0.062(0.091)
	30	0.491(0.189)	0.558(0.192)	0.028(0.060)	0.515(0.185)	0.558(0.192)	0.061(0.090)
15	20	0.143(0.095)	0.111(0.094)	0.010(0.023)	0.215(0.113)	0.146(0.107)	0.033(0.045)
	30	0.142(0.097)	0.108(0.094)	0.009(0.022)	0.214(0.114)	0.144(0.108)	0.028(0.043)
20	20	0.091(0.069)	0.055(0.059)	0.006(0.017)	0.162(0.090)	0.091(0.076)	0.025(0.036)
	30	0.090(0.068)	0.053(0.057)	0.007(0.017)	0.162(0.090)	0.089(0.076)	0.025(0.035)
50	20	0.033(0.030)	0.010(0.016)	0.006(0.012)	0.088(0.050)	0.031(0.030)	0.030(0.029)
	30	0.030(0.028)	0.008(0.014)	0.005(0.010)	0.087(0.050)	0.028(0.030)	0.025(0.026)

5.3.2 Out-of-control performance

To evaluate our statistic's performance in detecting OOC profiles, we calculate the detection rates for different scenarios. To define each scenario, most of the current literature adds a shift to the estimates of the non-linear IC model parameters ($\hat{\boldsymbol{\theta}}$). Here we use the shifts or bias of these parameters' estimates shown in Table 5.2. These biases were estimated based on an analysis on the real data that is the basis for the simulated model shown in Equation 5.9. Instead of adding shifts to the non-linear model parameters arbitrarily, to simulate the assignable causes, the shifts were added to the regressors not fit in the IC model, and subsequently the bias of $\hat{\boldsymbol{\theta}}_{\mathbf{0}}$ were estimated through Equation 5.7. Here we use these asymptotic biases or shifts in $\hat{\boldsymbol{\theta}}_{\mathbf{0}}$ which are shown in Table 5.2 as OOC scenarios in our simulations. These are reflected in the profiles shown in Figure 5.4.

For each scenario, we ran 10000 simulations and calculate the detection rates for the T^2 charts and our χ^2 chart. Figures 5.5 and 5.6 show the detection rates' averages for each scenario analyzed when the number of profiles used in Phase I (m) is 15 and 50 respectively. Our proposed statistic appears to be more robust with respect to the number of profiles used in Phase I given that its performance is independent of m. When $m = 50$, we observe that on average the three methods perform very similar. However, when $m = 15$, the T^2 charts show high percentage of false alarms.

Table 5.2: Out-of-control behavior reflected on the asymptotic bias of the non-linear IC model parameters estimates ($\hat{\boldsymbol{\theta}}_{\mathbf{0}}$), based on shifts on the regressors not fit in the IC model.

	Shifts in average log. weight			Shifts in min. temperature			Shifts in min. oxygen			Shifts in min. pH		
δ	bias($\hat{\theta}_1$)	bias($\hat{\theta}_2$)	bias($\hat{\theta}_3$)	bias($\hat{\theta}_1$)	bias($\hat{\theta}_2$)	bias($\hat{\theta}_3$)	bias($\hat{\theta}_1$)	bias($\hat{\theta}_2$)	bias($\hat{\theta}_3$)	bias($\hat{\theta}_1$)	bias($\hat{\theta}_2$)	bias($\hat{\theta}_3$)
-3	-1.0377	-0.0017	-0.0019	-0.9511	-0.153	0.0083	-0.2051	0.2541	0.0012	-0.7506	0.1422	0.0014
-2	-0.6865	-0.0017	-0.0019	-0.6287	-0.1026	0.0049	-0.1314	0.1688	2e-04	-0.4951	0.0942	3e-04
-1	-0.3353	-0.0017	-0.0019	-0.3064	-0.0522	0.0015	-0.0578	0.0835	-8e-04	-0.2396	0.0463	-8e-04
0	0.0159	-0.0017	-0.0019	0.0159	-0.0017	-0.0019	0.0159	-0.0017	-0.0019	0.0159	-0.0017	-0.0019
1	0.3671	-0.0017	-0.0019	0.3382	0.0487	-0.0053	0.0896	-0.087	-0.0029	0.2714	-0.0497	-0.0029
2	0.7183	-0.0017	-0.0019	0.6605	0.0991	-0.0086	0.1632	-0.1722	-0.0039	0.5269	-0.0977	-0.004
3	1.0695	-0.0017	-0.0019	0.9829	0.1496	-0.012	0.2369	-0.2575	-0.0049	0.7824	-0.1457	-0.0051

The OOC scenarios considered previously were the result from adding shifts to one regressor not fit in the IC model at a time. Another set of OOC scenarios was created by performing a two-level full factorial design on the shifts added to the regressors not fit in the IC model, and subsequently the bias of $\hat{\boldsymbol{\theta}}_{\mathbf{0}}$ was estimated through Equation 5.7 for each scenario. Five of those scenarios, the ones that resulted closer to the IC mean profile, are considered in this analysis and are shown in Figure 5.7. Scenario 3 shown in the figure is the IC state of the process, scenarios 4 and 5 are closer to the IC state, and scenarios 1 and 2 are farther away

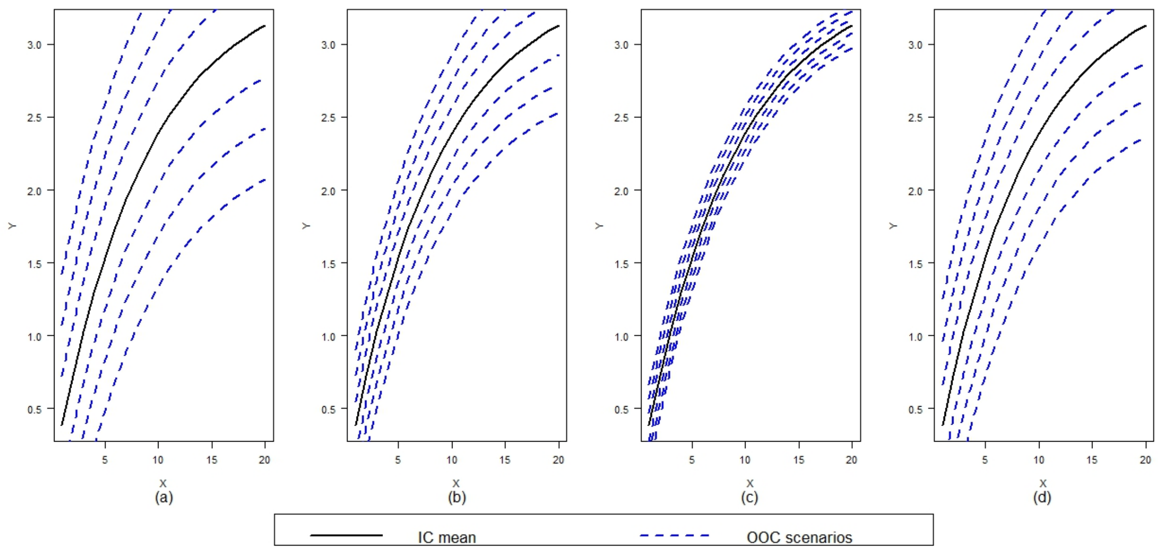

Figure 5.4: Four OOC scenarios: (a) shifts in average, (b) shifts in minimum temperature, (c) shifts in minimum oxygen, and (d) shifts in the minimum pH.

from it.

Figure 5.8 shows the detection rate performance of our method compared to the classical T^2 charts for these five scenarios for different number of profiles used in Phase I data (m). We observe that as m increases the performance of the T^2 charts tend toward a χ^2. For small m, our method shows better detection performance in the sense that it does not show high percentage of false alarms for scenarios closer to the IC state (scenarios 3 and 4). Moreover, as observed previously, our method shows more consistency given that it appears to be independent of the number of profiles used in Phase I.

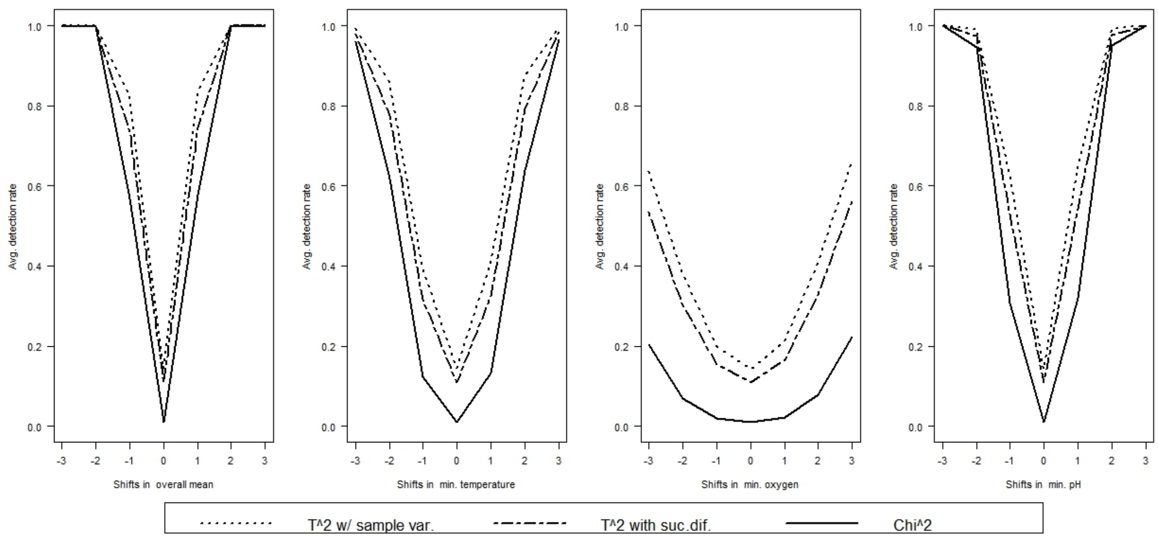

Figure 5.5: Average detection rate performance of the three methods for different OOC scenarios when $m = 15$.

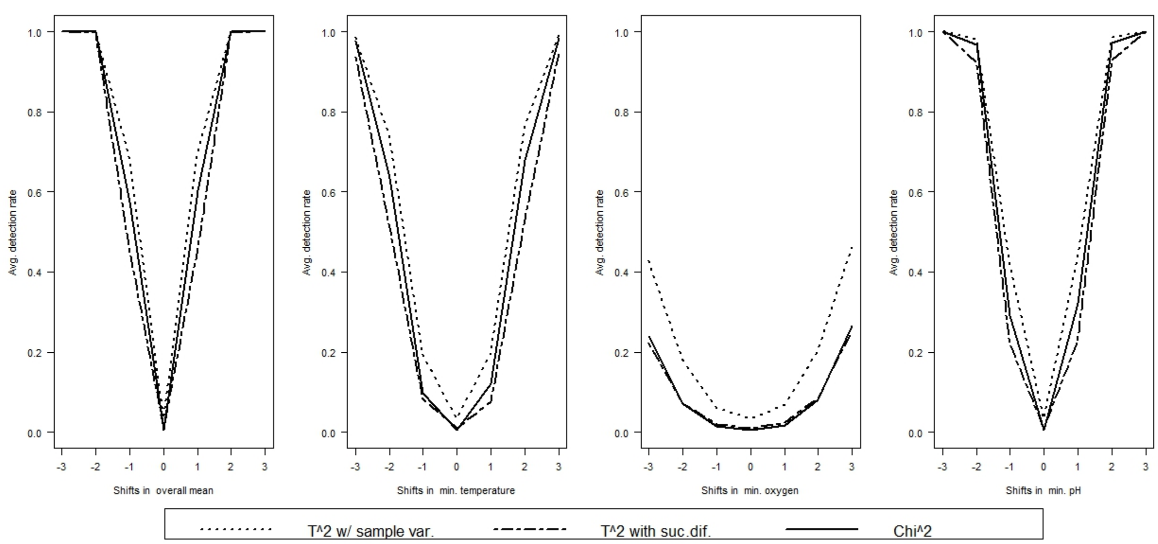

Figure 5.6: Average detection rate performance of the three methods for different OOC scenarios when $m = 50$.

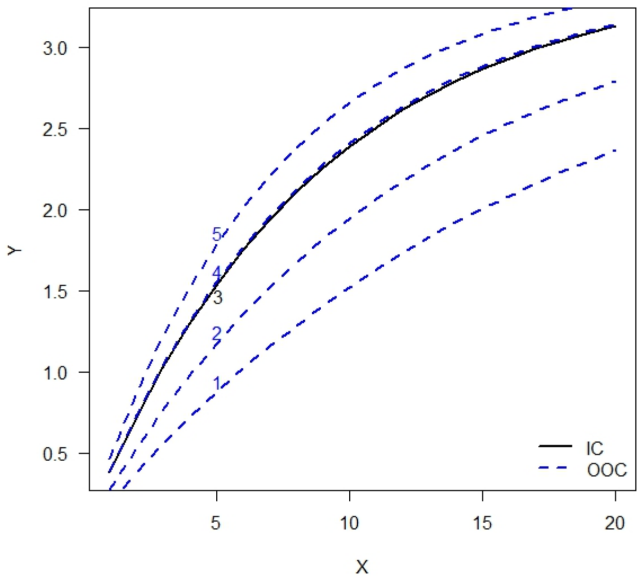

Figure 5.7: OOC scenarios.

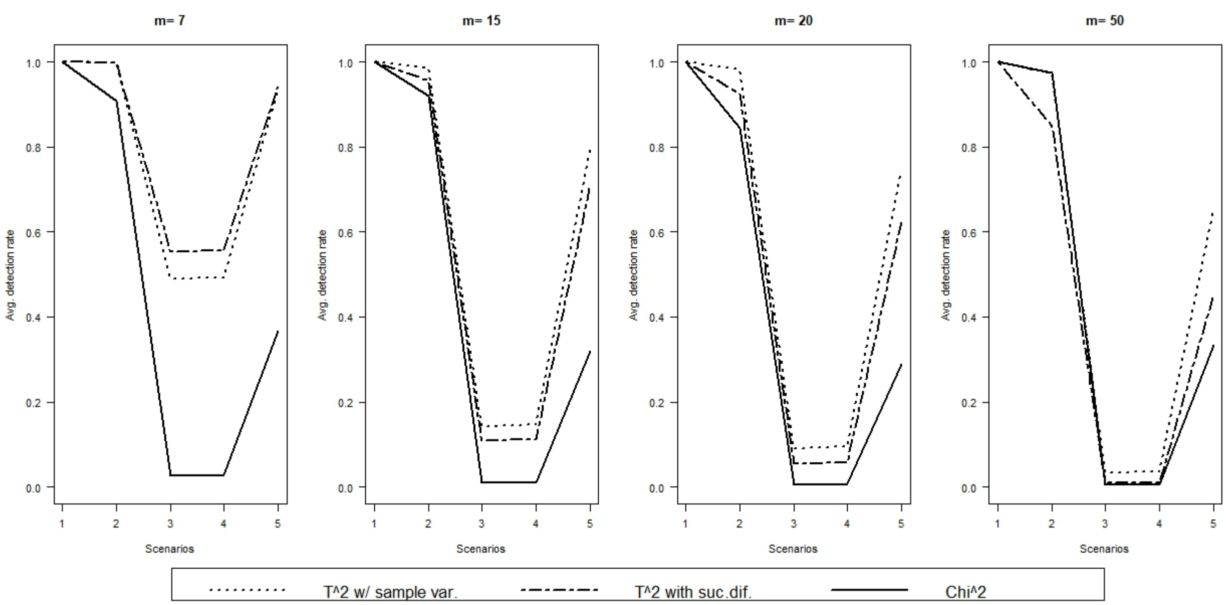

Figure 5.8: Detection rates performance of the three methods for different OOC scenarios.

5.3.3 Comparison with control chart on the predictive values

In this section we compare our χ^2 control chart and the T^2 charts to the Shewhart-type control chart on the predictive mean profile under two models: the parametric Gompertz model analyzed in Chapter 3 and the GP model proposed in Chapter 4. We evaluate the false alarm probabilities and detection rates for each method and for the same scenarios analyzed in Sections 5.3.1 and 5.3.2.

Table 5.3 shows the false alarm rates for the Shewhart-type control chart under both models. Although this chart with Bonferroni correction based on any of the models shows slightly higher percentage of false alarm when $m = 7$, we observe that in general it has good performance given that it does not have high percentage of false alarm when the number of profiles used in Phase I is small. For any number of profiles used in Phase I, the chart based on the parametric Gompertz model shows slightly better performance. Independent of the model used, this chart shows similar results to the χ^2 and better false alarm rate performance than the T^2 charts shown in Table 5.1, specially when the number of profiles used in Phase I is small.

Figure 5.9 shows a comparison of the detection performance between the Shewhart-type control chart with Bonferroni correction (under both models) and the χ^2 and T^2 charts analyzed in Sections 5.3.2. The Shewhart charts on the predictive mean profile and χ^2 chart presented earlier in this chapter appear to have a robust behavior with respect to the number of profiles used in Phase I, even as small as 7. Both Shewhart and χ^2 charts seem to have

Table 5.3: Average false alarm rates (and their standard deviations in parentheses) for different scenarios using Shewhart-type control chart on the predictive values based on a parametric model and a non-parametric model.

		$\alpha = 0.01$				$\alpha = 0.05$			
Number of profiles used in Phase I (m)	Number of obs. per profile (n)	Gompertz model	Gompertz model with Bonferroni correction	GP model	GP model with Bonferroni correction	Gompertz model	Gompertz model with Bonferroni correction	GP model	GP model with Bonferroni correction
7	20	0.198(0.137)	0.030(0.044)	0.258(0.161)	0.060(0.069)	0.507(0.188)	0.084(0.083)	0.565(0.196)	0.130(0.112)
	30	0.249(0.156)	0.033(0.047)	0.322(0.182)	0.068(0.076)	0.595(0.195)	0.087(0.085)	0.655(0.199)	0.140(0.119)
15	20	0.154(0.094)	0.017(0.025)	0.184(0.106)	0.027(0.034)	0.456(0.145)	0.057(0.053)	0.492(0.150)	0.076(0.064)
	30	0.196(0.110)	0.018(0.026)	0.232(0.120)	0.029(0.035)	0.543(0.154)	0.057(0.054)	0.582(0.154)	0.077(0.065)
20	20	0.146(0.085)	0.015(0.022)	0.167(0.092)	0.021(0.028)	0.447(0.132)	0.052(0.047)	0.472(0.134)	0.064(0.054)
	30	0.187(0.097)	0.015(0.022)	0.211(0.105)	0.022(0.028)	0.534(0.137)	0.051(0.047)	0.561(0.140)	0.065(0.054)
50	20	0.129(0.062)	0.011(0.016)	0.137(0.064)	0.013(0.018)	0.426(0.100)	0.041(0.034)	0.438(0.102)	0.047(0.036)
	30	0.164(0.071)	0.010(0.016)	0.177(0.074)	0.013(0.018)	0.510(0.105)	0.039(0.033)	0.527(0.106)	0.046(0.037)

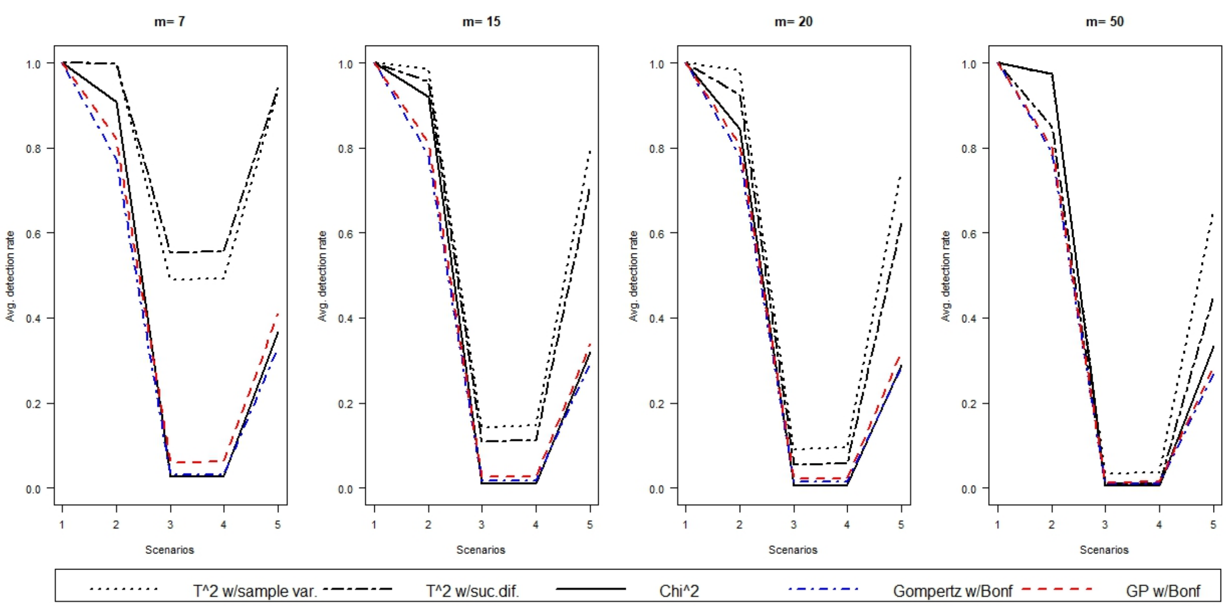

Figure 5.9: Average detection rate performance of the three methods for different OOC scenarios for different values of m.

similar good and robust performance. The advantage of the Shewhart chart on the predictive mean profile is that we can monitor the process while the profile is evolving, and hence, do not need to wait until the end of the profile formation to determine the state of the process.

5.4 Summary and conclusions

In this chapter we propose a χ^2 statistic for monitoring profiles as an alternative to the T^2 statistic. This χ^2 statistic is based on a non-linear mixed model and can be used to test if a given profile is different from the rest of the profiles. In the monitoring phase, "the rest of the profiles" represent the IC profiles. This test considers the estimated asymptotic variance of the fixed parameters obtained using the actual data, not the parameters estimates themselves. Our results based on false alarm rates and detection rates show that the performance of our proposed statistic is robust even for small number of profiles used in Phase I, while the classical T^2 control charts works well when the number of profiles in Phase I is at least 50.

Even though the χ^2 chart is conceptually more robust as shown in Section 5.2, and shows good performance even for small sample of profiles used in Phase I, its focus is on detecting signals based on the statistical significance of the estimated parameters. In addition, we would need to wait until the profile ends for determining the state of the process. To address these issues we show a methodology in Chapters 3 and 4 for applying a Shewhart control chart based on a predictive mean profile in which the prediction band would act as control limits. In this chapter we compare our χ^2 control chart and the classical T^2 charts to the Shewhart-type control chart proposed in previous chapters. Both χ^2 and Shewhart charts show similar robust performance with respect to the number of profiles used in Phase I. The T^2 charts show high percentage of false alarms when there is a small sample of profiles from

Phase I.

A Shewhart-type control chart where the central line is the predictive mean profile and the prediction band becomes the control limits has two clear advantages. The first is that it has more practical significance since it allows us to monitor the observed data directly instead of the estimated parameters. The second advantage is that there is no need to wait until the profile ends to monitor the process, it allows us to monitor the process as the profile is being formed. In addition, as we show in this chapter, the Shewhart chart on the predictive values shows comparable results to the χ^2 chart that is conceptually robust and shows better performance than the classical T^2 charts.

Chapter 6

Summary and Next Steps

The goal of this research is to provide alternatives to study two main concerns in the current literature on profile monitoring in Phase II. First, most of them do not monitor within profiles which does not allow the quality engineer to make corrections at early stage of the profile. Second, most of the current approaches focuses on the parameters or features of the model instead of the quality characteristic itself.

Chapter 3 introduces a practical situation in which the two main concerns outlined above are critical for its quality monitoring. We propose a two-stage parametric approach to first model the average non-linear behavior and then, based on its residuals, to model the non-constant variance function. Based on this model, we propose the use of a Shewhart-type control chart to monitor the profiles over time and the campaigns. For this specific case study, we illustrate a statistical process control where the real issue is practical significance since the economic consequences are important for the quality engineers. This application gave us the insight for proposing more generalized procedures to deal with these two main concerns outlined earlier.

One of the goals of this dissertation is to show a procedure for monitoring within the profiles based on a predictive mean profile. We started this analysis with a specific application in Chapter 3. However, not always the researchers have a parametric model that expresses the complex relationships between the variables. Chapter 4 shows that when the profiles have replications, the Gaussian Process modeling is effective in the prediction of the mean profile since it will converge to the true unknown model. GP model has a huge advantage over any other non-parametric technique because it naturally provides both the prediction mean and variance.

We proposed to use both as a basis for monitoring within the profiles through a Shewhart type control chart. However, one of the challenges of monitoring a profile as its data are being collected is the multiple testing issue.

To deal with this issue, we explored two approaches to deal with profiles that consist of many measurements to which we called "over-sampled" profiles. Based on our results, the *sub-grouping approach* works best. This approach suggests to calculate the average of each subgroup and apply our method on only these averages. When the number of subgroups is as high as to 1000, which means that we can have observations up to 5000, our Shewhart chart with Bonferroni varying correction works well.

In Chapter 5 we propose a χ^2 statistic as an alternative to the T^2 statistic. Our statistic tests if a given profile is behaving different from the rest of the profiles which in the monitoring phase these become the IC profiles. This statistic is based on a non-linear mixed model that uses the covariance matrix of the parameters created by the model. Currently, most of the techniques use different estimators for this covariance matrix that are based on the parameters estimates themselves. With our proposed statistic, we show that the quality characteristic is the basis for these covariance matrix and hence of the monitoring scheme. Our results show that our statistic is more robust even for very small number of profiles used in Phase I.

In our simulations, we have proven that a Shewhart-type control chart based on the preditive mean profile and prediction band has good properties for monitoring profiles in Phase II. Its performance is comparable to the T^2 chart when the number of profiles used in Phase I is higher enough. However, when the number of profiles used in Phase I is as small as 7, the Shewhart control chart works by far better than the T^2 charts, and as robust as the χ^2 chart proposed in Chapter 5.2. Additionally, the Shewhart-control chart has the advantage that it can monitor the profile as the profile data are being collected. The quality manager does not need to wait until the profile ends to determine the state of the process.

Bibliography

B. Ankenman, B. L. Nelson, and J. Staum. Stochastic kriging for simulation metamodeling. *Operations Research*, 58:371–382, 2010.

M. Araneda, J. Hernandez, E. Gasca-Leyva, and M. Vela. Growth modelling including size heterogeneity: Application to the intensive culture of white shrimp (p. vannamei) in freshwater. *Aquacultural Engineering*, 56:1–12, 2013.

D. Bates and D. Watts. *Nonlinear Regression Analysis and Its Applications.* John Wiley and Sons, Hoboken, NJ. USA., 2007.

R. Bellio, J. Jensen, and P. Seiden. Applications of likelihood asymptotics for nonlinear regression in herbicide bioassays. *Biometrics*, 56:1204–1212, 2000.

M. Binois and R. B. Gramacy. hetgp: Heteroskedastic Gaussian process modeling and design under replication. *R package version 1.0.2*, 2017.

M. Binois, R. B. Gramacy, and M. Ludkovski. Practical heteroskedastic Gaussian process modeling for large simulation experiments. *Journal of Computational and Graphical*

Statistics, 0:1–41, 2018. doi: 10.1080/10618600.2018.1458625. URL https://doi.org/10.1080/10618600.2018.1458625.

P. Q. A. D. Boeing Commercial Airplane Group, Material Division. Advanced quality system tools, AQS D1-9000-1.

R. Carroll and D. Ruppert. *Transformation and Weighting in Regression*. 1988.

S. I. Chang and S. Yadama. Statistical process control for monitoring non-linear profiles using wavelet filtering and B-spline approximation. *International Journal of Production Research*, 48(4):1049–1068, 2010.

B. Colosimo and M. Pacella. A comparison study of control charts for statistical monitoring of functional data. *International Journal of Production Research*, 48 (6):1575–1601, 2010.

M. Davidian and R. Carroll. Variance function estimation. *Journal of the American Statistical Association*, 82:1079–1091, 1987.

M. Davidian, R. Carroll, and W. Smith. Variance functions and the minimum detectable concentration in assays. *Biometrika*, 75(3):549–556, 1988.

Y. Ding, L. Zeng, and S. Zhou. Phase I analysis for monitoring nonlinear profiles in manufacturing processes. *Journal of Quality Technology*, 38 (3):199–216, 2006.

R. Ganesan, T. K. DAS, and V. Venkataraman. Wavelet-based multiscale statistical process monitoring: A literature review. *IIE Transactions*, 36(9):787–806, 2004.

M. Gardner, L.-C. Lu, R. Gyuresik, J. Wortman, B. Hornung, H. Heinish, E. Rying, S. Rao, and J. David. Equipment fault detection using spatial signatures. *IEEE Transactions on Components, Packaging, and Manufacturing Technology-Part C*, 20:295–304, 1997.

R. Gnanadesikan and J. R. Kettenring. Robust estimates, residuals, and outlier detection with multiresponse data. *Biometrics*, 28(1):81–124, 1972.

R. Goedhart, M. Schoonhoven, and R. Does. Guaranted in-control performance for the Shewhart X and $\bar{X}$ control charts. *Journal of Quality Technology*, 49(2):155–171, 2017.

P. W. Golberg, C. K. Williams, and C. M. Bishop. Regression with input-dependent noise: A Gaussian process treatment. *Advances in Neural Information Processing Systems*, 10: 493–499, 1998.

D. S. Holmes and A. E. Mergen. Improving the performance of the T^2 control chart. *Quality Engineering*, 5(4):619–625, 1993.

J. E. Jackson. *A user's guide to principal components.* Wiley series in probability and mathematical statistics. Applied probability and statistics. New York : Wiley, c1991., 1991. ISBN 0471622672.

J. E. Jackson and G. S. Mudholkar. Control procedures for residuals associated with principal component analysis. *Technometrics*, 21(3):341–349, 1979.

W. A. Jensen and J. B. Birch. Profile monitoring via nonlinear mixed models. *Journal of Quality Technology*, 41 (1):18–34, 2009.

W. A. Jensen, L. A. Jones-Farmer, C. W. Champ, and W. H. Woodall. Effects of parameter estimation on control chart properties: a literature review. *Journal of Quality Technology*, 38:349–364, 2006.

W. A. Jensen, J. B. Birch, and W. H. Woodall. High breakdown estimation methods for phase I multivariate control charts. *Quality and Reliability Engineering International*, 23 (5):615–629, 2007.

W. A. Jensen, J. B. Birch, and W. H. Woodall. Monitoring correlation within linear profiles using mixed models. *Journal of Quality Technology*, 40(2):167, 2008.

M. K. Jeong, J.-C. Lu, X. Huo, B. Vidakovic, and D. Chen. Wavelet-based data reduction techniques for process fault detection. *Technometrics*, 48(1):26–40, 2006.

J. Jin and J. Shi. Automatic feature extraction of waveform signals for in-process diagnostic performance improvement. *Journal of Intelligent Manufacturing*, 12(3):257–268, 2001.

L. Kang and S. L. Albin. On-line monitoring when the process yields a linear profile. *Journal of quality Technology*, 32(4):418, 2000.

K. Kersting, C. Plagemann, P. Pfaff, and W. Burgard. Most likely heteroscedastic Gaussian process regression. In *Proceedings of the 24th International Conference on Machine Learning*, ICML '07, pages 393–400, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-793-3.

K. Kim, M. A. Mahmoud, and W. H. Woodall. On the monitoring of linear profiles. *Journal of Quality Technology*, 35(3):317, 2003.

M. Lázaro-Gredilla and M. K. Titsias. Variational heteroscedastic Gaussian process regression. In *Proceedings of the 28th International Conference on International Conference on Machine Learning*, ICML'11, pages 841–848, USA, 2011. Omnipress. ISBN 978-1-4503-0619-5.

B. Leung. The oversampling technique for analog to digital conversion: A tutorial overview. *Analog Integrated Circuits and Signal Processing*, 1(1):65–74, 1991.

M. J. Lindstrom and D. M. Bates. Nonlinear mixed effects models for repeated measures data. *International Biometric Society*, 46(3):673–687, 1990.

P. Medawar. The growth, growth energy, and ageing of the chicken's heart. *Proceedings of Royal Society B*, pages 332–355, 1940.

R. Noorossana, M. Eyvazian, A. Amiri, and M. A. Mahmoud. Statistical monitoring of multivariate multiple linear regression profiles in phase I with calibration application. *Quality and Reliability Engineering International*, 26(3):291–303, 2010.

R. Noorossana, A. Saghaei, and A. Amiri. *Statistical analysis of profile monitoring*, volume 865. John Wiley & Sons, 2011.

P. Qiu. *Introduction to statistical process control*. CRC Press, 2013.

P. Qiu and D. Xiang. Univariate dynamic screening system: An approach for identifying individuals with irregular longitudinal behavior. *Technometrics*, 56(2):248–260, 2014. doi: 10.1080/00401706.2013.822423. URL `https://doi.org/10.1080/00401706.2013.822423`.

P. Qiu, C. Zou, and Z. Wang. Nonparametric profile monitoring by mixed effects modeling. *Technometrics*, 52(3):265–277, 2010.

N. Quadrianto, K. Kersting, M. D. Reid, T. S. Caetano, and W. L. Buntine. Kernel conditional quantile estimation via reduction revisited. In *2009 Ninth IEEE International Conference on Data Mining*, pages 938–943, Dec 2009.

V. Quevedo, S. Vegas, and G. Vining. A tutorial on an iterative approach for generating Shewhart control limits. *Quality Engineering*, 28 (3):305–312, 2016.

C. Rasmussen and C. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006. ISBN ISBN 026218253X.

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

G. Seber and C. Wild. *Nonlinear Regression*. John Wiley and Sons, Hoboken, NJ. USA., 2003.

W. A. Shewhart. Economic quality control of manufactured product. *Bell System Technical Journal*, 9(2):364–389, 1930.

G. Stegmann, R. Jacobucci, J. R. Harring, and K. J. Grimm. Nonlinear mixed-effects modeling programs in r. *Structural Equation Modeling: A Multidisciplinary Journal*, 25 (1):160–165, 2018.

J. Sullivan and W. H. Woodall. A comparison of multivariate control charts for individual observations. *Journal of Quality Technology*, 28:398–408, 1996.

X. Tian, P. Leung, and E. Hochman. Shrimp growth functions and their economic implications. *Aquacultural Engineering*, 12(2):81–96, 1993.

J. A. Vargas. Robust estimation in multivariate control charts for individual observations. *Journal of Quality Technology*, 35(4):367–376, 2003.

G. Vining and R. Myers. Combining taguchi and response philosophies: A dual response approach. *Journal of Quality Technology*, 22(1):38–45, 1990.

E. Walker and S. P. Wright. Comparing curves using additive models. *Journal of Quality Technology*, 34:118–129, 2000.

A. Wang, K. Wang, and F. Tsung. Statistical surface monitoring by spatial-structure modeling. *Journal of Quality Technology*, 46:359–376, 2014.

J. D. Williams, W. H. Woodall, J. B. Birch, and J. H. Sullivan. Distribution of Hotelling's T^2 statistic based on the successive differences estimator. *Journal of Quality Technology*, 38(3):217–229, 2006.

J. D. Williams, W. H. Woodall, and J. B. Birch. Statistical monitoring of nonlinear product and process quality profiles. *Quality and Reliability Engineering International*, 23:925–941, 2007.

J. D. Williams, J. B. Birch, and A. S. Abdel-Salam. Outlier robust nonlinear mixed model estimation. *Statistics in Medicine*, 34(8):1304–1316, 2015.

W. H. Woodall. Current research on profile monitoring. *Production*, 17(3):420–425, 2007.

W. H. Woodall and F. W. Faltin. *An Overview and Perspective on Control Charting*, volume Chapter 2, pages 7–20. Marcel Dekker, New York, NY, 1996.

W. H. Woodall and F. W. Faltin. Rethinking control chart design and evaluation. 31(4): 596–605, 2019.

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

S. Wright. Reviews. [Review of the books *The Biology of Population Growth*, by R. Pearl, and *The Natural Increase of Mankind*, by J. S. Sweeney]. *Journal of the American Statistical Association*, 21(156):493–497, 1926. URL `http://www.jstor.org/stable/2276990`.

L. Yongxiang, Q. Zhou, X. Huang, and L. Zeng. Pairwise estimation of multivariate Gaussian process models with replicated observations: Application to multivariate profile monitoring. *Technometrics*, 60:70–78, 2018.

R. Yu, P. Leung, and P. Bienfang. Predicting shrimp growth: Artificial neural network versus nonlinear regression models. *Aquacultural Engineering*, 34(1):26–32, 2006.

L. Zhang, K. Wang, and N. Chen. Monitoring wafers' geometric quality using an additive Gaussian process model. *IIE Transactions*, 48:1–15, 2016.

Y. Zhang, Z. He, C. Zhang, and W. H. Woodall. Control charts for monitoring linear profiles with within-profile correlation using Gaussian process models. *Quality and Reliability Engineering International*, 30:487–501, 2014.

C. Zou, Y. Zhang, and Z. Wang. A control chart based on a change-point model for monitoring linear profiles. *IIE Transactions*, 38(12):1093–1103, 2006.

C. Zou, F. Tsung, and Z. Wang. Monitoring general linear profiles using multivariate exponentially weighted moving average schemes. *Technometrics*, 49(4):395–408, 2007.

C. Zou, F. Tsung, and Z. Wang. Monitoring profiles based on nonparametric regression methods. *Technometrics*, 50(4):512–526, 2008.