Influence of Biological Conditions to Temporal Gene Expressions Based on Variance Analysis

Dianliang Deng* Kian Razeghi Jahromi† Zaiying Zhou‡

Abstract

Temporal gene expression data are of particular interest to researchers as they contain rich information in characterization of gene functions and have been widely used in biomedical studies and cancer early detection. In contrast to the rich literature on how to estimate the gene expressions over the time under a given condition, few researchers consider identifying the different effects of multiple conditions on the gene expression profiles. Besides its intrinsic effect, a gene has various expression patterns under different biological conditions and these conditions result in the variation of gene expression variance. In this paper, we will investigate the effects of conditions to the gene expressions and then classify the conditions according to the variational variance functions of gene expressions. We propose a non-linear regression model with log-normal distribution to characterize the variance functions of genes under the given conditions. Then, based on the parameter estimates, a chi-square test is proposed to test the equality of variance functions for different conditions. Furthermore, the Mahalanobis distance is used for the classification of conditions. The proposed methods are applied to the dataset of 18 genes in *P. aeruginosa* expressed in 24 biological conditions. The simulation studies show that our methods are well performed for the classification of conditions for the temporal gene expressions.

Key Words: Temporal gene expressions, log-normal distribution, Mahalanobis distance, variance analysis, Wald statistic.

1. Introduction

The high throughput gene expression techniques, such as, oligonucleotide and DNA microarray, serial analysis gene expression (SAGE), etc, make it possible now to quickly generate huge amount of time series data on gene expression under various conditions (Bjarnason *et al*. (2003); Cho *et al*. (1998); Spellman *et al.* (1998); Yuan and Lin (2007); Zhu *et al*. (2007)). A general goal common to many of these time course experiments is to collect gene expression time series in multiple biological conditions such as different cancer tumor types or different treatments, identify genes that exhibit different temporal expression profiles across multiple biological conditions. These data usually have several main features: containing large scale of data set, having many genes, being measured over many time levels, involving multiple conditions.

In the data analysis of temporal gene expressions, most literature focuses on detecting temporally differentially expressed genes between two experiment conditions. Hong and Li (2006) identified genes with different time-course expression profile using functional hierarchical models. Yuan and Kendziorski (2006) proposed a hidden Markov modeling (HMM) method to efficiently distinguish differentially expressed genes at each time point and classify genes based on their temporal expression patterns. This method can be used not only to compare two or more biological conditions over the short and long time series, but also to pinpoint the genes that have different expression profiles across conditions. Fang *et al.* (2012) proposed the non-linear regression model vis spline method to characterize the relative change rate of genes to classify the gene expressions.

*Department of Mathematics and Statistics, University of Regina, Sask., S4S 0A2, Canada
†Department of Mathematics and Statistics, University of Regina, Sask., S4S 0A2, Canada
‡Department of Mathematics, Anhui Normal University, Wuhu, China

Besides its intrinsic effect, a gene has various expression patterns under different biological conditions and these conditions result in the variation of gene expression variance. In fact, it is the dispersions of gene expressions not their expectations that reflect the effects of biological conditions because under different conditions, the mean functions of gene expressions do not have a big change but the variance functions may differ from each other tremendously. For example, for a contrast experiment with a control group and a treatment group, the gene profiles for two groups may have a similar expression but have totally different variations. The genes in the control group have a small dispersion but the variances of genes in treatment group could vary largely. However, to our knowledge, there is no discussion about the classification of biological conditions.

The goal of this article is to investigate the influence of conditions to the gene expressions and then classify the conditions according to the variance analysis of data from temporal gene expressions. The interest is the dispersion function over time for all genes together under a given condition and then compare the differences among all dispersion functions obtained under all conditions. We need to estimate the variance function under each condition. Afterward we assess whether two conditions have a similar effect to genes by testing the similarity of two variance functions, and conduct the classification to all conditions by using the obtained estimators of variance functions.

Toward these end, we need first to estimate the variance functions. In fact, the variance function estimation plays an important role in many contexts. Excepts their own interest, variance function estimates are needed to construct confidence intervals/bands for the mean function and compute weighted least squares estimates of the mean function. Relative to mean function estimation, the literature on variance function estimation is sparse. The existing methods include kernel, polynomial and wavelet procedure. Müller and Stadtüller (1987, 1993) considered difference based kernel estimators of the variance function. Hall and Carroll (1989) proposed kernel estimators of the variance function based on the squared residuals from a rate optimal estimator of the mean function. Ruppert *et al.* (1997) and Fan and Yao (1998) used local polynomial smoothing of the squared residuals from an optimal estimators of the mean function. Moreover, Brown and Livine (2007) established asymptotic normality for a class of difference-based kernel estimators of variance function and Wang *et al.* (2008) derived the minimax rate of convergence. More recently, Cai and Wang (2008) considered a wavelet thresholding approach to adaptive variance function estimation in heteroscedastic nonparametric regression. However, the shortcoming in the aforementioned methods is that the variance function is estimated locally and there is no explicit global expression for estimator of the variance function over the time and thus it is very difficult to make the comparison for obtained estimators. Therefore in current paper we will develop an order-dependent thresholding approach to estimate the variance function in parametric regression model and give a global form of estimator for the variance function. Further by choosing the common order-dependent threshold knots for all estimators of the variance functions, we can compare whether two variance functions have a similar behavior and make the classification to all variance functions.

The rest of this paper is organized as follows. The method for estimation of variance function is proposed in Section 2. In Section 3, the statistic to test the equality of variance functions for two biological conditions is derived and the simulation studies are conducted in Section 4. The proposed strategy is illustrated with a data set of 18 genes in *P. Aeruginosa* expressed in 24 conditions in Section 5 and the concluding remarks are given in Section 6.

2. Estimation of Variance Function

Let $Y_i(t)$ be the value of the i-th gene curve at time t and write

$$Y_i(t) = \mu_Y(t) + \epsilon_i(t) \tag{1}$$

where $\epsilon_i(t)$ is the random noise with zero mean and variance $\sigma^2(t)$. Also, let $\boldsymbol{Y}_i(\boldsymbol{t}_i) = (Y_i(t_{i1}), ..., Y_i(t_{ip_i}))^T$ be the vector of p_i observations for gene i and $\boldsymbol{t}_i = (t_{i1}, ..., t_{ip_i})^T$ be the corresponding vector of times at which these measurements are made for genes $i = 1, 2, ..., g$. Also, it is assumed that the expected value of $\boldsymbol{Y}_i(\boldsymbol{t}_i)$ is $\mu_Y(\boldsymbol{t}_i) = (\mu_Y(t_{i1}), ..., \mu_Y(t_{ip_i}))^T$ and covariance matrix is $\Sigma_{\boldsymbol{Y}}$ with (r, s) entry $\sigma_{rs} = \text{cov}(Y_i(t_r), Y_i(t_s))$ for $(r, s = 1, 2, ..., p_i)$

To estimate the variance function of this model, we define

$$X_i(t) = (Y_i(t) - \mu_Y(t))^2 \tag{2}$$

and

$$\boldsymbol{X}_i(\boldsymbol{t}_i) = (X_i(t_{i1}), X_i(t_{i2}), ..., X_i(t_{ip_i}))^T \tag{3}$$

Furthermore, we need to determine a reasonable distribution for $\boldsymbol{X}_i(\boldsymbol{t}_i)$. The reasonable assumption is to consider that $X_i(t_{ij})$ for $j = 1, ..., p_i$, and $i = 1, ..., g$ relatively follows the chi-square distribution. In this situation we are dealing with multivariate data sets, we should therefore consider the Wishart distribution which is a multivariate distribution and it is a well-known distribution for interpreting the covariance matrices. However, it is not easy to find the estimation of covariance matrix under this assumption when we are dealing with a high-dimensional data set. On the other hand, the known results indicate that the log-normal distribution offers a good approximation for the chi-square distribution (see Jouini *et al.* 2011). Also, the logarithm of the log-normal random variable has a normal distribution. As a result, this fact can convince us to use the log-normal distribution for estimating the variance function. Therefore, $\boldsymbol{X}_i(\boldsymbol{t}_i)$ can be considered as multivariate random variable with log-normal distribution. That is, if set $\boldsymbol{W}_i(\boldsymbol{t}_i) \equiv [W_i(t_{i1}),\ W_i(t_{i2}),\ ...,\ W_i(t_{ip_i})]^T = [\log(X_i(t_{i1})),\ \log(X_i(t_{i2})),\ ...,\ \log(X_i(t_{ip_i}))]^T$, then $\boldsymbol{W}_i(\boldsymbol{t}_i)$ has a multivariate normal distribution. Let $E(\boldsymbol{W}_i(\boldsymbol{t}_i)) = \mu_W(\boldsymbol{t}_i) = [\mu_W(t_{i1}),\ ..., \mu_W(t_{ip_i})]^T$ and Σ_{W_i} is

$$\begin{pmatrix} d_{11} & \cdots & d_{1p_i} \\ \vdots & \ddots & \vdots \\ d_{p_i1} & \cdots & d_{p_ip_i} \end{pmatrix}$$

Then the density function of multivariate random variable $\boldsymbol{X}_i(\boldsymbol{t}_i)$ is

$$f_{X_i}(\boldsymbol{x}) = \frac{1}{(2\pi)^{p_i/2}|\Sigma_{W_i}|^{1/2}\prod_{r=1}^{p_i} x_r} \exp\left[-\frac{1}{2}(\log \boldsymbol{x} - \mu_{W_i}(\boldsymbol{t}_i))^T\Sigma_{W_i}^{-1}(\log \boldsymbol{x} - \mu_{W_i}(\boldsymbol{t}_i))\right], \quad 0 < x_r < \infty, r = 1, ..., p_i$$

where $\log \boldsymbol{x} = [\log x_1, \cdots, \log x_{p_i}]^T$ is a p_i-component vector. Moreover, the mean of $\boldsymbol{X}_i(\boldsymbol{t}_i)$ is $\mu_X(\boldsymbol{t}_i) = [\mu_X(t_{i1}),\ \cdots, \mu_X(t_{ip_i})]^T$ where $\mu_X(t_{ir}) = \exp(\mu_W(t_{ir}) + \frac{1}{2}d_{rr}), r = 1, ..., p_i$ (see Kotz *et al.* 2004) and the covariance matrix is

$$\Sigma_X = \begin{pmatrix} \sigma'_{11} & \cdots & \sigma'_{1p_i} \\ \vdots & \ddots & \vdots \\ \sigma'_{p_i1} & \cdots & \sigma'_{p_ip_i} \end{pmatrix}$$

where $\sigma'_{rs} = \left[\exp[(\mu_W(t_{ir}) + \mu_W(t_{is})) + \frac{d_{rr}+d_{ss}}{2}]\right] \times [\exp(d_{rs}) - 1], r, s = 1, ..., p_i$.

To find the variance function $\sigma^2(t)$ it is enough to compute $E[(\boldsymbol{Y}_i(t) - \mu(t))^2] = E[X_i(t)]$. In other words, $\sigma^2_Y(t) = E[X_i(t)] = \mu_X(t)$. Since MLEs of parameters have invariant properties, finding the maximum likelihood estimator for $\sigma^2_Y(t) = \mu_X(t)$ is equivalent to finding the ML estimators for $\mu_W(t)$ and Σ_W. Before giving the ML estimator for $\mu_W(t)$ and Σ_W we consider following assumptions:

(A1) $\mu_W(t)$ is approximated by using a linear combination of a set of truncated power basis functions. Given a sequence of K interior knots $0 < \tau_1 < \tau_2 < \ldots < \tau_K < \tau$ where τ is the end time of observations. The regression basis functions of order Q are $1,\ t,\ t^2,\ \ldots,\ t^Q,\ (t-\tau)^Q_+,\ \ldots, (t-\tau_K)^Q_+$ and we can determine $\mu_W(t) = \boldsymbol{B}^T(t)\boldsymbol{\beta}$, where $\boldsymbol{B}(t)$ is the vector of $q(= 1 + Q + K)$ basis functions:

$$\boldsymbol{B}(t) = \left(1,\ t,\ t^2,\ \ldots,\ t^Q,\ (t-\tau_1)^Q_+,\ \ldots, (t-\tau_K)^Q_+\right)^T$$

(A2) We assume $\Sigma_{W_i} = \sigma^2 \boldsymbol{G}_i(\boldsymbol{\beta}, \alpha)\boldsymbol{R}_i(\rho)\boldsymbol{G}_i(\boldsymbol{\beta}, \alpha)$ where

$$\boldsymbol{R}_i(\rho) = \begin{pmatrix} 1 & \rho^{|t_{i1}-t_{i2}|} & ... & \rho^{|t_{i1}-t_{ip_i}|} \\ \rho^{|t_{i2}-t_{i1}|} & 1 & \cdots & \rho^{|t_{i2}-t_{ip_i}|} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{|t_{ip_i}-t_{i1}|} & \rho^{|t_{ip_i}-t_{i2}|} & \cdots & 1 \end{pmatrix}$$

and

$$\boldsymbol{G}_i(\boldsymbol{\beta}, \alpha) = \begin{pmatrix} \exp(\frac{1}{2}\alpha\boldsymbol{B}^T(t_1)\boldsymbol{\beta}) & 0 & \cdots & 0 \\ 0 & \exp(\frac{1}{2}\alpha\boldsymbol{B}^T(t_2)\boldsymbol{\beta} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \exp(\frac{1}{2}\alpha\boldsymbol{B}^T(t_{p_i})\boldsymbol{\beta}) \end{pmatrix}$$

Specially, when observing times are equally spaced then $R_i(\rho)$ is expressed as follows

$$\boldsymbol{R}_i(\rho*) = \begin{pmatrix} 1 & \rho* & \cdots & \rho*^{p_i-1} \\ \rho* & 1 & \cdots & \rho*^{p_i-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho*^{p_i-1} & \rho* & \cdots & 1 \end{pmatrix}$$

where $\rho* = \rho^{|t_{i2}-t_{i1}|}$. In particular, when difference between any two adjacent times is one then $\rho = \rho*$.

(A3) $\mu_Y(t)$ can be considered as the average expression of g gene profiles. There exist many approaches in the literature that can be used to recover the mean curve $\mu_Y(t)$, including kernel, local polynomial, smoothing splines, regression splines and wavelet-based methods among others. We can choose one of existing methods to estimate the function $\mu_Y(t)$, denote by $\hat{\mu}_Y(t)$ the estimator of $\mu_Y(t)$. Here, $\mu_Y(t)$ is estimated by fitting local lines in one dimensional based on the pooled data from all gene expressions, which minimizes

$$\sum_{i=1}^{g}\sum_{r=1}^{p} K\left(\frac{t_{ir}-t}{h}\right)(y_{ir} - \alpha_{0t} - \alpha_{1t}(t - t_{ir}))^2$$

with respect to α_{0t} and α_{1t} where K and h are kernel and bandwidth, respectively. The local estimator of $\mu_Y(t)$ is given by $\hat{\mu}_Y(t) = \hat{\alpha}_{0t}$.

Now let $\boldsymbol{w}_{\text{obs}} = (\boldsymbol{w}_1(\boldsymbol{t}_1), \boldsymbol{w}_2(\boldsymbol{t}_2), ..., \boldsymbol{w}_g(\boldsymbol{t}_g))^T$ be the observed data from the matrix of the random variables $\boldsymbol{W}_{\text{obs}} = (\boldsymbol{W}_1(\boldsymbol{t}_1), \boldsymbol{W}_2(\boldsymbol{t}_2), ..., \boldsymbol{W}_g(\boldsymbol{t}_g))$. Based on the previous assumptions the likelihood function for the parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}, \alpha, \sigma, \rho)$ is

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{w}_{\text{obs}}) &= \mathcal{L}(\boldsymbol{\beta}, \alpha, \sigma^2, \rho|\boldsymbol{w}_{\text{obs}}) \\ &= \prod_{i=1}^{g}\left[(2\pi)^{-\frac{p}{2}}|\Sigma_{W_i}|^{-1/2}\exp\left\{-\frac{1}{2}(\boldsymbol{w}_i - \boldsymbol{B}^T\boldsymbol{\beta})^T\Sigma_{W_i}^{-1}(\boldsymbol{w}_i - \boldsymbol{B}^T\boldsymbol{\beta})\right\}\right] \qquad (4)\end{aligned}$$

In general, the gene expression profiles are recorded simultaneously with the equal number of equally spaced observed times. Therefore we can assume that $p_i = p$ and the all gene expression profiles share the common observing time points $\boldsymbol{t} = (t_1, t_2, ..., t_p)$ for $i = 1, ..., g$. In this situation, we have that $\Sigma_{W_i} = \Sigma_W = \sigma^2\boldsymbol{G}(\boldsymbol{\beta}, \alpha)\boldsymbol{R}(\rho)\boldsymbol{G}(\boldsymbol{\beta}, \alpha)$ where

$$\boldsymbol{R}(\rho) = \begin{pmatrix} 1 & \rho & \cdots & \rho^{p-1} \\ \rho & 1 & \cdots & \rho^{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{p-1} & \rho & \cdots & 1 \end{pmatrix}$$

$\boldsymbol{B} = (\boldsymbol{B}(t_1,), \boldsymbol{B}(t_2), ..., \boldsymbol{B}(t_p))$ and $\boldsymbol{G}(\boldsymbol{\beta}, \alpha) = \text{diag}(\exp(\frac{1}{2}\alpha\boldsymbol{B}^T(t_1)\boldsymbol{\beta}, ..., \exp(\frac{1}{2}\alpha\boldsymbol{B}^T(t_p)\boldsymbol{\beta})$. Thus, the likelihood (4) can be simplified as:

$$\begin{aligned}&\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{w}_{\text{obs}}) \\ &= (2\pi\sigma^2)^{-\frac{gp}{2}}\prod_{i=1}^{g}|\boldsymbol{GRG}|^{-1/2}\exp\left\{\sum_{i=1}^{g}-\frac{1}{2\sigma^2}(\boldsymbol{w}_i - \boldsymbol{B}^T\boldsymbol{\beta})^T\boldsymbol{G}^{-1}\boldsymbol{R}^{-1}\boldsymbol{G}^{-1}(\boldsymbol{w}_i - \boldsymbol{B}^T\boldsymbol{\beta})\right\} \\ &= (2\pi\sigma^2)^{-\frac{gp}{2}}|\boldsymbol{G}|^{-g}|\boldsymbol{R}|^{-g/2}\exp\left\{\sum_{i=1}^{g}-\frac{1}{2\sigma^2}(\boldsymbol{w}_i - \boldsymbol{B}^T\boldsymbol{\beta})^T\boldsymbol{G}^{-1}\boldsymbol{R}^{-1}\boldsymbol{G}^{-1}(\boldsymbol{w}_i - \boldsymbol{B}^T\boldsymbol{\beta})\right\}\end{aligned}$$

and the log likelihood function of parameter $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2, \alpha, \rho)$ is

$$\begin{aligned}l(\boldsymbol{\theta}|\boldsymbol{w}_{\text{obs}}) &= \log[\mathscr{L}(\boldsymbol{\theta}|\boldsymbol{w}_{\text{obs}})] = -\frac{gp}{2}\log(2\pi\sigma^2) \\ &- g\log|\boldsymbol{G}| - \frac{g}{2}\log(|\boldsymbol{R}|) - \frac{1}{2\sigma^2}\sum_{j=1}^{g}(\boldsymbol{w}_i - \boldsymbol{B}^T\boldsymbol{\beta})^T\boldsymbol{G}^{-1}\boldsymbol{R}^{-1}\boldsymbol{G}^{-1}(\boldsymbol{w}_i - \boldsymbol{B}^T\boldsymbol{\beta})\end{aligned} \qquad (5)$$

From the log likelihood function, the maximum likelihood estimator $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \hat{\alpha}, \hat{\rho})$ is defined by

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} l(\boldsymbol{\theta}|\boldsymbol{w}_{\text{obs}})$$

and based on the normal equations for parameters $(\boldsymbol{\beta}, \sigma^2, \alpha, \rho)$, the algorithm for the computation of MLEs of parameters $(\boldsymbol{\beta}, \sigma^2, \alpha, \rho)$ and the derivation for expected information matrix of MLEs are given in Appendix.

Finally, based on invariant properties of MLEs we can obtain the MLE of $\sigma_Y^2(\boldsymbol{t}) = (\sigma_Y^2(t_1), ..., \sigma_Y^2(t_p))^T$ as follows

$$\widehat{\sigma_Y^2}(\boldsymbol{t}) = \hat{\mu}_X(\boldsymbol{t}) = (\exp(\hat{\mu}_W(t_1) + \hat{d}_{11}/2), \cdots, \exp(\hat{\mu}_W(t_p) + \hat{d}_{pp}/2))^T \qquad (6)$$

where $\hat{\mu}_W(t_r)$ is the r^{th} element of $\hat{\mu}_W(\boldsymbol{t}) = \boldsymbol{B}^T(t)\hat{\boldsymbol{\beta}}$ and $\hat{d}_{rr}$ is the r^{th} diagonal element of

$$\widehat{\Sigma}_W = \hat{\sigma}^2\boldsymbol{G}(\hat{\boldsymbol{\beta}}, \hat{\alpha})\boldsymbol{R}(\hat{\rho})\boldsymbol{G}(\hat{\boldsymbol{\beta}}, \hat{\alpha}) \qquad (7)$$

Further, from (6) and (7), and note that the diagonal elements of Σ_W do not depend on the parameter ρ, the MLE $\widehat{\sigma_Y^2}(t)$ can be written as

$$\widehat{\sigma_Y^2}(t) = \exp\{\boldsymbol{B}^T(t)\hat{\boldsymbol{\beta}} + \frac{\hat{\sigma}^2}{2}\exp(\hat{\alpha}\boldsymbol{B}^T(t)\hat{\boldsymbol{\beta}})\}. \tag{8}$$

3. Model-Based Test for Equality of Variance Functions of Two Conditions

In this section we are going to construct an asymptotic test for the equality of variance functions for two conditions using Wald statistics and Fisher information matrix. The hypothesis which we are interested in is

$$H_0 : \sigma_{Y_r}^2(t) = \sigma_{Y_s}^2(t) \qquad \text{vs.} \qquad H_1 : \sigma_{Y_r}^2(t) \neq \sigma_{Y_s}^2(t) \tag{9}$$

for all t. From equation (8), the variance function $\sigma_Y^2(t)$ is the function of the parameters $\boldsymbol{\beta}, \alpha, \sigma^2$. Since the common set of truncated power basis function $\boldsymbol{B}(t)$ is chosen for all gene expressions, each gene expression curve can be written as

$$\sigma_{Y_r}^2(t) = \exp\{\boldsymbol{B}^T(t)\boldsymbol{\beta}_r + \frac{\sigma_r^2}{2}\exp(\alpha_r\boldsymbol{B}^T(t)\boldsymbol{\beta}_r)\}. \tag{10}$$

for $r = 1, 2, ..., p$.

Therefore, from (10) the hypotheses (9) are equivalent to the following:

$$H_0 : (\boldsymbol{\beta}_r,\ \sigma_r^2,\ \alpha_r) = (\boldsymbol{\beta}_s,\ \sigma_s^2,\ \alpha_s) \ \text{ vs. } \ H_1 : (\boldsymbol{\beta}_r,\ \sigma_r^2,\ \alpha_r) \neq (\boldsymbol{\beta}_s,\ \sigma_s^2,\ \alpha_s).$$

The above hypothesis can be tested by using the Wald type statistic:

$$\boldsymbol{\chi}_{rs}^2 = (\widehat{\theta}_r^* - \widehat{\theta}_s^*)^T \widehat{M}_{rs}^{-1}(\widehat{\theta}_r^* - \widehat{\theta}_s^*) \tag{11}$$

where $(\widehat{\theta}_r^* - \widehat{\theta}_s^*) = (\widehat{\boldsymbol{\beta}}_r - \widehat{\boldsymbol{\beta}}_s, \widehat{\sigma^2}_r - \widehat{\sigma^2}_s, \widehat{\alpha}_r - \widehat{\alpha}_s)^T$ and

$$\begin{aligned} M_{rs} &= M_{rs}(\theta_r^*, \theta_s^*) = \operatorname{cov}(\widehat{\theta}_r^* - \widehat{\theta}_s^*) \\ &= \operatorname{cov}(\hat{\theta}_r^*) + \operatorname{cov}(\hat{\theta}_s^*) = I^{-1}(\theta_r^*) + I^{-1}(\theta_s^*) \\ &= \begin{pmatrix} \boldsymbol{C}_r^{-1} + \boldsymbol{C}_s^{-1} & 0 \\ 0 & \boldsymbol{D}_r^{-1} + \boldsymbol{D}_s^{-1} \end{pmatrix} \end{aligned}$$

with

$$\begin{aligned} \boldsymbol{C}_r^{-1} + \boldsymbol{C}_s^{-1} &= \left[\frac{g}{\sigma_r^2}\boldsymbol{B}_r^T\boldsymbol{G}^{-1}(\alpha_r)\boldsymbol{R}^{-1}(\rho_r)\boldsymbol{G}^{-1}(\alpha_r)\boldsymbol{B}_r\right]^{-1} \\ &+ \left[\frac{g}{\sigma_s^2}\boldsymbol{B}_s^T\boldsymbol{G}^{-1}(\alpha_s)\boldsymbol{R}^{-1}(\rho_s)\boldsymbol{G}^{-1}(\alpha_s)\boldsymbol{B}_s\right]^{-1} \end{aligned}$$

$$\begin{aligned} D_r^{-1} + D_s^{-1} &= \begin{pmatrix} \frac{g(p-1)(1-\rho_r^2)}{(1-\rho_r^2)^2} & -\frac{g\rho_r\,\mathrm{tr}(\boldsymbol{\Xi}_r+\boldsymbol{\Xi}_r\mathbf{K}_p)}{4(1-\rho_r^2)} \\ -\frac{g\rho_r\,\mathrm{tr}(\boldsymbol{\Xi}_r+\boldsymbol{\Xi}_r\mathbf{K}_p)}{4(1-\rho_r^2)} & \frac{g\,\mathrm{tr}(\boldsymbol{\Xi}_r\boldsymbol{\Xi}_r+R^{-1}\boldsymbol{\Xi}_rR^{-1}\boldsymbol{\Xi}_r)}{4} \end{pmatrix}^{-1} \\ &+ \begin{pmatrix} \frac{g(p-1)(1-\rho_s^2)}{(1-\rho_s^2)^2} & -\frac{g\rho_s\,\mathrm{tr}(\boldsymbol{\Xi}_s+\boldsymbol{\Xi}_s\mathbf{K}_p)}{4(1-\rho_s^2)} \\ -\frac{g\rho_s\,\mathrm{tr}(\boldsymbol{\Xi}_s+\boldsymbol{\Xi}_s\mathbf{K}_p)}{4(1-\rho_s^2)} & \frac{g\,\mathrm{tr}(\boldsymbol{\Xi}_s\boldsymbol{\Xi}_s+R^{-1}\boldsymbol{\Xi}_sR^{-1}\boldsymbol{\Xi}_s)}{4} \end{pmatrix}^{-1} \end{aligned}$$

and $\hat{M}_{rs} = M(\hat{\theta}_r^*, \hat{\theta}_s^*)$. Furthermore, the derivation of expressions for $\boldsymbol{C}_r, \boldsymbol{C}_s, D_r$ and D_s is given in Appendix. Also, under the null hypothesis $H_0 :\ (\boldsymbol{\beta}_r, \sigma_r^2, \alpha_r) = (\boldsymbol{\beta}_s, \sigma_s^2, \alpha_s)$, $\boldsymbol{\chi}_{rs}^2$ asymptotically has chi-square distribution with $N_B + 3$ degree of freedoms (N_B is number of bases in the model).

4. Simulation Studies

To check the accuracy of ML estimator for estimates $\mu(t)$ introduced in second section, simulation studies were designed. To demonstrate the performance of MLE for the parameters $\boldsymbol{\beta}$, α, σ^2, and ρ in the proposed model, random samples are generated from multivariate normal random variables with the following three different mean functions:

$$\begin{aligned}
\text{Model 1} \quad &: \quad \mu_W(t) = \cos(4\pi t) - 2\left(t - \frac{1}{2}\right)^2 + 1 \\
\text{Model 2} \quad &: \quad \mu_W(t) = \sin(4\pi t) - 2\left(t - \frac{3}{4}\right)^2 + 1 \\
\text{Model 3} \quad &: \quad \mu_W(t) = \sin(2\pi t) - \cos(2\pi t)
\end{aligned}$$

and covariance matrix $\Sigma_W = \sigma^2 \boldsymbol{GRG}$ with

$$\boldsymbol{G} = \text{diag}[\exp(\frac{1}{2}\alpha\mu_W(t_1)), ..., \exp(\frac{1}{2}\alpha\mu_W(t_p))]$$

and

$$\boldsymbol{R} = \begin{pmatrix} 1 & \rho & \dots & \rho^{p-1} \\ \rho & 1 & \dots & \rho^{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{p-1} & \rho & \dots & 1 \end{pmatrix}$$

We consider 25 equally spaced time points between 0 and 1. Meanwhile, $\rho = 0.965$, $\alpha = 2$, and $\sigma^2 = 0.2$. Let $\boldsymbol{w}$ be the matrix of random numbers

$$\boldsymbol{w} = \begin{pmatrix} w_{11} & w_{12} & \dots & w_{1p} \\ \vdots & \dots & \dots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ w_{g1} & w_{g2} & \dots & w_{gp} \end{pmatrix} = (\boldsymbol{w}_1, \cdots, \boldsymbol{w}_g)^T$$

with $p = 25$ and $g = 20, 40$ and 60. The first thing we are interested in finding is the difference between the sample mean and real mean. The sample mean of $\boldsymbol{w}$ is $\bar{\boldsymbol{w}} = \frac{1}{g}\mathbf{1}^T\boldsymbol{w} = (\bar{w}_1,\ \bar{w}_2,\ \dots,\ \bar{w}_p)$ where

$$\bar{w}_r = \frac{1}{g}\sum_{i=1}^{g} w_{ir}, \quad r = 1, 2, ..., p.$$

Figures 1-3 demonstrate the difference between true mean function $\mu_W(t)$ and sample mean $\bar{\boldsymbol{w}}$. From Section 2 we considered using $\boldsymbol{B}^T(t)\boldsymbol{\beta}$ to approximate the true mean function $\mu_W(t)$, where $\boldsymbol{B}(t) = (1,\ t,\ t^2,\ t^3,\ t^4,\ t^5,\ (t-0.25)_+^5,\ (t-0.5)_+^5,\ (t-0.75)_+^5)^T$ and $\boldsymbol{\beta}$ is the vector of unknown parameters. Since we know $\boldsymbol{W} \sim MVN(\mu_W(t), \Sigma_W)$, it is possible to use the introduced normal equations to estimate $\boldsymbol{\beta}$, σ^2, α and ρ. However, it is not likely to derive the closed forms for the estimators of these four parameters from the normal equations. we use the iterative algorithm to solve the maximum likelihood equations for the parameters $\boldsymbol{\beta}$, σ^2, α and ρ. The details for the iterative algorithm are given in the appendix.

The results for the estimates of parameters are given in the following table.

Model	Estimate for $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9)$	Estimate for σ^2	Estimate for α	Estimate for ρ
1	1.4, 1.5, -64.5, -0.1084, 1553, -2297.4, 4391.5, -4161.5, 4.0395	0.2086	1.9990	0.9679
2	0.09, -.50, 285.03, -2390, 6688, -6285.2, 7181, 258.191, -9.12	0.1906	1.9092	0.9630
3	-0.987, 5.591, 30.517, -96.533 41.186, 34.004, -53.88, -62.99, -26.88	0.2130	2.0863	0.9665

Table 1: The estimates of parameters $\boldsymbol{\beta}, \sigma^2, \alpha$ and ρ for three models

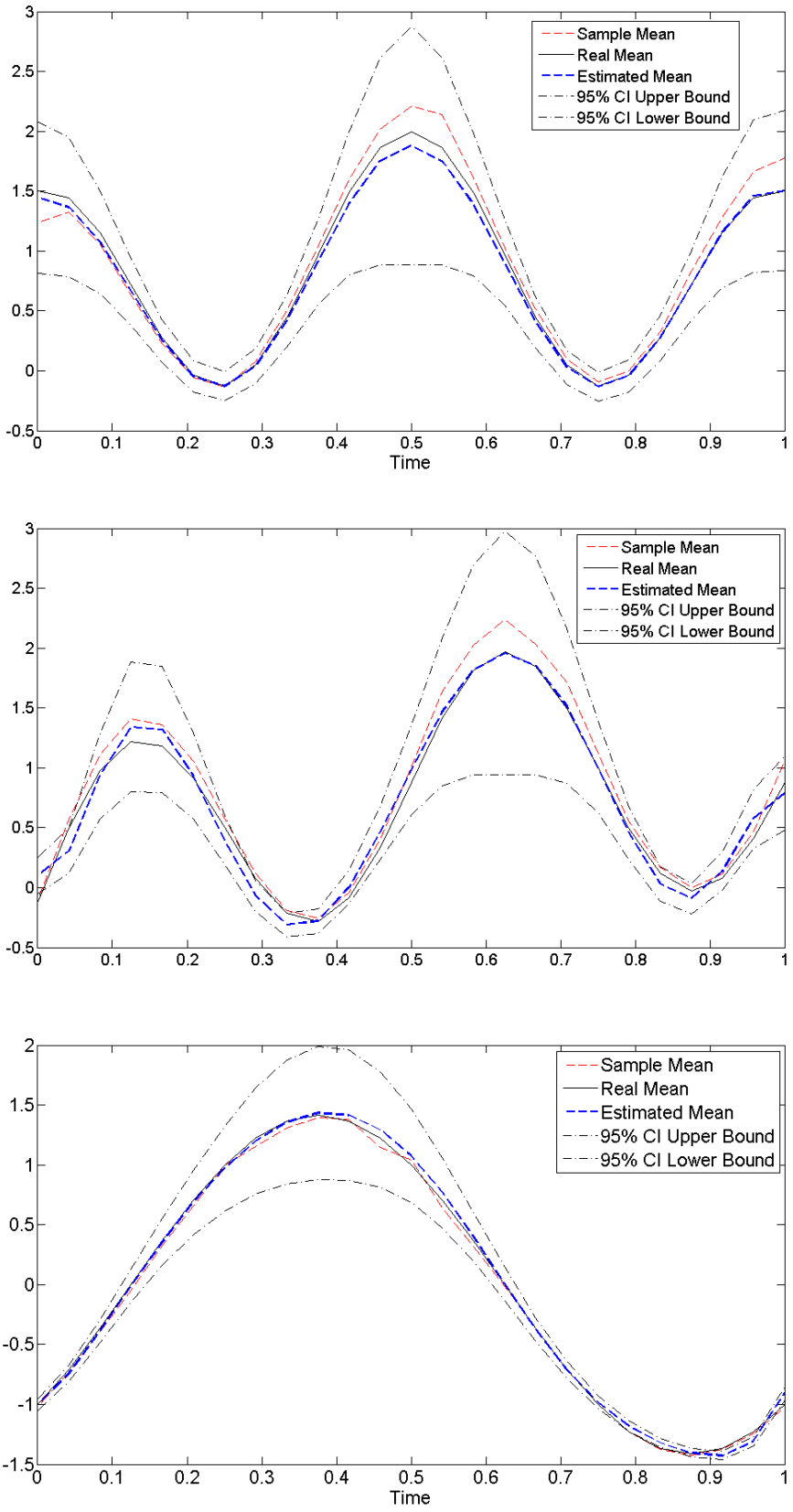

Figure 1: Comparison between real means, estimated means, and sample means for model 1, model 2 and model 3, respectively

From simulation results given in Table 1, the estimates for σ^2, α and ρ are very close to the true values of the parameters, which demonstrates that the proposed method performs well in the estimation of model parameters. Furthermore, as Figures 1-3 show, the estimated mean function and the true mean function are close to each other. Moreover, comparing with the sample mean, the estimated mean is much closer to the true mean and more smooth. Therefore the proposed procedure has a good performance to estimate the mean function in the model.

Next, to evaluate the power of the chi-square test in equation (10), a constant number a is added to the original mean function in each model. As a result, the second mean function for each model is defined as follows:

$$\text{Model 1} \quad : \quad \mu_W(t,a) = \cos(4\pi t) - 2\left(t - \frac{1}{2}\right)^2 + 1 + a$$

$$\text{Model 2} \quad : \quad \mu_W(t,a) = \sin(4\pi t) - 2\left(t - \frac{3}{4}\right)^2 + 1 + a$$

$$\text{Model 3} \quad : \quad \mu_W(t,a) = \sin(2\pi t) - \cos(2\pi x) + a$$

where $a \in \{0,\ 0.05,\ 0.1,\ 0.15,\ 0.2\}$. The L_2 distance between $\mu_W(t)$ and $\mu_W(t,a)$ is $(\int_0^1 (\mu_W(t) - \mu_W(t,a))^2 dt)^{\frac{1}{2}} = a$. The results given in Tables 2-4 present the empirical levels and powers when the nominal level is 0.05 for the given models (Number of replications is 10000). As results show, the chi-square test $\boldsymbol{\chi}^2$ in equation (10) maintains the nominal level ($\alpha = 0.05$) well although it is a little aggressive for small sample size (g=20) and is powerful enough to test the similarity of variance functions in the gene expression data.

L_2 distance of mean functions	Power with 20 samples	Power with 40 samples	Power with 60 samples
0.00	0.1060	0.0501	0.0453
0.05	0.1450	0.1202	0.1590
0.10	0.3010	0.4001	0.5730
0.15	0.6160	0.8603	0.9971
0.20	0.9060	0.9961	1.0000

Table 2: The power analysis of chi-square test $\boldsymbol{\chi}^2$ (model 1)

L_2 distance of mean functions	Power with 20 samples	Power with 40 samples	Power with 60 samples
0.00	0.0840	0.0450	0.0482
0.05	0.1560	0.1620	0.1932
0.10	0.3230	0.4320	0.7450
0.15	0.5786	0.8331	0.9861
0.20	0.9170	0.9561	0.9980

Table 3: The power analysis of chi-square test $\boldsymbol{\chi}^2$ (model 2)

L_2 distance of mean functions	Power with 20 samples	Power with 40 samples	Power with 60 samples
0.00	0.0870	0.0470	0.0491
0.05	0.1210	0.1630	0.2230
0.10	0.3320	0.4210	0.5620
0.15	0.5950	0.8600	0.9500
0.20	0.8540	0.9510	0.9900

Table 4: The power analysis of chi-square test $\boldsymbol{\chi}^2$ (model 3)

5. An Application to the Temporal Gene Expressions in *P. Aeruginosa*

We now consider the analysis of the data set of 18 genes in *P. aeruginosa* expressed in 24 conditions (see Table 5). For each condition, each gene was measured every 30 minutes for

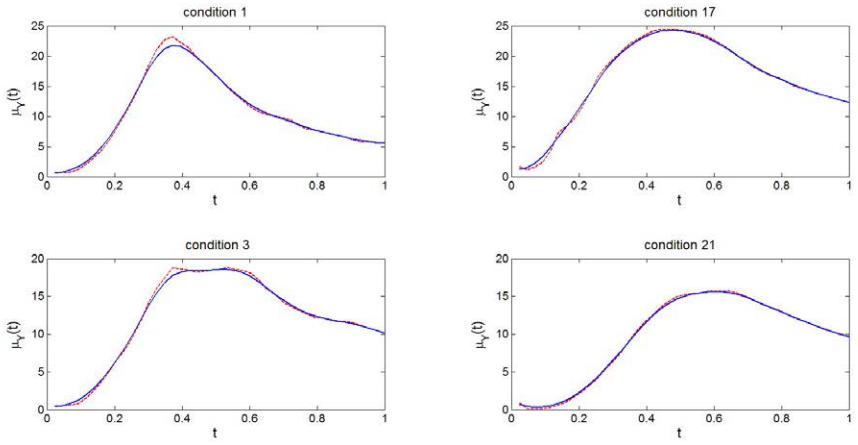

Figure 2: Estimates of $\mu_Y(t)$ for conditions 1, 3, 17, and 21 with sample mean (dash line) and estimated spline curve(solid line).

21 hours and has 43 observations. Without lost of generality we can rescale time points as $\mathbf{t}/43$ assume $0 \leq t \leq 1$. Three interior knots, $0 < 0.25 < 0.5 < 0.75 < 1$, are selected and the basis functions $\mathbf{B}(t)$ is:

$$\boldsymbol{B}(t) = \left(1,\ t,\ t^2,\ t^3,\ t^4, (t-0.25)_+^4,\ (t-0.5)_+^4,\ (t-0.75)_+^4\right).$$

As discussed in Section 2, to estimate the mean function and variance function for each condition, we have to find $W_i(t_j) = \log(Y_i(t_i) - \mu_Y(t_j))^2$ for $i = 1, 2, ...18; j = 1, 2, ...43$. For the assumed reason, the kernel smoothing method (see equation (3)) is first used to estimate the mean function for each condition. Figure 4 shows the sample means with estimated means using kernel smoothing for four different conditions. We then found the ML estimates of $\boldsymbol{\beta} = (\beta_1,\ \ldots,\ \beta_8)^T$, σ^2, α and ρ using the iterative algorithm which is explained in Appendix. Figure 5 shows a comparison between $\mu_W(t) = \boldsymbol{B}(t)\boldsymbol{\beta}$ and sample mean of $W_i(t_j)$ for four selected conditions. Also, Figure 6 demonstrates the estimated square root $\sigma(t)$ of variance function ($\sigma^2(t)$, which is obtained based on equation (10)) for same conditions as Figure 4.

Moreover, based on the chi square test which is defined in previous section, we found that when the significance level is 0.05, there is no significant difference among variances of conditions 1, 3, 6, 8, 9, 11, 12, 13, 14, 15, 19 and 22. Also, variances under conditions 4, 7 ,and 17 are same. Figure 7 indicates the classification tree for 24 conditions based on minimum Mahalanobis distance.

6. Concluding Remarks

In this paper, a non-linear regression model with multivariate log normal distribution is derived to estimate the variance functions in the high dimensional data. In addition, chi square test is proposed to classify the curves of variance functions and the proposed method is applied to study the influence of biological conditions on temporal gene expressions. The simulation studies show that the proposed statistic test maintains the nominal levels and has the powerful advantage.

The proposed estimation of variance function in this paper is based on B-splines, by which the curve of variance function can be represented by finite number of parameters and

Code	Name	Protein	Ratio	Remarks
A6	PA5283	Probable transcriptional regulator	99.68 %	48 % similar to putative transcriptional regulator (Bacillus subtilis)
B3	PA2975 (rluC)	Ribosomal large subunit pseudouridine synthase C	99.68 %	Transcription, RNA processing &degradation
B4	PA4991	Hypothetical protein	100 %	Unknown
B5	PA5237	Conserved hypothetical protein	100 %	87 % similar to hypothetical yigC gene product of E. coli
C4	PA0287 (gpuP)	3-guanidinopropionate transport protein	100 %	Transport of small molecules
D1	PA3115 (fimV)	Motility protein FimV	100 %	Membrane proteins; Motility & Attachment
D2	PA3879 (narL)	Two-component response regulator NarL	99.67 %	74 % similar to E.coli NarL protein
D3	PA0894	Hypothetical protein	99.02 %	Unknown
E5	PA1875	Probable outer membrane protein precursor	100 %	41 % similar to alkaline pro-tease secretion protein AprF
E6	PA0573	Hypothetical protein	100 %	Unknown
F2	PA3902	Hypothetical protein	100 %	Unknown
F3	PA3212	Probable ATP-binding component of ABC transporter	100 %	65 % similar to putative amino acid abc transporter, ATP-binding protein (Helicobacter pylori J99)
F5	PA2997 (nqrC)	Na+translocating NADH: ubiquinone oxidoreductase subunit Nrq3	100 %	Energy metabolism
G2	PA0649 (trpG)	Anthranilate synthase component II	100 %	Energy metabolism; Biosynthesis of co-factors, prosthetic groups & carriers; Amino acid biosynthesis & metabolism
G5	PA1748	Probable enoyl-CoA hydratase/isomerase	98.2 %	61 % similar to putative enoyl-coA hydratase EchA3 (Mycobacterium tuberculosis)
G6	PA3771	Probable transcriptional regulator	99.22 %	54 % similar to a region of putative regulatory protein (Streptomyces coelicolor)
H3	PA1841	Hypothetical protein	100 %	43 % similar to hypothetical yeaK gene product of (E. coli)
σ70	σ70	σ factor		As a control

Table 5: 18 genes in *P. aeruginosa* expression

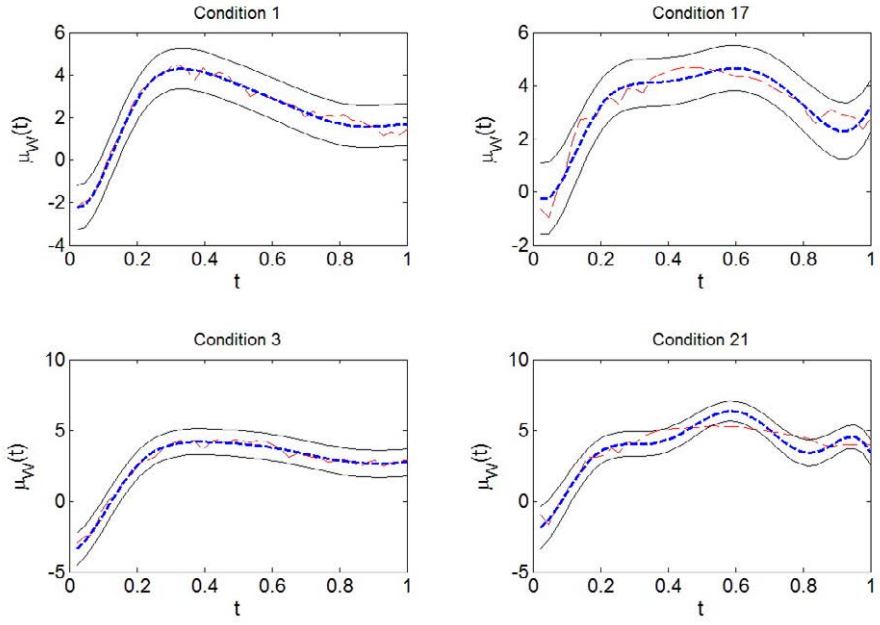

Figure 3: Estimates of $\mu_W(t)$ for conditions 1, 3, 17, and 21. In these four figures blue dashed line, red lines and solid lines denote the smoothing $\boldsymbol{B}(t)\boldsymbol{\beta}$, the sample mean and the 95% confidence upper limit and lower limit for $\mu_W(t)$, respectively.

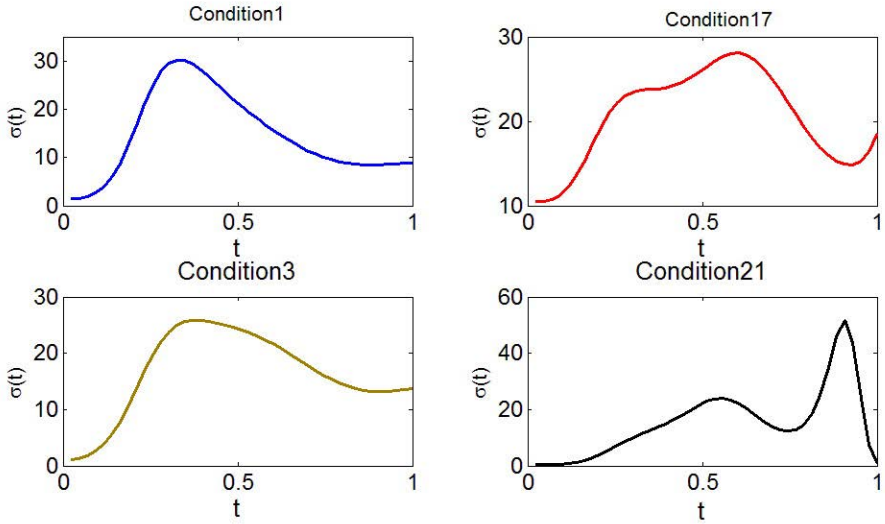

Figure 4: Estimated standard deviation function $\hat{\sigma}(t)$ for conditions 1, 3, 17, 21

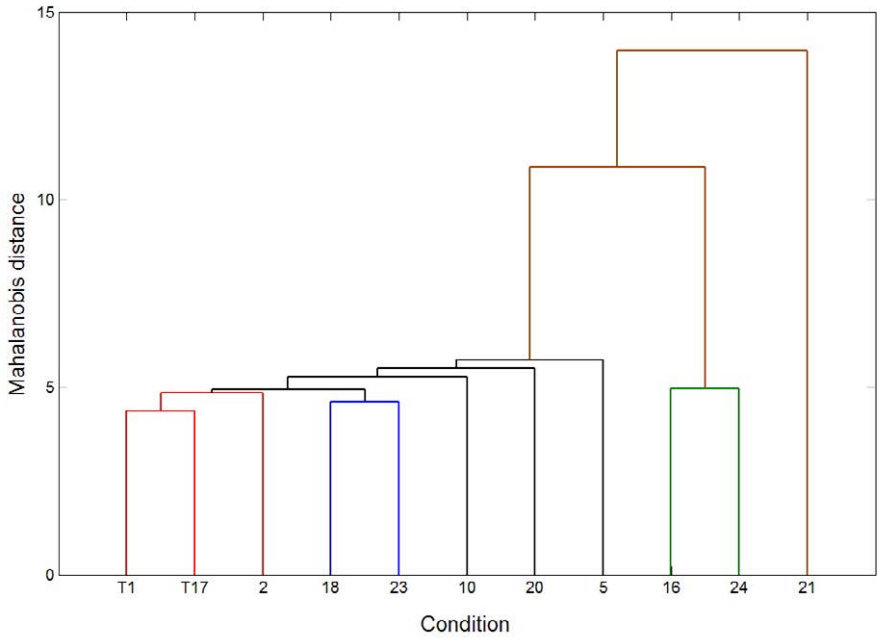

Figure 5: Classification chart based on minimum Mahalanobis distance(T1=[1, 3, 6, 8, 9, 11, 12, 13, 14, 15, 19, 22] and T17=[4, 7, 17])

then the curves can be compared by testing the equality of the corresponding parameters. The method suggested in this paper can be applied to the gene expression data with time varying covariates. In the future study, the structured nonparametric methods may be useful to estimate the time-dependent variance functions. Since the trajectory of temporal gene expression can be considered as a random element in Hilbert space, the corresponding statistical inferences in Hilbert space should be developed to analyze the temporal gene expression data.

Acknowledgement

The research of D. Deng is partially supported by Natural Sciences and Engineering Research Council of Canada.

Appendix

In this appendix, we give the derivation of the likelihood normal equations for parameters $\theta = (\boldsymbol{\beta}, \alpha, \sigma, \rho)$. At first we note the facts that

$$\boldsymbol{G}^{-1} = \mathrm{diag}(\exp(-\frac{1}{2}\alpha\mu(t_1), ..., \exp(-\frac{1}{2}\alpha\mu(t_p))),$$

$$|\boldsymbol{R}| = (1-\rho^2)^{p-1}$$

and

$$\boldsymbol{R}^{-1} = \frac{1}{1-\rho^2}[\boldsymbol{I}_p + \rho^2\boldsymbol{K}_p - \rho(\boldsymbol{H}_p + \boldsymbol{H}_p^T)]$$

where $\boldsymbol{K}_p = \mathrm{diag}(0, \mathbf{1}_{p-2}, 0)$ ($\mathbf{1}_{p-2} = (1, \ldots, 1)_{1\times(p-2)}$) and

$$\boldsymbol{H}_p = \begin{pmatrix} 0 & 0 \\ \boldsymbol{I}_{p-1} & 0 \end{pmatrix}_{p\times p},$$

Now, from the log likelihood function (5) and after tedious calculation, the likelihood normal equations for parameters $\boldsymbol{\beta}, \sigma^2, \alpha$ and ρ are given as follows:

$$\frac{1}{\sigma^2}\sum_{j=1}^{g}\Big[\boldsymbol{B}\boldsymbol{G}^{-1}\boldsymbol{R}^{-1}\boldsymbol{G}^{-1}(\boldsymbol{W}_i-\boldsymbol{B}^T\boldsymbol{\beta})+\alpha\boldsymbol{B}\boldsymbol{E}_i\boldsymbol{G}^{-1}\boldsymbol{R}^{-1}\boldsymbol{G}^{-1}(\boldsymbol{W}_i-\boldsymbol{B}^T\boldsymbol{\beta})\Big]-g\alpha\boldsymbol{B}\mathbf{1}=0$$

$$\sum_{j=1}^{g}\left[\frac{p}{\sigma^2}-\frac{1}{\sigma^4}(\boldsymbol{W}_i-\boldsymbol{B}^T\boldsymbol{\beta})^T\boldsymbol{G}^{-1}\boldsymbol{R}^{-1}\boldsymbol{G}^{-1}(\boldsymbol{W}_i-\boldsymbol{B}^T\boldsymbol{\beta})\right]=0$$

$$\sum_{j=1}^{g}\Big[(p-1)\rho(1-\rho^2)-\frac{\rho}{\sigma^2}(\boldsymbol{W}_i-\boldsymbol{B}^T\boldsymbol{\beta})^T\boldsymbol{G}^{-1}(\boldsymbol{I}_p+\boldsymbol{K}_p)\boldsymbol{G}^{-1}(\boldsymbol{W}_i-\boldsymbol{B}^T\boldsymbol{\beta})$$
$$+\frac{1+\rho^2}{\sigma^2}(\boldsymbol{W}_i-\boldsymbol{B}^T\boldsymbol{\beta})^T\boldsymbol{G}^{-1}\boldsymbol{H}_p\boldsymbol{G}^{-1}(\boldsymbol{W}_i-\boldsymbol{B}^T\boldsymbol{\beta})\Big]=0$$

$$\sum_{j=1}^{g}\left[\mathbf{1}^T\boldsymbol{\Xi}\mathbf{1}-\frac{1}{\sigma^2}\mathbf{1}^T\boldsymbol{\Xi}\boldsymbol{E}_i\boldsymbol{G}^{-1}\boldsymbol{R}^{-1}\boldsymbol{G}^{-1}(\boldsymbol{W}_i-\boldsymbol{B}^T\boldsymbol{\beta})\right]=0$$

where

$$\begin{aligned}\mathbf{1} &= (1,1,...,1)^T_{p\times 1}\\ \boldsymbol{B} &= (\boldsymbol{B}(t_1),...,\boldsymbol{B}(t_p)),\\ \boldsymbol{\Xi} &= \mathrm{diag}(\boldsymbol{B}^T\boldsymbol{\beta}) = \mathrm{diag}(\boldsymbol{B}^T(t_1)\boldsymbol{\beta},...,\boldsymbol{B}^T(t_p)\boldsymbol{\beta}),\\ \boldsymbol{E}_i &= \mathrm{diag}(\boldsymbol{W}_i)-\boldsymbol{\Xi} = \mathrm{diag}(\boldsymbol{W}_i(t_1)-\boldsymbol{B}^T(t_1)\boldsymbol{\beta},...,\boldsymbol{W}_i(t_p)-\boldsymbol{B}^T(t_p)\boldsymbol{\beta})\end{aligned}$$

and therefore from these normal equations, the maximum likelihood estimates for the parameters $\boldsymbol{\beta}, \sigma^2, \rho$ and α can be iteratively computed. Now let $\theta^{(k)} = (\boldsymbol{\beta}^{(k)}, \sigma^{2(k)}, \rho^{(k)}, \alpha^{(k)})$ be the vector of estimated parameters at the kth iteration. Then the estimates of parameters at the $(k+1)$th iteration are

$$\begin{aligned}\boldsymbol{\beta}^{(k+1)} &= \Big[g\mathbf{1}^T\boldsymbol{\Xi}(\boldsymbol{\beta}^{(k)})\overline{\boldsymbol{E}}(\boldsymbol{\beta}^{(k)})\boldsymbol{G}^{-1}(\boldsymbol{\beta}^{(k)},\alpha^{(k)})\boldsymbol{R}^{-1}(\rho^{(k)})\boldsymbol{G}^{-1}(\boldsymbol{\beta}^{(k)},\alpha^{(k)})\boldsymbol{B}^T\Big]^{-1}\\ &\quad\Big[\sum_{i=1}^{g}\mathbf{1}^T\boldsymbol{\Xi}(\boldsymbol{\beta}^{(k)})\overline{\boldsymbol{E}}(\boldsymbol{\beta}^{(k)})\boldsymbol{G}^{-1}(\boldsymbol{\beta}^{(k)},\alpha^{(k)})\boldsymbol{R}^{-1}(\rho^{(k)})\boldsymbol{G}^{-1}(\boldsymbol{\beta}^{(k)},\alpha^{(k)})\boldsymbol{W}_i - g\sigma^{2(k)}\mathbf{1}^T\boldsymbol{\Xi}(\boldsymbol{\beta}^{(k)})\Big]\\ \sigma^{(k+1)} &= \Big[\mathrm{tr}\{\boldsymbol{\Xi}(\boldsymbol{\beta}^{(k)})\}\Big]^{-1}\mathrm{tr}\Big\{\boldsymbol{G}^{-1}(\boldsymbol{\beta}^{(k)},\alpha^{(k)})\boldsymbol{R}^{-1}(\rho^{(k)})\boldsymbol{G}^{-1}(\boldsymbol{\beta}^{(k)},\alpha^{(k)})\boldsymbol{A}(\boldsymbol{\beta}^{(k)})\boldsymbol{\Xi}(\boldsymbol{\beta}^{(k)})\Big\}\\ 1-\rho^{2(k+1)} &= \mathrm{tr}\Big\{\boldsymbol{G}^{-1}(\boldsymbol{\beta}^{(k)},\alpha^{(k)})(\boldsymbol{I}_p+\boldsymbol{K}_p)\boldsymbol{G}^{-1}(\boldsymbol{\beta}^{(k)},\alpha^{(k)})\boldsymbol{A}(\boldsymbol{\beta}^{(k)})\Big\}/[(p-1)\sigma^{(k)}]\\ &\quad -(1+\rho^{2(k)})\,\mathrm{tr}\Big\{\boldsymbol{G}^{-1}(\boldsymbol{\beta}^{(k)},\alpha^{(k)})\boldsymbol{H}_p\boldsymbol{G}^{-1}(\boldsymbol{\beta}^{(k)},\alpha^{(k)})\boldsymbol{A}(\boldsymbol{\beta}^{(k)})\Big\}/[\rho^{(k)}(p-1)\sigma^{2(k)}]\\ \alpha^{(k+1)} &= \Big[g\sigma^{2(k)}\boldsymbol{B}\mathbf{1}-\sum_{i=1}^{g}\boldsymbol{B}\boldsymbol{E}_i(\boldsymbol{\beta}^{(k)})\boldsymbol{G}^{-1}(\boldsymbol{\beta}^{(k)},\alpha^{(k)})\boldsymbol{R}^{-1}(\rho^{(k)})\boldsymbol{G}^{-1}(\boldsymbol{\beta}^{(k)},\alpha^{(k)})(\boldsymbol{W}_i-\boldsymbol{B}\boldsymbol{\beta}^{(k)})\Big]^{-1}\\ &\quad\Big[g\boldsymbol{B}\boldsymbol{G}^{-1}(\boldsymbol{\beta}^{(k)},\alpha^{(k)})\boldsymbol{R}^{-1}(\rho^{(k)})\boldsymbol{G}^{-1}(\boldsymbol{\beta}^{(k)},\alpha^{(k)})(\bar{\boldsymbol{W}}-\boldsymbol{B}\boldsymbol{\beta}^{(k)})\Big]\end{aligned}$$

where

$$\overline{\boldsymbol{W}}=\frac{1}{g}\sum_{i=1}^{g}\boldsymbol{W}_i,\quad \overline{\boldsymbol{E}}=\frac{1}{g}\sum_{i=1}^{g}\boldsymbol{E}_i,\quad \boldsymbol{A}(\boldsymbol{\beta})=\frac{1}{g}\sum_{i=1}^{g}(\boldsymbol{W}_i-\boldsymbol{B}^T\boldsymbol{\beta})(\boldsymbol{W}_i-\boldsymbol{B}^T\boldsymbol{\beta})^T$$

Now to obtain the test statistic for assessing the similarity of two gene conditions, we need to derive the expected information matrix for the estimates $\widehat{\boldsymbol{\beta}}$. Since $\widehat{\boldsymbol{\beta}}$ is ML estimator of $\boldsymbol{\beta}$ it is asymptotically distributed as multivariate normal distribution and its variance is equal to Crammer Rao lower bound (Casella and Berger, 2001). Meanwhile,

$\widehat{\sigma^2}$, $\widehat{\alpha}$ and $\widehat{\rho}$ have same property as $\widehat{\boldsymbol{\beta}}$. Moreover, after the complicated derivation, the expected information matrix for the estimators $\widehat{\theta} = (\widehat{\boldsymbol{\beta}},\ \widehat{\sigma^2},\ \widehat{\rho},\ \widehat{\alpha})$ has the following form,

$$I(\theta) = \begin{pmatrix} \boldsymbol{C} & 0 \\ 0 & \boldsymbol{A} \end{pmatrix}$$

where $\boldsymbol{C} = \frac{g}{\sigma^2}\boldsymbol{B}^T\boldsymbol{G}^{-1}\boldsymbol{R}^{-1}\boldsymbol{G}^{-1}\boldsymbol{B}$ and

$$\boldsymbol{A} = \begin{pmatrix} \frac{g(p-1)(1-\rho^2)}{(1-\rho^2)^2} & -\frac{g\rho(p-1)}{2\sigma^2(1-\rho^2)} & -\frac{g\rho\,\mathrm{tr}(\boldsymbol{\Xi}+\boldsymbol{\Xi}\mathbf{K}_p)}{4(1-\rho^2)} \\ -\frac{n\rho(p-1)}{2\sigma^2(1-\rho^2)} & \frac{gp}{2\sigma^4} & \frac{g}{2\sigma^2}\,\mathrm{tr}(\boldsymbol{\Xi}) \\ -\frac{g\rho\,\mathrm{tr}(\boldsymbol{\Xi}+\boldsymbol{\Xi}\mathbf{K}_p)}{4(1-\rho^2)} & \frac{g}{2\sigma^2}\,\mathrm{tr}(\boldsymbol{\Xi}) & \frac{g\,\mathrm{tr}(\boldsymbol{\Xi}\boldsymbol{\Xi}+R^{-1}\boldsymbol{\Xi}R^{-1}\boldsymbol{\Xi})}{4} \end{pmatrix}$$

Furthermore, from the above expression of information matrix, the ML estimator $\widehat{\boldsymbol{\beta}}$ is asymptotically independent of the ML estimators $\widehat{\sigma^2}$, $\widehat{\alpha}$ and $\widehat{\rho}$.

REFERENCES

Bjarnason, J., Southward, C. M. and Surette, M. G. (2003). Genomic profling of iron-responsive genes in salmonella enterica serovar typhimurium by high-throughput screening of a random promoter library. *Journal of Bacteriology*, **185**, 4973-4982.

Brown, L. D. and Levine, M. (2007). Variance estimation in nonparametric regression via the difference sequence method. *The Annals of Statistics*, **35**, 2219-2232.

Cai, T. and Wang, L. (2008). Adaptive variance function estimation in heteroscedastic nonparametric regression. *The Annals of Statistics*, **36**, 2025-2054.

Casella, G. and Berger, R. L. (2001). Statistical Inference. Duxbury Press, 2nd edition.

Cho, R. J., Campbell, M. J., Winzeler, E. A., Steinmetz, L., Conway, A., Wodicka, L., Wolfsberg, T. G., Gabrielian, A. E., Landsman, D., Lockhart, D. J. and Davis, R. W. (1998). A genome-wide transcriptional analysis of the mitotic cell cycle. *Molecular cell*, **2**, 65-73.

Fan, J. and Yao, Q. (1998). Effcient estimation of conditional variance functions in stochastic regression. *Biometrika*, **85**, 645-660.

Fang, H., Deng, D., Tian, G., Shen, L., Duan, K. and Song, J. (2012). Analysis for temporal gene expressions under multiple biological conditions. *Statistics in Biosciences*, **4**, 282-299.

Hall, P. and Carroll, R. J. (1989). Variance function estimation in regression: The effect of estimating the mean. *Journal of the Royal Statistical Society. Series B (Methodological)*, **51**, 3-14.

Hong, F. and Li, H. (2006). Functional hierarchical models for identifying genes with different time-course expression profles. *Biometrics*, **62**, 534-544.

Jouini, W., Le Guennec, D., Moy, C. and Palicot, J. (2011). Log-normal approximation of chi-square distributions for signal processing. *In General Assembly and Scientific Symposium*, pages 14. IEEE.

Kotz, S., Balakrishnan, N. and Johnson, N. L. (2004). *Continuous Multivariate Distributions, Models and Applications*. John Wiley & Sons.

Mller, H. G. and Stadtmller, U. (1987). Estimation of heteroscedasticity in regression analysis. *The Annals of Statistics*, **15**, 610-625.

Mller, H. G. and Stadtmller, U. (1993). On variance function estimation with quadratic forms. *Journal of Statistical Planning and Inference*, **35**, 213-231.

Ruppert, D., Wand, M. P., Holst, U. and HSJER, O. (1997). Local polynomial variance-function estimation. *Technometrics*, **39**, 262-273.

Spellman, P. T., Sherlock, G., Zhang, M. Q., Iyer,V. R., Anders, K., Eisen, M. B., Brown, P. O., Botstein, D. and Futcher, B. (1998). Comprehensive identification of cell cycle-regulated genes of the yeast saccharomyces cerevisiae by microarray hybridization. *Molecular biology of the cell*, **9**, 3273-3297.

Wang, L., Brown, L. D., Cai, T. and Levine, M. (2008). Effect of mean on variance function estimation in nonparametric regression. *The Annals of Statistics*, **36**, 646-664.

Yuan, Y. and Kendziorski, C. (2006). Hidden markov models for microarray time course data in multiple biological conditions. *Journal of the American Statistical Association*, **101**, 1323-1332.

Yuan, M. and Lin, Y. (2007). Model selection and estimation in the gaussian graphical model. *Biometrika*, **94**, 19-35.

Zhu, H., Tang, Y., Ivanciu, L., Centola, M., Lupu, C., Taylor, F. B. and Lupu, F. (2007). Temporal dynamics of gene expression in the lung in a baboon model of e. coli sepsis. *BMC Genomics*, **8**, p58.