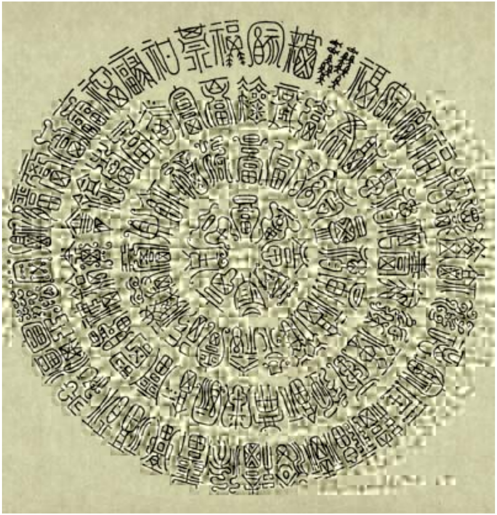
Nonparametric Profile Monitoring By Mixed Effects Modeling

王 (Zhaojun WANG)
(zjwang@nankai.edu.cn)

School of Mathematical Science
Nankai University
Tianjin 300071, China

北京大学
2010年7 7日



Thanks to ...

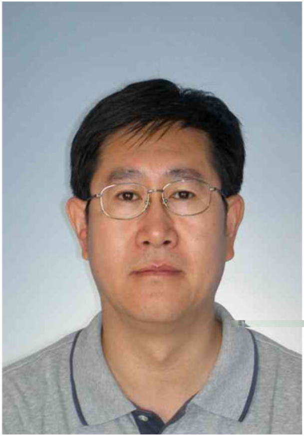
- 北京大学;
- **Technometrics;**
- 邱培华, 邹长亮;
- NFSC of China

Thanks to ...

- 北京大学;
- **Technometrics;**
- 邱培华, 邹长亮;
- NFSC of China


WHO AM I ?

王



王

Outline

1. A Motivation Example
2. Brief Introduction of Control Chart
3. Brief Review of Literature for Monitoring Linear Profile
4. Our Proposed Method For Monitoring General Profile Based on Mixed Effect Model
5. Conclusions

A Motivation Example

A deep reactive ion etching (DRIE) process is a key operation in a Micro-Electro-Mechanical System (MEMS) fabrication to form desired patterns on semiconductor wafers. The desired pro le is the one with smooth and vertical sidewalls (the center on in Fig 1). The positive and negative pro les are due to over- and under-etching.

Statistical Process Control (SPC)

SPC is a powerful collection of problem-solving tools useful in achieving process stability and improving capability through the reduction of variability. SPC can be applied to any process. Its seven major tools (**magnificent seven**) are

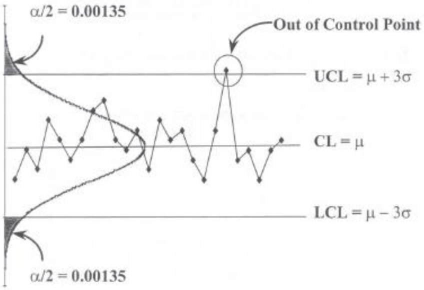
1. Histogram or stem-and-leaf plot
2. Check sheet
3. Pareto chart
4. Cause-and-e ect diagram
5. Defect concentration diagram
6. Scatter diagram
7. **Control Chart**

(Montgomery, 2005, Fifth Edition, P148)

Brief Introduction of Control Chart

How works (1)

For the Shewhart-X chart:



The Shewhart X Chart and Some Concepts

1. **The first control chart—**

Some Problems

1. The e ect of nonnormal (Robust or nonparametric)
2. the e ect of unknown process parameters and ^{2} (Estimation & Self-starting)
3. How to use the information in the past observations (Runs Rules, Adaptive)
4. How to detect the small or moderate shift (Cumulative Sum, EWMA)
5. How to detect the interval shift (Dual CUSUM, Combined Shewhart-CUSUM or EWMA)
6. How to detect the shift in variance (R, S, MR)

Other control charts (1)

- **Shewhart** Charts with/without **Runs Rules** $T(k; m; a; b)$: k of the last m standardized sample means fall in the interval $(a; b)$. (Champ & Woodall 1987, *Technometrics*)
- **Exponentially Weighted Moving Average** (EWMA) (Robert 1959, Lucas & Saccucci 1990, *Technometrics*)

$$Y_n = (1 - \ )Y_{n \ 1} + \ (X_n - \ ); Y_0 = 0;$$

- **Cumulative Sum** (CUSUM) (Page 1954, *Biometrika*; Hawkings & Olwell 1998)

$$S_n^+ = \max\{0; S_{n \ 1}^+ + \frac{X_n - }{=\sqrt{n}} - k\}; \ S_0^+ = 0;$$

$$S_n \ = \min\{0; S_{n \ 1} + \frac{X_n - }{=\sqrt{n}} + k\}; \ S_0 \ = 0:$$

Other control charts (1)

Brief Introduction of Control Chart

Other control charts (1)

- **Shewhart** Charts with/without **Runs Rules** $T(k; m; a; b)$: k of the last m standardized sample means fall in the interval $(a; b)$. (Champ & Woodall 1987, *Technometrics*)
- **Exponentially Weighted Moving Average** (EWMA) (Robert 1959, Lucas & Saccucci 1990, *Technometrics*)

$$Y_n = (1 - \quad)Y_{n \ 1} + \quad (X_n - \quad); Y_0 = 0:$$

- **Cumulative Sum** (CUSUM) (Page 1954, *Biometrika*; Hawkings & Olwell 1998)

$$S_n^+ = \max\{0; S_{n \ 1}^+ + \frac{X_n - }{=\sqrt{n}} - k\}; \ S_0^+ = 0;$$

$$S_n \ = \min\{0; S_{n \ 1} + \frac{X_n - }{=\sqrt{n}} + k\}; \ S_0 \ = 0:$$

Other control charts (2)

- **Nonparametric Methods**: Hodges-Lehmann, Bootstrap, Sign statistics, Wilcoxon rank statistics,...(Chakraborti et al 2001, *JQT*)
- **Adaptive Control Chart** (王 2002，应用概率统计)
    - Variable Sample Size (VSS)
    - Variable Sampling Interval (VSI) (Reynolds et al 1988, *Tech.*)
    - Variable Sample Size and Sampling Interval (VSSI)
    - VSI at Fixed Time (VSIFT, Reynolds 1996, *JQT*)
    - SPRT control chart (Stoumbos & Reynolds, 1997, *JQT*, 2001, *Nonlinear Ananlysis*)
    - Adaptive EWMA (Capizzi & Masarotto 2003, *Technometrics*)
    - For correlated data (Reynolds, Arnold, & Baik (1996), *JQT*)
- **Self-starting** Control Chart (Quesenberry 1991, *JQT*)
- **Dual CUSUM** (Zhao, Tsung, Wang 2005, *IIE Transactons*)

Other control charts (2)

- **Nonparametric Methods**: Hodges-Lehmann, Bootstrap, Sign statistics, Wilcoxon rank statistics,...(Chakraborti et al 2001, *JQT*)
- **Adaptive Control Chart** (王 2002，应用概率统计)
  - Variable Sample Size (VSS)
  - Variable Sampling Interval (VSI) (Reynolds et al 1988, *Tech.*)
  - Variable Sample Size and Sampling Interval (VSSI)
  - VSI at Fixed Time (VSIFT, Reynolds 1996, *JQT*)
  - SPRT control chart (Stoumbos & Reynolds, 1997, *JQT*, 2001, *Nonlinear Ananlysis*)
  - Adaptive EWMA (Capizzi & Masarotto 2003, *Technometrics*)
  - For correlated data (Reynolds, Arnold, & Baik (1996), *JQT*)
- **Self-starting** Control Chart (Quesenberry 1991, *JQT*)
- **Dual CUSUM** (Zhao, Tsung, Wang 2005, *IIE Transactons*)

Other control charts (2)

- **Nonparametric Methods**: Hodges-Lehmann, Bootstrap, Sign statistics, Wilcoxon rank statistics,...(Chakraborti et al 2001, *JQT*)
- **Adaptive Control Chart** (王 2002，应用概率统计)
  - Variable Sample Size (VSS)
  - Variable Sampling Interval (VSI) (Reynolds et al 1988, *Tech.*)
  - Variable Sample Size and Sampling Interval (VSSI)
  - VSI at Fixed Time (VSIFT, Reynolds 1996, *JQT*)
  - SPRT control chart (Stoumbos & Reynolds, 1997, *JQT*, 2001, *Nonlinear Ananlysis*)
  - Adaptive EWMA (Capizzi & Masarotto 2003, *Technometrics*)
  - For correlated data (Reynolds, Arnold, & Baik (1996), *JQT*)
- **Self-starting** Control Chart (Quesenberry 1991, *JQT*)
- **Dual CUSUM** (Zhao, Tsung, Wang 2005, *IIE Transactons*)

Other control charts (2)

- **Nonparametric Methods**: Hodges-Lehmann, Bootstrap, Sign statistics, Wilcoxon rank statistics,...(Chakraborti et al 2001, *JQT*)
- **Adaptive Control Chart** (王 2002，应用概率统计)
  - Variable Sample Size (VSS)
  - Variable Sampling Interval (VSI) (Reynolds et al 1988, *Tech.*)
  - Variable Sample Size and Sampling Interval (VSSI)
  - VSI at Fixed Time (VSIFT, Reynolds 1996, *JQT*)
  - SPRT control chart (Stoumbos & Reynolds, 1997, *JQT*, 2001, *Nonlinear Ananlysis*)
  - Adaptive EWMA (Capizzi & Masarotto 2003, *Technometrics*)
  - For correlated data (Reynolds, Arnold, & Baik (1996), *JQT*)
- **Self-starting** Control Chart (Quesenberry 1991, *JQT*)
- **Dual CUSUM** (Zhao, Tsung, Wang 2005, *IIE Transactons*)

Kim, Mahmoud & Woodall (2003, *JQT*)

Model: $y_{ij} = B_0 + B_1 x_i \ + \ _{ij};\ i = 1; \cdots ; n;$ where $B_0 = A_0 + A_1 x;\ B_1 = A_1;\ x_i \ = x_i - x$ (code data).

$$\begin{cases} Z_I(j) = \ \ b_{0j} + (1 - \ )Z_I(j-1);\ Z_I(0) = B_0 \\ UCL = B_0 + L_I \ \sqrt{\frac{}{(2 \ \ )n}};\ LCL = B_0 - L_I \ \sqrt{\frac{}{(2 \ \ )n}} \end{cases}$$

$$\begin{cases} Z_S(j) = \ \ b_{1j} + (1 - \ )Z_S(j-1);\ Z_S(0) = B_1 \\ UCL = B_1 + L_S \ \sqrt{\frac{}{(2 \ \ )S_{xx}}};\ LCL = B_1 - L_S \ \sqrt{\frac{}{(2 \ \ )S_{xx}}} \end{cases}$$

$$\begin{cases} Z_E(j) = \max \ \ \ln(\mathrm{MSE}_j) + (1 - \ )Z_E(j-1); \ln \ _0^2 \ ; \\ Z_E(0) = \ln \ _0^2; \\ UCL = L_E \ \sqrt{\frac{}{1}\mathrm{Var}[\ln(\mathrm{MSE}_j)]}; \end{cases}$$

Brief Review for Monitoring Linear Profile

MEWMA (Zou, Tsung, & Wang *Tech* 2007)

- Model:

$$\mathbf{Y}_j = \mathbf{X}_j\beta + \varepsilon_j, \quad j = 1, 2, \ldots.$$

- Signal:

$$U_j = \mathbf{W}_j' \Sigma^{-1} \mathbf{W}_j > L \frac{}{2-},$$

where $\mathbf{W}_j = \mathbf{Z}_j + (1-)\mathbf{W}_{j-1}$, $\mathbf{Z}_j(\beta) = (\hat{\beta}_j - \beta)/$,
$Z_j(\ ) = {}^{-1}\ {}^{2}\ (n-p)\hat{\sigma}_j^2/{}^2; n-p$; $\hat{\beta}_j = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_j$,
$\hat{\sigma}_j^2 = \frac{1}{n-p}(\mathbf{Y}_j - \mathbf{X}\hat{\beta}_j)'(\mathbf{Y}_j - \mathbf{X}\hat{\beta}_j)$.

- The Estimator of Change point:

$$\hat{} = \arg\max_{0\ t<k}\{lr(t,n; k,n)\}.$$

MEWMA (Zou, Tsung, & Wang *Tech* 2007)

- Diagnostic Statistics: $(i = 2; \dots ; p)$

$$T_{test} = \frac{\mathrm{p}\,\overline{(k-\mathrm{b})n}\,(\tilde{}^{(1)}_{\mathrm{b};k} - {}^{(1)})}{\tilde{}_{\mathrm{b};k}}; \quad {}^{2}_{test} = \frac{[(k-\mathrm{b})n - p]\tilde{}^{2}_{\mathrm{b};k}}{2};$$

$$F^{(i)}_{test} : (k-\mathrm{b}) \quad \tilde{}^{(i)}_{\mathrm{b};k} - \quad {}^{(i)} \quad {}^{2} \cdot \quad m_{ii}\tilde{}^{2}_{\mathrm{b};k} > F \quad (p-1; (k-\mathrm{b})n - p; \mathbf{R})$$

where m_{ii}s are diagonal elements of $\mathbf{M} = (\mathbf{X}^{0}\mathbf{X}) \quad {}^{1}$, $\mathbf{R} = \mathrm{diag}\{m_{11}^{\frac{1}{2}}; \cdots ; m_{pp}^{\frac{1}{2}}\}\mathbf{M}\mathrm{diag}\{m_{11}^{\frac{1}{2}}; \cdots ; m_{pp}^{\frac{1}{2}}\}$, is the correlation matrix for $\hat{\beta}$ (For multivariate F, see Kotz *et al*.(2000). *Continuous Multivariate Deistributions*).

The Chart based on Change point for unknown parameters(Zou, Zhang, & Wang 2006, *IIE Transactions*)

- Model: The coded model (Kim et al 2003) with unkown parameters, but there are m IC historical data.
- Standardized LRT statistic:

$$slr(k_1 n; kn) = \frac{lr(k_1 n; kn) - E[lr(k_1 n; kn)]}{\sqrt{Var[lr(k_1 n; kn)]}};$$

where

$$lr(k_1 n; kn) = -2(l_0 - l_1) = kn \ln[b^2_{kn}(b^2_{k_1 n})^{\frac{k_1}{k}}(b^2_{k_2 n})^{\frac{k_2}{k}}]:$$

Brief Review for Monitoring Linear Profile

The Chart based on Change point for unknown parameters(Zou, Zhang, & Wang 2006, *IIE Transactions*)

- The plotted statistic for LRT chart is

$$slr_{\max;m;k} = \max_{m \quad k_1 < k} slr(k_1 n; kn):$$

If $slr_{\max;m;m+t} > h_{m;t}$, an out-of-control signal is given.

- The plotted statistic for EWMA chart is

$$Y_{max}(m;t) = \max_{m \quad j < m+t} Y_j(m;t);$$

where $Y_j(m;t) = \max \;\; 0; \;\; \cdot slr(jn;kn) + (1 - \;\;) \cdot Y_{j \;\; 1}(m;t) \;\;$, $Y_{m \;\; 1}(m;t) = 0$. If $Y_{max}(m;t) > h_{m;t}$, then an alarm is triggered.

The Chart based on Change point for unknown parameters(Zou, Zhang, & Wang 2006, *IIE Transactions*)

- The Control Limit: (Hawkins, Qiu, & Kang, 2003, *JQT*)
  For a given false alarm probability (FAP) , the control limit of our EWMA, $h_{m;t}(\ )$ can be obtained by solving the following equations:

$$Pr\ \ Y_{max}(m;t) > h_{m;t}(\ )\ Y_{max}(m;i) \leq h_{m;i}(\ ); 1 \leq i < t\ \ =\ \ ;\ t >$$

$$Pr\ \ Y_{max}(m;1) > h_{m;1}(\ )\ \ =\ \ :$$

- The estimation of change-point is

$$\mathrm{b} = \underset{m \quad k_1 < k}{arg} \max\{slr(k_1 n; kn)\}:$$

The Chart based on Change point for unknown parameters(Zou, Zhang, & Wang 2006, *IIE Transactions*)

- The diagnostic Statistics:

$$l_{lr}(\hat{\ }) = kn \ln \left[1 + \frac{k_1 k_2 (y_{k_1 n} - y_{k_2 n})^2}{k(k_1 \mathrm{b}^2_{k_1 n} + k_2 \mathrm{b}^2_{k_2 n})} \right];$$

$$_{lr}(\hat{\ }) = kn \ln \left[\frac{k_1 \mathrm{b}^2_{k_1 n} + k_2 \mathrm{b}^2_{k_2 n}}{k} (\mathrm{b}^2_{k_1 n})^{\frac{k_1}{k}} (\mathrm{b}^2_{k_2 n})^{\frac{k_2}{k}} \right];$$

$$S_{lr}(\hat{\ }) = kn \ln \left[1 + \frac{k_1 k_2 (\frac{1}{k_1} S_{xy(k_1 n)} - \frac{1}{k_2} S_{xy(k_2 n)})^2}{nS_{xx}[k(k_1 \mathrm{b}^2_{k_1 n} + k_2 \mathrm{b}^2_{k_2 n}) + k_1 k_2 (y_{k_1 n} - y_{k_2 n})^2]} \right]$$

can be used to detect the shifts in intercept (B_0), standard deviation , and slope (B_1).

Brief Review for Monitoring Linear Profile

The Self-starting chart with unknown parameters(Zou et al 2007 *JQT*)

- Model: Kang & Albin's (2000) model with unknown parameters, but there are m IC historical data.
- The Standardized Recursive Residuals for the future samples are defined by $(i = 1, 2, \cdots, n; \quad j = m + 1, m + 2, \cdots)$

$$e_{ij} = \frac{y_{(j-1)n+i} - z_i' \hat{\beta}_{(j-1)n+i-1}}{\sqrt{S_{(j-1)n+i-1}\left(1 + z_i'(\mathbf{X}'_{(j-1)n+i-1}\mathbf{X}_{(j-1)n+i-1})^{-1} z_i\right)}}$$

- $\{w_{ij}\}$ is defined as

$$w_{ij} = \Phi^{-1}\left[T_{(j-1)n+i-3}\left(e_{ij}\right)\right].$$

The Self-starting chart with unknown parameters (Zou et al 2007 *JQT*)

- The Plotted Statistics:

$$\mathrm{EWMA}_{\mathrm{IS}}(j) = \sqrt{n} w_j + (1 - \quad)\mathrm{EWMA}_{\mathrm{IS}}(j-1);$$

$$\mathrm{EWMA} \quad (j) = \max \quad 0; \frac{sw_j - 1}{\sqrt{2}} + (1 - \quad)\mathrm{EWMA} \quad (j-1) \quad ;$$

where $w_j = \frac{1}{n} \sum_{i=1} w_{ij}$ and $sw_j = \frac{1}{n \quad 1} \sum_{i=1} (w_{ij} - w_j)^2$

Nonparametric Regression (Zou, Tsung, and Wang, 2008, *Technometric*)

- **Model**

$$y_{ij} = g(x_{ij}) + "_{ij}; \quad i = 1; \cdots ; n \quad j = 1; 2; \cdots ;$$

where $x_1 \leq x_2 \leq \cdots \leq x_n$ and x_i varies in the interval $[0; 1]$, g is a known linear or nonlinear function, $"$ is iid from $N(0; \ ^2)$.

- **Hypothesis**

$$H_0 : g = g_0 \qquad = \ _0 \qquad \longleftrightarrow \qquad H_1 : g \neq g_0 \qquad = \ _0.$$

Brief Review for Monitoring Linear Profile

Nonparametric Regression (Zou et al. 2008, *Technometric*)

- **The vector-matrix form**

$$lr = \frac{1}{{}^{2}_{0}} \; (\mathbf{Y} - \mathbf{G}_0) \;\; - (\mathbf{Y} - \mathbf{W}\mathbf{Y}) \quad ;$$

  where $\mathbf{G}_0 = (g_0(x_1); g_0(x_2); \cdots ; g_0(x_n))^T$,
  $\mathbf{Y} = (y_1; y_2; \cdots ; y_n)^T$, $\mathbf{W} = (\mathbf{W}_n(x_1); \mathbf{W}_n(x_2); \cdots ; \mathbf{W}_n(x_n))^T$,
  $\mathbf{W}_n(x_i) = (W_{n1}(x_i); W_{n2}(x_i); \cdots ; W_{nn}(x_i))^T$, and $\mathbf{A}$ means
  $\mathbf{A}^T \mathbf{A}$.

- **the null distribution of** lr

  Fan et al (2001, Ann Statist) proved that the asymptotic null distribution of lr are independent of nuisance functions and approximately distributed as a scaled ^{2}.

- **Our difficulty**

  The small sample distribution of lr depends on g_0.

Brief Review for Monitoring Linear Profile

Nonparametric Regression (Zou et al. 2008, *Technometric*)

- **The transformed data and GLR**

  Transform each pro le data set, $\{y_i; x_i\}_{i=1}^{n}$, to $\{y_i - g_0(x_i); x_i\}_{i=1}^{n}$, the hypothesis is equivalent to

$$H_0 : g = 0; \quad = \quad _0 \leftrightarrow H_1 : g \neq 0; \quad = \quad _0.$$

  The GLR statistics are

$$lr_z = \mathbf{Z} \quad - (\mathbf{Z} - \mathbf{WZ}) \quad = \mathbf{Z}^T \mathbf{V} \mathbf{Z}^T;$$

  where $\mathbf{V} = \mathbf{W}^T + \mathbf{W} - \mathbf{W}$, $z_i = (y_i - g_0(x_i))/ \ _0$ and $\mathbf{Z} = (z_1; z_2; \cdots; z_n)^T$. Under H_0 and some conditions,

$$lr_z \xrightarrow{L} N(\ _z; \ ^2_z);$$

$$_z = \frac{2}{h} \ K(0) - \frac{1}{2} \int K^2(t)dt \ ; \ ^2_z = \frac{8}{h} \int \ K(t) - \frac{1}{2} K * K(t) \ ^2 dt.$$

Nonparametric Regression (Zou et al. 2008, *Technometric*)

- **For Monitoring the profile**

$$\mathbf{Z}_j = (\mathbf{Y}_j - \mathbf{G}_0) = {}_0;$$

where $\mathbf{Y}_j = (y_{j1}; y_{j2}; \cdots ; y_{jn})^T:$

- **For Monitoring the variance**
  Based on (Fan et. al. 2001, *Ann Statist*)

$$\hat{b}_j^2 = \frac{1}{n}(\mathbf{Z} - \mathbf{WZ}) \quad = \quad {}_j^2 + O_p(n^{\ \frac{1}{2}}) + O_p(n^{\ 1} h^{\ \frac{1}{2}});$$

the monitoring statistics for variance are

$$\tilde{}_j = \quad {}^1 \quad n\hat{b}_j^2; \mathbf{I} - \mathbf{V}) \quad ;$$

where $(\cdot; \mathbf{A})$ is the null CDF of $n\hat{b}_j^2$.

Brief Review for Monitoring Linear Profile

Nonparametric Regression (Zou et al. 2008, *Technometric*)

- **The calculation of $(\cdot; \mathbf{A})$**

  The null distribution of nb_j^2 is the linear combination of independent $^2_1$-variates with coe cients given by the eigenvalues of $(\mathbf{I} - \mathbf{V})$ (For proof, see Box 1954(*Ann Math Statist*), for algorithm, see Imhof 1961 (*Biometrika*))
  由于我们不必利用其精确形式，故仅利用与其三阶段矩匹配的即Œ (Imhof 1961, Azzalini & Bowman 1993, *JRSS*)

The EWMA charting statistics

$$\mathbf{E}_j = \ \mathbf{U}_j + (1 - \ )\mathbf{E}_{j\ 1} \quad j = 1; 2; \cdots ;$$

where $\mathbf{U}_j = (\mathbf{Z}_j^T; \tilde{\ }_j)^T$, $\mathbf{\Sigma} = diag(\mathbf{V}; 1)$.
The chart signals if

$$Q_j = \mathbf{E}_j^0 \mathbf{\Sigma} \mathbf{E}_j > L \frac{\quad}{2 - \ } :$$

Brief Review for Monitoring Linear Profile

Some remarks

- The form of our proposed NEWMA chart is similar to that in Zou, Tsung, and Wang (2007, *Technometrics*)
- We do not considered their correlation between $\mathbf{Z}_j$ and $\sim_j$ in the matrix $\mathbf{\Sigma}$.
- If g_0 is unknown, we can use the Phase I in-control data to estimate it.
- The normal distribution of $''_{ij}$ is used for motivating our proposed method only and is not necessary in asymptotic theory (only need to satisfy $E[''_{ij}] = 0$ and $E(|''_{ij}|^4) < \infty$). However, it's useful for evaluating $(\cdot; A)$.

Monitoring Profile Based on Nonparametric Regression

Guideline on design for the nonparametric regression

- **The smoothing parameter**: =0.2
- **The kernel function** $K(\cdot)$: The GLR test does not depend on the structures of the smoothing procedure (Fan et al. (2001)). Our simulated results also show this point of view. We use

$$K_E(u) = \frac{3}{4}(1 - u^2)I(|u| \leq 1);$$

- **Bandwidth**: The data-driven bandwidth methods (CV or GCV) may not be appropriate for our on-line monitoring problem. We use the empirical bandwidth:

$$h_E = c \times \left(\frac{1}{n} \sum_{i=1}^{n} (x_i - x)^2 \right)^{\frac{1}{2}} n^{\frac{1}{5}};$$

The diagnostic aids

- **Identify the change point**: The change point is estimated as

$$\mathrm{b} = \arg\max_{0 \quad t<k} \{lr(tn; kn)\}:$$

Under some conditions, such as $0 < \lim_{k\,!\,1} \; =k = \; < 1$, and $nh^5 = O(1)$, for the following two types of OC models:
(i) $g(u) = g_0(u) + \; _n(u)$, where $_n(u)$ has a continuous second derivative and its rate satis es $nh \int_0^1 \, {}^2_n(u) f(u) du \to \infty$ as $n \to \infty$. (ii) $\; = \; _0$, where the size of shift satis es $n^{\frac{1}{2}} | \; - 1| \to \infty$ as $n \to \infty$. We have

$$|\mathrm{b} - \;| = O_p(1):$$

The diagnostic aids

- **Detect whether the shift in variance occurs**: The statistics and acceptance region are

$$ {}_1{}^{1}\left(\frac{}{2}; \mathbf{I}-\mathbf{V}; k-b\right) < \sum_{j=b+1}^{k} \mathbf{Z}_j^T(\mathbf{I}-\mathbf{V})\mathbf{Z}_j < {}_1{}^{1}\left(1-\frac{}{2}; \mathbf{I}-\mathbf{V}; k-b\right) $$

where ${}_1(\cdot; \mathbf{A}; l)$ is the CDF of the sum of l independent random variables $\mathbf{Z}^T\mathbf{A}\mathbf{Z}$ and correspondingly ${}_1{}^{1}(\ ; \mathbf{A}; l)$ is the percentile of the ${}_1(\cdot; \mathbf{A}; l)$ distribution.

Brief Review for Monitoring Linear Profile

The diagnostic aids

- **Detect the shift in the regression function**: The test statistic

$$\frac{1}{(k-b)}\left(\sum_{j=b+1}^{k}\mathbf{Z}_j\right)^T \mathbf{V}\left(\sum_{j=b+1}^{k}\mathbf{Z}_j\right) > \ ^{1}(1- \ ;\mathbf{V})$$

  is used to detect if the regression function shifted, where $^{1}(\ ;\mathbf{A})$ is the percentile of the distribution of quadratic form $\mathbf{Z}^T\mathbf{AZ}$.

- **Detect the changed part of regression curve**: We suggest plotting the nonparametric smoothing curve of the average of $(k-b)$ sample pro les, say $\frac{1}{k \ b}\sum_{j=b+1}^{k}\mathbf{WY}_j$, and the IC pro le model together.

Monitoring Profile Based on Mixed Effect Model

Background and Motivation(Qiu, Zou, Wang, 2009, *Technometrics*)

In the literature, most pro le monitoring control charts require a fundamental assumption that random errors within a pro le are i.i.d., which is often invalid in applications. As an example, within-pro le data in the deep reactive ion etching (DRIE) example exhibit obvious serial correlation over time.

The Phase I Model

The Nonparametric Mixed E ect (NME) model is

$$y_{ij} = g(x_{ij}) + f_i(x_{ij}) + "_{ij}; \text{ for } j = 1;2;:::;n_i; \ i = 1;2;:::;m;$$

where g is the population pro le function (i.e., the xed-e ects term), f_i is the random-e ects term describing the variation of the i-th individual pro le from g, $\{x_{ij}; y_{ij}\}_{j=1}^{n_i}$ is the i-th sample collected for the i-th pro le, and $"_{ij}$s are i.i.d. random errors with mean 0 and variance ^{2}.

The Estimation in NME Model

- **For $g(s)$ and $f_i(s)$, $s \in [0, 1]$:** LLME is obtained by minimizing the following penalized local liner kernel likelihood function:

$$\sum_{i=1}^{m} \left\{ \frac{1}{\sigma^2} \sum_{j=1}^{n_i} [y_{ij} - \mathbf{z}_{ij}^T(\beta + \alpha_i)]^2 K_h(x_{ij} - s) + \alpha_i^T \mathbf{D}^{-1} \alpha_i + \ln|\mathbf{D}| + n_i \ln(\sigma^2) \right\},$$

where $K_h(\cdot) = K(\cdot/h)/h$, K is a symmetric density kernel function, h is a bandwidth, $\mathbf{z}_{ij}^T = (1, x_{ij} - s)$, β is a deterministic two-dimensional coefficient vector, and $\alpha_i \sim (0, \mathbf{D})$ is a two-dimensional vector of the random effects.
(LLME is proposed by Wu and Zhang (2002, *JASA*) by combining linear mixed effects modeling and local linear kernel smoothing for the longitudinal data.)

The Iterative procedure for evaluating the estimation

Step 1. Set the initial values for $\mathbf{D}$ and σ^2, denoted as $\mathbf{D}_{(\mathbf{0})}$ and $\sigma^2_{(0)}$.
Step 2. At the k-th iteration, for $k \geq 0$, compute estimates of β and α_i by solving the so-called mixed-model equation (cf., Davidian and Giltinan 1995; Wu and Zhang 2002), and the resulting estimates are denoted as

$$\hat{\beta}^{(k)} = \{\sum_{i=1}^{m} \mathbf{Z}_i^T \mathbf{\Sigma}_i \mathbf{Z}_i\}^{-1}\{\sum_{i=1}^{m} \mathbf{Z}_i^T \mathbf{\Sigma}_i \mathbf{y}_i\}$$

$$\hat{\alpha}_i^{(k)} = \{\mathbf{Z}_i^T \mathbf{K}_i \mathbf{Z}_i + \sigma^2_{(k)}[\mathbf{D}_{(k)}]^{-1}\}^{-1} \mathbf{Z}_i^T \mathbf{K}_i (\mathbf{y}_i - \mathbf{Z}_i \hat{\beta}^{(k)});$$

where $\mathbf{Z}_i = (\mathbf{z}_{i1}, \cdots, \mathbf{z}_{in_i})^T$, $\mathbf{y}_i = (y_{i1}, \cdots, y_{in_i})^T$,
$\mathbf{\Sigma}_i = (\mathbf{Z}_i \mathbf{D}_{(k)} \mathbf{Z}_i^T + \sigma^2_{(k)} \mathbf{K}_i^{-1})^{-1}$ and
$\mathbf{K}_i = \mathrm{diag}\{K_h(x_{i1} - s), \cdots, K_h(x_{in_i} - s)\}$.

The Iterative procedure for evaluating the estimation

Step 3. Based on $\hat{\beta}^{(k)}$ and $\hat{\alpha}_i^{(k)}$, update $\mathbf{D}$ and σ^2 by

$$\mathbf{D}_{(k+1)} = \frac{1}{m}\sum_{i=1}^{m} \hat{\alpha}_i^{(k)}[\hat{\alpha}_i^{(k)}]^T$$

$$\sigma^2_{(k+1)} = \frac{1}{m}\sum_{i=1}^{m} \frac{1}{n_i}[\mathbf{y}_i - \mathbf{Z}_i(\hat{\beta}^{(k)} + \hat{\alpha}_i^{(k)})]^T \mathbf{K}_i [\mathbf{y}_i - \mathbf{Z}_i(\hat{\beta}^{(k)} + \hat{\alpha}_i^{(k)})].$$

Step 4. Repeat Steps 2-3 until the following condition is satisfied:

$$\|\mathbf{D}_{(k)} - \mathbf{D}_{(k-1)})\|_1 / \|\mathbf{D}_{(k-1)}\|_1 \leq \varepsilon;$$

where ε is a pre-specified small positive number (e.g., $\varepsilon = 10^{-4}$), and $||\mathbf{A}||_1$ denotes the sum of absolute values of all elements of $\mathbf{A}$. Then, the algorithm stops at the k-th iteration.

The Estimation in NME Model

- **The initial value**: A simple but e ective method is to set $\mathbf{D}_{(0)}$ to be the identity matrix and

$$ {}^{2}_{(0)} = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{n_i}\sum_{j=1}^{n_i}[y_{ij} - \hat{g}^{(P)}(x_{ij})]^2; $$

  where $\hat{g}^{(P)}(x_{ij})$ is the standard local linear kernel estimator constructed from the pooled data

- **The Properties:** Under some conditions, we have
  (i) $\hat{g}(s_1) = g(s_1)\{1 + O_p[m^{\ \frac{1}{2}} + O(h^2)]\}$;
  (ii)
  $\mathrm{b}(s_1; s_2) = \ (s_1; s_2)\{1 + O_p[h^2 + (nh)^{\ \frac{1}{2}} + m^{\ \frac{1}{2}} + (mnh^3)^{\ \frac{1}{2}}]\}$;
  (iii) $\mathrm{b}^2 = \ {}^2\{1 + O_p[h^2 + (nh)^{\ \frac{1}{2}} + m^{\ \frac{1}{2}} + (mnh^3)^{\ \frac{1}{2}}]\}$.

Phase II Nonparametric Pro le Monitoring

For Phase II Monitoring, we have to face two major issues:

- Amount of computation for on line monitoring;
- The response is observed at di erent design points in di erent pro les.

To overcome the above di culties, at any point $s \in [0; 1]$, we consider the following weighted local likelihood:

$$WL(a; b; s; \quad ; t) = \sum_{i=1}^{t} \sum_{j=1}^{n} [y_{ij} - a - b(x_{ij} - s)]^2 K_h(x_{ij} - s) \frac{(1 - \quad)^{t \quad i}}{{}^2(x_{ij})};$$

where is a weighting parameter and ${}^2(x) = \quad (x; x) + \quad {}^2$ is the variance function of the response.

Phase II Nonparametric Profile Monitoring

The local linear kernel estimator of $g(s)$, defined as the solution to a of the minimization problem $\min_{a,b} WL(a, b, s, \lambda, t)$, has the expression

$$\hat{g}_{t,h,\lambda}(s) = \sum_{i=1}^{t}\sum_{j=1}^{n_i} U_{ij}^{(t,h,\lambda)}(s) y_{ij} \Big/ \sum_{i=1}^{t}\sum_{j=1}^{n_i} U_{ij}^{(t,h,\lambda)}(s),$$

where

$$U_{ij}^{(t,h,\lambda)}(s) = \frac{(1-\lambda)^{t-i} K_h(x_{ij}-s)}{\sigma^2(x_{ij})}\left[m_2^{(t,h,\lambda)}(s) - (x_{ij}-s) m_1^{(t,h,\lambda)}(s) \right],$$

$$m_l^{(t,h,\lambda)}(s) = \sum_{i=1}^{t} (1-\lambda)^{t-i} \sum_{j=1}^{n_i} (x_{ij}-s)^l K_h(x_{ij}-s)/\sigma^2(x_{ij}), \quad l = 0, 1, 2.$$

when $\lambda = 0$, the resulting estimation is similar to the GEE estimation (Lin and Carrol, 2000, *JASA*)

Monitoring Profile Based on Mixed Effect Model

Phase II Monitoring Statistics

For testing the hypothesis $H_0 : g = 0 \leftrightarrow H_1 : g \neq 0;$ a natural statistic would be

$$T_{t;h;} = c_{0;t;} \int \frac{[\hat{g}_{t;h;}(s)]^2}{{}^2(s)} {}_1(s)ds;$$

where $c_{t_0;t_1;} = a^2_{t_0;t_1;} = b_{t_0;t_1;}$; $a_{t_0;t_1;} = \sum_{i=t_0+1}^{t_1} (1 -)^{t_1 \; i} n_i; b_{t_0;t_1;} = \sum_{i=t_0+1}^{t_1} (1 -)^{2(t_1 \; i)} n_i;$ and ${}_1$ is some pre-speci ed density function.

In practice, we suggest using the following discretized version:

$$T_{t;h;} \approx \frac{c_{0;t;}}{n_0} \sum_{k=1}^{n_0} \frac{[\hat{g}_{t;h;}(s_k)]^2}{{}^2(s_k)};$$

where $\{s_k; k = 1; \ldots; n_0\}$ are some pre-speci ed i.i.d. design points.

Monitoring Profile Based on Mixed Effect Model

The Asymptotic Properties of $T_{t;h;}$

Theorem 1 Under some conditions, we have

(i) If $n_i h$ is bounded (for each i), then

$$(T_{t;h;} - \sim_h) = \sim_h \xrightarrow{L} N(0;1);$$

where $\sim_h = \frac{\int [K(u)]^2 du}{h} \int \frac{\Gamma_1(x)}{\Gamma_2(x)} dx; \quad \sim_h^2 = \frac{2 \int [K \quad K(u)]^2 du}{h} \int \frac{\Gamma_1^2(x)}{\Gamma_2^2(x)} dx:$

(ii) If $n_i h \to \infty$(for each i), then

$$\frac{1}{d_{0;t;}} T_{t;h;} \overset{D}{\sim} \frac{1}{n_0} \zeta^T \zeta;$$

where $d_{t_0;t;} = \sum_{i=t_0+1}^{t_1} (1 - \quad)$

The Asymptotic Properties of $T_{t;h;}$

- **Some Remarks**: The asymptotic distribution of $T_{t;h;}$ depends on whether $n_i h$ is bounded; $\boldsymbol{\Omega}$ may not be positive de nite.
- **The OC model** is

$$y_{ij} = \begin{array}{ll} g_0(x_{ij}) + f_i(x_{ij}) + "_{ij}; & \text{if} \quad 1 \leq i \leq \\ g_1(x_{ij}) + f_i(x_{ij}) + "_{ij}; & \text{if} \quad i > \end{array}$$

where is an unknown change point, and $g_1(x) = g_0(x) + \;(x)$ is the unknown OC regression function.

- **Some notations**:

$$= \int \quad (u) + \frac{h^2}{2}{}_1 \;{}^{00}(u) \;{}^2 \frac{{}_1(u)}{{}^2(u)} du; \qquad {}_1 = \int K(t)t^2 dt;$$

$$_1 = \int {}^2(u) \frac{{}_1(u) \;\;(u;u)}{{}^2(u)} du; \qquad {}_2 = \int [\;{}^{00}(u)]^2 \;{}_1(u) du:$$

Monitoring Profile Based on Mixed Effect Model

The OC Asymptotic Properties of $T_{t;h;}$

Theorem 2 Under some conditions, we have

(i) If $n_i h$ is bounded for each i, $c_{0;t;}\ nh\ _1 \to 0$, then

$$(T_{t;h;}\ - \tilde{\ }_h - c_{0;t;}\quad) = \tilde{\ }_h \xrightarrow{L} N(0;1);$$

(ii) If $n_i h$ is bounded for each i, $\ _2 \to 0$, then $T_{t;h;}$ has nontrivial power (i.e., greater than the nominal level) when $\ \propto c_{0;t;}^{4=9}$ and $h = O(c_{0;t;}^{2=9})$.

(iii) If $n_i h \to \infty$ for each i, then

$$\frac{1}{d_{0;t;}} T_{t;h;}\ \overset{D}{\sim} \frac{1}{n_0} \zeta^T \zeta;$$

where $\zeta \sim N_{n_0}(\delta;\Omega)$, where $\delta = [\ (s_1);:::;\ (s_{n_0})]^T$.

Some Remarks

- **The size of m and n_i**: To attain the desirable IC distribution properties, we recommend to use IC data with $n_i \geq 40$ and $m \geq 80$.
- **The choice of bandwidth**: For the Phase I model, the CV method by combining leave-one-subject-out and leave-one-point-out (Wu and Zhang 2002) is used; For the Phase II model, we suggest using the following empirical bandwidth formula,

$$h_E = \begin{cases} c_1 n^{-\frac{1}{5}} \left[\sum_{i=1} (x_i - \bar{x})^2 / n\right]^{\frac{1}{2}} & \text{for balanced design} \\ c_2[\bar{n}(2 - \ )/\ ]^{-1/5} \sqrt{\mathrm{Var}(x)} & \text{for random design;} \end{cases}$$

where $\bar{n}$ and $\mathrm{Var}(x)$ are the averaged number of design points and the variance of design points within a profile, respectively, constants $c_1; c_2 \in [1; 2]$.

Some Remarks

- **The choice of** : Traditionally, a larger leads to a quicker detection of larger shifts. However, in the mixed-e ect modeling, this is not true, the reason is the common pro le g is needed to estimate. From Theorems 1 and 2, the cannot be chosen too large. Empirically, we suggest choosing $\in [0{:}02; 0{:}1]$.
- **The choice of** $\{s_k; k = 1; 2; : : : ; n_0\}$: Based on our numerical experience, selection of $\{s_k; k = 1; 2; : : : ; n_0\}$ does not a ect performance of our chart much, as long as n_0 is not too small and s_ks cover all the key parts of g_0. ($n_0 \geq 40$)

Monitoring Profile Based on Mixed Effect Model

Simulation Study

- **Parameters**: $m = 500; n = 200$. IC ARL=200, Kernel $K(x) = 0{:}75(1 - x^2)I(-1 \leq x \leq 1)$, $n_i = 20$, $x_{ij} \sim U(0;1)$, $s_k = (k - 0{:}5){=}n_0$, $n_0 = 40$, $\ = 30$, $c_2 = 1{:}5$, $\ = 0{:}1$.
- **Competitor**: The control chart based on xed-e ects modeling for monitoring nonparametric pro les as an alternative method, denoted as FENPC. In this case, $f_i = 0;\ ^2(x) = \ ^2$.
- **IC model**:

$$\text{(I)} : f_i(x_{ij}) = 0; \quad \text{(III)} : f_i(x_{ij}) = b\ \ _i \cos(2\ \ x_{ij});$$

$$\text{(II)} : f_i(x_{ij}) = b\ \ _i x_{ij}; \quad \text{(IV)} : [f_i(x_{i1}); : : : ; f_i(x_{in})]^T \sim b \cdot \mathrm{N}_n(\mathbf{0}; \mathbf{\Sigma});$$

where $\ _i; i = 1; 2; : : : ;$ are independent standard normal random variables, $\mathbf{\Sigma} = (\ _{jk})$ and $\ _{jk} = 0{:}2^{jx_{ij}\ \ x_{ik}j}$, b is a constant.

- **OC model**:

$$\text{(i)} : g_1(x) = 2\ \ (x - 0{:}5); \quad \text{(ii)} : g_1(x) = \ \ \sin(2\ \ (x - 0{:}5)){:}$$

Simulated Results For IC ARL

Table: IC ARL and SDRL values of charts MENPC and FENPC.

	b	Model (I) ARL	Model (I) SDRL	Model (II) ARL	Model (II) SDRL	Model (III) ARL	Model (III) SDRL	Model (IV) ARL	Model (IV) SDRL
	0.25	205	203	196	197	198	199	206	208
MENPC	0.50	205	203	201	200	195	194	208	204
	1.00	205	203	193	190	194	194	206	205
	0.25	199	197	110	109	170	172	38.3	34.0
FENPC	0.50	199	197	29.8	29.2	105	104	21.5	20.0
	1.00	199	197	8.48	8.29	35.5	34.6	15.1	14.2

Simulated Results For OC ARL

		OC Model (i)		OC Model (ii)	
		MENPC	FENPC	MENPC	FENPC
IC model (II)	0.20	130 (1.36)	139 (1.48)	85.3 (0.83)	100 (0.98)
	0.30	80.5 (0.78)	98.0 (0.99)	40.5 (0.32)	52.2 (0.46)
	0.40	48.6 (0.42)	62.6 (0.59)	22.3 (0.15)	29.0 (0.21)
	0.60	20.7 (0.13)	28.4 (0.20)	10.6 (0.05)	13.1 (0.06)
	0.80	12.1 (0.06)	16.0 (0.09)	6.81 (0.03)	8.57 (0.03)
	1.20	6.64 (0.02)	8.43 (0.03)	4.06 (0.02)	5.14 (0.02)
	1.60	4.60 (0.02)	5.82 (0.02)	2.93 (0.01)	3.71 (0.01)
	2.00	3.51 (0.01)	4.49 (0.01)	2.33 (0.01)	2.96 (0.01)
IC model (III)	0.20	131 (1.38)	162 (1.73)	68.3 (0.64)	121 (1.25)
	0.30	81.0 (0.79)	121 (1.26)	31.2 (0.24)	65.7 (0.60)
	0.40	48.1 (0.42)	81.2 (0.76)	17.6 (0.11)	34.2 (0.25)
	0.60	21.4 (0.14)	33.3 (0.24)	9.05 (0.04)	14.4 (0.06)
	0.80	12.4 (0.06)	17.7 (0.09)	6.02 (0.02)	9.14 (0.03)
	1.20	6.59 (0.03)	9.04 (0.03)	3.70 (0.01)	5.39 (0.02)
	1.60	4.51 (0.02)	6.10 (0.02)	2.68 (0.01)	3.92 (0.01)
	2.00	3.43 (0.01)	4.71 (0.01)	2.20 (0.01)	3.15 (0.01)

Monitoring Profile Based on Mixed Effect Model

Case Study

- **Data set** (Walker and Wright 2002, *JQT*): This dataset is from a manufacturing process of particle boards, whose density properties are critical to their machinability and therefore require careful control and monitoring. Density readings are collected by a laser device at xed vertical depths of a board. So, observations from a single board can be regarded as a pro le. In the dataset, there are 24 pro les. In each pro le, design points are xed at $x_j = 0:002 \times j$, for $j = 0; : : : ; 313$.
  The data exhibit a signi cant amount of positive autocorrelation.
- **Existing work**: Walker and Wright (2002) compare multiple VDP pro les using additive modeling and smoothing splines. Williams *et al.* (2007) focus on the Phase I analysis using nonlinear regression curve estimation.
- **IC data** As Williams et al. (2007) Pointed out the 15th pro le is an outlier, the remaining 23 pro les are used as the IC data.

Monitoring Profile Based on Mixed Effect Model

Case Study

Figure 1: The resulting estimation of g.

Case Study

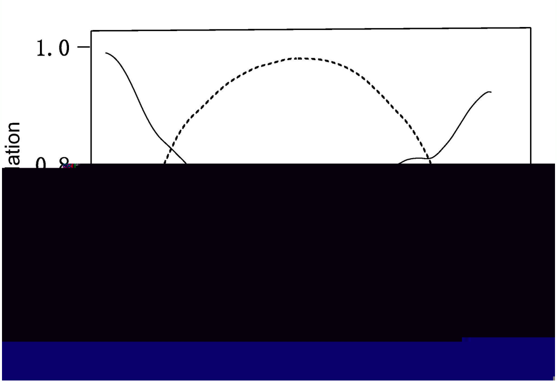


Figure 2: The estimated correlation functions $(1;\cdot)$, and $(157;\cdot)$.

Case Study

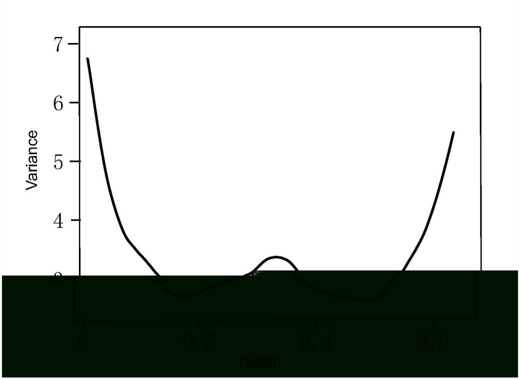


Figure 3: The estimated variance functions ${}^2(x)$.

Case Study

Figures 1-3 show us there is strongly correlation within-pro le. To demonstrate our proposed method, we assume g shifts from $g_0 = \hat{g}$ to $g_1(x) = x$ and $f_1(x) = \cos(2\ x)$, respectively, at time $\ = 23$, and choose $s_k = x_k; k = 1; 2; \cdots ; 314$. The IC ARL is xed at 370. The simulated OC ARL is 9.33 and 14.30, respectively.

Conclusions & Considerations

Conclusions

- 本报告简单地总结了关于Linear Pro le, Nonlinear Pro le及混合效应Pro le的监 › 方法，尤其是较详细地介绍了基于非参数回归及非参混合效应模型的监 › 方法.
- 最近也有作 讨论利用异常点的检测方法来监 › Pro le，请见*IIE Transactions*, 2009.

Considerations

- 我们仍Œ以把单指标模型、半参数模型和变系数模型的一些方法引入Pro le的监 › .
- 对于高维的生物芯片，如何监 › 其质量.

Thank you all for your attention!

Any question or comment?

Welcome to Nankai!!

Thank you all for your attention!

Any question or comment?

Welcome to Nankai!!

Thank you all for your attention!

Any question or comment?

Welcome to Nankai!!