# On-line statistical monitoring of batch processes using Gaussian mixture model

**Tao Chen** *, **Jie Zhang** **

* *School of Chemical and Biomedical Engineering, Nanyang Technological University, Singapore 637459 (e-mail: chentao@ntu.edu.sg).*
** *School of Chemical Engineering and Advanced Materials, Newcastle University, Newcastle upon Tyne, NE1 7RU, U.K. (e-mail: jie.zhang@ncl.ac.uk)*

**Abstract:** The statistical monitoring of batch manufacturing processes is considered. It is known that conventional monitoring approaches, e.g. principal component analysis (PCA), are not applicable when the normal operating conditions of the process cannot be sufficiently represented by a Gaussian distribution. To address this issue, Gaussian mixture model (GMM) has been proposed to estimate the probability density function of the process nominal data, with improved monitoring results having been reported for continuous processes. This paper extends the application of GMM to on-line monitoring of batch processes, and the proposed method is demonstrated through its application to a batch semiconductor etch process.

*Keywords:* Batch processes, mixture model, principal component analysis, probability density estimation, multivariate statistical process monitoring.

## 1. INTRODUCTION

Batch processing is of great importance in many industrial applications due to its flexibility for the production of low-volume, high-value added products. With increasing commercial competition it is crucial to ensure consistent and high product quality, as well as process safety. These requirements have resulted in wide acceptance of the technique of multivariate statistical process monitoring (Martin et al., 1999; Qin, 2003). The basis of the monitoring schemes is historical data that has been collected when the process is running under normal operating conditions (NOC). This data is then used to establish confidence bounds for the monitoring statistics, e.g. Hotelling's $T^2$ and squared prediction error (SPE), to detect the onset of process deviations. The primary objective of process monitoring is to identify abnormal behavior as early as possible, in addition to keeping an acceptably low false alarm rate.

As a result of the multi-way characteristic of batch process data, special tools are required for the modelling and monitoring purposes, including multi-way principal component analysis (MPCA) (Nomikos and MacGregor, 1995b), hierarchical PCA (Rannar et al., 1998) and multi-way partial least squares (MPLS) (Nomikos and MacGregor, 1995a). The methods for on-line monitoring of batch process can be classified into two categories. The first does not require measurements on the entire batch duration to be available. Techniques that are within this class include hierarchical and two-dimensional dynamic PCA (Rannar et al., 1998; Lu et al., 2005). In the other category, the entire batch data is required for the calculation of the monitoring statistics, whilst the data from a new batch is available only up to the current time. Therefore the future data must be predicted in some way (Nomikos and MacGregor, 1995b). In this paper the latter of the two approaches is considered, and the details will be discussed subsequently in Section 2.

However, the afore reviewed conventional process monitoring methods are based on a restrictive assumption that the NOC can be represented by a multivariate Gaussian distribution. Specifically the confidence bounds for $T^2$ and SPE are calculated by assuming the PCA/PLS scores and prediction errors are Gaussian distributed. This assumption may be invalid when the process data is collected from a complex manufacturing process. To address this issue, Gaussian mixture model (GMM) (Chen et al., 2006; Choi et al., 2004; Thissen et al., 2005), which is capable of approximating any probability density function (*pdf*), has been proposed for the monitoring of continuous processes, as well as batch-wise monitoring of batch processes.

The major contribution of this paper is to extend the application of GMM to on-line monitoring of batch processes. As the first step MPCA is applied to the nominal batch data to extract the low-dimensional representation of the process. The challenge with on-line monitoring is that the scores and SPE must be predicted based on available process measurements up to the current time step. Clearly the predicted scores and SPE are not identical to the values that are calculated from the entire batch duration, and thus the predictions may not conform to the nominal distribution even if the process is running normally. We follow the approach of Nomikos and MacGregor (1995b) to pass the nominal batches through the monitoring procedure and collect the predicted scores and SPE at each time step. Then GMM is employed to estimate the joint *pdf* of these predicted scores and SPE from MPCA at each time step, as opposed to the traditional $T^2$ and SPE where the process data is assumed to be Gaussian distributed.

The rest of this paper is organized as follows. Section 2 gives a summary of the PCA and GMM tools for process monitoring, followed by the discussion of the on-line monitoring strategy in Section 3. Section 4 demonstrates the application of the on-line monitoring techniques to a batch semiconductor manufacturing process. Finally Section 5 concludes this paper.

## 2. PCA AND GAUSSIAN MIXTURE MODEL FOR PROCESS MONITORING

This section presents a brief overview of the PCA and GMM techniques. A number of issues related to the application to process monitoring are discussed, including model selection and the construction of confidence bound.

### *2.1 PCA*

Principal component analysis (PCA) (Jolliffe, 2002) is a general multivariate statistical projection technique for dimension reduction, where the original data is linearly projected onto low-dimensional space such that the variance is maximized. Formally the $D$-dimensional data $\mathbf{x}$ is represented by a linear combination of the $Q$-dimensional scores $\mathbf{t}$ plus a noise vector $\mathbf{e}$: $\mathbf{x} = \mathbf{W}\mathbf{t} + \mathbf{e}$, where $\mathbf{W}$ are the eigenvectors of the sample covariance matrix having the $Q$ largest eigenvalues ($Q \leq D$). Consequently normal process behavior can be characterized by the first $Q$ principal components, which capture the main source of data variability.

The proper number of principal components can be selected using a number of criteria, including variance ratio, cross-validation and the "broken-stick" rule (Jolliffe, 2002). This is essentially a model selection problem. The "broken-stick" rule is adopted in this paper due to its low computation and good results reported in the literature (Nomikos and MacGregor, 1995b). According to this rule, the $q$-th principal component should be retained if the percentage of variance explained by it exceeds the corresponding $G$ value given by

$$G(q) = \frac{100}{C} \sum_{i=q}^{C} \frac{1}{i} \tag{1}$$

where $C = \min(D, N)$.

In statistical process monitoring, the next step is to define the monitoring statistics and the corresponding confidence bounds. Traditionally two metrics are used: $T^2 = \mathbf{t}^{\mathrm{T}}\mathbf{\Lambda}^{-1}\mathbf{t}$ and SPE as $r = \mathbf{e}^{\mathrm{T}}\mathbf{e}$, where $\mathbf{\Lambda}$ is a diagonal matrix comprising the $Q$ largest eigenvalues.

As discussed previously, the first issue with $T^2$ and SPE is that the corresponding confidence bounds are calculated based on restrictive Gaussian distribution. Secondly two separate metrics are required for process monitoring. Practically the process is identified as deviating from normal operation if either $T^2$ or SPE moves outside the confidence bounds. This empirical solution could potentially increase the false alarm level [1] . The technique of GMM is suitable for addressing the two issues simultaneously. In our previous work (Chen et al., 2006) we have demonstrated that a unified monitoring statistic can be obtained by estimating the joint *pdf* of the PCA scores and log-SPE using GMM, i.e. the *pdf* of a $(Q+1)$-dimensional vector $\mathbf{z} = (\mathbf{t}^{\mathrm{T}}, \log r)^{\mathrm{T}}$. The logarithm operator is used to transform the non-negative SPE onto the whole real axis on which the GMM is defined.

In this paper the methodology in (Chen et al., 2006) is followed to establish the confidence bounds for process monitoring based on PCA and GMM techniques. GMM is described in detail in the next subsection.

### *2.2 Gaussian mixture model*

As a general tool for *pdf* estimation, Gaussian mixture model (GMM) has been used in a wide variety of problems in applied statistics and pattern recognition. A GMM is a weighted sum of $M$ component densities, each being a multivariate Gaussian with mean $\boldsymbol{\mu}_i$ and covariance matrix $\mathbf{\Sigma}_i$:

$$p(\mathbf{z}|\boldsymbol{\theta}) = \sum_{i=1}^{M} \alpha_i G(\mathbf{z};\, \boldsymbol{\mu}_i, \mathbf{\Sigma}_i) \tag{2}$$

where the weights satisfy the constraint: $\sum_{i=1}^{M} \alpha_i = 1$. A GMM is parameterized by the mean vectors, covariance matrices and mixture weights: $\boldsymbol{\theta} = \{\alpha_i, \boldsymbol{\mu}_i, \mathbf{\Sigma}_i; i = 1, \ldots, M\}$.

Given a set of training data $\{\mathbf{z}_n, n = 1, \ldots, N\}$, the parameters can be estimated by maximizing the likelihood function: $L(\boldsymbol{\theta}) = \prod_{n=1}^{N} p(\mathbf{z}_n|\boldsymbol{\theta})$. In the context of process monitoring, $\mathbf{z}_n$ is the $(Q + 1)$-dimensional vector of PCA scores and log-SPE: $\mathbf{z}_n = (\mathbf{t}_n^{\mathrm{T}}, \log r_n)^{\mathrm{T}}$. The maximization is typically implemented iteratively using the expectation-maximization (EM) algorithm (Dempster et al., 1977).

The number of mixture components, $M$, must be selected prior to the training of a GMM. This is a model selection problem that can be addressed using a number of methods, including cross-validation and Bayesian information criterion (BIC) (Schwarz, 1978). BIC is widely applied in model selection problems for its effectiveness and low computational cost. According to BIC the model is selected such that $L - (H/2) \log N$ is the largest, where $L$ is the log-likelihood of the data and $H$ is the total number of parameters within the model. The motivation of BIC is that a good model should be able to sufficiently explain the data (the log-likelihood) with low model complexity (the number of parameters). In this study BIC is adopted for the selection of number of mixtures.

One of the advantages of the GMM for process monitoring is that it provides the likelihood value as the single statistic for the construction of confidence bounds, as opposed to the confidence bounds for two statistics (i.e. the $T^2$ and SPE) in conventional process monitoring techniques. In

[1] Suppose 95% confidence bound is used, and thus by definition the false alarm rate is 5% for both $T^2$ and SPE. The probability of either $T^2$'s bound or SPE's bound being exceeded, when the process is runing normally, will be equal to or greater than 5%.

practice a single monitoring statistic simplifies the plant operators' decision effort, and it may be more sensitive to some subtle process faults (Chen et al., 2006).

On the basis of the *pdf* $p(\mathbf{z}|\boldsymbol{\theta})$ for the normal operating data, the $100\beta\%$ confidence bound is defined as a likelihood threshold $h$ that satisfies the following integral (Chen et al., 2006):

$$\int_{\mathbf{z}:p(\mathbf{z}|\boldsymbol{\theta})>h} p(\mathbf{z}|\boldsymbol{\theta})d\mathbf{z} = \beta \tag{3}$$

To determine the confidence bound, we can calculate the likelihood of all the nominal data, and then find $h$ that is less than the likelihood of $100\beta\%$ (e.g. 99%) of the nominal data (Thissen et al., 2005). This approach is applicable to most continuous processes where the number of nominal data points can be up to several thousand; however it may be unreliable when the nominal data is very limited as in batch process monitoring. The estimation of the confidence bound based on limited batches would be very sensitive to the data, and thus a small perturbation in the data would result in very different estimation of the $h$.

To address this issue, we resort to numerical Monte Carlo simulation to approximate the integral in Eq. (3) (Chen et al., 2006). Specifically we generate $N_s$ random samples, $\{\mathbf{z}^j, j = 1, \ldots, N_s\}$, from $p(\mathbf{z}|\boldsymbol{\theta})$. These samples serve as the "pseudo data" (since the real data is not sufficient) to represent the normal process behavior. Thus the Monte Carlo samples, in conjunction with nominal process data, are used to calculate the confidence bound $h$. Then a new batch $\mathbf{z}$ is considered to be faulty if $p(\mathbf{z}|\boldsymbol{\theta}) < h$ (or equivalently $-p(\mathbf{z}|\boldsymbol{\theta}) > -h$). The number of Monte Carlo samples required ($N_s$) to approximate the confidence bounds is dependent on the dimension of $\mathbf{z}$, and it can be determined heuristically.

## 3. MONITORING OF BATCH PROCESSES

To analyze the three-way batch data ($N \times J \times K$) ($N$, $J$ and $K$ denote the number of batches, process variables at each time instance, and time steps, respectively), multi-way analysis methods have been proposed to unfold the data array into a two-way matrix on which conventional PCA is then performed (Nomikos and MacGregor, 1995b). This study unfolds the data array into a large matrix ($N \times JK$) such that each batch is treated as a "data point". This two-way matrix is then pre-processed to zero mean and unit standard deviation on each column, prior to the application of PCA to extract the scores $\mathbf{t}_n$ and SPE $r_n$, $n = 1, \ldots, N$. Then a Gaussian mixture model is developed for the joint vector $\mathbf{z}_n = (\mathbf{t}_n^{\mathrm{T}}, \log r_n)^{\mathrm{T}}$, followed by the calculation of confidence bound using Monte Carlo simulation.

### *3.1 On-line monitoring*

In the on-line monitoring stage, it is necessary to project the new batch onto the PCA space to obtain the scores and SPE, and then to calculate the likelihood value under the GMM to identify possible process anomaly. The issue is that, at time step $t$, the batch measurements are only available up to the current time. It is possible to develop multiple PCA and GMM models at each time step; however this strategy requires excessive computation and computer memory. A more reasonable and widely accepted method is to predict the scores and SPE using the available measurements.

More specifically, let $\bar{\mathbf{x}}_{1:t}$ be the vector of a new batch with available measurements from time step 1 to $t$. Note $\bar{\mathbf{x}}_{1:t}$ is a vector of order $Jt$. According to Nomikos and MacGregor (1995b), the least square prediction of the scores is:

$$\bar{\mathbf{t}}_{1:t} = \left(\mathbf{W}_{1:t}^{\mathrm{T}}\mathbf{W}_{1:t}\right)^{-1}\mathbf{W}_{1:t}^{\mathrm{T}}\bar{\mathbf{x}}_{1:t} \tag{4}$$

where $\mathbf{W}_{1:t}$ is the sub-matrix of $\mathbf{W}$ having the rows corresponding to time step 1 to $t$. In Eq. (4) the matrix to be inverted is well conditioned due to the orthogonality of the loading $\mathbf{W}$. Since the future measurements are not available, the prediction error can only be calculated up to time step $t$:

$$\bar{\mathbf{e}}_{1:t} = \bar{\mathbf{x}}_{1:t} - \mathbf{W}_{1:t}\bar{\mathbf{t}}_{1:t} \tag{5}$$

The SPE is then obtained as $\bar{\mathbf{e}}_{1:t}^{\mathrm{T}}\bar{\mathbf{e}}_{1:t}$. It was suggested to use the "instantaneous" SPE associated with the latest on-line measurements for process monitoring (Nomikos and MacGregor, 1995b), i.e. $\bar{\mathbf{e}}_t^{\mathrm{T}}\bar{\mathbf{e}}_t$, which is expected to increase the sensitivity of fault detection method. However the instantaneous SPE leads to an excessive number of false alarms in the case study of this paper (see details in Section 4). This phenomenon could be due to the non-Gaussian distribution of the process data. The SPE calculated from Eq. (5), which in a sense is a smoothed version of the instantaneous SPE, may be a more appropriate monitoring metric. We will discuss this issue through the application study in Section 4.

Clearly the predicted scores and SPE from Eqs. (4)(5), based on current available measurements, are not identical to the values that are calculated should the entire batch be available. As a result the predicted scores and SPE may not conform to the *pdf* developed based on the entire duration of nominal batches, even if the process being monitored is running normally. This is a serious issue particularly in the initial stage of a batch processing, when only a small number of measurements are available to calculate the scores and SPE. We follow the standard approach in on-line batch process monitoring (Nomikos and MacGregor, 1995b) to pass each of the nominal batches through the monitoring procedure to collect the predicted scores and SPE at each time step from Eqs. (4)(5), and then apply GMM to estimate the joint *pdf* of these predicted scores and log-SPE at each time step, and to establish the confidence bounds as presented in Section 2. Essentially we propose to replace the confidence bounds for $T^2$ and SPE in (Nomikos and MacGregor, 1995b), where the process data is assumed to be Gaussian distributed, with more powerful Gaussian mixture model. For on-line monitoring of a new batch, the scores and SPE are calculated from Eqs. (4)(5), and the likelihood value is calculated under the GMM for the current time step. If this likelihood value is lower than the confidence bound, the process under monitoring is considered to be in a faulty condition.

Table 1. Variables used for the monitoring of the semiconductor process.

| | | | |
|---|---|---|---|
| 1 | Endpoint A detector | 7 | RF impedance |
| 2 | Chamber pressure | 8 | TCP tuner |
| 3 | RF tuner | 9 | TCP phase error |
| 4 | RF load | 10 | TCP reflected power |
| 5 | RF Phase error | 11 | TCP Load |
| 6 | RF power | 12 | Vat valve |

## 4. CASE STUDY

The manufacture of semiconductors is introduced as an example of the on-line monitoring of batch processes. This study focuses specifically on an Al-stack etch process performed on the commercially available Lam 9600 plasma etch tool (Wise et al., 1999). Data from 12 process sensors, listed in Table 1, was collected during the wafer processing stage which run for 80 s. A sampling interval of 1 s was used in the analysis. Thus for each batch, the data is of the order $(12 \times 80)$. A series of three experiments, resulting in three distinct data groups, were performed where faults were intentionally introduced by changing specific manipulated variables (TCP power, RF power, pressure, plasma flow rate and Helium chunk pressure). There are 107 normal operating batches and 20 faulty batches. Twenty batches were randomly selected from the normal batches to investigate the effect of false alarms. The remaining 87 nominal batches were used to build the MPCA and GMM models.

### *4.1 Off-line analysis*

According to MPCA, the three-way nominal data array $(N \times J \times K = 87 \times 12 \times 80)$ is unfolded into a large two-way matrix of the order $(87 \times 960)$, which is then mean-centered and scaled to unit standard deviation on each column. Then PCA is applied to the pre-processed data, where two principal components are retained according to the broken-stick rule (Jolliffe, 2002). Considering there are 960 columns in the unfolded matrix, it is not surprising to find that two principal components explain only 45.50% of the total variance (similar results can be found in the literature, e.g. (Nomikos and MacGregor, 1995b)).

Figure 1 gives the scatter plot of the PCA scores corresponding to the first two principal components. It is clear that the nominal data exhibits the characteristic of multiple groups, and it cannot be adequately approximated by a single multivariate Gaussian distribution. As a result, the 99% confidence bound does not capture the region of NOC accurately. In addition to the normal testing batches, 17 out of 20 faulty batches are within the confidence bound, resulting in 17 missing errors. Clearly more complex models are required to represent the nominal behavior of the process.

To develop a GMM for the PCA scores and log-SPE, the appropriate number of mixtures must be selected. According to the BIC, the GMM with three mixture components is utilized for the analysis of the semiconductor process. Once the GMM is developed, the 95% and 99% confidence bounds is calculated using Monte Carlo simulation presented in Section 2.2, where the number of random samples is heuristically determined to be 10,000. Despite the large sample size, the CPU time for the Monte Carlo simulation was only 0.03 s (Matlab implementation under Windows XP with Pentium 2.8 GHz CPU). In the literature the 95% is treated as "warning bound" and 99% "action bound". Throughout this section the process is classified as faulty if the 99% confidence bound is violated.

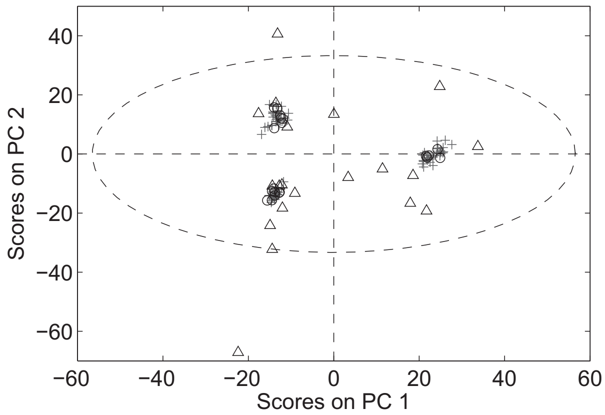


Fig. 1. Bivariate scores plot for principal components 1 and 2 with 99% confidence bound (− − − −): nominal (+), normal (∘) and faulty (△).

Table 2. Off-line monitoring results.

| | $T^2$ | SPE | $T^2$ + SPE | GMM |
|---|---|---|---|---|
| False alarms | 0 | 0 | 0 | 0 |
| Missing errors | 17 | 7 | 7 | 4 |

Table 2 summarizes the off-line batch-wise monitoring results for both conventional PCA and the GMM approach. Both methods incur no false alarms in this example. The large number of missing errors from $T^2$, as depicted in Figure 1, is the result of over-estimation of the confidence bound. It appears that SPE is more sensitive to the fault and it attains seven missing errors. By combining $T^2$ and SPE in the way that the process is identified as faulty if either metric is exceeded, the number of missing errors is still seven. Table 2 clearly indicates that GMM outperforms the conventional PCA in terms of smaller number of missing errors through the direct estimation of the joint *pdf* of the PCA scores and log-SPE.

### *4.2 On-line monitoring*

The on-line monitoring results are given in Table 3. A normal testing batch is considered to be a false alarm if it is identified as faulty within the batch duration. A missing error means a faulty batch is not detected during the entire duration. Similar to the off-line monitoring, $T^2$ fails to detect most of the faulty batches because the scores do not conform to a multivariate Gaussian distribution. A comparison between Table 3 (a) and (b) suggests that the instantaneous SPE can detect more faulty batches than the smoothed SPE; however the increased sensitivity is at the cost of dramatically decreased robustness. The number of false alarms for instantaneous SPE is excessively large (13 out of total 20 batches), and thus the smoothed SPE is adopted for the rest of this paper. Table 3 indicates that the GMM approach gives better results than the conventional MPCA in terms of smaller number of false alarms and missing errors.

Table 3. On-line monitoring results. (a) SPE is calculated based on process measurements at current time step (instantaneous SPE); (b) SPE is calculated based on process measurements from batch beginning to current time step (smoothed SPE).

(a)

| | $T^2$ | SPE | $T^2$ + SPE | GMM |
|---|---|---|---|---|
| False alarms | 0 | 13 | 13 | 8 |
| Missing errors | 17 | 2 | 2 | 0 |

(b)

| | $T^2$ | SPE | $T^2$ + SPE | GMM |
|---|---|---|---|---|
| False alarms | 0 | 3 | 3 | 1 |
| Missing errors | 17 | 4 | 4 | 2 |

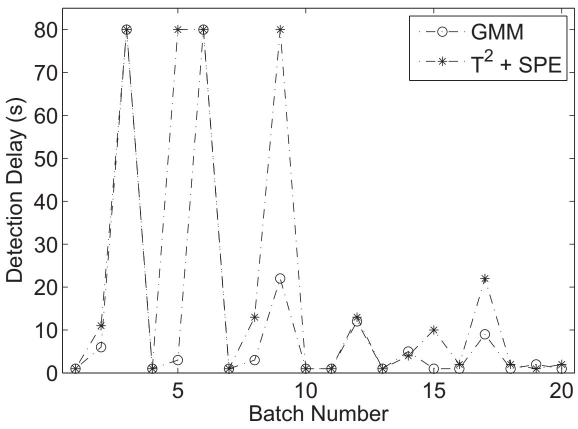


Fig. 2. Delay in the detection of the faulty batches.

It should be noted that the number of missing errors in on-line monitoring is not the only index to evaluate the monitoring performance. Of greater practical importance is the time delay between the occurrence and the detection of the fault. Figure 2 illustrates the detection delay of the 20 faulty batches using MPCA and GMM. To facilitate the calculation of average delay for comparison, the detection delay is artificially set to the batch duration (i.e. 80 s) if a faulty batch is not detected by the monitoring system. Essentially this is to assume that the abnormal behavior will be identified in some way (e.g. the presence of off-specified product) when the batch finishes. In practice plant operators are often not able to identify the fault until much later than the end of batch duration. On average, the detection delay for GMM is 11.6 s that is significantly shorter than 20.3 s obtained by the PCA method. Since the process is operating relatively fast, the reduction of delay in 9 s (equivalently 9 time steps) may not be sufficient for the operators to take appropriate actions in practice. Nevertheless if the proposed approach is applied to monitor a slow process, for example batch fermentation that takes several days to complete where data is sampled every half day, a shorter detection delay of 9 time steps would provide significant advantage in terms of reduced operational cost and improved process safety and product quality.

Figure 3 illustrates the on-line monitoring charts of a normal batch, which is false-alarmed by conventional PCA. Since the value of on-line SPE increases with time, we

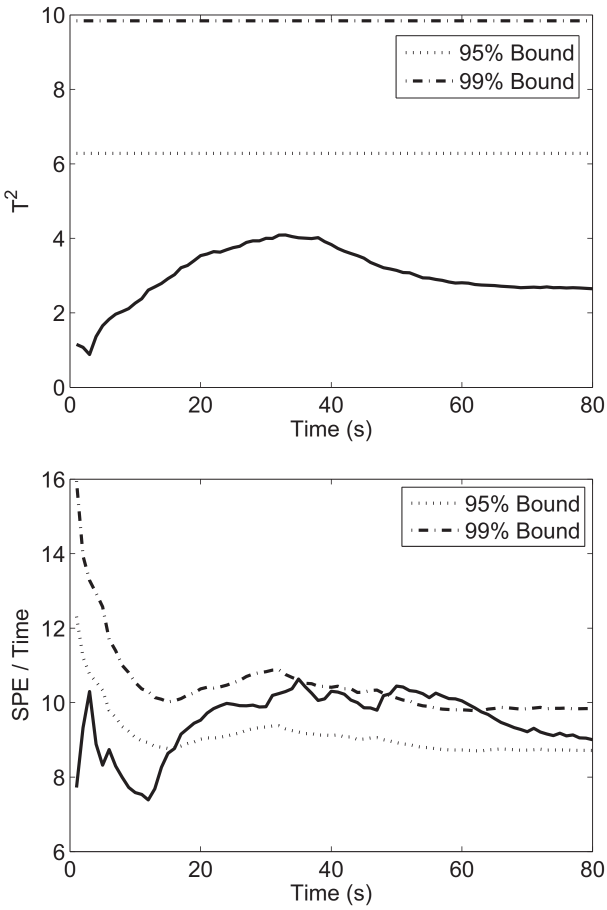


Fig. 3. On-line monitoring of a normal batch using $T^2$ and SPE.

plot SPE divided by time for better illustration in the figure. The $T^2$ indicates that this batch is under normal operation; however $T^2$ is not a reliable index for the monitoring of this process as discussed previously. The SPE metric appears to be susceptible to process disturbance; it exceeds the 95% confidence bound from 17 s and is over the 99% bound between 50 s to 60 s, despite the fact that the process is running normally. Figure 4 shows the GMM based monitoring chart, where the negative likelihood value is plotted. The GMM approach correctly recognizes that this batch is within the region of NOC during the whole batch duration.

Figure 5 and 6 give the on-line monitoring charts of a faulty batch (batch 5 as in Figure 2), using conventional PCA and the GMM approach, respectively. Both $T^2$ and SPE fails to detect this fault. In contrast, the likelihood value from the GMM is becoming outside the 99% confidence bound since time 3 s.

## 5. CONCLUSIONS

This paper extends the GMM technique for the modelling and on-line performance monitoring of batch manufacturing processes. The handling of the unobserved future batch measurements is discussed for the purpose of on-line monitoring. The GMM provides a probabilistic approach to estimating the *pdf* of the nominal process data and therefore enables more accurate calculation of the

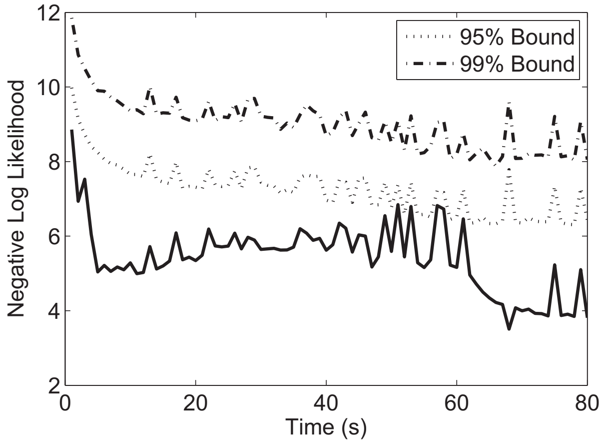


Fig. 4. On-line monitoring of a normal batch using GMM.

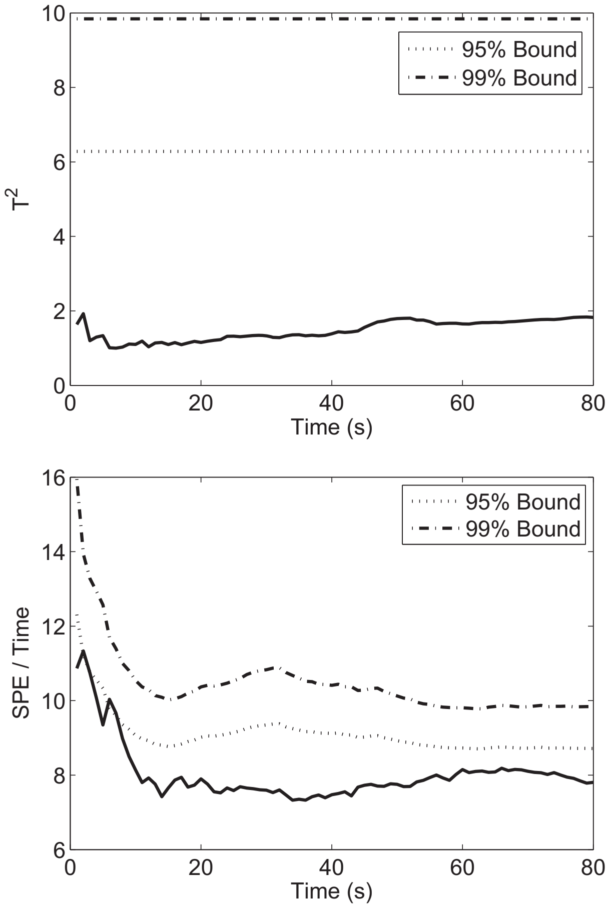


Fig. 5. On-line monitoring of a faulty batch using $T^2$ and SPE.

confidence bounds. The case study confirms that through accurate modelling of the process historical data collected from NOC, GMM is a promising approach to maintaining a low rate of both false alarms and missing errors in process performance monitoring.

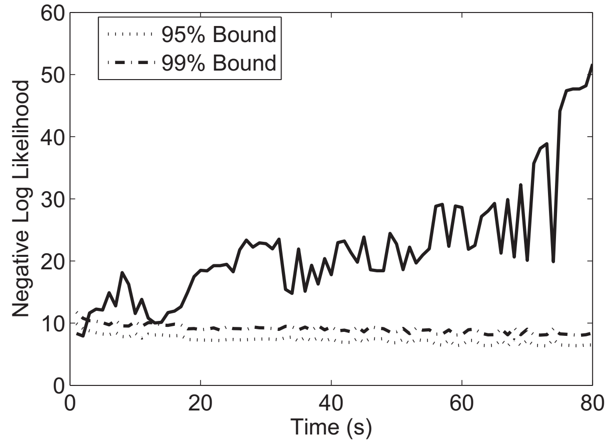


Fig. 6. On-line monitoring of a faulty batch using GMM.

## REFERENCES


Chen, T., Morris, J., and Martin, E. (2006). Probability density estimation via an infinite Gaussian mixture model: application to statistical process monitoring. *Journal of the Royal Statistical Society C (Applied Statistics)*, 55, 699–715.

Choi, S.W., Park, J.H., and Lee, I.B. (2004). Process monitoring using a Gaussian mixture model via principal component analysis and discriminant analysis. *Computers and Chemical Engineering*, 28, 1377–1387.

Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of Royal Statistical Society B*, 39, 1–38.

Jolliffe, I.T. (2002). *Principal Component Analysis*. Springer, 2nd edition.

Lu, N., Yao, Y., and Gao, F. (2005). Two-dimensional dynamic pca for batch process monitoring. *AIChE Journal*, 51, 3300–3304.

Martin, E.B., Morris, A.J., and Kiparrisides, C. (1999). Manufacturing performance enhancement through multivariate statistical process control. *Annual Reviews in Control*, 23, 35–44.

Nomikos, P. and MacGregor, J.F. (1995a). Multi-way partial least squares in monitoring batch processes. *Chemometrics and Intelligent Laboratory Systems*, 30, 97–108.

Nomikos, P. and MacGregor, J.F. (1995b). Multivariate SPC charts for monitoring batch processes. *Technometrics*, 37, 41–59.

Qin, S.J. (2003). Statistical process monitoring: basics and beyond. *Journal of Chemometrics*, 17, 480–502.

Rannar, H., MacGregor, J.F., and Wold, S. (1998). Adaptive batch monitoring using hierarchical PCA. *Chemometrics and Intelligent Laboratory Systems*, 41, 73–81.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461–464.

Thissen, U., Swierenga, H., de Weijer, A.P., Wehrens, R., Melssen, W.J., and Buydens, L.M.C. (2005). Multivariate statistical process control using mixture modelling. *Journal of Chemometrics*, 19, 23–31.

Wise, B.M., Gallagher, N.B., Butler, S.W., White, D.D., and Barna, G.G. (1999). A comparison of principal component analysis, multiway principal component analysis, trilinear decomposition and parallel factor analysis for fault detection in a semiconductor etch process. *Journal of Chemometrics*, 13, 379–396.