

Multivariate statistical process control using mixture modelling

U. Thissen¹, H. Swierenga², A. P. de Weijer², R. Wehrens¹, W. J. Melssen¹ and L. M. C. Buydens^{1*}

¹Analytical Chemistry, Institute for Molecules and Materials (IMM), Radboud University Nijmegen, Toernooiveld 1, NL-6525 ED Nijmegen, The Netherlands

²Teijin Twaron Research Institute, Postbus 9600, NL-6800 TC Arnhem, The Netherlands

Received 17 December 2004; Revised 26 January 2005; Accepted 28 February 2005

When performing process monitoring, the classical approach of multivariate statistical process control (MSPC) explicitly assumes the normal operating conditions (NOC) to be distributed normally. If this assumption is not met, usually severe out-of-control situations are missed or in-control situations can falsely be seen as out-of-control. Combining mixture modelling with MSPC (MM-MSPC) leads to an approach in which non-normally distributed NOC regions can be described accurately. Using the expectation maximization (EM) algorithm, a mixture of Gaussian functions can be defined that, together, describe the data well. Using the Bayesian information criterion (BIC), the optimal set of Gaussians and their specific parametrization can be determined easily. Artificial and industrial data sets have been used to test the performance of the combined MM-MSPC approach. From these applications it has been shown that MM-MSPC is very promising: (1) a better description of the process data is given compared with standard MSPC and (2) the clusters found can be used for a more detailed process analysis and interpretation. Copyright © 2005 John Wiley & Sons, Ltd.

KEYWORDS: model-based clustering; process monitoring; kernel density estimation (KDE)

1. INTRODUCTION

Both univariate statistical process control (SPC) and multivariate SPC (MSPC) charts have been used successfully in many chemical applications [1–5]. These charts are well-known tools for monitoring process behaviour, e.g. the univariate Shewhart chart or the multivariate principal component analysis (PCA)-based combination of Hotelling's T^2 and the Q -plot. The aim of these plots is to detect whether a process deviates significantly from a predefined level, the so-called normal operating conditions (NOC). For this purpose, parametric hypothesis tests are used, explicitly assuming a normal distribution.

However, if (M)SPC is used in cases where this assumption is incorrect, inaccurate evaluations of the process quality are made. This leads to situations where significant deviations from the NOC can be missed while proper process behaviour can be seen as out-of-control. For instance, a non-normally distributed NOC can occur for non-linear, dynamic or multistate processes. For these cases, other (non-linear) methods than PCA might be used for monitoring. Alternatively, a separate MSPC procedure could be set up for each

working point. In the past, non-parametric methods have been proposed for describing the NOC more accurately in cases where the normality assumption is not valid. Several of these approaches have been based on the (principles of the) Parzen window method [6,7], which is a kernel density estimation (KDE) method. With KDE, each training point is designated as a unit centre, and an identical basis function (the kernel, usually a Gaussian function) is constructed at each centre. A data set is then described by the total density of each data point for the set of kernels. A certain limit density describes the total shape of the data. For example, Doymaz *et al.* [8] use the KDE approach for defining the NOC, while Martin and Morris [9] modified this approach in order to select a non-parametric kernel based on bootstrapping. A possible drawback of the KDE approach is that differences in local densities cannot be modelled well owing to the identical spherical kernels used. Furthermore, the width of the kernel has to be determined using a method such as cross-validation, which can be computationally intensive. However, other studies describe approaches that can use basis functions with different covariance matrices to describe the local character of the data. For instance, Johnston and Kramer [10] use variable elliptical basis functions for process monitoring (fault detection) and fault identification. Their method requires two parameters to be optimized (e.g. with cross-validation): the number of basis functions and the overlap parameter. In Reference [11] also, fault

*Correspondence to: L. M. C. Buydens, Analytical Chemistry, Institute for Molecules and Materials (IMM), Radboud University Nijmegen, Toernooiveld 1, NL-6525 ED Nijmegen, The Netherlands. E-mail: l.buydens@science.ru.nl
Contract/grant sponsor: Dutch Technology Foundation STW; Contract/grant number: NCH4501.

detection and identification are performed on the basis of a separate PCA model for each cluster. This means that for each cluster the T^2 and Q are determined. The clusters have been defined on the basis of a two-dimensional Gaussian smoothing transformation performed on the first two principal components, using a heuristic rule to find the number of clusters. Finally, Choi *et al.* [12] use a Gaussian mixture modelling approach that is trained with the expectation maximization (EM) algorithm. In this way the NOC data can be described by a (limited) set of Gaussians that can each be parametrized differently [13,14]. Based on the cluster membership of an observed data point for each Gaussian, the corresponding local PCA model is selected to perform fault detection. However, treating the NOC as a set of independent clusters can be a drawback of these approaches. The possibility occurs that non-normally distributed process states can be divided into separate clusters and treated differently. Furthermore, and especially if no detailed cluster or process information is present, the main aim is to find those observations that are different from the total NOC space regardless of individual underlying Gaussian functions.

In this paper the standard MSPC approach is used in combination with the EM-based mixture modelling (MM-MSPC). However, the approach followed is fundamentally different. First the data are subjected to PCA, after which two monitoring parameters are derived: (1) a monitoring parameter representing significant and systematic variation of the process (T^2 for MSPC and an overall density parameter for MM-MSPC) and (2) the Q -statistic representing residual variation, which can be used to monitor possible changes in the correlation structure between parameters. In MM-MSPC an observation will be seen as out-of-control if the density derived from the total set of Gaussians is smaller than some threshold density or if the Q -statistic exceeds its limit. The number of Gaussians used to describe the data is determined with the Bayesian information criterion (BIC). In comparison with the mixture modelling approaches from the literature as discussed above, the local monitoring parameters are replaced by global ones and PCA is performed once: before describing the in-control data by the mixture of Gaussians. Additionally, it will be shown that (sets of) cluster(s) can be used for more detailed process analysis. This proposed use of mixture modelling for process monitoring and process analysis is demonstrated on both simulated problem cases and a real-world industrial dry-jet wet spinning process of Twaron fibres.

2. MULTIVARIATE STATISTICAL PROCESS CONTROL CHARTS

When many process parameters have to be monitored simultaneously, the use of univariate SPC has some drawbacks (see e.g. References [1–4]). Using MSPC, multiple parameters are taken into account simultaneously by aggregating them into two statistics using a PCA model: the Hotelling's T^2 -statistic containing all the significant and systematic information of the original variables [15], and the Q -statistic (also called the squared prediction error) containing orthogonal (residual) variance to the T^2 -statistic.

The moments and critical limits of these statistics are based on the NOC and are used to describe proper process functioning. Obviously, this description is crucial for any process and therefore it is a critical condition for MSPC to work accurately. Furthermore, the derivations of the T^2 -statistic and the Q -statistic depend strongly on the assumption of the NOC data being distributed normally, which is not necessarily the case.

3. MIXTURE MODELLING

The theory of mixture modelling, or model-based clustering, has been described extensively in References [13,14] and therefore will only be discussed briefly in this section. When using mixture modelling, the data are described by a mixture of probability distributions (clusters). Usually Gaussians are used (Figure 1). The division of the data into clusters might then be used for classification purposes, where a specific class can be described by multiple Gaussians (see e.g. Reference [16]). A class is a group of data that is labelled on the basis of prior knowledge (e.g. known process features), while a cluster groups data on the basis of only numerical similarities.

The first step in the clustering procedure is to estimate two types of parameters from the data: (1) the moments of the Gaussians (μ_c and Σ_c , the mean and the covariance matrix of Gaussian c respectively) and (2) the mixture parameters (τ_c , the proportion of Gaussian c in the mixture, i.e. the fraction of objects in cluster c). For one data point, x_i , the density from Gaussian c can be calculated as follows (in which k denotes the dimension of the data):

$$\phi_{i,c}(x_i | \mu_c, \Sigma_c) = \frac{\exp\left[-\frac{1}{2}(x_i - \mu_c)^T \Sigma_c^{-1} (x_i - \mu_c)\right]}{(2\pi)^{k/2} |\Sigma_c|^{1/2}} \quad (1)$$

The total density for that specific data point is a weighted sum of the densities of G individual Gaussians (the mixture):

$$\phi_{i,\text{tot}} = \sum_{c=1}^G \tau_c \phi_{i,c} \quad (2)$$

Later in this paper, $\phi_{i,\text{tot}}$ will be referred to as DMM, the density estimated from mixture modelling. The likelihood

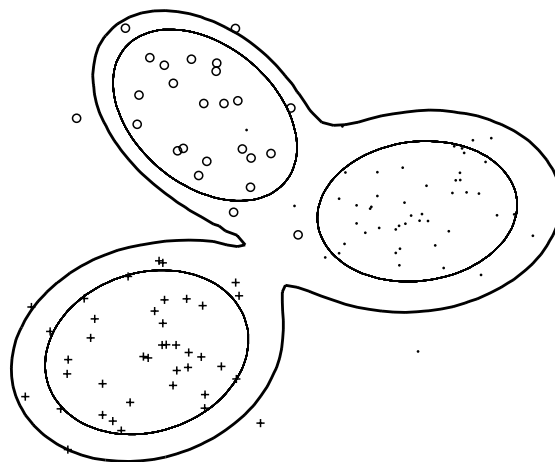


Figure 1. For a two-dimensional data set the density contours of three Gaussians (thin lines) and a mixture of the three (thick line) are shown.

over all n objects as a function of the Gaussian moments and the mixture parameters can be used to show how well the total data are described by the mixture:

$$L(\tau, \mu, \Sigma | x) = \prod_{i=1}^n \phi_{i,\text{tot}} \quad (3)$$

The optimal parameter values correspond to the ones that maximize the likelihood. However, when calculating the likelihood, it is necessary that the class label of each data point is known. In other words, μ_c , Σ_c and τ_c have to be estimated from the data underlying Gaussian c (i.e. cluster c). If the sample labels are unknown, these labels also have to be estimated (i.e. the data are incomplete). In order to find the class labels, the best Gaussian parameters and the mixture parameters, the expectation maximization (EM) algorithm is used [13,14].

The first step in this algorithm is to give a first estimation of the class label for each object using hierarchical clustering. Next the EM algorithm iterates between an M-step, which optimizes the maximum likelihood parameters given an estimate of class membership, and an E-step, which adapts the estimate of class membership given the estimates from the M-step:

$$z_{ic} = \frac{\hat{\tau}_i \phi_c(x_i | \mu_c, \Sigma_c)}{\sum_{j=1}^G \hat{\tau}_j \phi_j(x_i | \mu_j, \Sigma_j)} \quad (4)$$

The parameter z_{ic} can be seen as the conditional probability of object i belonging to class c . The iteration between the E- and M-steps is done until convergence. Note that by definition the sum of all conditional class probabilities equals *one*. After convergence, class labels can be derived, if necessary, from the z_{ic} values: objects are assigned to the class having the highest z_{ic} value. The uncertainty of designating a certain data point i to a cluster is assessed by $1 - \max_c(z_{ic})$.

To limit the number of parameters that need to be estimated, it is possible to apply restrictions to the shapes, volumes and sizes of the individual Gaussians. Several examples are shown in Table I. A selection between different distribution types can be made on the basis of prior knowledge of the data. On the one hand, if it is known that the data consist of several groups that have the same spherical shape and volume, the extremely restricted EII distribution can be used. On the other hand, if there is no prior knowledge or if the data are known to consist of diverse groups, ellipsoids could be used with different volumes, shapes and orientations (i.e. completely unrestricted, VVV). Obviously, other situations can occur that are in between the EII and the VVV situations. For this reason, these extreme two cases are used

in the following example. For the unrestricted case the parameters that need to be estimated are (1) the mixing proportion of each Gaussian ($G - 1$ parameters), (2) the k -dimensional mean of each of the G clusters (kG parameters) and (3) the parameters of the (symmetrical) covariance matrix for each cluster ($\frac{1}{2}G(k^2 + k)$ parameters). For the most restricted case (EII), only the mixture proportions, means and one value for the diagonal have to be estimated: $G - 1 + kG + 1 = G(k + 1)$. To illustrate, if EII is used with $G = 5$ clusters and $k = 12$, then 64 parameters have to be estimated. However, if VVV is used on the same data set with the same number of clusters, then 454 parameters have to be assessed.

The user has to select the number of clusters to consider, as well as the parametrization types. To pick the best combination, several criteria can be used. An often used criterion is the Bayesian information criterion (BIC) [20]:

$$\text{BIC} = 2 \ln L - n_p \ln n \quad (5)$$

In this equation, L is the likelihood, n is the number of objects and n_p is the total number of parameters to be estimated. The BIC shows a trade-off between a good fit quality and the number of parameters required to obtain the fit. A high BIC value is obtained when the data are described accurately (a high value of L) with only a few parameters. A standard strategy to obtain the best result is the following. Based on prior knowledge, for all suitable parametrizations, the data are described with a range of clusters. The best combination of parametrization and cluster number is the solution with the maximal BIC value. If more than one combination shows a maximum, the simplest solution is selected (i.e. the most restricted parametrization with as few clusters as possible).

4. SOFTWARE

The MCLUST package is used to perform mixture modelling [21]. The original software has been written for S-plus (<http://www.insightful/splus>) and can be obtained from the worldwide web (<http://www.stat.washington.edu/mclust>). In this paper, version 1.7.0 (April 2003) for R (<http://www.r-project.org>) was used.

5. COMBINING MIXTURE MODELLING WITH MSPC FOR PROCESS MONITORING

As discussed above, describing the NOC data is a crucial first step for process monitoring. For instance, a problematic situation occurs if the method describing the NOC incorrectly assumes it to be distributed normally. In this case the

Table I. Characteristics of some parametrizations of probability distributions used. Additionally, the total number of parameters (n_p) to be estimated is shown with $\alpha = G(k + 1) - 1$ and $\beta = \frac{1}{2}(k^2 + k)$, where k is the dimension of the data and G is the number of clusters. This table is based on References [17–19]

	Distribution	Cluster volume	Cluster shape	Cluster orientation	Number of parameters (n_p)
EII	Spherical	Equal	Equal	—	$\alpha + 1$
VII	Spherical	Variable	Equal	—	$\alpha + k$
EEE	Ellipsoidal	Equal	Equal	Equal	$\alpha + \beta$
VVV	Ellipsoidal	Variable	Variable	Variable	$\alpha + G\beta$
EEV	Ellipsoidal	Equal	Equal	Variable	$\alpha + G\beta - (G - 1)k$
VEV	Ellipsoidal	Variable	Equal	Variable	$\alpha + G\beta - (G - 1)(k - 1)$

Table II. Characteristics of the two-dimensional normally distributed clusters in the two artificial data sets. Both clusters 1 are spherical (equal variance in both directions and a covariance of zero) whereas both clusters 2 have different variances in both directions. Cluster 2 of data set 2 is oriented in the x_2 -direction with a covariance of zero (see Figure 2)

	Cluster	Mean	Variances	Covariance	Samples
Data set 1	1	(-3, -3)	0.25	0	1000
	2	(0,0)	4; 5	4	500
Data set 2	1	(-2, -2)	1	0	1000
	2	(2,0)	1; 9	0	1000

NOC would be represented incorrectly. The description of the NOC is of crucial importance to evaluate if future measurements are a part of the NOC and, consequently, if these represent good performance. Therefore this section compares the approaches and the performances of standard MSPC and mixture modelling for describing the NOC. Next the approach is explained to perform process monitoring when combining mixture modelling and MSPC.

First, as an illustration, two artificial data sets are used (representing the NOC) consisting of objects drawn from a two-dimensional mixture of two Gaussians (Table II). It should be stressed that even though the separate clusters in both data sets are normally distributed, the two data sets as a whole are not. Obviously this is an extreme situation where usually no standard MSPC is performed. However, this situation can be used well to illustrate the differences between the MSPC approach for NOC description and the mixture modelling approach.

Comparing the mixture modelling approach and the standard MSPC approach, relevant differences can be observed between the way the NOC is described (Figure 2). For mixture modelling, using the BIC values, the proper model type and number of clusters can be identified: two clusters

differing in volume, shape and orientation (VVV). The resulting mixture density contour of the data, representing the 95-percentile, agrees well with the contour of the data points. Using the standard MSPC approach, the data cannot be described accurately using the 95% confidence interval. These solutions found are affected by parts of data that have a higher density (data set 1) or by the different positions of subgroups (data set 2). For both data sets the MSPC-based 95% confidence interval describing the NOC includes regions for which no NOC data are present: the NOC is described too conservatively. When performing process monitoring, this obviously leads to many false negatives (falsely ignoring out-of-control situations).

Hence the previous example shows that, in principle, MSPC can benefit from a non-parametric NOC description. Therefore the only difference between MM-MSPC (i.e. the combination of mixture modelling and MSPC) and MSPC is the derivation of the parameter describing the systematic process variance, which is the T^2 -statistic for MSPC. For MM-MSPC this new parameter is based on Equation (2) and it is the weighted sum of the densities of the individual Gaussians in the mixture. For process monitoring, this parameter will be called DMM, the mixture modelling density. This parameter is calculated from the mixture model based on the first A significant principal components. In contrast to MSPC, the limit value for this parameter cannot be calculated from a specific distribution but has to be derived numerically. This is done by taking a certain percentile density value from the training set (usually the 95-percentile). Note that this estimate is robust against outliers especially if large quantities of data are available, which is often the case. Using DMM, an observation is seen to represent an out-of-control situation if its density is *smaller* than the limit density. In that case it is likely that the measurement does not belong to the total mixture of probability distributions. For both MSPC and MM-MSPC also the Q -statistic is calculated, which is the same for both methods. This can be seen

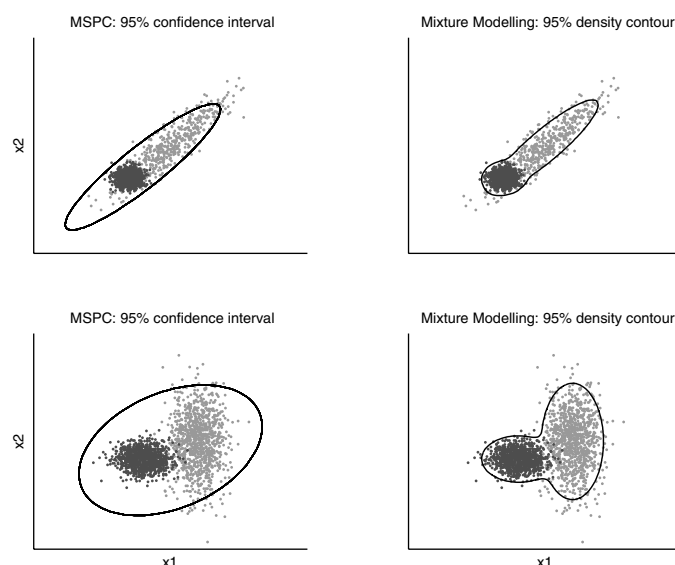


Figure 2. Performance of mixture modelling and the standard MSPC approach for describing non-normally distributed data. The upper and lower rows of subplots relate to data sets 1 and 2 respectively.

easily from the fact that the Q -statistic for both methods is based on the same set of principal components, representing the residual information (from principal component $A + 1$ until the final principal component).

6. INDUSTRIAL PROCESS DATA

Twaron fibres are used for life protection gear (bullet-proof vests and helmets), reinforcement of tyres, cut protective clothing, hoses, optical fibre cables and brake linings. These Twaron fibres are produced by a dry-jet wet spinning process [22,23]. This process starts with a mixture of poly(*p*-phenylene terephthalamide) (PPTA) and concentrated sulphuric acid (100%). The solvent (sulphuric acid) is solidified by cooling and mixed with PPTA powder to form a solid solution. In an extrusion-like process the material is intensively mixed and, via a filter, the spinning dope is extruded from spinneret holes into fine filaments. Coagulation of the filaments (precipitation) takes place in the spinning bath, which consists of diluted sulphuric acid. The sulphuric acid is removed from the filaments during coagulation and washing and finally the yarn is dried. The process results in high-tenacity and high-modulus yarns. As a final step, the yarn produced is wounded onto a bobbin. In this production process, 23 process parameters have been measured. These include various temperatures, revolutions of pumps and mixer, power consumptions and pressures at different stages in the process. For proprietary reasons the process parameters cannot be specified. The value of each process parameter is an average value at a certain time interval. This interval reflects the time required to produce a full bobbin of yarn. Furthermore, complementary to the process parameters, also a quality parameter was measured representing the counted number of yarn fluffs on a bobbin. When a fluff occurs, the yarn itself is not broken but some of the filaments are. A higher fluff count represents a lower yarn quality.

In order to test MSPC and MM-MSPC, the process parameters have been divided into four types of data sets: a training set, test set 1, test set 2 and a remaining data set. The division of the process parameters into one of these data sets has been done on the basis of the quality parameter. Figure 3 shows the distribution of the quality parameter which is used to divide the data into the separate sets. The resulting data sets are (1) a training set (385 objects \times 23 variables) representing high-quality product which is located completely left in the histogram, (2) test set 1 (42 objects \times 23 variables) which also represents high-quality product and is also located on the left of the histogram, (3) test set 2 (95 objects \times 23 variables) representing out-of-control samples which was selected on the basis of expert knowledge and is located on the right of the histogram, and (4) the so-called 'grey set' (428 objects \times 23 variables) which is an intermediate set between the NOC (training set and test set 1) and test set 2. The total quality of this set is unknown, but it contains both in-control and out-of-control measurements. Thus both the training set and test set 1 are sampled from the high-quality data (i.e. the NOC) with sizes of 90% and 10% respectively. The division of the NOC into the training set and test set 1 is performed randomly. In order to decrease the influence of this random selection on the analysis, the random selection and the corresponding analysis are performed 10 times. As a result, MSPC and MM-MSPC are performed 10 times to account for the different realizations of the training set and test set 1.

7. RESULTS AND DISCUSSION

This section shows the results when applying both MSPC and MM-MSPC for industrial process monitoring. The first step for both methods is to derive a PCA model on auto-scaled data. This means that the data have zero mean and unit standard deviation. It appears that during the 10-fold analysis (based on the 10 random representations of the

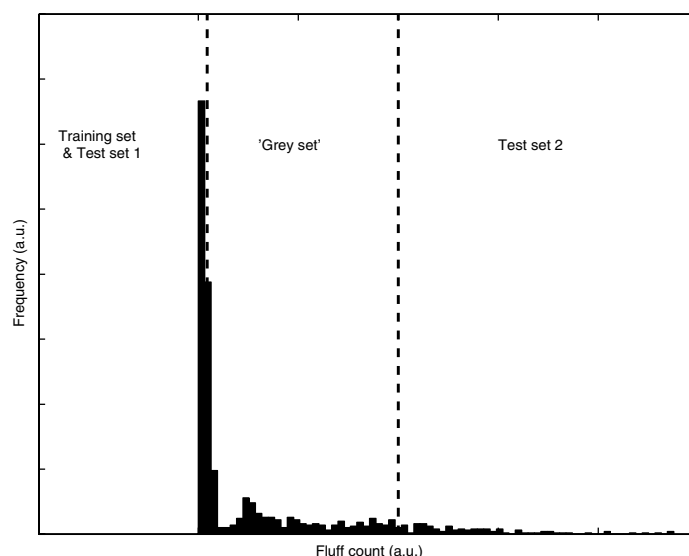


Figure 3. Histogram of the quality parameter. The NOC is represented by the training set and test set 1. Test set 2 is used to represent the worst 10% of the total data. The quality of the 'grey set' is undefined. Note that the lowest fluff count possible is zero. Transitions from one data group to another are indicated by broken vertical lines.

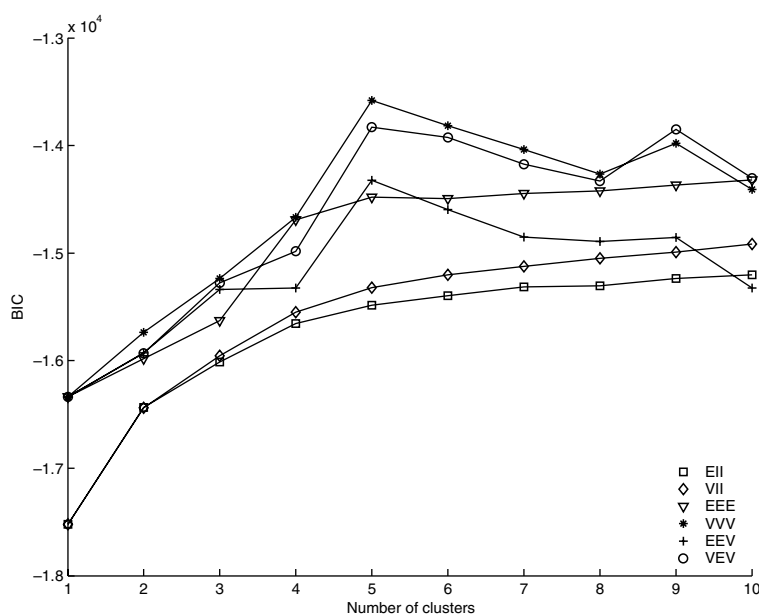


Figure 4. BIC values for all considered combinations of model type and number of clusters.

training set and test set 1) the results do not change or change only slightly for each representation. The number of significant PCs (12) each explaining at least 2% of variance, the model type (VVV) and the number of clusters (5) do not vary, while the explained variance of these 12 significant principal components varies between 91.9% and 92.5%. The data from the significant PCs are used for MSPC to calculate T^2 , while MM-MSPC performs the mixture modelling to calculate the DMM parameter. Figure 4 shows the distribution of the BIC values for all combinations of model types and clusters considered. The five-cluster solution for VVV clearly performs best. It is not surprising that model VVV is selected to be the best one: there are enough data points available to estimate the cluster parameters and it allows the most free cluster description, which can be seen from the fact that no restrictions are placed on the distribution, volume,

shape and orientation of the Gaussians (for more details see Table I). In practice, and especially for 'real-world' data, clusters can be very different from each other, therefore requiring a flexible description of the total data.

For one of the 10 random selections considered, Figure 5 shows how both MSPC and MM-MSPC describe the data using a 95% confidence interval and the 95% density contour respectively. In contrast to Figure 2, the quality of the data description cannot be seen fully from this figure, because it only shows two of the 12 significant PCs. In the dimensions shown, both methods include empty regions in their description, although the effect is perhaps exaggerated, since only two PCs are used for visualization. It can also be seen that the irregular shape of the two large groups with their tails is followed by the MM-MSPC approach. The MSPC approach is based on the total mean and the total covariance

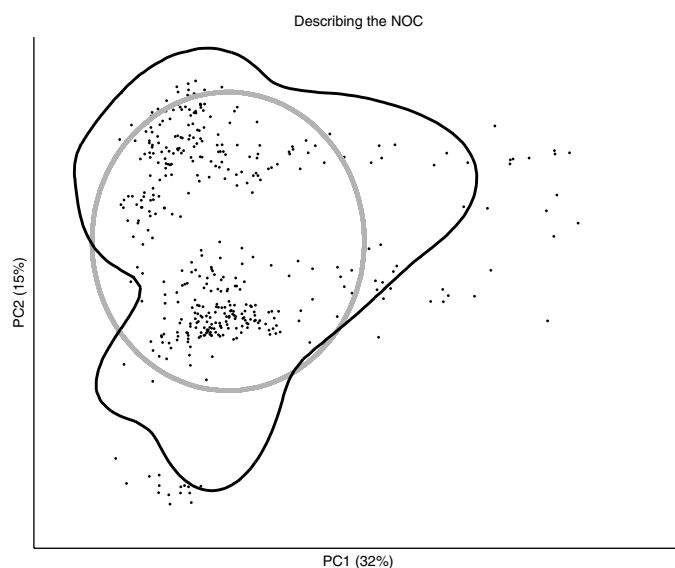


Figure 5. Describing the NOC on the basis of the training data. The MSPC description is shown in grey while the black contour shows the MM-MSPC result.

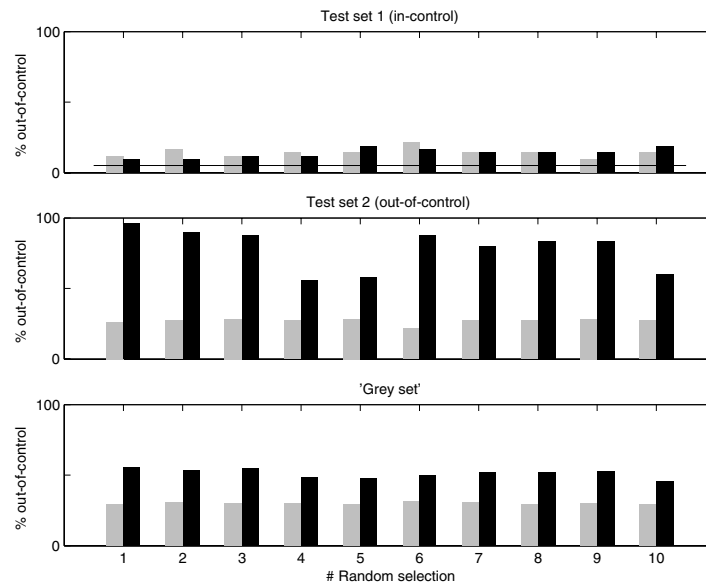


Figure 6. Percentages of out-of-control objects found on the basis of T^2 and Q for MSPC (grey) and DMM and Q for MM-MSPC (black). For test set 1 the ideal percentage (5%) is indicated with a horizontal line. For test set 2 the optimal result is 100%, while for the 'grey set' no such criterion can be defined.

matrix, while the shape of the data is not taken into account explicitly.

With the help of the calculated limits, the number of out-of-control situations was evaluated for test sets 1 and 2 and the grey set. Figure 6 shows the percentages of out-of-control objects found from the 10-fold analysis. The percentages presented are based on a combination of out-of-control points on (1) for MSPC either the T^2 -statistic or the Q -statistic or (2) for MM-MSPC either the DMM statistic or the Q -statistic. For all cases the Q -statistic is the same for both methods, and for test set 1 on average it contributes 2.9% (MSPC) and 1.4% (MM-MSPC). For test set 2, 7.7% and 1.1% are contributed for MSPC and MM-MSPC respectively. For the grey set the Q -statistic adds most for MSPC (21.8%), while for MM-MSPC the contribution is 4.9%.

From the figure it can be seen that MM-MSPC outperforms MSPC, because approximately the same number of out-of-control points is found for test set 1, while (many) more out-of-control points are found in test set 2. For test set 1 the mean error rates are $14.3\% \pm 3.2\%$ (MSPC) and $14.0\% \pm 3.4\%$ (MM-MSPC). For test set 2 the mean results are $27.7\% \pm 0.7\%$ (MSPC) and $78.0\% \pm 14.5\%$ (MM-MSPC). It appears that the individual MM-MSPC results are always better than the MSPC results but generally show a higher standard deviation. This is caused by MM-MSPC results ranging from 57.9% up to 95.8%, while the individual MSPC results range from 26.3% to 28.4%. These differences represent the differences in the training set and test set 1 due to the 10-fold random sampling. For MM-MSPC, different representations of the training set can result in the situation where different clusters are obtained owing to the absence or presence of a subgroup of extreme data points. Apparently, MSPC suffers less from this situation. This is probably caused by the fact that the MSPC-based NOC is described too conservatively, which means that the NOC borders are too loose. This can also be seen from the MSPC

results of test set 2: only an average of 27.7% of the out-of-control situations are correctly found.

Additionally, the number of training objects that are seen to be out-of-control on the basis of MSPC (with limits adapted according to Reference [24]) has been evaluated. These results are not shown, but it appears that MSPC finds more than the expected 5% out-of-control training values ($13.6\% \pm 0.5\%$). This is caused by the fact that in standard MSPC no accurate description of the NOC data is possible. Using MM-MSPC, this is not the case, because DMM_{lim} is set manually to the 95-percentile training value. This means that exactly 5% of the training points will be seen as out-of-control. However, in MSPC also a limit value can be set manually to the 95-percentile on the basis of the training set, but this will lead to a situation where even fewer true out-of-control points will be found.

When evaluating the grey data set for both MSPC and MM-MSPC, it can be seen that MM-MSPC finds $45.4\% \pm 4.5\%$ out-of-control values on the basis of DMM, while the Q -parameter contributes another 5.9%. On the basis of T^2 , MSPC only finds $8.2\% \pm 0.3\%$ out-of-control points, while the Q -parameter adds 21.8%. Note that for the grey data set no optimal result is known. Nevertheless, this set can be used to compare the two methods as described further. Figure 7 shows the results of evaluating the objects from the grey region by both MSPC and MM-MSPC for one of 10 examples. Objects found in-control in MSPC have a lower quality (more fluffs) than the out-of-control objects, which clearly is incorrect. MM-MSPC out-of-control objects have higher fluff counts than in-control objects, as one would expect. MM-MSPC in control objects on average have a lower number of fluffs than MSPC in-control objects; the reverse is true for the two classes of out-of-control objects. Although the samples for this grey area cannot be considered as simple in-control or out-of-control objects, these

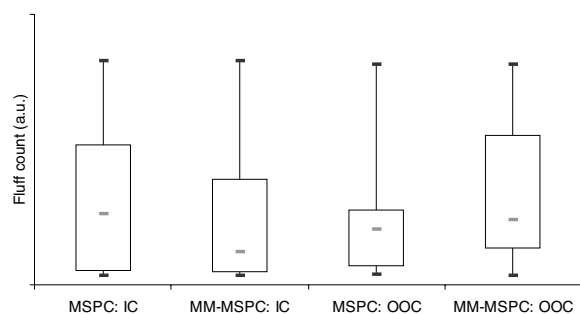


Figure 7. Box plot of the fluff counts for the out-of-control (OOC) and in-control (IC) situations found when evaluating the grey data set. The box plot shows the minimal and maximal values of each set together with the 25-, 50- and 75-percentiles.

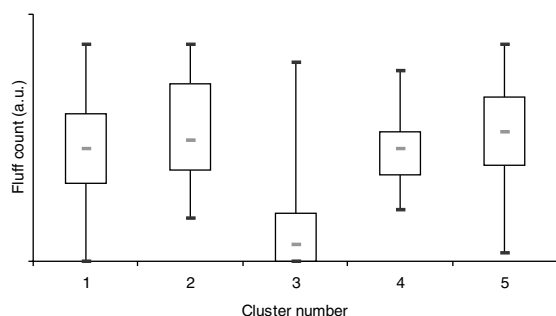


Figure 8. Boxplot showing the number of fluff counts per cluster for one random selection (of 10) of the training set.

results indicate that MM-MSPC performs much better than MSPC for this data set.

In addition to process monitoring, the evaluation of MM-MSPC clusters might also be useful for more detailed process analysis. Note, however, that a certain class (characterized by distinct process features) may be represented by more than one cluster. This means that a class which is not distributed normally can be described by two or more Gaussians (i.e. a sum of normally distributed clusters). Therefore sometimes several separate in-control process states can occur which can each be described by one or a combination of Gaussians.

Figure 8 shows the number of fluff counts per cluster after classifying one of the 10 examples of the training set. It can be seen that one of the clusters (number 3) contains the lowest fluff counts. When mapping test set 2 (all true out-of-control objects) into these five clusters, no test objects are mapped into cluster 3. On validation of these results by process experts, the most important deviating parameters for this cluster could, at least partly, be related to known relations between these parameters and high-quality fibres (low fluff counts). Therefore it follows that the NOC can consist of different working points in terms of process parameters while resulting in products meeting their quality requirements but showing some variance in quality. Thus it appears that the clusters can be divided into two groups: one group with relatively high fluff counts and one group with opposite characteristics. The differences found for the two groups can give indications in practice for further process improvement. Unfortunately, owing to proprietary reasons, the physical relations between these groups and low fluff counts cannot be given. Finally, when performing a cluster analysis on the

basis of the other randomly selected samples, similar observations can be made (not shown here).

If it is known that the data consist of different process states (operating points) and each state can be assigned to one cluster, standard MSPC could also be performed on the separate clusters. However, if many process states are present, MSPC has to be performed for each state separately. Unfortunately, this situation is one of the disadvantages of SPC compared with MSPC. Furthermore, in a convex NOC space as defined by MSPC, the shortest route between any two measurements is always located inside the NOC. A convex space is a space which contains all line segments between any pair of its points (e.g. an ellipse). A concave space is the opposite of a convex space. For concave spaces (as defined by MM-MSPC) this means that the shortest route from one in-control situation to another can possibly be seen as out-of-control. In our case this was a situation that was hardly encountered: the transitions between most in-control measurements were also in-control. Transitions from one cluster to another occurred inside the NOC region.

8. CONCLUSIONS

This paper shows the combination of mixture modelling, or model-based clustering, and MSPC as a method for process monitoring. This approach, MM-MSPC, enables the identification of different clusters and therefore allows non-normally distributed NOC data to be described as a mixture. Using two artificial data sets and an industrial data set, it could be concluded that both methods perform similarly when evaluating in-control measurements. However, on the one hand, MM-MSPC was much better able to correctly find the out-of-control situations. On the other hand, MM-MSPC is also more sensitive than MSPC to data points that are removed from areas already sampled with a low density. Nevertheless, almost all MM-MSPC results were better than the MSPC ones. When analysing a data set with unknown quality, the results of MM-MSPC also appeared to be much more likely than the MSPC results. In addition, indications have been given that groups of the clusters found can give indications for further process analysis. This has been based on the fact that different groups reflect significantly different characteristics. These results could be validated partly by expert knowledge.

Acknowledgements

Thanh N. Tran (Radboud University Nijmegen, The Netherlands), Stan Gielen (Dutch Foundation for Neural Networks, Nijmegen, The Netherlands), Sijmen de Jong (Unilever, Vlaardingen, The Netherlands) and Age Smilde (University of Amsterdam, The Netherlands) are thanked for constructive discussions on this work. This research was supported by the Dutch Technology Foundation STW (NCH4501).

REFERENCES

1. Kresta JV, MacGregor JF, Marlin TE. Multivariate statistical monitoring of process operating performance. *Can. J. Chem. Eng.* 1991; **69**: 35–47.

2. Kourti T, MacGregor JF. Process analysis, monitoring and diagnosis, using multivariate projection methods. *Chemometrics Intell. Lab. Syst.* 1995; **28**: 3–21.
3. Nomikos P, MacGregor JF. Multivariate SPC charts for monitoring batch processes. *Technometrics* 1995; **37**: 41–59.
4. Nijhuis A, de Jong S, Vandeginste BGM. Multivariate statistical process control in chromatography. *Chemometrics Intell. Lab. Syst.* 1997; **38**: 51–62.
5. Rius A, Ruisánchez I, Callao MP, Rius FX. Reliability of analytical systems: use of control charts, time series models and recurrent neural networks. *Chemometrics Intell. Lab. Syst.* 1998; **40**: 1–18.
6. Parzen E. On estimation of a probability density function and mode. *Ann. Math. Statist.* 1962; **33**: 1065–1076.
7. Webb A. *Statistical Pattern Recognition*. Wiley: Chichester, 2002.
8. Doymaz F, Chen J, Romagnoli JA, Palazoglu A. A robust strategy for real-time process monitoring. *J. Process Control* 2001; **11**: 343–359.
9. Martin EB, Morris AJ. Non-parametric confidence bounds for process performance monitoring charts. *J. Process Control* 1996; **6**: 349–358.
10. Johnston LPM, Kramer MA. Probability density estimation using elliptical basis functions. *AIChE J.* 1994; **40**: 1639–1649.
11. Chen J, Liu J. Mixture principal component analysis models for process monitoring. *Ind. Eng. Chem. Res.* 1999; **38**: 1478–1488.
12. Choi SW, Park JH, Lee I-B. Process monitoring using a Gaussian mixture model via principal component analysis and discriminant analysis. *Comput. Chem. Eng.* 2004; **28**: 1377–1387.
13. Fraley C, Raftery AE. Model-based clustering, discriminant analysis, and density estimation. *J. Am. Statist. Assoc.* 2002; **97**: 611–631.
14. McLachlan G, Peel D. *Finite Mixture Models*. Wiley: New York, 2000.
15. Jackson JE. *A User's Guide to Principal Components*. Wiley: New York, 1991.
16. Wehrens R, Simonetti AW, Buydens LMC. Mixture modelling of medical magnetic resonance data. *J. Chemometrics* 2002; **16**: 274–282.
17. Banfield JD, Raftery AE. Model-based Gaussian and non-Gaussian clustering. *Biometrics* 1993; **49**: 803–821.
18. Celeux G, Govaert G. Gaussian parsimonious clustering models. *Pattern Recogn.* 1995; **28**: 781–793.
19. Fraley C, Raftery AE. How many clusters? Which clustering method? Answers via model-based cluster analysis. *Comput. J.* 1998; **41**: 578–588.
20. Schwarz G. Estimating the dimension of a model. *Ann. Statist.* 1978; **6**: 461–464.
21. Fraley C, Raftery AE. MCLUST: software for model-based clustering, density estimation and discriminant analysis. *Tech. Rep. 415*, Department of Statistics, University of Washington, 2002.
22. Hearle JWS. *High Performance Fibers*. CRC Press: Boston, MA, 2001.
23. Picken SJ, Boerstel H, Northolt MG. Processing rigid polymers to high performance fibers. In *Encyclopedia of Materials: Science and Technology*, vol. 8, Buschow KHJ, Cach RW, Flemings MC, Ilschner B, Kramer EJ, Mahajan S (eds). Elsevier: Oxford, 2001; 7883–7887.
24. Tracy ND, Young JC, Mason RL. Multivariate control charts for individual observations. *J. Qual. Technol.* 1992; **24**: 88–95.