INTERNATIONAL JOURNAL OF CHEMICAL REACTOR ENGINEERING

Volume 2 2004 *Article* A14

A Hybrid Methodology For On-Line Process Monitoring

Rakesh Kumar* Avinash M. Jade†

Valadi K. Jayaraman‡ Bhaskar D. Kulkarni**

*Chemical Engineering Division, National Chemical Laboratory, Pune, India, rhgarg@yahoo.com

†Chemical Engineering Division. National Chemical Laboratory, Pune, India, amjade@rediffmail.com

‡Chemical Engineering Division. National Chemical Laboratory, Pune, India, jayaram@che.ncl.res.in

**Chemical Engineering Division. National Chemical Laboratory, Pune, India, bdk@ems.ncl.res.in

ISSN 1542-6580

A Hybrid Methodology For On-Line Process Monitoring

Rakesh Kumar, Avinash M. Jade, Valadi K. Jayaraman, and Bhaskar D. Kulkarni

Abstract

A hybrid strategy of using (i) locally linear embedding for nonlinear dimensionality reduction of high dimensional data and (ii) support vector machines for classification of the resultant features is proposed as a robust methodology for process monitoring. Illustrative examples substantiate the methodology vis-à-vis current practice.

KEYWORDS: Process monitoring, locally linear embedding, abnormality detection, support vector domain distribution, novelty detection, single class classification

1. INTRODUCTION

Last few years have experienced an explosive growth in the amount of data collected on different experimental systems due to availability of sophisticated instrumentation. New applications that require the storage and retrieval of huge amounts of data are emerging. They include examples such as protein matching in biomedical applications, fingerprint recognition, meteorological predictions, satellite image repositories, genomic data, text categorization etc. Most problems of interest in practice involve data with a large number of measurements (or dimensions). For example in many chemical process plants, sensors provide a large amount of measurements (features). This information overload can be a significant problem for plant operators responsible for insuring the safety and economic operation of the plant, particularly during abnormal situations resulting from disturbances, faults (sensor, equipment failure etc), human error, and/or unanticipated operating conditions. Thus abnormality detection in process plant constitutes a very vital aspect of safe and optimal operation of complex chemical plants. A number of methods have been proposed for batch process monitoring.

The early work, based on multiway principal component analysis (MPCA) (Wold et. al., 1987), was developed for batch process monitoring by Nomikos and MacGregor (1994). Multiway partial least square (MPLS) was then developed to correlate the process data and the product quality data (Nomikos and MacGregor, 1995). Various researchers have proposed several variants to the original methodology based on MPCA (e.g. Rannar et al., 1998). Louwerse and Smilde (2000) used PARAllel FACtor analysis (PARAFAC) and Tucker three-way models for monitoring batch processes. Boque and Smilde (1999) used multivariate statistical procedures based on multiway covariates regression models. Nonlinear principal component analysis (NLPCA) based on principal curves and neural networks produced independent principal components to unfold batch process data and get the nonlinear batch trajectory (Dong and McAvoy, 1996b). Martin and Morris (1996) and Martin et al. (1996) introduced a control chart based on a non-parametric method. Wise et al. (1999), Westerhuis et al. (1999) and Dahl, Piovoso, and Kosanovich (1999) applied and compared several alternatives for multivariate statistical analysis of batch process data. Combination of the orthonormal function approximation and the MPCA is proposed to analyze and monitor batch processes at the different operating time (Chen and Liu, 2000). Dynamic PCA and dynamic PLS models have also been used for on-line batch process monitoring (Chen and Liu, 2002). The performance of statistical process monitoring of batch processes can be enhanced by incorporating external information in model development (Ramaker et al., 2002). Some of these approaches have been evaluated by Van Sprang et al. (2002). Also there are number of approaches, which has been used for fault detection and diagnosis of the batch as well as continuous process using artificial neural networks (ANN) and combination of ANN with fuzzy and wavelets (Chen et. al., 1999; Dong and McAvoy, 1996a; Dong and McAvoy, 1996b; Fan et. al., 1993; Farell et. al., 1994; Hoskins et. al., 1991; Kavuri & Venkatasubramanian ,1993, 1994; Rengaswamy & Venkatasubramanian, 2000; Ruiz et. al., 2000,2001; Scenna, 2000; Ungar et. al., 1990; Venkatasubramanian & Chan, 1989; Venkatasubramanian et. al., 1990; Wang et. al., 1999; Watanabe et. al., 1989; Zhao et. al., 1997).

Many of the techniques elaborated above are based on the use of clustering and dimensionality reduction valid for linear structures (Devijver and Kittler, 1982; Duda et al., 2001; Jain et al., 2000; Mardia et al. 1979). Many real life data sets however contain essential nonlinear structures that are imperceptible to linear methods (Bailer-Jones et al., 1998; McClurkin et al., 1991; Murase and Nayar, 1995). A number of techniques to perform nonlinear mappings have been proposed in literature. They include: non-linear PCA (Malthouse, 1998), multi-dimensional scaling (MDS) (Borg and Groenen,1997), Sammon mapping (Sammon, 1969), singular value decomposition (SVD), self-organizing map (SOM) (Kohonen, 1995), generative topographic mapping (Bishop et al.,1998), principal curves and surfaces (Hastie and Stuetzle, 1989), auto-encoder neural networks (DeMers and Cottrell, 1993), mixtures of linear models (Tipping and Bishop, 1999) etc. All of these methods while extremely useful in general, have some or the other specific limitation. Thus for instance: there is no single and unique solution to nonlinear PCA while MDS and Sammon mapping give a point-to-point mapping but cannot provide the underlying mapping function. Consequently they cannot accommodate new data points (Sammon, 1969; Mao and Jain, 1995) and the entire procedure has to be repeated from start using all data points. Multi-dimensional scaling and neural networks are hard to train and time-consuming, as are principal curves and surfaces. The latter, as well as

the generative topographic mapping, need large data sets to estimate their many parameters. Mixtures of localized linear models require the user to set a number of parameters, which are highly specific to each data set and determine how well the model fits the data.

Recently, several entirely new approaches have been devised to address these problems. These methods combine the advantages of PCA and MDS viz. computational efficiency; few free parameters; non-iterative global optimization of a natural cost function—with the ability to recover the intrinsic geometric structure of a broad class of nonlinear data manifolds. These algorithms can be local or global. Local approaches such as locally linear embedding (LLE) (Roweis and Saul, 2000), Laplacian eigenmaps (Belkin and Niyogi, 2002) attempt to preserve the local geometry of the data; essentially, they seek to map nearby points on the manifold to nearby points in the low-dimensional representation. Global approaches such as Isomap (Tenenbaum et al., 2000) attempts to preserve geometry at all scales, mapping nearby points on the manifold to nearby points in low-dimensional space, and faraway points to faraway points. Thus isomap preserves the neighborhood of each object, as well as the 'geodesic' distances between all pairs of objects. The global approach tends to give a more faithful representation of the data's global structure, and its metric-preserving properties are better understood theoretically. The local approaches have two principal advantages: (1) computational efficiency: they involve only sparse matrix computations which may yield a polynomial speedup; (2) representational capacity: they may give useful results on a broader range of manifolds, whose local geometry is close to Euclidean, but whose global geometry may not be.

LLE recovers global nonlinear structure from locally linear fits. Unlike clustering methods for local dimensionality reduction, LLE maps its inputs into a single global coordinate system of lower dimensionality, and its optimizations does not involve local minima. LLE is based upon reconstruction of data, preserving local neighborhoods, and thus also the clusters which may be present in the database. Therefore clustering algorithms such as Support Vector Domain Distribution (SVDD) (Tax and Duin, 1999) should show superior clustering ability for LLE data representation than other representations like PCA. In the present work, we illustrate these advantages of LLE combined with SVDD to make the abnormality detection scheme more robust.

2. LLE ALGORITHM

The LLE algorithm is based on simple geometry. Consider the data set $\{\vec{X}\}_{i=1,2.....P} \in \Re^D$, sampled from some smooth underlying manifold. For a well sampled (i.e. there is enough data) manifold, we expect each data point and its neighbors to lie on or close to a locally linear patch of the manifold. We can thus approximate the non-linear manifold in the vicinity of $\vec{X}_i$ by a linear hyperplane passing through its nearest neighbors. In the simplest formulation of LLE, one identifies *N* nearest neighbors for every data point, as measured by Euclidean distance (Other notions of "closeness" are also possible, such as all points within a certain radius, or by using more sophisticated rules based on local metrics.) and then minimize the reconstruction error as measured by a cost function

$$\varepsilon(W) = \sum_i \left| \vec{X}_i - \sum_j W_{ij} \vec{X}_j \right|^2 \qquad (1)$$

subject to two constraints:

a) Each data point $\vec{X}_i$ is reconstructed only from its neighbors, enforcing $W_{ij} = 0$ if $\vec{X}_j$ does not belong to this set and

b) $\sum_j W_{ij} = 1$ for every i.

The weights W_{ij} signify the contribution of the j^{th} data point to the i^{th} reconstruction. The optimal weights W_{ij} subject to these constraints are found by solving a least squares problem, as discussed in Appendix.

The constrained weights that minimize these reconstruction errors characterize intrinsic geometric properties of each neighborhood, as opposed to properties that depend on a particular frame of reference. This is due to the fact that for any particular data point, the weights are invariant to rotations, rescalings, and translations of that data point and its neighbors. The invariance to rotations and rescalings results from the form of Equation (1); the invariance to translations is imposed by the sum-to-one constraint (b). Since the data lie on or near a smooth nonlinear manifold of dimensionality d<<D, there exists a linear mapping— comprising a translation, rotation, and rescaling—that maps the high dimensional coordinates of each neighborhood to global internal coordinates on the manifold. Thus reconstruction weights W_{ij}, invariant to such transformations, should characterize the local geometry, both in the original data space and local patches on the manifold. In particular, the same weights W_{ij} that reconstruct the i^{th} data point in D dimensions should also reconstruct its embedded manifold coordinates in d dimensions. Based on this idea each high dimensional observation $\vec{X}_i$ is mapped to a low dimensional vector $\vec{Y}_i$ representing global internal coordinates on the manifold. This is accomplished by choosing *d* dimensional coordinates $\vec{Y}_i$ to minimize the reconstruction errors as measured by embedding cost function:

$$\Phi(Y) = \sum_i \left| \vec{Y}_i - \sum_j W_{ij} \vec{Y}_j \right|^2 \tag{2}$$

The embedding cost defines a quadratic form in the vectors $\vec{Y}_i$.

$$\Phi(Y) = \sum_{ij} M_{ij} (\vec{Y}_i . \vec{Y}_j) \tag{3}$$

Here *M* is $P \times P$ matrix:

$$M_{ij} = \delta_{ij} - W_{ij} - W_{ji} + \sum_k W_{ki} W_{kj} \tag{4}$$

where δ_{ij} is 1 if $i = j$ and 0 otherwise.

To ensure the uniqueness of the solution the following two constraints are imposed: translation invariance by requiring the coordinates to be centered on the origin i.e. $\sum_i \vec{Y}_i = 0$ and we constrain the embedding vectors to have unit covariance,

$$\frac{1}{P} \sum_i \vec{Y}_i . \vec{Y}_i^T = \mathrm{I} \tag{5}$$

where I is the $d \times d$ identity matrix.

These constraints do not affect the generality of the solutions as $\Phi(Y)$ is invariant to translation, rotations and homogeneous rescalings. The additional constraint that the covariance is equal to the identity matrix expresses an assumption that reconstruction errors for different coordinates in the embedding space should be measured on the same scale.

The optimal embedding $\vec{Y}_{i=1,2.....P} \in R^d$ is given by eigenvectors associated with the smallest d+1 eigenvalues of the matrix M (Horn and Johnson, 1990). The bottom eigenvector of this matrix is discarded, as it is a vector composed of all ones, with zero as eigenvalue. Discarding this eigenvector enforces the constraint that the embeddings have zero mean, as the components of other eigenvectors must sum to zero, by virtue of orthogonality.

The bottom d+1 eigenvectors (corresponding to smallest d+1 eigenvalues) of the matrix M can be determined without performing a full matrix diagonalization (Bai et al., 2000). Moreover, the matrix M can be stored and manipulated as the sparse symmetric matrix

$$M = (\mathrm{I} - W)^T (\mathrm{I} - W) \tag{6}$$

giving substantial computational savings for large values of P.

Although the reconstruction weights for each data point are computed from its local neighborhood independently, the embedding coordinates are computed by an $P \times P$ eigensolver, a global operation that couples all data points in connected components of the graph defined by the weight matrix. The different dimensions in the embedding space can be computed successively; this is done simply by computing the bottom eigenvectors from Equation (2) one at a time.

The nearest neighbor parameter N is a measure of the "quality" of input-output mapping (i.e. how well the high-dimensional structure is represented in the embedded space). If N is set too small, the mapping will not reflect any global properties; if it is too high, the mapping will lose its nonlinear character and behave like traditional PCA, as the entire data set is seen as local neighborhood. N is selected based on the residual variance (Kouropteva et al., 2002). It is defined as $1 - \rho^2_{E_x E_y}$ where ρ is the standard linear correlation coefficient, computed over all entries of E_x and E_y; E_x and E_y are the matrices of Euclidean distances (between pairs of points) in X and Y (as computed above), respectively. The lower the residual variance is, the better the high-dimensional data are represented in the embedded space. Hence, the optimal value for N, N_{opt} can be determined as

$$N_{opt} = \arg\min_N (1 - \rho^2_{E_x E_y}) \tag{7}$$

A few techniques like linear interpolations and training a neural network or RBF network (Vlachos et al., 2002) are available for mapping a new (previously unseen) sample. In the present work we have however preferred a simple strategy of concatenating the new sample with given samples and repeating the whole LLE procedure for on-line implementation. This preference is based on our observation that the LLE algorithm takes only few seconds of time to run (as LLE only involves sparse matrix computations), retaining non-linear mapping even for query point. Approaches like Neural or RBF networks are hard to train and linear interpolations may lose the non- linearity of data.

The procedure as described above leads to nonlinear dimensionality reduction of data. We shall now briefly describe the classification using SVDD.

3. SUPPORT VECTOR DOMAIN DISTRIBUTION

Support vector domain distribution (Tax and Duin, 1999) avoids solving the harder density estimation problem and uses the simple task of finding the support vectors of the multivariate distribution. The objective of classification of data domain is that the given set of data should be represented in a unique

minimal volume spherical domain enclosing all or nearly all the training points. The effect of outliers is reduced by using slack variables ξ_i to allow for data points outside the sphere and task is to minimize the volume of the sphere and number of data points outside the sphere.

$$F(R,\vec{a},\xi_i) = R^2 + C\sum_i \xi_i \qquad i = 1, \quad .....P \tag{8}$$

with constraints

$$(\vec{x}_i - \vec{a})^T(\vec{x}_i - \vec{a}) \le R^2 + \xi_i \qquad \forall \xi_i \ge 0 \tag{9}$$

P is the number of objects in training set and $\vec{a}$ is the center of the sphere. The parameter C characterizes the trade off between the volume of sphere and number of data points that lie outside.
Combining (7) & (8), we formulate the Lagrangian as,

$$L(R,\vec{a},\alpha_i,\xi_i) = R^2 + C\sum_i \xi_i - \sum_i \alpha_i(R^2 + \xi_i - (\vec{x}_i \cdot \vec{x}_i - 2\vec{a}\cdot\vec{x}_i + \vec{a}\cdot\vec{a})) - \sum_i \gamma_i \xi_i \tag{10}$$

with Lagrange multipliers $\alpha_i \ge 0$ & $\gamma_i \ge 0$.

After replacing dot products by kernel, the dual formulation amounts to the maximization of

$$L = \sum_i \alpha_i K(\vec{x}_i,\vec{x}_i) - \sum_{i,j} \alpha_i \alpha_j K(\vec{x}_i,\vec{x}_j) \tag{11}$$

with constraints

$$0 \le \alpha_i \le C \tag{12}$$

$$\sum_i \alpha_i = 1 \tag{13}$$

Only for some set of points the equality in Equation (9) is satisfied. These points lie on the boundary of sphere and are called as support vectors for which the coefficients α_i are non-zero. These points completely describe the sphere. The radius of the sphere is calculated as the distance of support vector for which $\alpha_i < C$ from the center of the sphere. The outliers or abnormal points are the bound objects for which $\alpha_i = C$. Having completed the training process a test point $\vec{z}$ is declared as an outlier, if the distance of the point to the center of the sphere is larger than the radius:

$$K(\vec{z},\vec{z}) - 2\sum_i \alpha_i K(\vec{z},\vec{x}_i) + \sum_{i,j} \alpha_i \alpha_j K(\vec{x}_i,\vec{x}_j) > R^2 \tag{14}$$

Different kernel functions can be used to get different domain description boundaries. The most popular kernels are polynomial kernel and Gaussian RBF kernel.

Polynomial kernel $$K(\vec{x}_i, \vec{x}_j) = (1 + \vec{x}_i \cdot \vec{x}_j)^n \quad (15)$$

Gaussian RBF kernel $$K(\vec{x}_i, \vec{x}_j) = \exp\left(\frac{-\left\|\vec{x}_i - \vec{x}_j\right\|^2}{2\sigma^2}\right) \quad (16)$$

4. CASE STUDIES

We have illustrated the method of online abnormality detection in a process plant using LLE-SVDD method in the following sub -sections with the two case studies viz. acetone-butanol fermentation and a benchmark semi-batch reactor problem.

4.1 Case study 1: Batch Acetone-Butanol Fermentation

Acetone-butanol fermentation is considered here as a case study. The mathematical model is taken from Vortruba et al. (1986). The model for the batch culture of *clostridium acetobutylicum* has been formulated using experimental data for anaerobic solvent production. The model takes into account biochemical as well as physiological aspects of growth and metabolite synthesis. The same example has been considered by Singhal (2002) for evaluating different pattern matching techniques. In this example we use the model

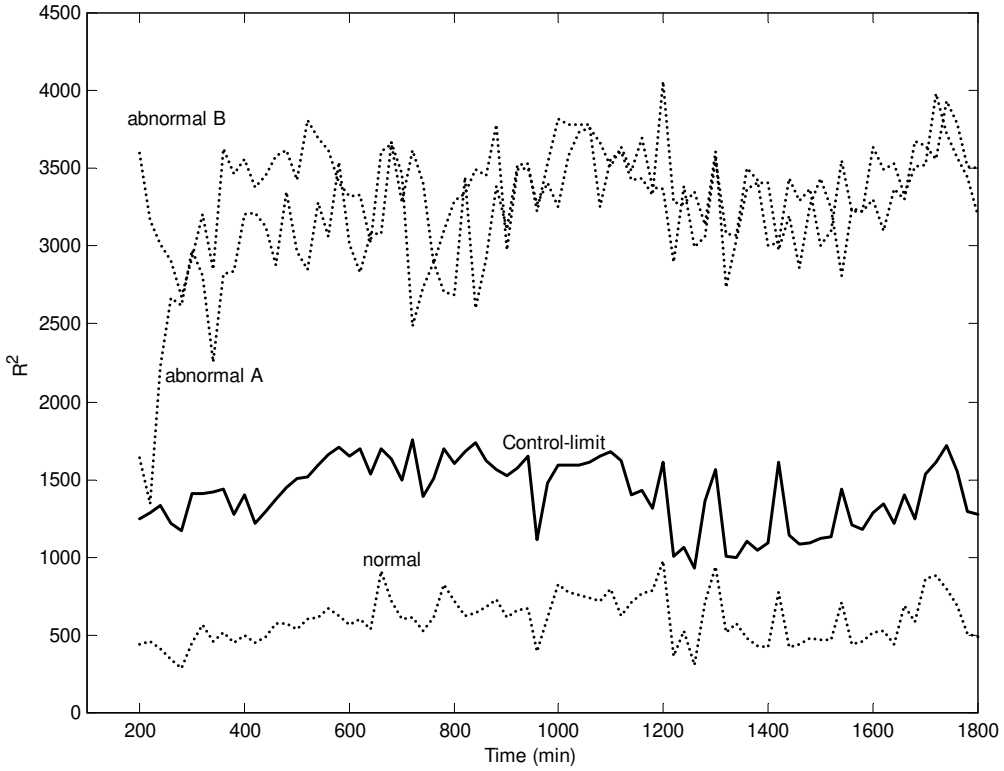

Figure 1. Abnormality detection in acetone-butanol fermentation using LLE-SVDD (showing abnormal batches A &B)

for online abnormality detection using SVDD along with nonlinear dimensionality reduction technique, LLE. The model consists of ten nonlinear differential equations. The parameters are the same as in Vortruba et al. (1986) and Singhal (2002). 55 normal batches were simulated by giving some variations in reactor cell concentration, substrate concentration and dimensionless cellular RNA concentration. This forms the historical database for the training of LLE-SVDD. In addition to this, two normal and six abnormal batches were simulated as test batches. Out of the six abnormal batches the first two correspond to abnormality due to slow substrate utilization, the next two correspond to abnormality due to dead inoculum and the remaining two correspond to abnormality due to increased cell sensitivity to butanol. The nine process variables used for monitoring are: reactor cell concentration, substrate concentration, butyric acid concentration, acetic acid concentration, butanol concentration, acetone concentration, ethanol concentration, CO_2 concentration and H_2 concentration. All the variables were measured after every 10 minutes. The total time required for a single batch is 30 hr.

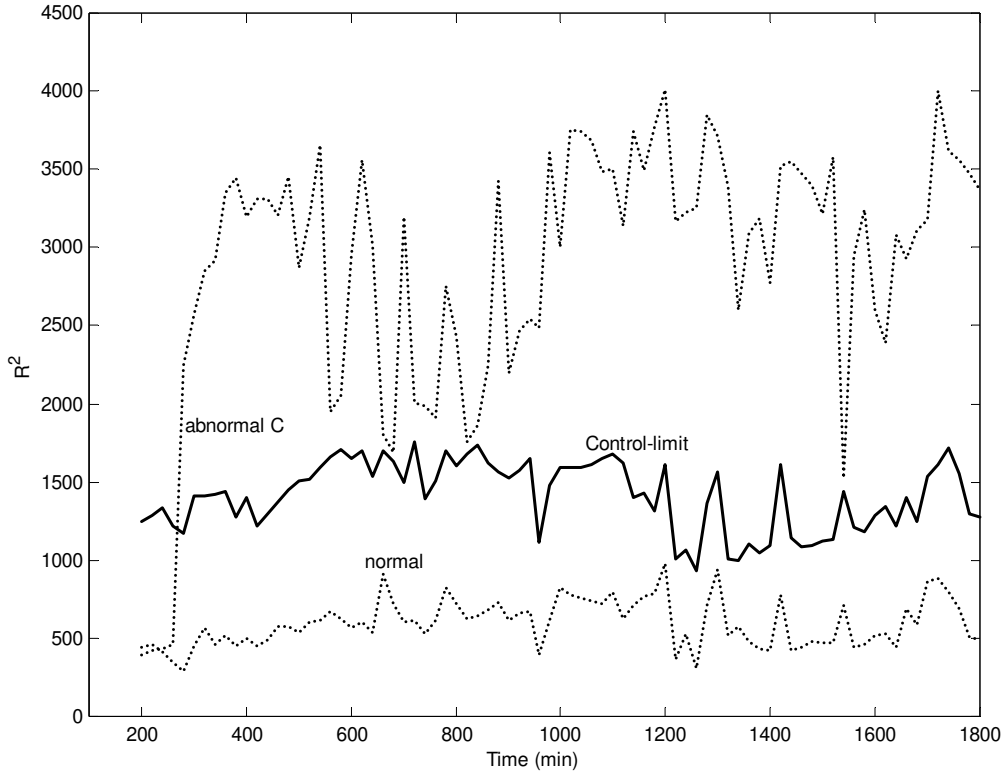

Figure 2. Abnormality detection in acetone-butanol fermentation using LLE-SVDD (showing abnormal batch C)

For this analysis, each nominal batch is unfolded, and can be represented as a data vector. For instance, at time τ, a batch can be written as,

$$[\ \vec{}^{T}\ \tau \quad \vec{}^{T}\ \tau - \Delta \quad \vec{}^{T}\ \tau - \ \Delta \qquad \vec{}^{T}\ \tau - \quad - \ \Delta\ ]$$

where

$$\vec{Q}(\tau) = \left[q_{1,}\ q_{2,}\quad \ldots\ q_{J}\right]^{T}$$

is the J-dimensional variables vector at time point τ. For this problem the number of monitoring variables J=9 and the length of moving window is *w*=20.

The training data for LLE-SVDD analysis consists of 55 normal batches with values of all the 9 selected variables at 20 consecutive sampling times. Thus data for LLE is a matrix of 55×180. We fixed the dimensionality of reduced space (*d*) into which LLE is projecting to 15. The nearest neighbor parameter *N* (as obtained from the residual variance criterion) changes very slightly while moving from one window to another; thus for sake of computational simplicity the value of *N* was fixed at an average value of 12. The reduced matrix with dimension of 55×15 was then used to train SVDD for novelty detection. The test data with reduced dimension was obtained using on-line LLE as described in section 2.

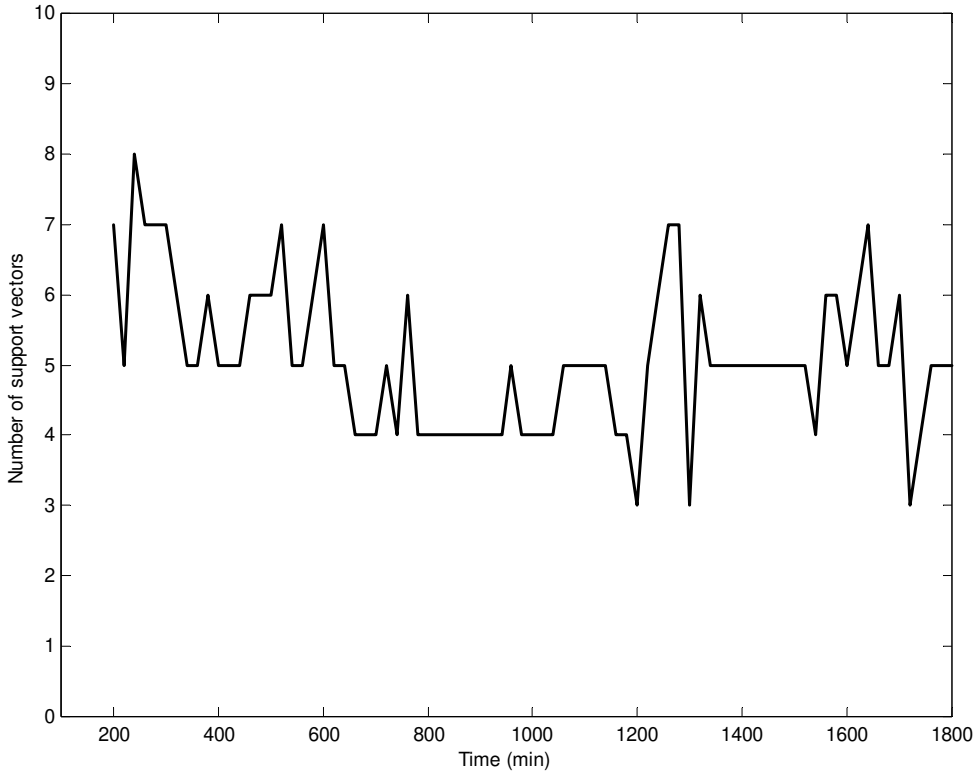

Figure 3. Number of support vectors obtained for Acetone-butanol fermentation problem

SVDD was carried out on this dataset by selecting an appropriate kernel, which is able to give the best description of the data domain. This classification was done by the following steps: (i) The quadratic optimization problem represented by Equation (11) was solved to get the support vectors and the corresponding Lagrange multipliers, (ii) These values were used to calculate the value of R^2 using Equation (14), (iii) The window was moved further by 10 minutes interval and the SVDD as described above was again carried out for the set of data vectors belonging to the new window. The procedure was repeated until the analysis covers the variables at the final time. Value of R^2 for support vectors act as the control limit for online testing of a new batch. The R^2 values are unique for each window and thus there exists a profile of R^2 for the nominal batches. As long as the R^2 value of the test batch lies below the SV line (solid line shown in Figure 1 &2), the batch is normal. LLE-SVDD identifies all the six abnormal test batches. In Figure 1 and 2, one test batch belonging to each case i.e. slow substrate utilization (abnormal batch A), dead

inoculum (abnormal batch B), increased cell sensitivity to butanol (abnormal batch C) is shown. Test batches A & B were detected as abnormal batch from the beginning of the batch, whereas batch C was detected as abnormal 240 minutes after the start of the batch. The polynomial kernel of order 2 was used with parameter, C= 1. SVDD parameters were obtained heuristically. The number of support vectors obtained throughout the batch duration is shown in figure 3 and on an average constitutes 9.24 % of total data. The number of outliers obtained for each window is zero. Computational requirement for SVDD algorithm is very low as it works on a fraction of data (support vectors) in training set. Dimensionality reduction with LLE requires more computations than that required by the conventional methods like PCA, but with an advantage of keeping the nonlinear features of the data intact.

An alternative way is to consider the data from the initial time point to current time point. Although this increases the computational load on SVDD, it is still manageable due to dimensionality reduction ability of LLE. The analysis was carried out by considering the following matrix for a batch at any given time instant τ :

$$[\vec{Q}^T(\tau) \quad \vec{Q}^T(\tau - \Delta t) \quad \vec{Q}^T(\tau - 2\Delta t) \quad \ldots \quad \vec{Q}^T(0)]$$

The data fed to LLE consists of variables from initial time point to the current time point. The data was first reduced by LLE before being processed by the classification algorithm. SVDD with polynomial kernel of order 2 along with parameter C=1 successfully identifies the normal and abnormal test batches as shown in Figure 4. (To avoid complexity, only two of the abnormal batches are shown in figure).

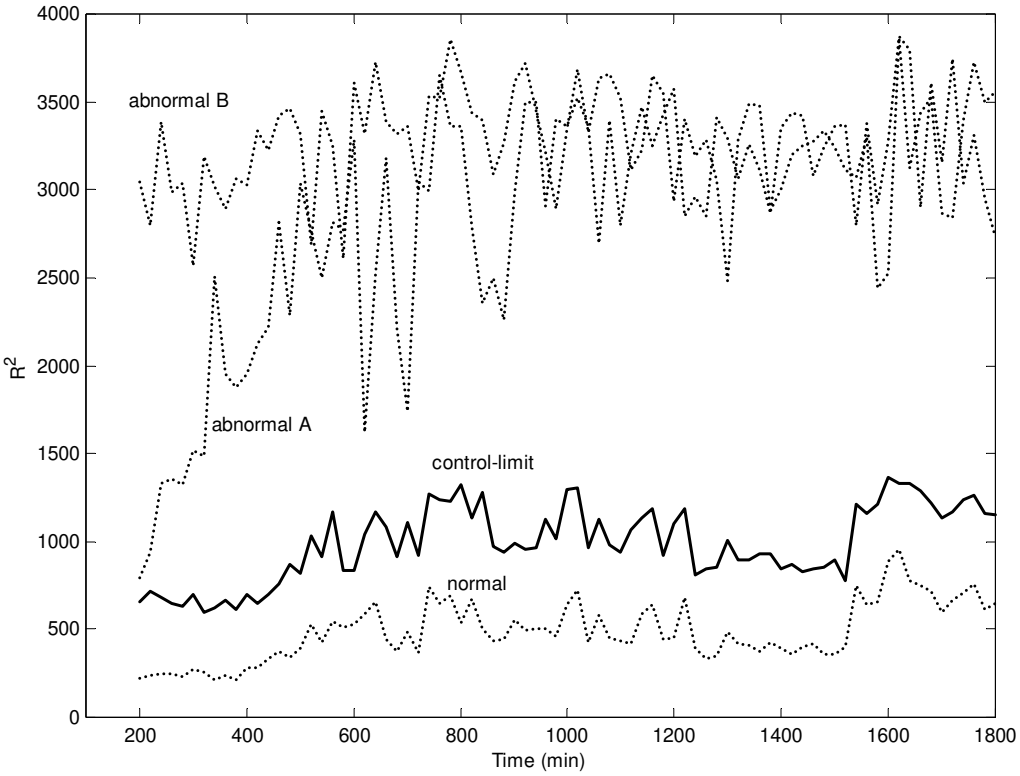

Figure 4. Abnormality detection in acetone-butanol fermentation using LLE-SVDD (Considering data from initial time point to the current one)

For testing the efficacy of the SVDD abnormality detection algorithm, we have compared the results with the recently proposed dynamic PCA (Chen and Liu, 2002). The criterion used for comparison of performance is time required by the method to detect the fault after it occurs. Q-chart for dynamic PCA

with three principal components (capturing 97% variance) with 95% confidence limit with two time lag windows is shown in Figure 5. As shown in figure the dynamic PCA identifies the normal batch. But it detects the abnormality of the batches (abnormal A, B & C) at much latter stages of the process, whereas LLE-SVDD identifies the abnormality of the batches from the beginning of the process (Figure1 & 2).

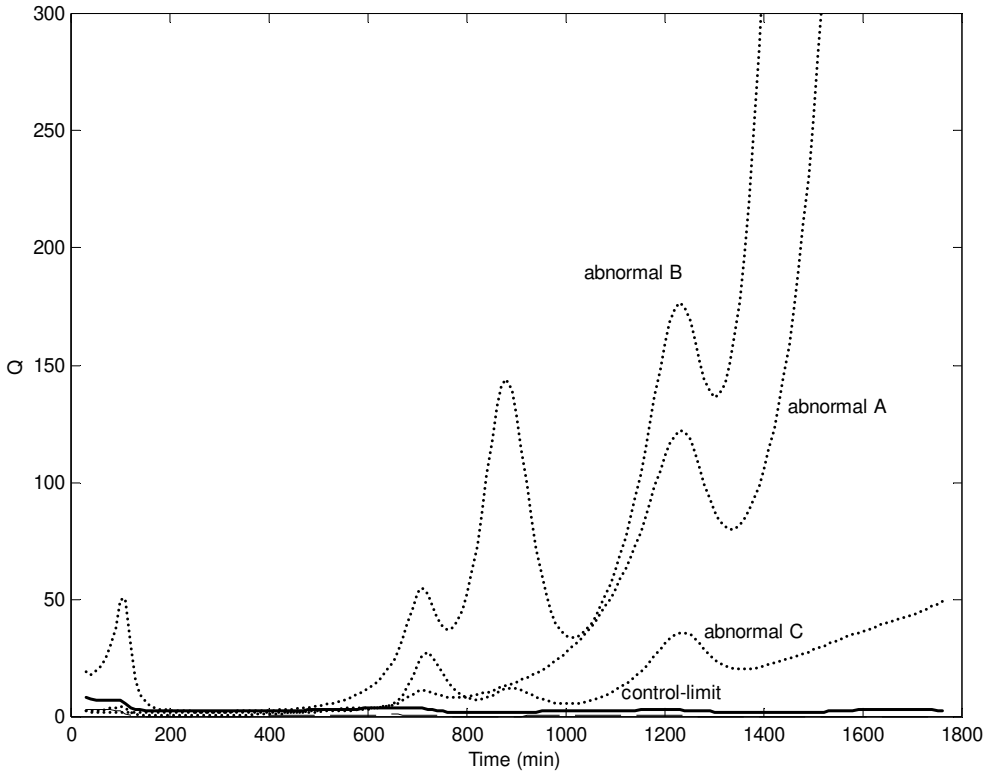

Figure 5. Q-chart for abnormality detection in acetone-butanol fermentation using dynamic PCA

4.2 Case Study 2: Semi-Batch Reactor for SBR production

This example is a simulated study of a semi batch reactor for the production of styrene-butadiene rubber (SBR) (Nomikos and MacGregor, 1994). This problem has been used as benchmark for evaluating the performance of various process monitoring methods (Nomikos and MacGregor, 1994; Chen and Liu, 2002). The reference data set contains 50 normal batches with some variations in the base conditions like impurities in the initial charge of organic phase and in the butadiene monomer feed to the reactor. The batch was divided into 200 time intervals and nine different variables were chosen for the purpose of monitoring. Apart from this reference set, three test batches were simulated: first is the normal batch (test batch A), second an abnormal batch with an initial organic impurity contamination in the butadiene feed, 30% above that of the base case (test batch B); and the third with the same problem, but this batch having contamination, 50 % above the normal level, started halfway through the operation (test batch C). The numerical data sets for all the variable measurements for the 50 nominal and three test batches (one normal and two abnormal) were obtained from Nomikos and MacGregor (1994). The three test batches are presented in the form of figures in their paper. It can be clearly seen from these trajectories that it would be difficult to differentiate normal and abnormal batches by visual observations alone, requiring a rational abnormality detection methodology.

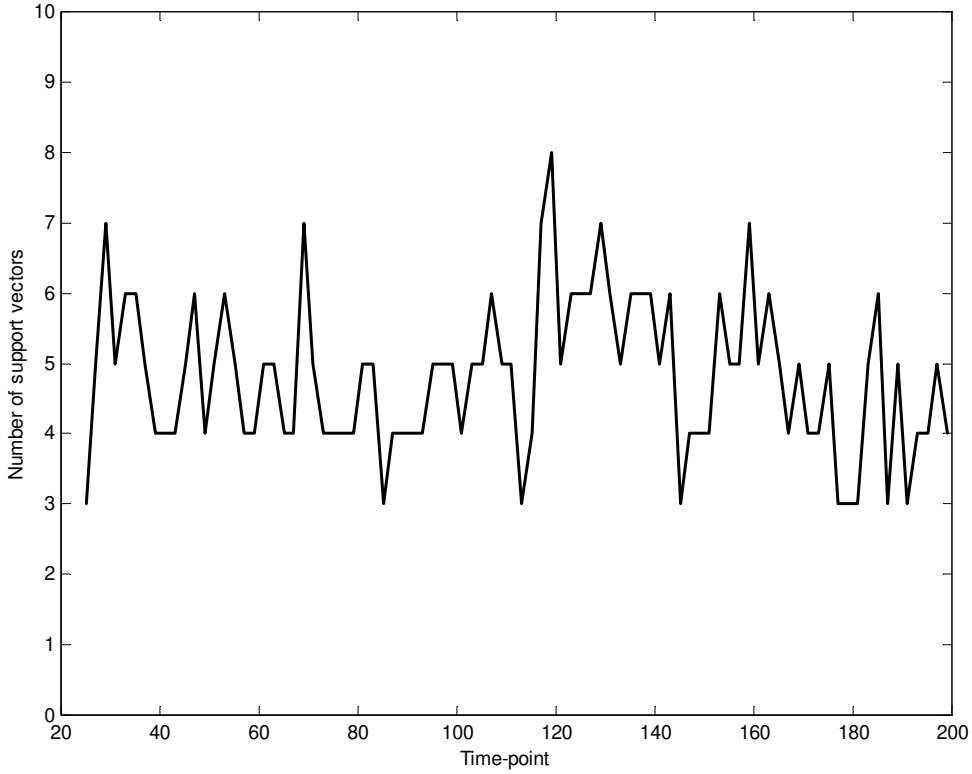

Figure 6. Number of support vectors obtained for Semi-batch reactor

LLE-SVDD analysis was done for SBR data using moving window of length 25. Thus training data for LLE-SVDD analysis consists of 50 normal batches with values of all the 9 selected variables at 25 consecutive sampling times. LLE reduces dimensionality of training data from 225 to 15 using the nearest neighbor parameter N equal to 12. The reduced matrix with dimension of 50×15 is then used to train SVDD. Thus LLE-SVDD analysis is done for each time-point and the window is moved after each time-point till the completion of the batch. Polynomial kernel of order 7 was used with parameter, C=0.5. The number of support vectors along the batch duration is shown in figure 6 and on an average constitutes 9.63 % of total data. For this problem too, we obtained zero outlier for each window during the complete batch duration. The support vectors along with their Lagrange multipliers can be used to calculate the value of R^2 for the test batch. The calculated R^2 values for support vectors on boundary of the hypersphere are shown in figure 7 as solid line, which acts as the control limit for the online test batch. It is clear from figure 7, LLE-SVDD hybridization works well for all the test batches i.e. it successfully identifies the normality and abnormality of the test batches for online monitoring. For instance, R^2 value of test batch A lies below the control-limit throughout the batch, hence it is a normal batch (Figure 7). The test batch B on the other hand crosses the control-limit at the 29^{th} time point and is classified as abnormal from 30^{th} time point to the end of batch. Similarly from figure 7, it is clear that the test batch C is normal up to 108^{th} time point, but at the 109^{th} time point it shows abnormality as it crosses the control limit and remains abnormal up to the end of the batch. The example clearly brings out the simplicity and usefulness of the method. This example has been studied by various methods in the literature and the performance of the proposed hybrid method for the SBR data is compared to that of benchmark MPCA method (Nomikos and MacGregor, 1994). Again the criterion used for comparison of performance is time required by the method to detect the fault after it occurs. MPCA method detect abnormality of test batch A within first 15 time points, while it detect the test

batch C as abnormal before 110th time point (Nomikos and MacGregor, 1994). The results of both the methods are found to be comparable i.e. there is not significant difference between times taken by the hybrid method and the conventional MPCA method to detect the fault after its occurrence (Nomikos and MacGregor, 1994).

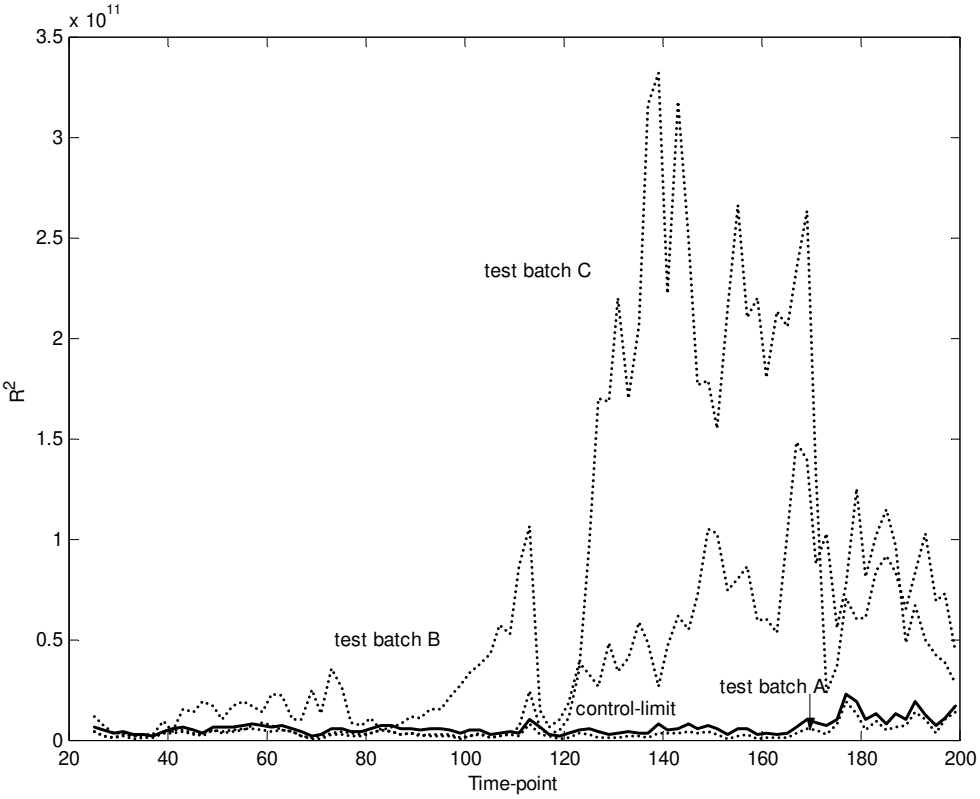

Figure 7. Abnormality detection in semi-batch reactor using LLE-SVDD

In working plants it may be very difficult to obtain large number of data pertaining to abnormal process conditions. As SVDD-LLE methodology requires only data belonging to normal conditions, the SVDD-LLE methodology is particularly advantageous as compared to many other existing techniques requiring large number of abnormal data. The trained SVDD algorithm can be completely characterized by a very small fraction (less than 10 %) of the total training data (i.e. support vectors) to define the distribution. This greatly reduces computational load during online testing. Another desirable feature of SVDD is that it requires solution of a quadratic optimization problem always leading to the unique global solution. On the other hand, some AI based methodologies solve a hard nonconvex optimization problem with the possibility of converging to one of the local minima. Additionally, the number of free parameters in SVDD does not depend explicitly on the input dimensionality, unlike other machine learning methods. The LLE part of the algorithm retains the relevant nonlinear features while reducing the input dimension rendering the hybrid methodology very attractive as compared to the existing methods.

5. CONCLUSION

The hybrid method using LLE and SVDD is illustrated with two case studies of acetone-butanol fermentation and a benchmark SBR problem. The results show that LLE along with SVDD can be a very powerful tool for online process monitoring. As most of the industrial processes are nonlinear in nature, nonlinear dimensionality reduction using LLE can be very useful in reducing the features of the data, which in turn reduces the time for abnormality detection technique like SVDD.

ACKNOWLEDGEMENTS

Financial assistance from ICI, UK is gratefully acknowledged. We are indebted to Professor J. F. MacGregor (McMaster University, Canada) for supplying data for the semi-batch reactor.

NOTATION

$\vec{a}$	center of hypersphere
C	parameter in SVDD
C_{jk}	local covariance matrix
D	original dimension
d	reduced dimension
E_x, E_y	matrix of Euclidean distances in X, Y
F	optimization function in SVDD
I	identity matrix
J	number of monitoring variables
K	kernel function
L	Lagrangian function
M	sparse matrix
N	number of nearest neighbors
n	order of polynomial kernel
P	number of data points
$\vec{Q}$	vector of monitoring variables
q	monitoring variable
R	radius of the hypersphere
t	time
w	length of moving window
W	weights for reconstruction
$\vec{X}$	original dataset
$\vec{x}$	datapoint
$\vec{Y}$	reduced dataset
$\vec{z}$	test point

Greek letters

α, γ	Lagarange multipliers
δ	parameter in Equation (4)
ε	reconstruction error
σ	width of Gaussian RBF kernel
Φ	embedding cost function
	reconstruction weight
$\vec{\psi}$	nearest neighbors
ρ	standard linear correlation coefficient
ξ	slack variables
τ	time point

Superscripts

T	transpose of matrix
-1	inverse of matrix

Subscripts

opt	optimal

REFERENCES

Bai, Z., Demmel, J., Dongarra, J., Ruhe, A., Van Der Vorst, H., "Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide", Society for Industrial and Applied Mathematics, Philadelphia, (2000).

Bailer-Jones, C.A.L., Irwin, M., Von Hippel, T., "Automated Classification of Stellar Spectra. II: Two-Dimensional Classification with Neural Networks and Principal Components Analysis", Monthly Notices of the Royal Astronomical Society, Vol. 298, 361-377 (1998).

Belkin, M., Niyogi, P., "Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering" in "Advances in Neural Information Processing Systems", T.G. Dietterich, S. Becker and Z. Ghahramani Eds., MIT Press, Cambridge, Vol. 14, 585-591 (2002).

Bishop, C.M., Svensén, M., Williams., C.K.I., "GTM: The Generative Topographic Mapping", Neural Computation, Vol.10, No. 1, 215-234 (1998).

Boque, R., Smilde, A. K., "Monitoring and Diagnosing Batch Processes with Multiway Covariates Regression Models", AIChE J., Vol. 45, No. 7, 1503-1520 (1999).

Borg, I., Groenen, P., "Modern Multidimensional Scaling: Theory and Applications", Springer-Verlag, Berlin, (1997).

Chen, J, Liu, J., "Post Analysis on Different Operating Time Processes using Orthonormal Function Approximation and Multiway Principal Component Analysis", J. Process Control, Vol. 10, No. 5, 411- 418 (2000).

Chen, J., Liu, K.-C., "On-line Batch Process Monitoring using Dynamic PCA and Dynamic PLS Models", Chem. Eng. Sci. Vol. 57, No. 1, 63-75 (2002).

Chen, B. H., Wang, X. Z., Yang, S. H., & Mcgreavy, C., "Application of Wavelets and Neural Networks to Diagnostic System Development. I. Feature Extraction", Comput. Chem. Engng. Vol. 23, No.7, 899-906 (1999).

Dahl, S. K., Piovoso, M. J., Kosanovich, K. A., "Translating Third-order Data Analysis Methods to Chemical Batch Processes", Chemometr. lntell. Lab. Syst., Vol. 46, No. 2, 161-180 (1999).

DeMers, D., Cottrell, G.W., "Nonlinear Dimensionality Reduction", in "Advances in Neural Information Processing Systems", C.L. Giles, S.J. Hanson and J.D. Cowan Eds., Morgan Kaufmann, San Mateo, CA, Vol. 5, 580-587 (1993).

Devijver, P.A., Kittler, J., "Pattern Recognition, A Statistical Approach", Prentice-Hall, London (1982).

Dong, D., McAvoy, T. J., "Nonlinear Principal Component Analysis — Based on Principal Curves and Neural Networks", Comput. Chem. Engng., Vol. 20, No. 1, 65-78 (1996a).

Dong, D., and McAvoy, T. J., "Batch Tracking via Nonlinear Principal Component Analysis", AIChE J.,Vol. 42, No. 8, 2199-2208 (1996b).

Duda, R.O., Hart, P.E., Stork, D.G., "Pattern Classification", John Wiley & Sons, New York, NY, 2nd edition, (2001).

Fan, J. Y., Nikolaou, M., & White, R. E., "An Approach to Fault Diagnosis of Chemical Processes via Neural Networks", AIChE J., Vol. 39, No. 1, 82-88 (1993).

Farell, A. E., Roat, S. D., "Framework for Enhancing Fault Diagnosis Capabilities of Artificial Neural Networks", Comput. Chem. Engng. Vol. 18, No. 7, 613-635 (1994).

Hastie, T., Stuetzle, W., "Principal Curves", J. Amer. Statist. Assoc. , Vol. 84, No. 406, 502-516 (1989).

Horn, R. A., Johnson, C. R., "Matrix Analysis", Cambridge University Press, Cambridge, 1990.

Hoskins, J. C., Kaliyur, K. M., Himmelblau, D. M., "Fault Diagnosis in Complex Chemical Plants using Artificial Neural Networks", AIChE J., Vol. 37, No. 1, 137-141 (1991).

Jain, A.K., Duin, R.P.W., Mao, J., "Statistical Pattern Recognition: a Review", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 1,4-37 (2000).

Kavuri, S. N., Venkatasubramanian, V.,"Using Fuzzy Clustering with Ellipsoidal Units in Neural Networks for Robust Fault Classification", Comput. Chem. Engng. Vol. 17, No. 8, 765-784 (1993).

Kavuri, S. N., & Venkatasubramanian, V.,"Neural Network Decomposition Strategies for Large Scale Fault Diagnosis", International Journal of Control Vol. 59 No. 3, 767-792 (1994).

Kohonen, T., "Self-organizing Maps", Springer Series in Information Sciences. Springer (1995).

Kouropteva, O., Okun, O., Pietikainen, M., "Selection of the Optimal Parameter Value for the Local Linear Embedding Algorithm", Proceedings of the First International Conference on Fuzzy Systems and Knowledge Discovery, Singapore, 359-363 (2002).

Louwerse, D. J., Smilde, A. K., "Multivariate Statistical Process Control of Batch Processes based on Three-Way Models", Chem. Eng. Sci., Vol. 55, No.7, 1225-1235 (2000).

Malthouse, E. C., "Limitations of Nonlinear PCA as Performed with Generic Neural Networks", IEEE Trans. Neural Networks, Vol. 9, No.1, 165-173 (1998).

Mao, J., Jain, A. K., "Artificial Neural Networks for Feature Extraction and Multivariate Data Projection", IEEE Trans. Neural Networks, Vol. 6, No. 2, 296-317 (1995).

Mardia, K.V., Kent, J.T., Bibby, J.M., "Multivariate Analysis", Academic Press, London, (1979).

Martin, E. B., Morris, A. J., "Non-Parametric Confidence Bounds for Process Performance Monitoring Charts", J. Process Control, Vol. 6, No. 6, 349-358 (1996).

Martin, E. B., Morris, A. J., Papazoglou, M. C., Kiparisassides, C.,"Batch Process Monitoring for Consistent Production", Comput. Chem. Engng., Vol. 20, No.S1, 599-604 (1996).

McClurkin, J.W., Optican, L.M., Richmond, B.J., Gawne, T.J., "Concurrent Processing and Complexity of Temporally Encoded Neuronal Messages in Visual Perception", Science, Vol. 253, 675-677 (1991).

Murase, H., Nayar, S. K., "Visual Learning and Recognition of 3-D Objects from Appearance", Int. J. Comp. Vision, Vol. 14, 5-24 (1995).

Nomikos, P., MacGregor, J. F., "Monitoring Batch Processes using Multiway Principal Component Analysis", AIChE J., Vol. 40, 1361-1375 (1994).

Nomikos, P., MacGregor, J. F., "Multi-Way Partial Least Square in Monitoring Batch Processes", Chemometr. lntell. Lab. Syst., Vol. 30, No.1, 97-108 (1995).

Ramaker, H.J., Van Sprang E.N.M., Gurden S.P., Westerhuis J.A., Smilde A.K., "Improved Monitoring of Batch Processes by Incorporating External Information", J. Process Control Vol. 12, No. 4, 569-576 (2002).

Rannar, S., MacGregor, J. F., Wold, S., "Adaptive Batch Monitoring using Hierarchical PCA", Chemometr. lntell. Lab. Syst., Vol. 41, No.1, 73-81 (1998).

Rengaswamy, R., Venkatasubramanian, V.,"A Fast Training Neural Network and its Updation for Incipient Fault Detection and Diagnosis", Comput. Chem. Engng. Vol. 24, No.2-7, 431-437 (2000).

Roweis, S.T., Saul, L.K., "Nonlinear Dimensionality Reduction by Locally Linear Embedding", Science Vol. 290, 2323-2326 (2000).

Ruiz, D., Nougués, J., Calderón, Z., Espuña A., Puigjaner L., "Neural Network Based Framework for Fault Diagnosis in Batch Chemical Plants", Comput. Chem. Engng., Vol. 24, No. 2-7, 777-784 (2000).

Ruiz, D., Cantón J., Nougués, J. M., Espuña A., Puigjaner, L., "On-line Fault Diagnosis System Support for Reactive Scheduling in Multipurpose Batch Chemical Plants", Comput. Chem. Engng., Vol. 25, No. 4-6, 829-837 (2001).

Sammon, J. W., "A Nonlinear Mapping for Data Structure Analysis", IEEE Transactions on Computer, Vol. 18, No. 5, 401-409 (1969).

Scenna N. J.,"Some Aspects of Fault Diagnosis in Batch Processes", Reliability Engineering & System Safety, Vol. 70, No.1, 95-110 (2000).

Singhal, A., "Pattern Matching in Multivariate Time Series Data", PhD Thesis, Univ. of California, Santa Barbara (2002).

Tax, D. M. J., Duin, R. P.W., "Support Vector Domain Distribution", Pattern Recognition Letters Vol. 20 No.11-13, 1191-1199 (1999).

Tenenbaum, J. B., de Silva, V., Langford, J. C., "A Global Geometric Framework for Nonlinear Dimensionality Reduction", Science, Vol. 290, 2319-2323 (2000).

Tipping, M.E., Bishop, C.M., "Mixtures of Probabilistic Principal Component Analyzers", Neural Computation, Vol. 11, No. 2, 443-482 (1999).

Ungar, L. H., Powell, B. A., Kamens, S. N.,"Adaptive Networks for Fault Diagnosis and Process Control", Comput. Chem. Engng. Vol. 14, No. 4-5, 561-572 (1990).

Van Sprang, E. N. M., Ramaker, H.-J., Westerhuis, J. A., Gurden, S. P., Smilde, A. K., "Critical Evaluation of Approaches for on-line Batch Process Monitoring", Chem. Eng. Sci. Vol. 57, No. 18, 3979-3991 (2002).

Venkatasubramanian, V., Chan, K., "A Neural Network Methodology for Process Fault Diagnosis", AIChE J., Vol. 35, No. 12, 1993-2002 (1989).

Venkatasubramanian, V., Vaidyanathan, R.,Yamamoto, Y.,"Process Fault Detection and Diagnosis using Neural Networks I: Steady State Processes", Comput. Chem. Engng. Vol. 14, No. 7, 699-712, (1990).

Vlachos, M., Domeniconi, C., Gunopulos, D., Kollios, G., Koudas, N., "Non-Linear Dimensionality Reduction Techniques for Classification and Visualization", in "Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining", (2002).

Vortruba, J., Volesky, B., Yerushalmi, L., "Mathematical Model of a Batch Acetone-Butanol Fermentation", Biotechnol. Bioeng., Vol. 28 247-255 (1986).

Wang, X. Z., Chen, B. H., Yang, S. H., Mcgreavy, C., "Application of Wavelets and Neural Networks to Diagnostic System Development. 2. An Integrated Framework and its Application",Comput. Chem. Engng. Vol. 23, No. 7, 945-954 (1999).

Watanabe, K., Matsura, I., Abe, M., Kubota, M., Himmelblau, D. M. , "Incipient Fault Diagnosis of Chemical Processes via Artificial Neural Networks", AIChE J., Vol. 35, No. 11, 1803-1812 (1989).

Westerhuis, J. A., Kourti, T., MacGregor, J. F., "Comparing Alternative Approaches for Multivariate Statistical Analysis of Batch Process Data", J. Chemometrics, Vol. 13, 397- 413 (1999).

Wise, B. M., Gallagher, N. B., Butler, S. W., White Jr., D. D., Barna, G. G., "A Comparison of Principal Component Analysis, Multiway Principal Component Analysis, Trilinear Decomposition and Parallel Factor Analysis for Fault Detection in a Semiconductor Etch Process", J. Chemometrics, Vol. 13, 379-396 (1999).

Wold, S., Geladi, P., Esbensen, K., Ohman, J., "Multi-way Principal Components and PLS Analysis", J. Chemometrics, Vol. 1, 41-56 (1987).

Zhao, J., Chen, B., Shen, J., "A Hybrid ANN-ES System for Dynamic Fault Diagnosis of Hydrocracking Process", Comput. Chem. Engng. Vol. 21, S929-S933 (1997).

APPENDIX

Optimal reconstruction weight calculation

Equation (1) has a closed form solution (Roweis and Saul, 2000), which determines the optimal weights W_{ij} that best reconstruct each data point from its neighbors. Suppose a particular data point $\vec{x}$ with neighbors $\vec{\psi}_{j\ \ldots 2N}$ and reconstruction weights $_{=1,2\ \ N}$ satisfying $\sum_{j} {}_{j} = 1$.

Thus the cost function (giving reconstruction error) to be minimized is:

$$\varepsilon = \left| \vec{x} - \sum_{j} {}_{j} \vec{\psi}_{j} \right|^2 \tag{17}$$

as $\sum_{j} {}_{j} = 1$

$$\varepsilon = \left| \sum_{j} {}_{j} (\vec{x} - \vec{\psi}_{j}) \right|^2 \tag{18}$$

Now introducing local covariance matrix

$$C_{jk} = (\vec{x} - \vec{\psi}_{j}).(\vec{x} - \vec{\psi}_{k}) \tag{19}$$

we have

$$\varepsilon = \sum_{jk} \quad _j \quad _k C_{jk} \tag{20}$$

This error can be minimized in closed form with constraint $\sum_j w_j = 1$. In terms of the inverse local covariance matrix, the optimal weights are given by:

$$_j = \frac{\sum_k C_{jk}^{-1}}{\sum_{lm} C_{lm}^{-1}} \tag{21}$$

Thus solution involves the inversion of local covariance matrix (symmetric and semi positive definite). If the covariance matrix is singular or nearly singular (e.g. when there are more neighbors than input dimensions [$N > D$]; or when the data points are not in general position), it can be conditioned by adding a small multiple of the identity matrix. This enforces a penalty on large weights, which exploit correlations beyond a certain level of precision in the data sampling process.