NTNU
Norwegian University of
Science and Technology

Condition Monitoring

Based on "black-box"- and First Principal Models

Tom-Rune Borgan

Master of Science in Engineering Cybernetics
Submission date: February 2011
Supervisor: Tor Engebret Onshus, ITK
Co-supervisor: Erling Lunde, Statoil

Norwegian University of Science and Technology
Department of Engineering Cybernetics

Problem Description

Problem Description

Condition monitoring based on combined black-box - and first principal models.

Statoil ASA has developed a tool for condition monitoring on oil- and gas processes. The tool is called Early Fault and Disturbance Detection (EFDD) and is a black-box equipment. The EFDD system contains, among others, a module called Principal Component Analysis (PCA). PCA is a linear data driven method used to support condition monitoring. To increase the performance of PCA a modification to the existing implementation is desired. The modification is based on using available models of the process being analyzed. The idea is that a priori knowledge about the process that may increase the performance of the PCA method.

The master thesis should focus on several models of heat exchanger and pump processes. The models should contain different degrees of complexity and serve to investigate the following problems.

1. Nonlinearities: By compensating known nonlinearities the linear PCA method would increase its performance.
2. Extrapolation: The black-box model is only valid for data region spanned by the configuration dataset. It is assumed that a modification with the use of models would extend the valid data region.
3. Instrumentation: Most processes have a sub-optimal instrumentation (lack of measurements). Investigate if the lack of measurements could be compensated by models.

The following approaches may be investigated, where at least A and B shall be verified.

A. Nonlinear expansion to PCA
B. MBPCA, model based PCA.
C. Nonlinear PCA, including Kernel PCA (KPCA) and neural networks.

The assignment should contain an overview of the different approaches with a discussion of their main properties. The assignment should explain different ways to take care of the problems mentioned above and document with the use of simulated experiments, or based on data from Statoil installations.

Assignment given: 10. September 2010
Supervisor: Tor Engebret Onshus, ITK

Assignment given: 31. August 2010
Supervisor: Tor Engebret Onshus, ITK

NTNU
Norwegian University of
Science and Technology

Condition monitoring of oil- and gas processes

Tom-Rune Borgan

Master of Science in Engineering Cybernetics
Submission data: February 2011
Supervisor: Tor Engebret Onshus, ITK

Norwegian University of Science and Technology
Department of Engineering Cybernetic

Summary and conclusions

This master thesis investigates Principal Component Analysis (PCA) methods used in the field of Early Fault and Disturbance Detection (EFDD). Statoil and ABB have in collaboration developed an application for EFDD that among others consist of PCA methods. The applications are used to run condition monitoring on oil- and gas processes, and are a software prototype currently being tested in Statoil. To make the method more robust and convincing to the operators it is desirable to improve the application.

This thesis will focus on the Principal Component Analysis (PCA) method and some extension to it based on Model Based PCA (MBPCA). The PCA in its simplest form have some severe restrictions due to linear and stationary data. The motivation will therefore be to see how PCA and extension based on MBPCA and Nonlinear PCA (NLPCA) methods operates when used on non-linear data.

For the PCA to describe a process adequately a certain amount of data is required. In the industry the process is often sort of instrumentation. The next motivation would be to investigate if lack of instrumentation could be replaced by some estimates and then approve the ability for the PCA analysis.

Another issue concerning instrumentation is the use of virtual tags. Virtual tags are mathematical functions based on already available measurements. The idea is based on process insight. If we know the process well and the cause of nonlinearities some additional nonlinear functions could be incorporated to increase the performance of the PCA method. To verify the factors mention above some of the methods would use data from a heat exchanger – and Centrifugal pump process, and some using only one of the processes.

The conclusion from the work is as follows:

Based on the simulations preformed it's evident that using MBPCA do improve the PCA method for fault detection, even if the model is not entirely correct to the real process. When it comes to using virtual tags the simulations on centrifugal pump increased the performance of the PCA method. NLPCA here based on Autoassociative Neural Networks did not perform as well as MBPCA but the method is harder to tune and therefore it would be wrong to brush aside this method. The improvement of using estimates for missing measurement gave small improvement.

Preface

This master thesis has been written as a part of the Master of Technology program at the Norwegian University of Science and Technology, department of Engineering Cybernetics. The process has been interesting and challenging but most of all given the author useful experience in a new discipline. First of all, I thank my supervisor, Professor Tor Engebret Onshus, for his comments and advice and ensuring the progress of this project. I also thank Erling Lunde at Statoil for his extraordinary support and guidance during the process. Finally, I thank my girlfriend Marianne, and our son, Kasper, for their love and support.

In this work MATLAB and SIMULINK software has been used, and a NLPCA toolbox for MATLAB.

Tom-Rune Borgan
Trondheim, Feb 2011

Figure list

Table list

Chapter 1

INTRODUCTION

1.1 Background

The industry today often consists of a big organization where a lot of decision making, maintenance planning and operational responsibilities are important daily tasks. To deal with this a process monitoring (PM) tools becomes evident. In collaboration between Statoil and ABB such a PM tool has been developed and implemented. The system is based on both a rule based - and a data driven PM method and the tool, as well as the overall concept, is referred to as Early Fault and Disturbance Detection (EFDD). Both the rule based – and the data driven methods have their own advantages and drawbacks.

The rule based method is based on having a database containing both nominal and faulty process signatures. This database needs to be updated as new scenarios occur and will therefore demand more from the user. This method will increase its performance over time because we will expect an increase in different scenarios in the database. A pure data driven method does not need any updating work and its approach will yield meaningful results upon installation, but a decreased performance over time. Through assessing these strengths and weaknesses there are reasons to believe that by combining them we will utilizes the overall PM system.

In this master thesis there will be focus on the data driven PM method called principal component analysis (PCA), model based PCA (MBPCA) and an expansion called nonlinear principal component analysis (NLPCA). All methods are powerful with several applications. The methods will be tested on an heat exchanger - and a pump processes. These processes are standard in the oil- and gas industry and a good PM method is therefore desirable. In order to test their performance on fault detection, different faults will be simulated using the heat exchanger- and pump models.

All the methods are data driven and therefore dependent on process data. To be able to describe the process adequately the number of data variables must be of a certain amount. In heat exchanger processes used for EFDD there are assumed some missing measurements. In this case there would be investigated if some of these missing measurements could be replaced with estimates, based on some real measurements and some process assumptions. There would be tested if this addition data would increase the performance of the PCA - , MBPCA - and NLPCA methods. A test is run to check if the additional data increases the performance of the PCA- MBPCA- and NLPCA methods.

A major drawback concerning PCA is non-linear data. The method is based on linear correlations between the data and this is not typical in the two processes being used. One approach to reduce this problem is the idea of virtual tags. Virtual tags are here based on mathematical functions of already measured data. The idea is based on process insight. If we

know the process well and the cause of nonlinearities some additional nonlinear functions could be incorporate to increase the performance of the PCA method. This idea will be tested for the centrifugal pump process.

Another approach to reduce the sensitivity of nonlinear data is the MBPCA. This method is based on analyzing the portion of the observed variance that cannot be predicted using a model of the process, and thus significantly enhances the attainable diagnostic resolution. This method will be tested for both processes.

The NLPCA is also a method that has been popular and has received increasing attentions for the last two decades. There are many sort of NLPCA and in this master thesis the NLPCA is based on autoassociative neural networks. This method is capable to handle linear and non-linear data. This method will also be tested on both processes.

As mention the methods will be tested on heat exchanger – and centrifugal processes. First there will be given a process description on both pump processes and heat exchanger and a process model are developed for both systems. All the results in this thesis will be based on model data and not real process data.

1.2 Problem description

The industry today is meeting an increased competition and safety demand and therefore having a top modern system to reduce downtime and discover severe faults before they cause dramatic consequences in a process is seen as a big advantage.

Early Fault and Disturbance Detection is a new software application developed by the combined efforts of ABB, Statoil and academic institutions. The overall goal is to support decision making: maintenance, repairs, production planning or optimization. The EFDD system consists of newly developed data-driven and generic process monitoring software [7]. State of the art fault detection algorithms are used for the purpose of monitoring a variety of processes, ranging from small tasks such as sensor integrity verification to plant-wide monitoring. There are several desirable attributes characteristics that an Early Fault and Disturbance Detection system should contain to work ideally [1]. In this master thesis the focus will be on early detection on different faults and not isolation of them. EFDD systems contains a module called Principal Component Analysis (PCA) (see chapter 13 in [11]) which is a linear analysis method used for condition monitoring. MBPCA [2] and NLPCA [20] are based on this method and would also be used on different scenarios. The use of PCA for EFDD system are being investigated on two processes, heat exchanger and pumps.

This master thesis investigates the PCA method in use for early fault and disturbance detection. The method will be tested on two processes, a plate heat exchanger and a centrifugal pump. The two processes are taken from an offshore installation. Several issues are presented to check the PCA - MBPCA - and NLPCA performance in these two processes. The main subjects are the lack of measurements, nonlinear- and non-stationary behaviour. Oil-and gas processes often have less measurements than is ideal. In these cases estimators or virtual measurements (measurements made from mathematical operators of existing measurements) could be incorporated. PCA will be compared to MBPCA with both linear and nonlinear models. PCA is a linear method and its ability to work with nonlinear data will be investigated. Normal PCA will also be compared to a Nonlinear PCA based on associative network.

1.2 Collection of PCA data

The PCA methods are data - driven and data must therefore be available. If the methods should be tested on real processes, two data sets are required. One set of data describing normal operation condition (NOC) taken from a process operating in a normal operating region. And one set containing known faults from the same process. It's often difficult to get access to this data from a running process. We also have to know when we have a potential fault data set and when we have a normal one. A high degree of process knowledge is therefore required.

To eliminate this problem it was used first principal models of the two processes that the PCA methods would be tested on. The advantage of using a model is the possibility to generate different kinds of faults that we want to test the system from and we have free access to manipulate different process parameters.

1.3 Description of plate heat exchanger

As mention in the problem description, one of the processes that are being used to investigate different PCA methods is a plate heat exchanger system. These processes are standard in oil –and gas industry and existing in several processes. The basis for the heat exchanger is a condensate cooling process. This is a plate heat exchanger and a flowchart of the process is given in figure 2. The process contains a control loop that controls the temperature on the condensate by manipulating the flow on the refrigerant side. A description of the measurements variables are listed in table 1. The heat exchanger is divided into two parts. HJ1 and HJ2. The reason for this kind of design could be many and will not be discussed further. In chapter 4 a model based on this process is made. The two parts HJ1 and HJ2 are modelled as one.

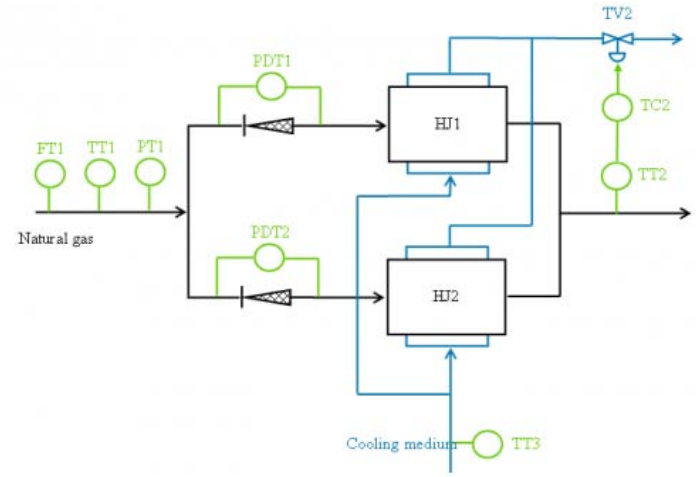


Figure 1 Flowchart plate heat exchanger pump (taken from an offshore installation)

Variable	Description
FT1	Inlet volumetric flow.
PDT1	Pressure drop across strainer.
PDT2	Pressure drop across strainer.
PT1	Gas inlet pressure.
TT1	Gas inlet temperature
TT2	Gas outlet temperature – of the joint cooler flow
TC2	Gas outlet temperature controller output, assumed to be identical to the valve position (TV2).
TT3	Cooling medium inlet temperature

Table 1 Measurement tags heat exchanger

Heat exchanger process is known to be nonlinear both to respect of flow and temperature. From the condensate cooling process in figure 3 there is a lack of output cold temperature and the flow on the cold side. Therefore when generating data from the heat exchanger model there is not being collected data from the cold side flow and cold side output temperature. So the basis data for PCA methods is the one in table 1 except for the pressures that are being neglected. To test the idea of about a priori knowledge, estimators are made for the missing flow and temperature on the cold side (see chapter 6.1 for details).

In a heat exchanger the problem of wax deposition is normal. The normal situation is that the wax deposition increases over the time the heat exchanger operates. This wax deposition affects the heat transfer coefficient and therefore reducing the efficiency of the heat exchanger. A detection of such a fault is wanted in an early state to be able to prepare maintenance. This is a considered as a fault and will be discussed further in chapter 5.

1.4 Description of centrifugal pump

The next process used for exploration of PCA methods is a centrifugal pump process. In this section a part of the condensate transfer process for export is investigated. Basis for the centrifugal pump is taken from an offshore installation that transferring condensate for export. In figure 2 a flowchart of the process is given and in table 2 a description of the measurements is presented. The process contains a control loop that controls the flow through the pump by manipulating the speed.

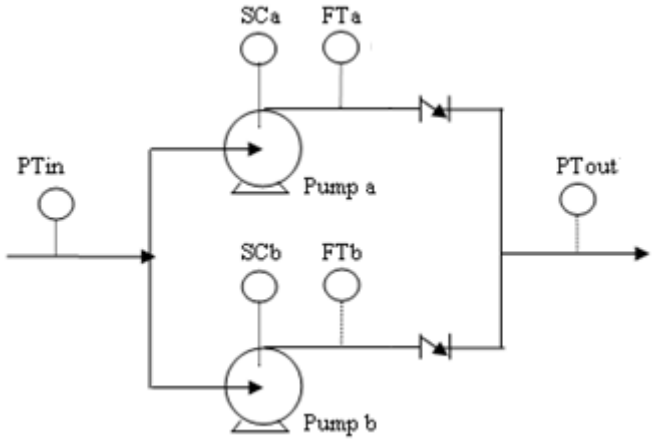


Figure 2 Flowchart Centrifugal pump (taken from an offshore installation)

Variable	Description
Ptin	Inlet pressure, common manifold
Sca/b	Output from the pump's speed controller (assumed to be equal to the actual speed)
FTa/b	The volumetric flow through the pump.
PTout	Common down stream manifold pressure

Table 2 Measurement tags Centrifugal pump

Also a centrifugal pump is known to be a nonlinear process. The main nonlinearity in the pump is the (near) quadratic pressure-flow and pressure-speed dependency. Only one pump is operating at a time therefore only one of the pump are modelled. For details of the pump model see chapter 6.2

1.5 Early Fault and Disturbance Detection

Early Fault and Disturbance Detection [EFDD] is a new software application developed by the combine efforts of ABB, Statoil and academic institutions. The overall goal is to support decision making: maintenance, repairs, production planning or optimization. This software utilizes both rule-based and pure data driven algorithms. The rule based method based on having a database with both nominal- and faulty process signatures. This method will demand more from the users then the data driven method in the sense that databases must be updated with new scenarios as they occur, but will increase its performance over time. The data driven method is opposite. A data driven approach will yield meaningful results upon installation, but the performance will not improve over time. This thesis will be focusing on the data-driven method.

1.5.1 Faults, failures, malfunctions

One of the main tasks of an EFDD system is to detect process abnormalities in a process and below follows a discussion of different sorts of process abnormalities.

“A fault is an unpermitted deviation of at least one characteristic property (feature) of the system from the acceptable, usual, standard condition” (page 9 in [20]).

Failure:

“A failure is a permanent interruption of a system’s ability to perform a required function under specified operating conditions” (page 9 in [20]).

Malfunction:

“A malfunction is an intermittent irregularity in the fulfilment of a system’s desired function” (see page 21 in [20]).

Figure 3 shows the relation of faults, failures and malfunctions. The fault may develop abruptly, like a step-function, or incipiently, like a driftlike function.

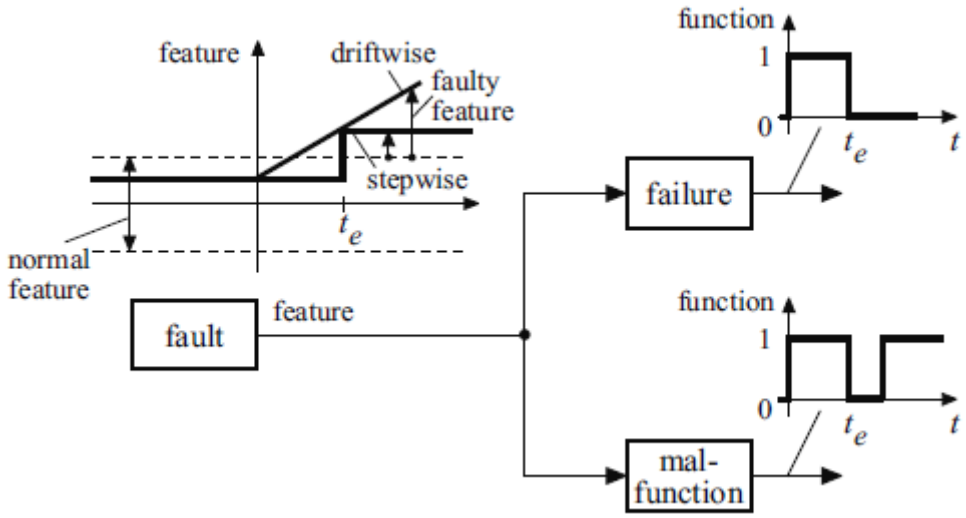


Figure 3 Development of the events "failure" or "malfunction" from a fault which causes a stepwise or driftwise change of a feature (taken from page 22 [10])

Faults may include

- Process unit failures
- Process unit degradation
- Control system failure
- Sensor failure
- Actuator failure
- Parameter drifts or grosses changes
- Operation beyond normal regimes

A more detailed discussion follows:

Process Unit Failure and Degradation

Structural changes result in changes in the information flow between different variables. This would obviously affect any models used in control.

Control, Actuator or Sensor Failure

Gross errors usually occur with sensors and actuators. These types of fault often propagate rapidly though a process due to a control system. Problems with sensors include measurement noise. Many sensor and actuator faults can be readily detected and remedied by means of a data validation system. This ensures that the data feed to afault detection system is valid.

Parameter Drifts or Gross Changes

These types of faults occur when an independent variable enters the system from the environment. This includes unusual process disturbances. This may affect control or operator decisions.

Operation beyond Normal Regimes

This is closely related to parameter changes as operating in a different process regime will result in grossly different relationships between variables.

Faults for analysis

The faults used in this report would be sensor failure, parameter drifts and operation beyond normal regimes. All this failure would be simulated for both processes. A more detailed description of the different faults is presented in chapter 5 and 6.

Chapter 2

PCA and extension

2.1 PCA Theory

Principle component analysis (PCA) is a powerful data driven method used in several applications, such as face recognition, image compression and predictive models for mention a few. The PCA method is one part of the EFDD system developed by Statoil and ABB and would be the main focus in this master thesis. PCA method task in EFDD system is used for detecting faults and disturbance in oil- and gas processes.

PCA is a non-parametric method that could effectively reveal hidden information of a process from sometimes confusing data sets (see chapter 13 in [9]). The idea behind PCA is to reduce the dimensionality of the original data set and still capture as much as possible of the variance present in the original data set. To achieve this goal the PCA is using projection technique based on linear algebra. It can project the original data set onto a lower dimensional space that contains the greatest variance of original data. This would give us a new set of variables; the principal components (PCs), which are uncorrelated to each other. This strength has become a popular pre-processing tool for multivariate statistical process monitoring (MSPM) schemes for detecting faults of multivariate process before they become critical.

However, PCA has some severe drawbacks for EFDD. Most processes in the real industry contain non-stationary, time-varying and nonlinear behaviour. These properties will often make a conventional PCA unsuitable, because PCA is a linear method and in addition, once static PCA is built, that means almost settled mean, variance, and covariance exists among variables. If a conventional PCA method is being used on these types of processes it may produce an excessive number of false alarms or miss faults.

2.2 PCA Calculation steps

First a data matrix $X \in \Re^{n \times m}$ that contains n samples of m process variables is collected under normal operation condition (NOC). The matrix X needs to be normalized to zero mean and unit variance. With the scale parameter vectors $\bar{x}$ and $s$as the mean and variance vectors respectively. The next step is to construct the covariance matrix R:

$$\underline{R} = \frac{1}{1-n}\underline{X}^T\underline{X} \tag{1}$$

Where X is the data matrix and then solve the eigenvalues λ_j of the matrix $\underline{R}$ and the eigenvectors p_j of

$$\left(\underline{R} - \lambda_j \, \mathrm{I}\right)p_j = 0 \quad j = 1, \dots, m \tag{2}$$

The eigenvectors with the corresponding most dominated eigenvalues are those of interest. This would explain most of the variance of the process. The transformation matrix $P \in \Re^{m \times a}$ is therefore generated choosing a eigenvectors or columns of p_j corresponding to a principal eigenvalues. Matrix P transforms the space of the measured variables into the reduced dimension space.

$$\underline{T} = \underline{X}\,\underline{P} \tag{3}$$

The columns of matrix P are called loadings and elements T are called scores.
Scores are the values of the original measured variables that have been transformed into the reduced dimension space. If all eigenvectors are selected the original data set could be transformed by:

$$\underline{X} = \underline{T}\,\underline{P}^T \tag{4}$$

If some eigenvectors are been neglected the original data space can be calculated as:

$$\underline{X} = \underline{T}\,\underline{P}^T + \underline{E} \tag{5}$$

,Where $\underline{E} = \underline{X} - \underline{\hat{X}}$. $\underline{T}\,\underline{P}^T$ is called the structure part and E is called the noise part.

This separation is done to reduce the dimensionality of a data set. Considering a large number of interrelated variables, while retaining as much as possible of the variation present in the original data set. This variation is captured in the T and P and is achieved by transforming the measured data to a new set of variables, the principle components, which are the most of the variation present in all of the original variables, ([9] chapter 13).

There are other methods for calculating the PCA [23]. Non-linear Iterative Partial Least Square method (NIPALS algorithm) and alternating least Squares Method (ARL algorithm) where the solution is (randomly) rotated in the factor space. These methods are much more computationally expensive and have little or no advantage over the covariance method and are not used here.

PCA algorithm:

1) The first step in a PCA is to put the input data in a matrix X.
   Let X = $[x_1\ x_2\ \dots\ x_m] \in R^{nxm}$ be the data matrix with n samples and m variables.

2) Find the mean for each column and subtract it from the corresponding columns.

3) Calculate the eigenvalues and eigenvectors for the autocorrelation matrix $A = 1/(n-1)X^T X$.

4) Choose how many eigenvalues to keep and set up the respective eigenvectors for these eigenvalues in a transformation matrix P = $[p_1\ p_2\ \dots\ p_r]$

5) Calculation of the new data matrix. $T = X'P[t_1\ t_2\ \dots\ t_r]$ The result data matrix T with all original data but a reduced number r<m of coordinates or variables, i.e. the principal components.

2.3 Select number of principal components

One of the main difficulties in using principal component analysis (PCA) is the selection of the number of principal components (PCs). If too few principal components are selected than required, a poor model will be obtained and an incomplete representation of the process will result. Different approaches to select PCs have been proposed in the past: (1) Akaike information criterion (AIC), (2) minimum description length (MDL), (3) imbedded error function (IEF), (4) cumulative percent variance (CPV), (5) scree test on residual percent variance (RPC), (6) average eigenvalue (AE), (7) parallel analysis (PA), (8) autocorrelation (AC), (9) cross validation based on the PRESS and R ratio, and (10) variance of reconstruction error (VRE). All these methods are discussed and evaluated in [12].

One of the most commonly used criteria for solving the number-of –components problem is the average eigenvalue method [15]. This criterion accepts all eigenvalues with values above the average eigenvalue and resets those below the average. The reason is that a PC contributing less than an "average" variable is insignificant. For covariance-based PCA the average eigenvalue is 1/m trace(covariance matrix),and for correlation-based PCA the average eigenvalue 1/m trace(correlation matrix), which is 1. Then all the eigenvalues above 1 will be selected as the principal eigenvalues to form the model.

The average eigenvalue method has a number of positive features that have contributed to its popularity. Perhaps the most important reason for its widespread use is its simplicity: You do not make any subjective decisions, but merely retain components with eigenvalues greater than one. It also has a positive side when small to moderate number of variables are being analyzed and the variable communalities are high. One of its drawbacks is that it could lead to retaining a certain number of components when the actual difference in the eigenvalues of successive components in only trivial. In short, the average eigenvalue method can be helpful when used judiciously, but the thoughtless application of this approach can lead to serious errors of interpretation.

Based on the simplicity and advantages when small to moderate numbers of variables are being analyzed the average eigenvalue method is here being used to give a hint of number of PCs. Since the number of variables used here are few, the PCA models would also be tested build on be PCs that are nearly 1 (0.9-1).

2.4 Model-based PCA

In model-based PCA (MBPCA) is an extension to PCA and used the portion of the observed variance that cannot be predicted using a model of the process, and thus significantly enhances the attainable diagnostic resolution. This method has shown good result in some applications such as Ethylene compressor monitoring [7] and Fault detection in NMOS Fabrication [2]. As described in chapter 2.4 the PCA method has some drawbacks do to nonlinearities.

The idea behind the MBPCA is to remove nonlinear data before given to PCA. The method makes use of models of the real system that is being monitored and subtracts the data made from the model from the real measurement process. In theory, when the model is perfect, the data not predicted by the model would be relatively insensitive to variations caused from non-linearity of the process or changes in operating conditions, thus significantly enhancing process diagnosis. However, since the model in use most likely consists of parametric and structural uncertainty the method is vulnerable. In figure 4 a example of MBPCA structure is presented.

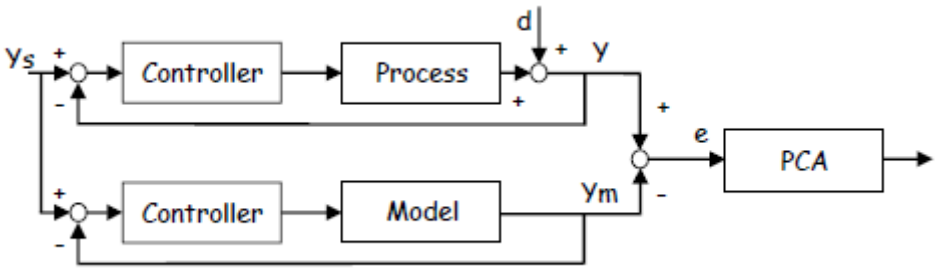


Figure 4 Block diagram illustrating the MBPCA principle (Taken from [7]).

2.5 Kernel based PCA

Another method for dealing with nonlinear data is a method called Kernel based PCA (KPCA). KPCA perform a nonlinear form of principal component. The only differance is that instead of calculate PCA on the data in space R^n it first map the data into a feature space F by some nonlinear mapping ϕ and then do PCA in space F [21].

2.6 Nonlinear PCA based on auto-associative neural network

As mentioned PCA is a method for dimensional reduction and, to work ideally, linear data is required. To be able to confront linear and nonlinear data, a nonlinear PCA (NLPCA) based on auto-associative neural networks may sometimes be preferable. Kramer [x] proposed such a method. This method is based on using an artificial neural network (ANN), which, in essence, is an identity mapping that consists of a total of 5 layers (see figure 2) for training procedures to generate nonlinear features. The networks are of a conventional type, featuring feedforward connections and linear or sigmoidal nodal transfer functions, trained by backpropagation.

NLPCA based on AA network algorithm

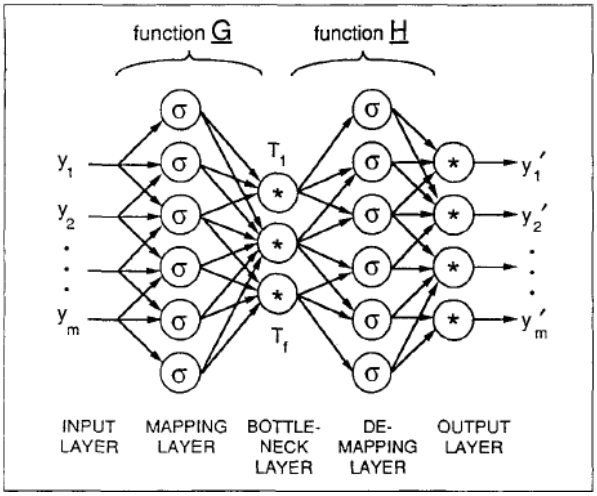


Figure 5 Network architecture for simultaneous determination of f nonlinear factors using an autoassociative network (Taken from [kramer])

In NLPCA, the mapping into feature space is generalized to allow arbitrary nonlinear functionalities. By analogy to Eq. (2.3), we seek a mapping in the form:

$$\underline{T} = \underline{\mathrm{G}}(\underline{\mathrm{X}}) \qquad (6)$$

Where $\underline{\mathrm{G}}$ is a nonlinear vector function, composed of f individual nonlinear vector functions; $\underline{\mathrm{G}}=\{G_1, G_{2,\ldots,} G_f\}$, analogous to the columns of P in Eq. (2.3), such that if $\mathrm{T_i}$represents the ith element of $\underline{T}$

$$\underline{\mathrm{T_i}} = \underline{\mathrm{G_i}}(\underline{\mathrm{X}}) \qquad (7)$$

By analogy to the linear case, $\mathrm{G_1}$ is referred to as the primary nonlinear factor, and $\mathrm{G_i}$is the ith nonlinear factor of $\underline{Y}$. The inverse transformation, restoring the original dimensionality of the data, analogous to Eq. (2.4), is implemented by a second nonlinear vector function $\underline{\mathrm{H}}=\{H_1, H_{2,\ldots,} H_m\}$:

$$X'_i = H_j(\underline{T}) \tag{8}$$

The loss of information is again measured by $\underline{E} = \underline{X} - \underline{X}'$, and analogous to PCA, the functions $\underline{G}$ and $\underline{H}$, a basis function approach is used has shown that functions of the following form are capable of fitting any nonlinear function $\underline{v} = f(\underline{u})$ to an arbitrary degree of precision:

$$v_k = \sum_{j=1}^{N_2} w_{jk2}\sigma\left(\sum_{i=1}^{N_1} w_{ij}u_i + \theta_j\right) \tag{9}$$

Where $\sigma(x)$ is any continuous and monotonically increasing function with $\sigma(x) \to 1$ as $x \to +\infty$ and $\sigma(x) \to 0$ as $c \to -\infty$. A suitable function is the sigmoid:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{10}$$

2.7 Adaptive PCA

All the different additions to PCA described so far have varied capability to handle nonlinearities, but they all require an extent of stationary data. As described in the theory of PCA, the model is made from defined normal operation conditions. From this interval a mean of the different variables is being calculated and later used when running online. If the mean changes too much from the NOC the PCA model would not be valid and may generate a false alarm in the use of early fault detection. To cope with this problem, the PCA must adapt the monitoring model in a continuous and automatic manner in compliance with present process condition. A rule based online recursive PCA (RPCA) is proposed in [3]. This method may be suitable for slow time-varying properties.

Chapter 3

PCA in Early fault detection

3.1 PCA Monitoring

To implement a monitoring and fault detection system based on PCA, two main tasks are necessary:

(a). OFF-LINE.

Gather a decided number of training data from a normal operation conditions (NOC). Then scale the training data by subtracting the mean $\bar{x}$ of the samples of each variable and divide it with the standard deviation s. Then find the eigenvalues and corresponding eigenvectors of the covariance matrix. Determine the number of principal components and the upper control limits for T^2 and Q statistics.

(b). ON-LINE

a) Obtain the next testing sample x, and scale it using the scale parameter vectors $\bar{x}$ and s.

b) Evaluate the $T^2$and Q statistics using the obtained PCA model. If one of these exceeds the upper limits, this measurement is considered an alarm. If there are some consecutive established numbers of alarms, an uncommon event has occurred.

c) Repeat from step 2.

When using PCA for early fault detection the use of multivariate control charts using Hotelling's T^2 and square prediction error (SPE) is most common [22]. Hotelling's T^2 that is a measure of the variation within the PCA model and SPE measure of lack of model fit. T^2 is the sum of normalized squared scores defined as:

$$\mathrm{T_i^2} = \mathrm{t_i}(\mathrm{T_k^T T_k})^{-1}\mathrm{t_i^T}$$

In this case $\mathrm{t_i}$ refers to the i^{th} row of T_k, the matrix of k scores vectors and the term in parentheses is a diagonal matrix of eigenvalues of the covariance matrix of X. The SPE statistic indicates how well each sample conforms to the PCA model and is a measure of the amount of variation not captured by the model and is defined as:

$$\mathrm{SPE_i} = \mathrm{e_i e_i^T} = \mathrm{x_i}(\mathrm{I} - \mathrm{P_k P_k^T})\mathrm{x_i^T}$$

Where $\mathrm{e_i}$is the i^{th} row of E. The columns of $\mathrm{P_k}$are the first k loadings vectors retained in the PCA model and I is the identity matrix of appropriate size (n by n)

This report will analysis SPE for process monitoring and the reason is that the SPE may have lower changes to give false alarms when monitoring the processes being investigated [13]. Both the heat exchanger- and the pump process are hardly normally distributed and stationary and since T^2 captures the non-stationary parts of the signals and is more likely to give an alarm then the SPE chart how looks much more stationary and random. The structure of the online process after defining PCA models is given in the figure 6.

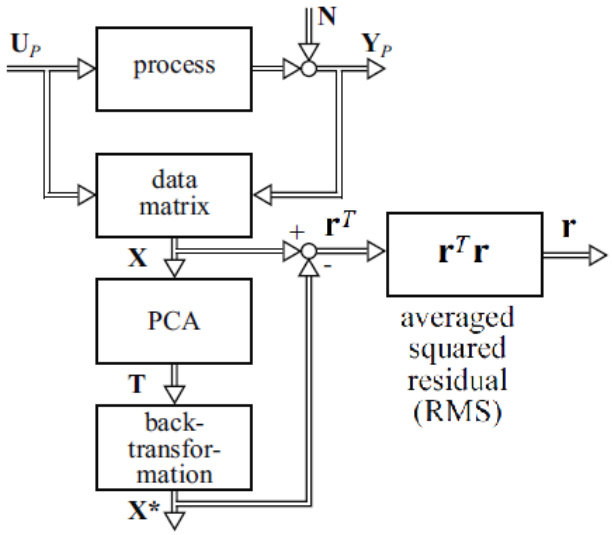


Figure 6 Structure for PCA for fault detection

3.2 Upper control limit

When monitoring SPE we need some assessment to the probability of fault. A normal situation is then to define an upper control limit (UCL). The value of this limit would depend on the process but often it's defined from the variance of the process. A popular rule is the Three-sigma rule [19]. The Three-sigma rule is applied to evaluate the contribution of the probability of fault to the predicted risk. It is known that, for a normal distribution, nearly all (99.7%) the values lie within 3 standard deviations of the mean. The UCL is therefore sat to 3 times the standard deviation of the SPE values. In addition to the UCL an indicator value for fault status could be used.

Since some values SPE values would violate the UCL a defined value of legal violations in a specific interval could be technique. This would then be a tuning parameter and depend on the process being monitored.

3.3 Implementation of PCA and NLPCA

The PCA method used in simulation chapter 6 is based on algorithm in chapter 2.2. Both the PCA algorithm and SPE plots are made in MATLAB script. The NLPCA method used are based on auto associative network and taken from a Nonlinear PCA toolbox for MATLAB [17].

Chapter 4

Models

As described earlier a plate heat exchanger and centrifugal pump are processes that are being used to investigate PCA performance in an EFDD system. In this chapter models and estimators on the two processes are given. These models will be used to simulate different kinds of faults and normal operation conditions that will give the PCA – MBPCA and NLPCA data. The models are implemented in MATLAB and SIMULINK.

4.1 Dynamic and static heat exchanger models

The heat exchanger process contains a dynamic model of the heat exchanger and a control loop that control loop that controls the output temperature on the hot-side by manipulating the cold side flow. The control loop is modelled by a PI – controller and a valve. In some of the experiments there would be used estimates for assumed missing measurements. These are based on different types of estimates concerning complexity.

A heat exchanger process can be described by physical laws concerning mass, energy and momentum. The process is known to be nonlinear and the main reason for the nonlinearity is the strong mass flow dependence and even temperature dependence of the heat transfer coefficients. The model being used here is based on what may be called “Direct lumping of the process” and taken from [5]. This means that the heat exchanger is divided into sections (cells), and a model for each section is computed and put in a network to form the overall heat exchanger (see figure 7).

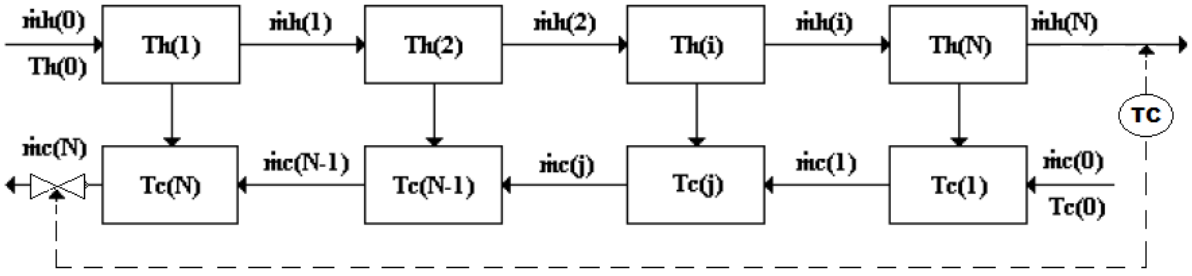


Figure 7 Elements in a plate heat exchanger with control loop

Process Assumptions:

- No pressure drops form inlet to outlets.
- Even mass distribution along the heat exchanger.
- Heat exchange with the ambient is ignored.
- $T_{ci},\ T_{hi},\ w_h$ are measurements from real process data, and w_c based on assumption of the pressure difference over the valve.
- Assumed densityρ_{hi} and p_{co}.
- The metal element below the two sections is ignored.
- PI controller with assumed structure and parameters.
- Assumed valve characteristic.

Models parameters are:

U	Total heat coefficient between the streams $[W/(m^2K)]$
h_h	Heat transfer coefficient at condensate (hot) side $[W/(m^2K)]$
h_c	Heat transfer coefficient at refrigerant (cold) side $W/(m^2K)]$
A	Heat exchanger area for heat conduction between the two stream $[m]$
$c_{p,h}$	Specific heat capacity for the mass at the condensate (hot) side $[J/(kg\ K)]$
$c_{p,c}$	Specific heat capacity for the mass at the refrigerant (hot) side $[J/(kg\ K)]$
ρ_h	Density condensate (hot) side $[kg/m^3]$
ρ_c	Density refrigerant (cold) side $[kg/m^3]$
C', y	Constant
k	Valve constant

Input data are:

m_h	The mass of the liquid at the condensate side $[kg]$
m_c	The mass of the liquid at the refrigerant side $[kg]$
w_h	Mass flow at the condensate side $[kg/s]$
w_c	Mass flow at the refrigerant side $[kg/s]$
$T_{h,i}$	Inlet temperature at the condensate side [℃]
$T_{c,i}$	Inlet temperature at the refrigerant side [℃]
p_{co}	Pressure before valve [bar]
p_{ct}	Pressure after valve [bar]
u	Output from controller

Output data are:

$T_{h,o}$	Outlet temperature at the condensate side [℃]
$T_{c,o}$	Outlet temperature at the refrigerant side [℃]

4.1.1 Dynamic model of plate heat exchanger

The heat exchanger is modelled using the two model equations (11) and (12) from [5].

$$\frac{dT_{h,j}}{dt} = \frac{w_h \Delta T_{h,j}}{m_{h,j}} - \frac{Q_j}{m_{h,j} c_{p,h,j}} \tag{11}$$

$$\frac{dT_{c,j}}{dt} = \frac{w_c \Delta T_{c,j}}{m_{c,j}} + \frac{Q_j}{m_{c,j} c_{p,c,j}} \tag{12}$$

$\Delta T_{h,j}$ and $\Delta T_{c,j}$ describes the temperature difference between the inlets and outlet on hot side and inlet and outlet on cold side respectively. j is the element index and are here set to 10. (10 element plate heat exchanger).

The coupling term Q_j is equal (13) and describes the energy transfer between the two sections .

$$Q_j = U_j A_j \Delta T_{lm,j} = w_h c_{p,h,j} \Delta T_{h,j} = -w_c c_{p,c,j} \Delta T_{c,j} \tag{13}$$

Where A is the Heat exchanger area for heat conduction between the two streams, U Coefficient of total heat conduction between the streams and the $\Delta T_{lm,j}$ term is the logarithmic mean temperature differences between the two sections and are given in (14). The heat conduction area is set to $0.5\ m^2$.

$$\Delta T_{lm,j} = \frac{\left(\Delta T_{1,j} - \Delta T_{2,j}\right)}{log\left(\Delta T_{1,j} / \Delta T_{2,j}\right)} \tag{14}$$

, where $\Delta T_{1,j} = T_{h,j,in} - T_{c,j,out}$ and $\Delta T_{2,j} = T_{h,j,out} - T_{c,j,in}$

The nonlinearities in a heat exchanger come from the strong massflow dependence and temperature dependence of the heat transfer coefficients. The model is here only based on the massflow dependence. To include this in the equations empirical relations from [5] has been used see (15).

$$Nu = \frac{hd}{k} = CPr^x Re^y = C\left(\frac{c\mu}{k}\right)^x \left(\frac{ud\rho}{\mu}\right)^y \tag{15}$$

By neglecting the temperature dependency and after some rearranging, (15) becomes.

$$h = C' \dot{m}^y \tag{16}$$

By assuming that equation (16) applies to both of the heat transfer coefficients, h_h and h_c, the overall heat transfer coefficient, U can be written as. The thermal resistance in the metal is here neglected.

$$U(t) = \frac{h_h(t)\, h_c(t)}{h_h(t) + h_c(t)} = \frac{C'\left(\dot{m}_h(t)\dot{m}_c(t)\right)^y}{\left(\dot{m}_h^y(t) + \dot{m}_c^y(t)\right)} \qquad (17)$$

C' and y are assumed to be the same for both h_h and h_c. This is done to keep the number of parameters that is later being assumed at a minimum.

Usually heat transfer coefficients are based on a distributed heat exchanger model. Therefore heat transfer coefficient used in lumped models U from equation (17) must be increased to give the same overall effectiveness when using few elements [8]. The constant y are assumed to be 0.67, and C' was adjusted to 5540 $W/(m^2K)$ to reach the same temperature output on hot side as for real data taken from an offshore installation.

4.1.2 PI-controller

The PI controller was made from the equation (18).

$$\mathrm{u} = \mathrm{Kp} + \mathrm{Kp/Tis} \qquad (18)$$

The parameters Kp and Ti is sat to 0.005 and 50 respectively. These are parameters selected by the author.

4.1.3 Valve model

The next model used in the plate heat transfer process is the valve. The valve equation used is given in equation (19)

$$w_c = ku\rho_c\sqrt{p_{co} - p_{ct}} \qquad (19)$$

The pressure before the valve p_{co} is simulated with a mean value of 1 bar. And the valve constant k is assumed by the author to be 0.3×10^{-3}. The density is sat to ρ_c 745 kg/m^3. u is the output from the PI – controller. To make the output pressure dependent on the valve opening the pressure p_{ct} was set as a constant multiplied by the output from the PI-controller.

In the equation given so far there are used assumptions about different constants.

Some of the constant are selected from trial and error from real measurements on an offshore installation. The constant most likely does not fit to a real heat exchanger.

4.1.4 Estimators of missing measurements

To support the PCA - and NLPCA method with more measurements, estimators was implemented. From the heat exchanger process there are no measurements of the flow on refrigerant liquid, but since we are controlling the refrigerant liquid we have access to the output u from the controller output. Based on the output u and assumption of a valve model we could estimate the flow.

4.1.5 Flow estimate cold side

The valve estimate was based on the same structure as the valve that serves as the real.

$$\hat{w}_c = \hat{k} u \hat{\rho}_c \sqrt{\hat{p}_{co} - \hat{p}_{ct}} \tag{20}$$

The constant $\hat{k}$ is here set to 0.25×10^{-3} and the pressure difference $\hat{p}_{co} - \hat{p}_{ct}$ is sat to 0.7 bar. The density is set to 1000 $kg/m^3$and the density $\hat{\rho}_c$ are assumed constant and equal 1000. u is the output from the PI - controller. The output from the controller is u and the reason why it's squared is to simulate an equal percentage characteristic.

4.1.6 Static estimate temperature cold side

A static estimation of the outlet temperature on the cold side is made to give more data to PCA methods. The estimate is based on equation (11) at steady state and the equation (13), and the estimate from (20).

$$\hat{T}_{c,o} = T_{c,i} + \frac{c_{p,h}}{c_{p,c}} \frac{w_h}{\hat{w}_c} - (T_{hi} - T_{ho}) \tag{21}$$

4.1.7 Estimate heat transfer coefficient

To increase the data quantity to the PCA methods further it's also made estimates on the heat transfer coefficient. These estimates are based on the estimates of flow and cold side output temperature and also the available measurements.

$$\hat{I}_h = \frac{UA}{c_{p,h}} = \frac{w_h \Delta T_h}{\Delta T_{lm}} \tag{22}$$

$$\hat{I}_c = \frac{UA}{c_{p,c}} = \frac{\hat{w}_c \Delta T_c}{\Delta T_{lm}} \tag{23}$$

4.1.8 Estimate temperature output hot side

To be able to use the MBPCA approach there was used a model of the hot side output temperature. The estimate use the same model equations as the model that serves as the real system eq. (11) and (12), but instead of using the logarithmic mean temperature differences between the two sections given in (14) the linear equation (24) was used.

$$\Delta T_{lm,j} = \frac{T_{h,in} + T_{h,out}}{2} - \frac{T_{c,in} + T_{c,out}}{2} \tag{24}$$

In chapter 6.1 there are used three different models when using MBPCA. The structure are the same but with some change in the values. The difference comes from equation (16). They all uses C' equal to 5400 $W/(m^2K)$ (a change from the real model by 150), but with different y.

MBPCA linear uses y =1

MBPCA incorrect non-liner y = 0.8

MBPCA correct non-liner y = 0.67

4.2 Implementation of heat exchanger in SIMULINK

As mentioned in the intro the models are simulated in MATLAB and SIMULINK. The model generating the data is simulated in SIMULINK and the PCA models are made in MATLAB code. In the figure 8 an overview of the model that serves as the real process and the estimators are given.

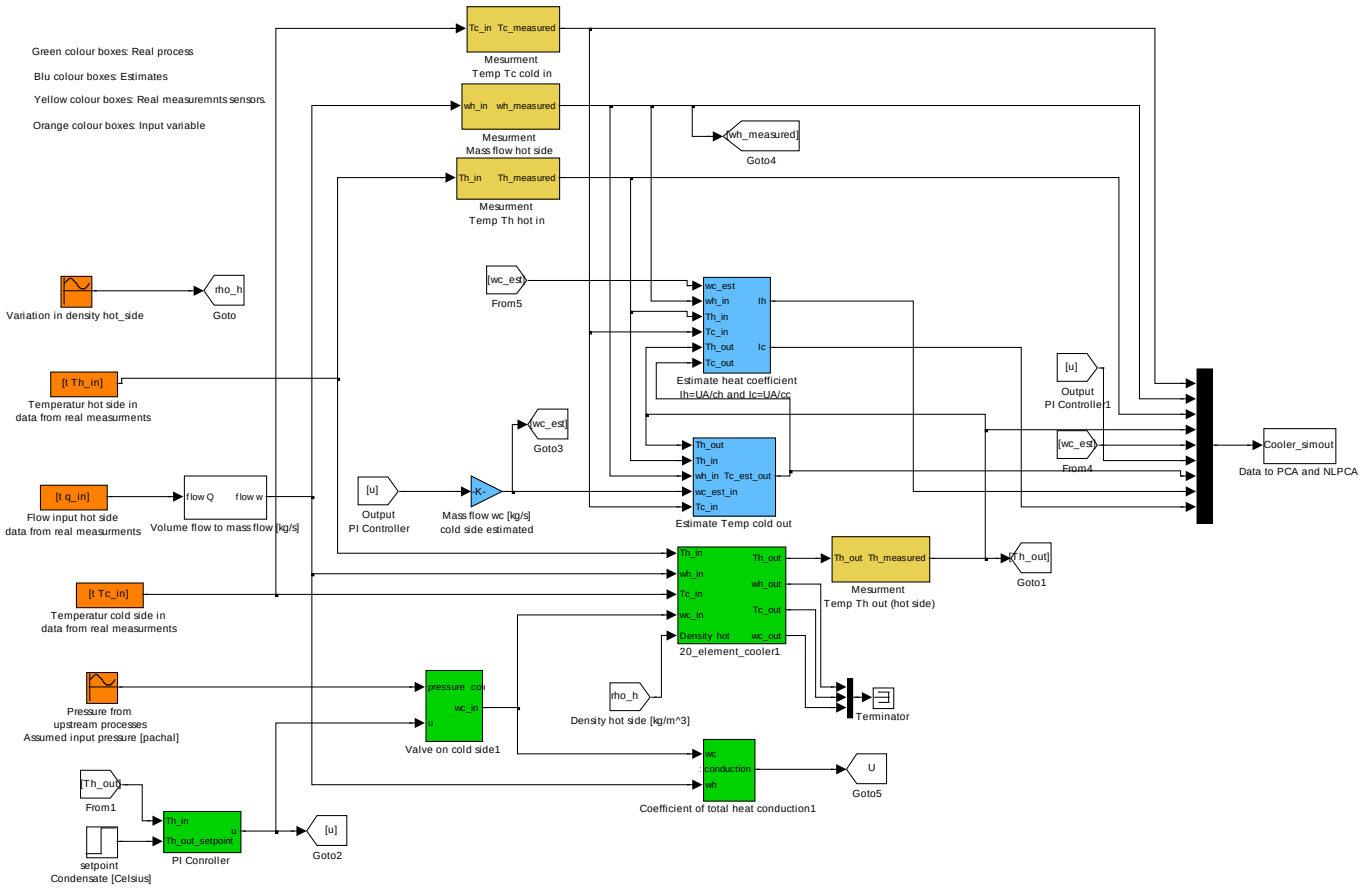


Figure 8 SIMULINK plate heat exchanger process for training data

4.2 Static centrifugal pump model

As for the heat exchanger process we need a process model of the centrifugal pump. The Centrifugal pump are modelled using a static equations made from a polynomial from [11]. There is also given a model of a control loop consisting of a PI – controller and an estimate of the flow for use in the MBPCA.

Centrifugal pump model

MODEL ASSUMPTION:

- Neglect pump dynamics
- Neglect the pressure losses by the casing of the pump

The model is made from an approximating polynomial derived from the Euler pulse moment equation and can be represented as the following:

$$p = k \cdot p_E - p_{HL} - P_D \tag{24}$$

where

p Pressure differential across the pump

k Correction factor. The factor is introduced to account for dimensional fluctuations, blade incongruity, blade volumes, fluid internal friction, and so on. The factor should be set to 1 if the approximating coefficients are determined experimentally.

P_E Euler pressure

p_{HL} Pressure loss due to hydraulic losses in the pump passages

p_D Pressure loss caused by deviations of the pump delivery from its nominal value

The Euler pressure can be approximated with the equation

$$P_E = \rho_{ref}\left(c_0 - c_1 \cdot q_p\right) \tag{25}$$

where

ρ_{ref} Fluid density [kg/m^3]

c_0, c_1 Approximating coefficients.

Q_P Pump volumetric delivery $[m^3/s]$

The pressure loss due to hydraulic losses in the pump passages, p_{HL}, is approximated with the equation

$$p_{HL} = \rho_{ref} \cdot c_2 \cdot q_p^2 \tag{26}$$

where

ρ_{ref} Fluid density $[kg/m^3]$

c_2 Approximating coefficient

q_P Pump volumetric delivery $[m^3/s]$

The pressure loss, p_D, is estimated with the equation

$$P_D = \rho_{ref} \cdot c_3 \left(q_D - q_p\right)^2 \tag{27}$$

where

ρ_{ref} Fluid density $[kg/m^3]$

c_3 Approximating coefficient

q_P Pump volumetric delivery $[m^3/s]$

q_D Pump design delivery (nominal delivery) $[m^3/s]$

The resulting approximating polynomial takes the form:

$$p = \rho_{ref} \left(k\left(c_0 - c_1 q_p\right) - c_2 q_p^2 - c_3 \left(q_D - q_p\right)^2\right) \tag{28}$$

Since the pump model will be simulated with different speed and density the affinity laws are used. This law makes it possible to establish pump characteristics for several velocities and densities given characteristics for one specific velocity and density. The affinity laws are said to be valid when the velocity is between 50 % to 120 % of the specific velocity n_{ref} [13].

$$q_{ref} = q \frac{n_{ref}}{n} \tag{29}$$

where q and ω are the instantaneous values of the pump delivery and angular velocity. Then the pressure differential across the pump at a different angular velocity and density is determined with the formula

$$p = p_{ref} \cdot \left(\frac{n}{n_{ref}}\right)^2 \cdot \frac{\rho}{\rho_{ref}} \qquad (30)$$

where p_{ref} is the pressure differential computed with equation (3.18) at pump delivery determined according to equation (3.19). The unknown coefficients are calculated from a pump curve given in appendix 1.

Control loop

The process contains a PI-controller that controls the flow thorough the pump by manipulating the speed. The PI – controller is taken from the Simulink library. The P- and I parameter are set by the author.

Flow estimate

Since MBPCA should be tested on the pump an estimate of the flow is made. This estimate will be dependent on the velocity and the pressure across the pump. The estimated also uses the affinity law from equation (3.19)

$$\hat{q}_p = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \qquad (31)$$

Where

$$a = -\left(c_{2\_est} + c_{3\_est}\right)$$

$$b = 2c_{3\_est}q_D - kc_{1\_est}q_p$$

$$c = kc_{0\_est} - c_{3\_est}q_D^2 - \frac{p}{\rho_{ref}}$$

To calculate the unknown coefficients c_0, c_1, c_2, and c_3 a pump characteristics is used for a specific fluid and a specific angular velocity. See appendix A for details

4.2.1 Implementation

All the process models are implemented in SIMULINK to form the pump process.

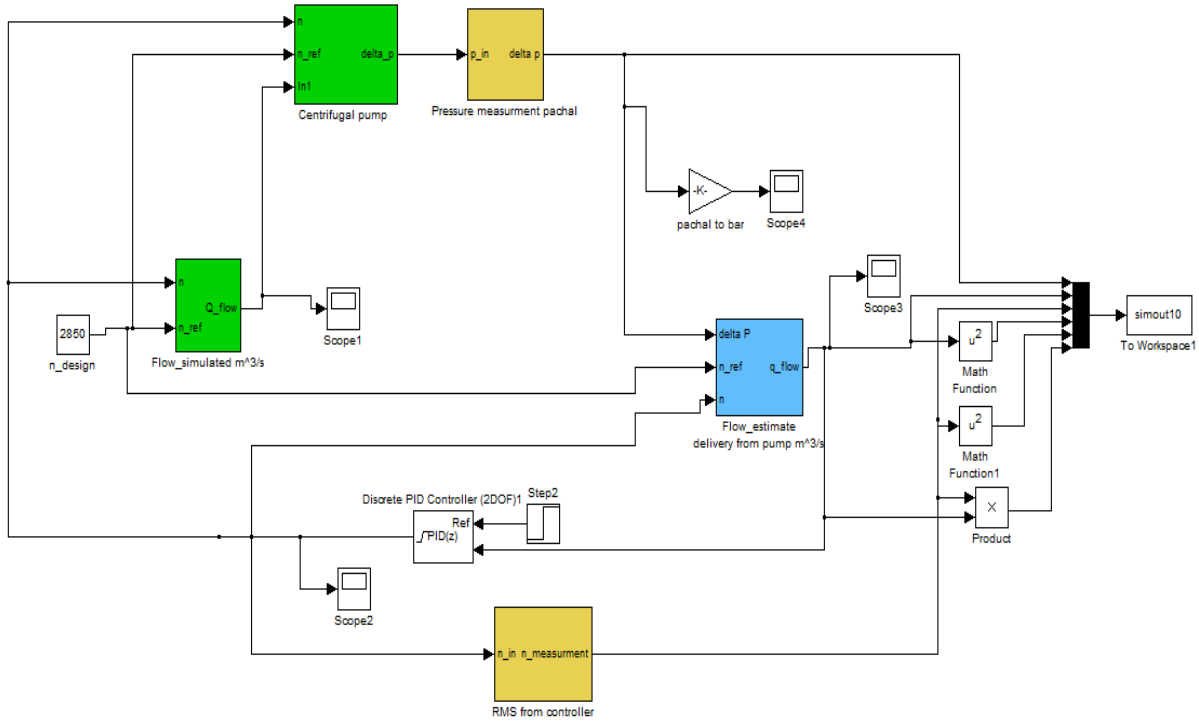


Figure 9 SIMULINK picture of Centrifugal pump process

Chapter 5

Experimental Setup

As mentioned in chapter 2 a plate heat exchanger- and a centrifugal pump are being monitored for early fault detection using PCA, MBPCA and NLPCA. The two processes both contain nonlinear behaviour that would be adequate for verifying the performance for the different method due to non-linear data. There would be performing to experiment series first on heat exchanger (see chapter 6.2) then on centrifugal pump (see chapter 6.3). Before the two simulation experiments series are taken place a setup description for both processes are given.

5.1 Simulation experiment heat exchanger

For the heat exchanger process there are 4 sections of simulations are taken place. In each section there is used three PCA -and one NLPCA model based on different number of measurements. The last three of the section contains experiment with MBPCA and the first one without MBPCA.

In the first section two of the PCA - and the NLPCA models use linear estimates datasets in addition to all other available data.

In the second section there is used the same measurements as for the first section except for the output hot temperature measurement. These measurements are used to form a linear MBPCA structure using the residual from a dynamic linear estimate subtracted from the real measurement.

The third section contains the same data setup as in section two, but here a nonlinear estimate of the output hot temperature is used. The nonlinearities are incorporated into the heat transfer coefficient, but are not of same character as the real one.

The fourth sections are the same as the third, but here the incorporated nonlinear heat transfer coefficient are exact the same as the real one.

For a summary of the methods and the dataset used see table 3

5.1.2 Generation of training data heat exchanger

The Normal Operating Training Dataset

Before simulation experiments are executed the PCA, MBPCA and NLPCA needs dataset for training. For accurate fault detection it is important that the processes have been richly excited over the period that the training data has been captured. This could be a major problem when training with data from real processes, due to change in production, operator's interference and fluctuating upstream and downstream- processes would frequently change the process behaviour. Since models of real processes are being used in this work the author has defined a normal operating region for the heat exchanger model (see table 4 for details). The normal operating training dataset is from now on called normal operation conditions (NOC).

Before forming the PCA- and NLPCA models a 23.5 hour simulation is executed. From this interval a sample is taken every minute giving total of 1400 samples for each variable. To remove noise from the signal and hopefully give more comfortable SPE charts and reduce calculation time for PCA and NLPCA the 1400 samples are divided into groups of 7 samples, giving totally 200 groups of samples. Instead of using the original samples the average in each of the grouped samples are used.

The number of samples in each group is a tuning parameter and could mask some faults if selected to height. Another consequence of the manipulation is to do with the online monitoring. By manipulating the data there will be a delay in the analysis because the group of samples needs to be filled before a SPE plot for the group is calculated. These are important considerations that have to be evaluated for each process being monitored.

In the heat exchanger experiments the input temperature variable and the hot side flow are data taken from a real offshore installation. All other input variables are simulated by the author. For a summary of the data inputs and their values see table x

5.1.2 Heat exchanger fault data set

In this section a discussion of different faults simulated in chapter 6 is presented. All the faults are simulated by adding on some bias step function to the measurements. The fault is triggered after the simulation has started and removed before it ends. All of the fault dataset used are the same as for the training dataset except for the fault simulated.

Sensor fault dataset

Sensor fault is a typical fault in the process industry and an early detection of such a fault is of usage interest. In the heat exchanger process there will be simulated a faults in the temperature measurement and cold side input.

Wax composition fault dataset

In heat exchanger process wax composition is a normal fault that would influence the efficiency of the exchanger capability to transport heat. The normal situation is that the wax composition increases over time. To detect such an error is valuable and could give maintenance personal information for planning maintenance stop. Since this type of fault typically evolves slowly over time a step function would not be very realistic, but will still be adequately for testing the different PCA methods

Novel Fault Data set

Novel means that the algorithm has not trained with the data for fault regions. The novel dataset does not need to be a fault, but only a change in production or other natural process drifting that was not present in the training dataset. Detecting such a fault would therefore be of varying interest depending on the process being monitored. For the heat exchanger process the novel fault dataset contains a change in process input flow on hot side.

5.2 Centrifugal pump training data set

For the centrifugal pump process two simulation experiments series are taken place.
The first one contains models build from a NOC with normal instrumentation and some virtual measurements. And the second series contain models build from a NOC where a MBPCA approach is used. The same measurements are used except for the flow measurement that is a residual of the flow (measurement subtracted with a flow estimate). (MBPCA approach). Both the two parts containing 3 different PCA models made from different numbers of datasets and one NLPCA model made of one dataset (see table 17). Both the NOC series operates in the same interval.

All the Centrifugal pump data variables used in the training dataset are defined by the author. In table 6 the range of values for the different variables are described.

5.2.1 Generation of training data heat exchanger

Before forming the PCA- and NLPCA models a 24 hour simulation is consistent to collect NOC data. From this interval the sampling rate is 1 minute giving a total of 1440 samples for each variable. To remove noise from the signal and thereby giving more comfortably SPE plots and in the same time reduce calculation time for PCA and NLPCA the 1440 samples are divided into groups of 10 samples giving 144 groups of samples. Instead of using the original samples the average in each of the grouped samples are used. The number of sample groups selected is a tuning parameter and will affect the sensitivity of fault detection on the two methods. Another problem with the data manipulation has to do with online analysis. If the EFDD are preformed online there will be a delay in the analysis, because the group of samples needs to be filled before a SPE plot for the group is calculated. These are important considerations that have to be evaluating for each process being monitored.

When the NOC datasets was available a construction of PCA and NLPCA models was taken place. The procedure is the same as for heat exchanger (see chapter 5.1.2). Then the SPE, UCL and fault indicator are calculated as for the heat exchanger (see chapter 5.1.3).

5.2.2 Centrifugal pump fault data set

To investigate the PCA and NLPCA for fault detection, known fault dataset need to be available. Since the simulations in chapter 6 is based on making data from mathematical models of the centrifugal pump the different faults are being simulated. All the faults that are being discussed below are simulated with a step-function (see chapter 2.2.1). In real processes many faults (some of them simulated here) may not act as step-function. But to reduce the scope of the simulations, only the step-function is being used. It should also be mention that only one fault is simulated in each dataset to reduce complexity.

The faults used in simulations in chapter 6 would be sensor failure, parameter change and operation beyond normal regimes. A more detailed description of the different faults is presented in chapter 6. Now a discussion of different faults simulated in chapter 6 are presented.

Sensor fault dataset

Sensor fault is often a source to failure in the process industry and would be an important failure to detect. In the pump process model there are two measurements that are serves as real, the pressure over the pump and the speed of the pump. Both these measurement faults would affect the hold process do to the height degree of correlation (see chapter 5 for detail process description). Both these measurements are being simulated with a fault in separate datasets. The fault dataset was generated by changing the value from the measurements by a step-function in the SIMULINK model.

Process malfunction fault dataset

During the lifetime of a centrifugal pump process wear and tear would appear. This is often a fault that growth slowly over time and a detection of such a fault would be valuable to know at an early state to make plans of a maintenance stop. To simulate such a fault one of the parameters in the pump model would be changed with a step-function during the simulation interval. A step change is not a ideal function to use to simulate such a fault but for simplicity this is done.

Novel Fault Data set

One data set is used as novel fault set. Novel means that the algorithm has not trained with the data for fault regions. The novel data is used to test the ability of the technique to detect unseen faults. The novel dataset don’t need to be a fault, but only a change in the process

operating conditions. These changes could come from other upstream- or downstream processes or locally do change in set point of the controller.

5.3 Procedure for making PCA and NLPCA

PCA

1. Normal operation data was generated executing a SIMULINK model on both the processes. All data variables (training data) for analysis were collected in a data array.

2. Training data was put into desired groups of interval and the mean value for each group was used further in the analysis. After this manipulation the data was centered (subtraction of mean and divided by standard deviation) and used in a PCA algorithm described in chapter 2.4.1. The output from the algorithm gave the number of variances on each principal component and the user could then select how many components that the model should be based on.

3. After selecting number of principal components the PCA algorithm calculated SPE plots and defined an upper control limit UCL with a value 3σ.

NLPCA based on ANN

1. Same as for point 1. in PCA

2. The same groups of data as for PAC in 2 was used for nonlinear PCA (NLPCA) algorithm based on auto-associative neural network [x]. Before running the algorithm the number of principal components in the hyper dimension must be selected by the user. The NLPCA then runs an optimization procedure to build a network that is reachable in the original plane, and then makes SPE plots and defining a UCL with a value of 3σ of the SPE plot.

5.3.1 SPE – and UCL plots and Fault indicator

From the SPE values there was selected an UCL (details see chapter 3.2) that serves as and upper limit for when the process is normal (without false). If the SPE plots is under this line the process is interprets as normal. This limit is here defined to be 3 times the standard deviation of the SPE plot (3σ) (see chapter 2 for details).

To have a supplement to the UCL there are used a fault indicator value. The indicators are the number of samples that violates the UCL during the selected interval. The intervals being used are of same length as the NOC dataset. Since some of the SPE samples always violate the UCL when using PCA models the NOC violations are ignored and the numbers of the indicator are the number of violations under operating subtracted from the NOC violations (made from the NOC). This is important to note when study the simulations in chapter 6.

5.4 Offline or Online Analysis

In the EFDD system it is desirable to run an online analysis from the processes. Some of the models especially NLPCA is relatively slow to train. This is not a serious problem to online analysis, because the models only need to be trained once to set a basis line for the normal operating and fault regions. On the other hand if the system that is being monitored often changes operation region, and therefore imply frequently demands of updating the models the use of NN models could be a problem. This has to be weighted with the type of process that is being monitored and the demand time required that faults must be detected. The analysis presented in the following section has been preformed offline in a batch-wise manner in a Matlab environment. This is because running the calculations online would add no value to the way the results are presented here. The results are focused on exploring the detection of faults, rather than taking appropriate action to correct them.

Chapter 6

6 Results

In this chapter all the result from the simulations are presented. The chapter is divided in two main parts, where the first one is containing results on the plate heat exchanger (chapter 6.1) and the second part centrifugal pump (chapter 6.2). Conclusions for the two processes are presented at the end of the chapter. Under each main parts there are given two simulation experiments for fault detection with different PCA- MBPCA and NLPCA models.

This section start with a presentation of the building process of the PCA and NLPCA for the heat exchanger and then the different faults simulations are given for the heat exchanger. Then the hold process is repeated with the use of the pump process. There are given a conclusion after both processes.

6.1 Heat exchanger results

There are used three PCA models containing different numbers of datasets and one NLPCA with one dataset (see table 3). When the MBPCA are used the T_{ho} is replaced with the residual $T_{ho} - \hat{T}_{ho}$. An overview of the simulations is given in table 4.

PCA Dataset	1: $[w_h\ T_{hi}\ T_{ho} T_{ci}\ u]$
PCA Dataset	2: $\left[w_h\ T_{hi}\ T_{ho} T_{ci}\ u\ \hat{w}_c\ \hat{T}_c\right]$
PCA Dataset	3: $\left[w_h\ T_{hi}\ T_{ho} T_{ci}\ u\ \hat{w}_c\ \hat{T}_c\ \hat{I}_h\ \hat{I}_c\right]$
NLPCA Dataset	2: $\left[w_h\ T_{hi}\ T_{ho} T_{ci}\ u\ \hat{w}_c\ \hat{T}_c\right]$

Table 3 PCA – and NLPCA models datasets for heat exchanger

Without MBPCA		
Fault 1	Fault 2	Fault 3
Sim.1	Sim.2	Sim.3
MBPCA linear model		
Fault 1	Fault 2	Fault 3
Sim.4	Sim.5	Sim.6
MBPCA incorrect non-linear model		
Fault 1	Fault 2	Fault 3
Sim.7	Sim.8	Sim.9
MBPCA correct non-linear model		
Fault 1	Fault 2	Fault 3
Sim.10	Sim.11	Sim.12

Table 4 Overview of the simulations on heat exchanger

The combined data from the experiments to obtain the normal operating region are shown in figure 10. For an explanation of the variables see table 6. It's clear that some variables have tight distribution (e.g variable 3: the temperature on input hot side), while others have large variances (e.g. variable 6- mass flow rate from estimated cold side). We can also see the presence of outliers (e.g. variable 7 – estimated temperature cold side out). In table 5 the different faults simulated on heat exchanger are given.

Fault	**Description**	**Value**	**Percent of NOC**
Fault 1	Sensor fault Temperature measurement cold side in	0.4℃	8 %
Fault 2	Wax composition	200 $W/(m^2K)$	0.83 %
Fault 3	Change in flow hot side	0.54 kg/s	17 %

Table 5 Different Faults simulated on heat exchanger

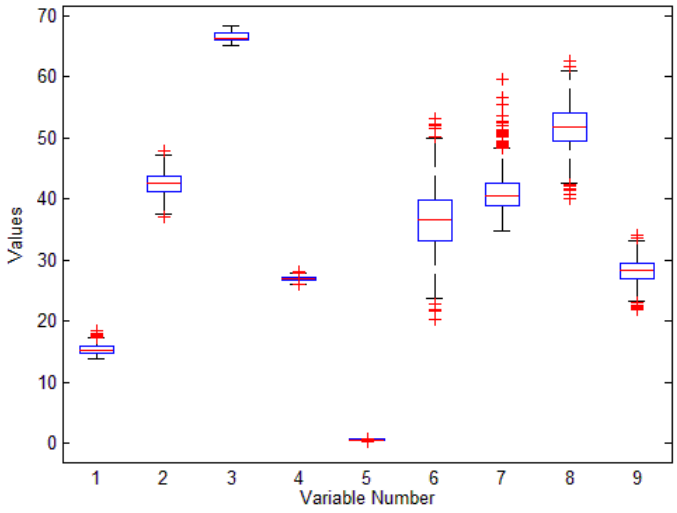


Figure 10 Box plot for heat exchanger normal operation region

Variable Name	Range	Noise Variance	Description
1 Tc-in	$14 - 18\ ℃$	0.1	Temperature cold side in
2 Wh	$36 - 48\ kg/s$	0.1	Flow hot side
3 Th-in	$65 - 68\ ℃$	0.1	Temperature hot side in
4 Th-out	$26 - 29\ ℃$	0.1	Temperature hot side out
5 u	$0.31 - 0.8$	-	Output controller
6 Wc-est	$20 - 53\ kg/s$	-	Flow estimate cold side
7 Tc-out-est	$35 - 60\ ℃$	-	Temperature estimate cold side out
8 Ih	$40 - 63$	-	Heat capacity hot side
9 Ic	$22 - 34$	-	Heat capacity cold side

Table 6 Variables selected for the Analysis

Training without MBPCA

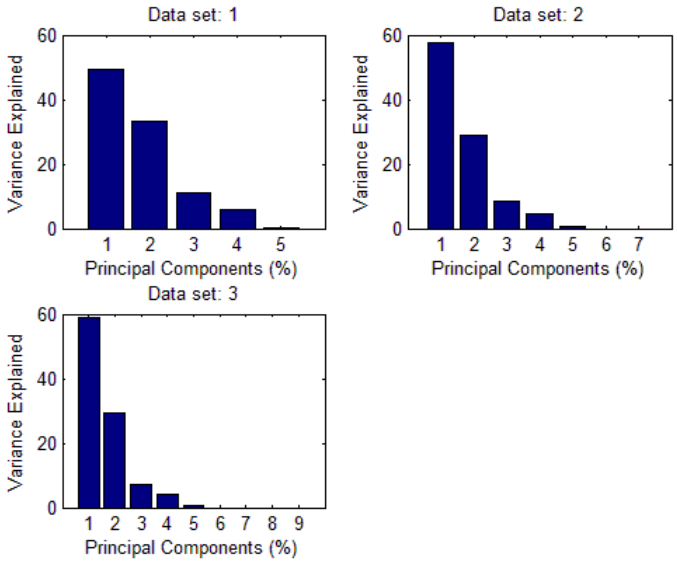


Figure 11 Variance from all variables in the different datasets heat exchanger.

When training without MBPCA we got the variance of each PCs given in figure 11. The number of principal components (PCs) was selected based on average eigenvalue method (see chapter 2.3). The average eigenvalue method proposes the use of 2 PCs for all the PCA models. Since the third PCs was far from 1 the propos was followed. In table 6 the overview of variance explained by the selected PCs are listed

PCA models	1.PCs	2. PCs	Variance explained
PCA Dataset:1	49.54 %	33.40 %	82.94 %
PCA Dataset:2	57.45 %	28.98	86.43 %
PCA Dataset:3	59.05%	29.31 %	88.36 %

Table 6 Principal components variance for the different PCA models heat exchanger

After the selection of principal components the SPE plot for normal operation region, upper control limit and number of legal violations was generated as described in chapter 3.2 (see figure 12 -14). The number of samples that violated the upper control limits was counted for each of the models and set as an acceptable number of violations meaning that if no more samples violate the UCL the process was defined to act normal (without faults). In table 6 the models, with their corresponding dataset and number of normal operation region violations (Legal violations), are presented. In NLPCA there is no opportunity to select number of PCs from a proposal. The PCs has to be selected before the training dataset and was here set to 6. The SPE - and UCL SPE values are given in figure 15.

SPE charts using trained data without MBPCA

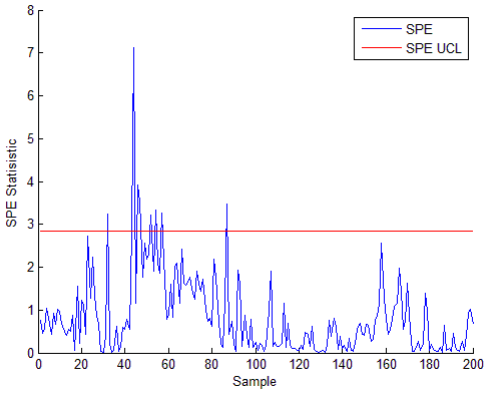


Figure 12 Normal operation region for PCA Dataset 1

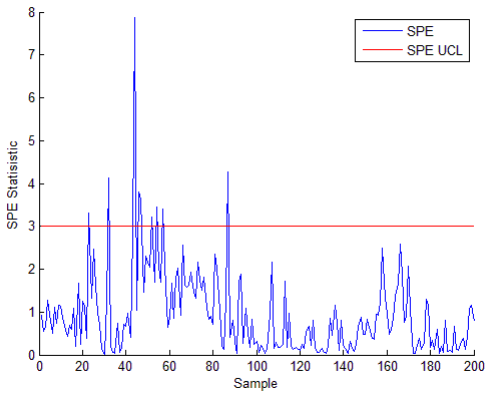


Figure 13 Normal operation region for PCA Dataset 2

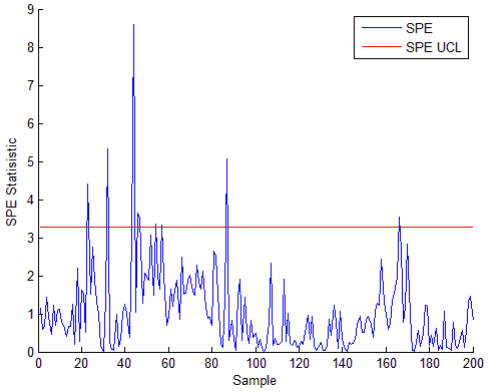


Figure 14 Normal operation region for PCA Dataset 3

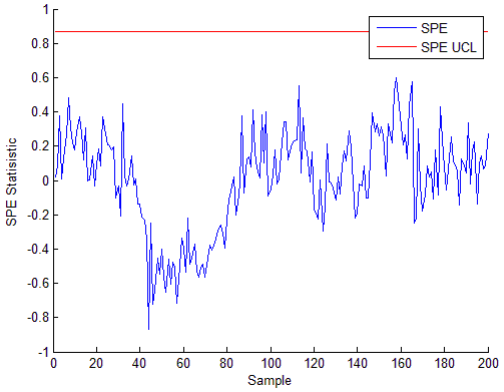


Figure 15 Normal operation region for NLPCA Dataset 2

Models	Legal violations
PCA Dataset 1: $[w_h\ T_{hi}\ T_{ho} T_{ci}\ u]$	9
PCA Dataset 2: $[w_h\ T_{hi}\ T_{ho} T_{ci}\ u\ \hat{w}_c\ \hat{T}_c]$	9
PCA Dataset: 3: $[w_h\ T_{hi}\ T_{ho} T_{ci}\ u\ \hat{w}_c\ \hat{T}_c\ \hat{I}_h\ \hat{I}_c]$	9
NLPCA Dataset 2: $[w_h\ T_{hi}\ T_{ho} T_{ci}\ u\ \hat{w}_c\ \hat{T}_c]$	0

Table 7 NOC violation in SPE plot

When all the models are made the system is ready to analyse data. The models are now being tested for different faults mention in chapter 5.1.2 and the overview of the different faults are given in table 5.

6.2 Simulations of fault without MBPCA

Simulation 1, Heat exchanger without MBPCA

FAULT 1

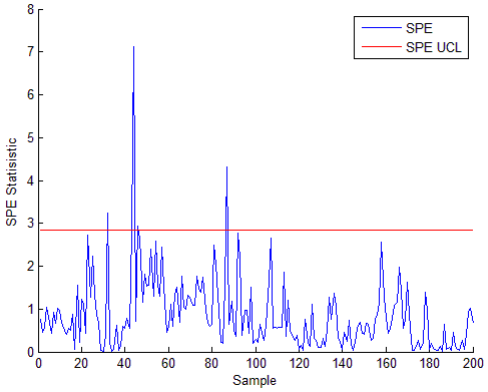


Figure 16 SPE plot PCA Dataset 1

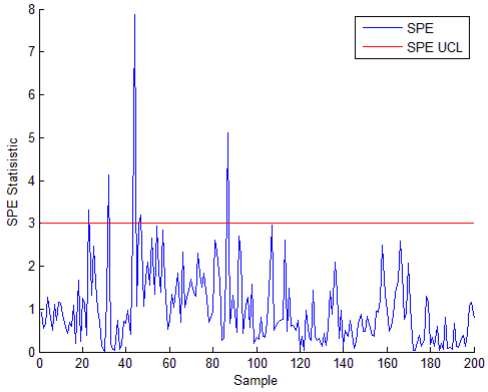


Figure 17 SPE plot PCA Dataset 2

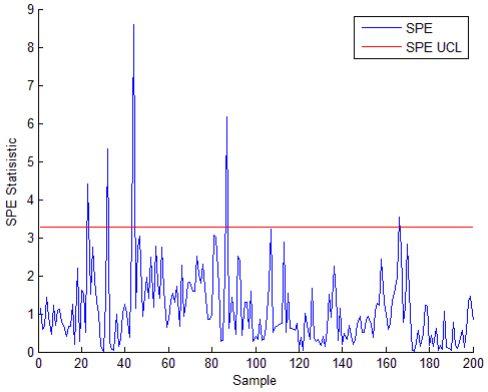


Figure 18 SPE plot PCA Dataset 3

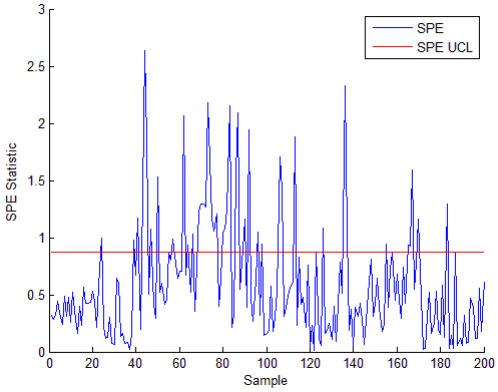


Figure 19 SPE plot NLPCA Dataset 2

In simulation 1, fault in temperature measurement on cold input side is simulated. A bias of 0.4 °C is added on the original measurement at sample 50 and is removed at sample 150. The plots of the SPE statistics together with the upper control limits (found from the normal operating data) are shown in figure 16-19.

Models	Counted Violations
PCA Dataset 1: $[w_h\ T_{hi}\ T_{ho} T_{ci}\ u]$	-4
PCA Dataset 2: $[w_h\ T_{hi}\ T_{ho} T_{ci}\ u\ \hat{w}_c\ \hat{T}_c]$	-4
PCA Dataset 3: $[w_h\ T_{hi}\ T_{ho} T_{ci}\ u\ \hat{w}_c\ \hat{T}_c\ \hat{I}_h\ \hat{I}_c]$	-4
NLPCA Dataset 2: $[w_h\ T_{hi}\ T_{ho} T_{ci}\ u\ \hat{w}_c\ \hat{T}_c]$	48

Table 8 Number of violating samples in 1 day interval

Table 8 shows how many samples that violate the upper control limits for PCA- and NLPCA models. There is only the NLPCA model that shows detection of the fault. But its also give some false alarms after sample 150 when the fault are removed.

Simulation 2, Heat exchanger without MBPCA

FAULT 2

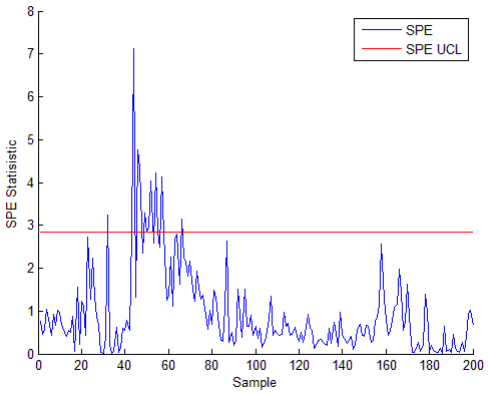


Figure 20 SPE plot PCA Dataset 1

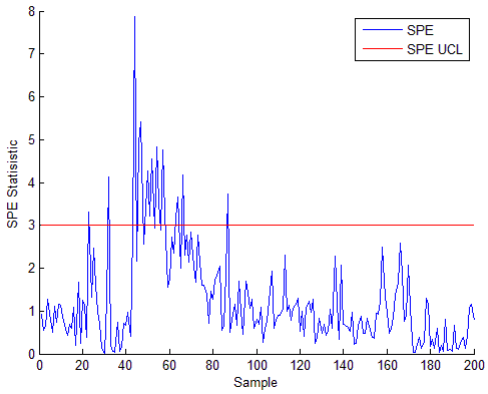


Figure 21 SPE plot PCA Dataset 2

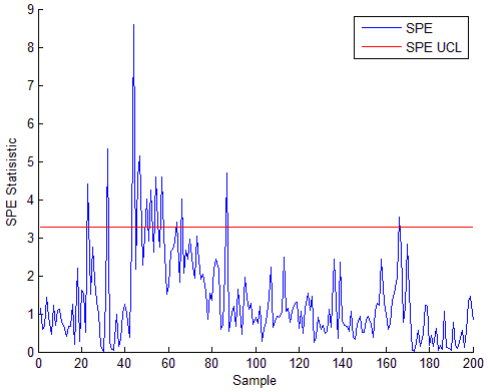


Figure 22 SPE plot PCA Dataset 3

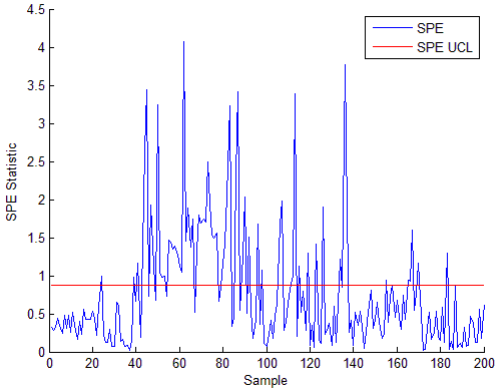


Figure 23 SPE plot NLPCA Dataset 2

In simulation 2, fault in the overall heat transfer coefficient are simulated. A bias of 200 $W/(m^2K)$ is added on the original measurement at sample 50 and is removed at sample 150. The plots of the SPE statistics together with the upper control limits (found from the normal operating data) are shown in figure 20-23

Models	Counted Violations
PCA Dataset 1: $[w_h\ T_{hi}\ T_{ho} T_{ci}\ u]$	4
PCA Dataset 2: $[w_h\ T_{hi}\ T_{ho} T_{ci}\ u\ \hat{w}_c\ \hat{T}_c]$	8
PCA Dataset 3: $[w_h\ T_{hi}\ T_{ho} T_{ci}\ u\ \hat{w}_c\ \hat{T}_c\ \hat{I}_h\ \hat{I}_c]$	5
NLPCA Dataset 2: $[w_h\ T_{hi}\ T_{ho} T_{ci}\ u\ \hat{w}_c\ \hat{T}_c]$	64

Table 9 Number of violating samples in 1 day interval

Table 9 shows how many samples that violate the upper control limits for PCA- and NLPCA models. The PCA models detect the fault poorly. NLPCA on the other hand shows a usable detection.

Simulation 3, Heat exchanger without MBPCA

FAULT 3

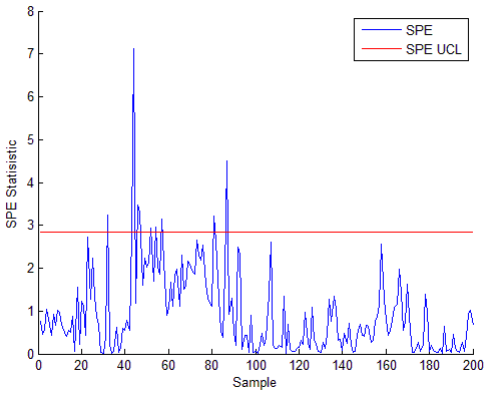


Figure 24 SPE plot PCA Dataset 1

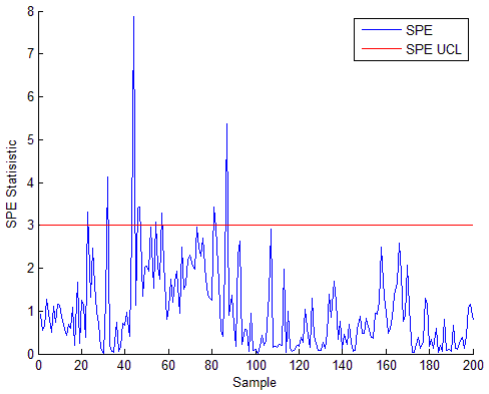


Figure 25 SPE plot PCA Dataset 2

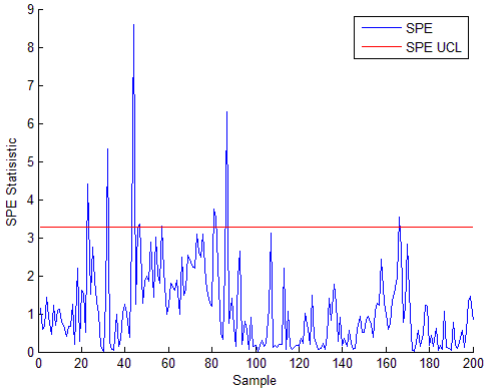


Figure 26 SPE plot PCA Dataset 3

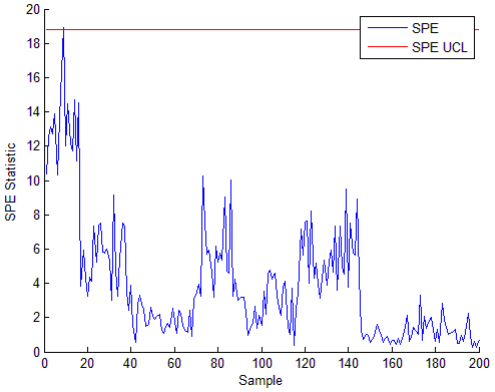


Figure 27 SPE plot NLPCA Dataset 2

In simulation 3, fault in the input flow on hot side is simulated temperature measurement on hot input side is simulated. A bias of 0.54 kg/s is added on the original measurement at sample 50 and is removed at sample 150. The plots of the SPE statistics together with the upper control limits (found from the normal operating data) are shown in figure 24-27.

Models	Counted Violations
PCA Dataset 1: $[w_h\ T_{hi}\ T_{ho} T_{ci}\ u]$	1
PCA Dataset 2: $[w_h\ T_{hi}\ T_{ho} T_{ci}\ u\ \hat{w}_c\ \hat{T}_c]$	0
PCA Dataset 3: $[w_h\ T_{hi}\ T_{ho} T_{ci}\ u\ \hat{w}_c\ \hat{T}_c\ \hat{I}_h\ \hat{I}_c]$	0
NLPCA Dataset 2: $[w_h\ T_{hi}\ T_{ho} T_{ci}\ u\ \hat{w}_c\ \hat{T}_c]$	1

Table 10 Number of violating samples in 1 day interval

Table 10 shows how many samples violations for the upper control limits for PCA- and NLPCA models. All the models detect the fault poorly in this simulation.

Training with MBPCA linear model

In this simulation a linear estimate is used for MBPCA structure. The estimate is based on a dynamic model of the heat exchanger but with the linear equations given I chapter 4.1.8

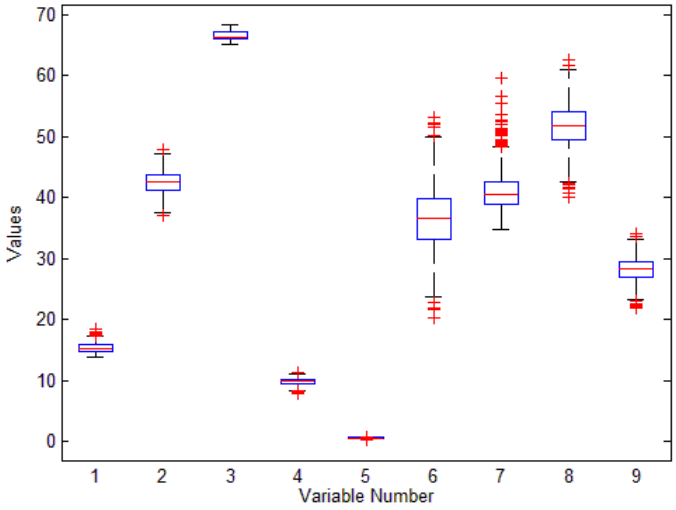


Figure 28 Box plot for heat exchanger normal operation region

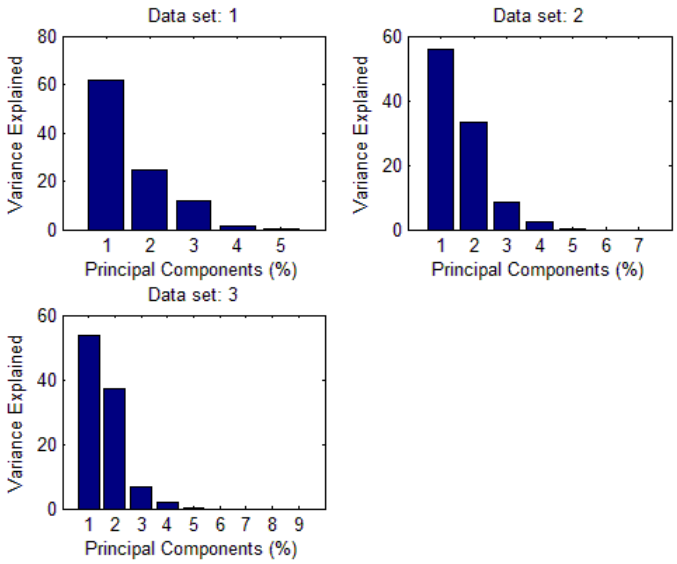


Figure 29 Variance from all variables in the different datasets heat exchanger.

When training with linear MBPCA we got the variance of each PCs given in figure 29. The number of principal components (PCs) was selected based on average eigenvalue method (see chapter 2.3). The average eigenvalue method proposes the use of 2 PCs for all the PCA models. Since the third PCs where far from 1 the propos was followed. In table 11 the overview of variance explained by the selected PCs are listed

PCA models	**1.PCs**	**2. PCs**	**Variance explained**
PCA Dataset:1	61.97 %	24.46%	86.44%
PCA Dataset:2	55.85 %	33 %	88.92 %
PCA Dataset:3	53.70%	37.31%	91 %

Table 11 Principal components variance for the different PCA models heat exchanger

After the selection of principal components the SPE plot for normal operation region, upper control limit and number of legal violations was generated as described in chapter 3.2 (see figure 30 -32). The number samples that violated the upper control limits was counted for each of the models and sat as an acceptable number of violations meaning that if no more samples violate the UCL the process was defined to act normal (without faults). In table 6 the models with there corresponding dataset and number of normal operation region violations (Legal violations) are presented (for details see chapter 3.2). In NLPCA there is no opportunity to select number of PCs from a proposal. The PCs has to be selected before the training dataset and was here set to 6. The SPE - and UCL SPE values are given in figure 33.

SPE charts using trained data with Linear MBPCA

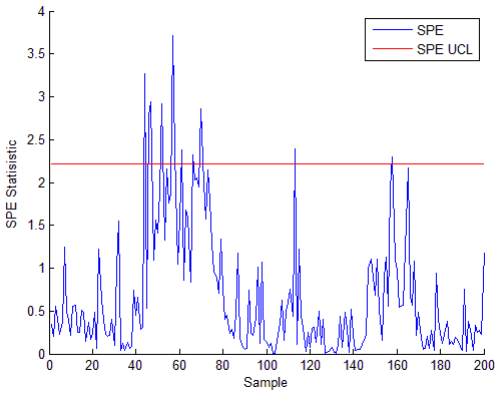


Figure 30 Normal operation region for PCA Dataset 1

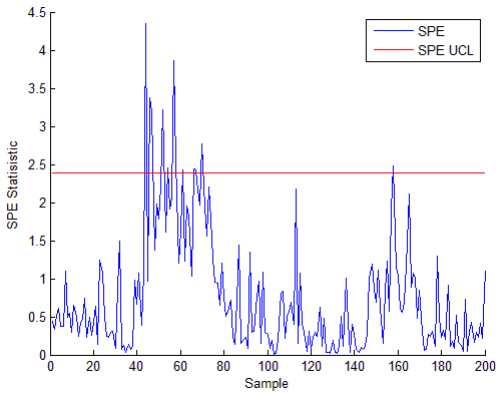


Figure 31 Normal operation region for PCA Dataset 2

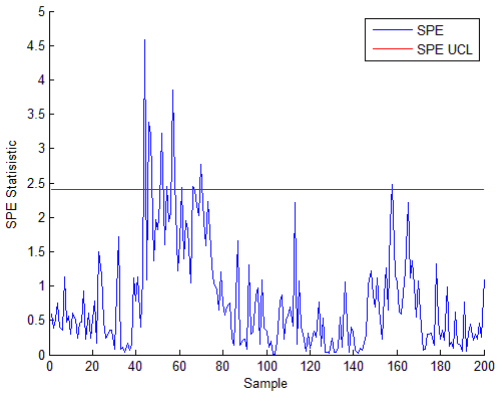


Figure 32 Normal operation region for PCA Dataset 3

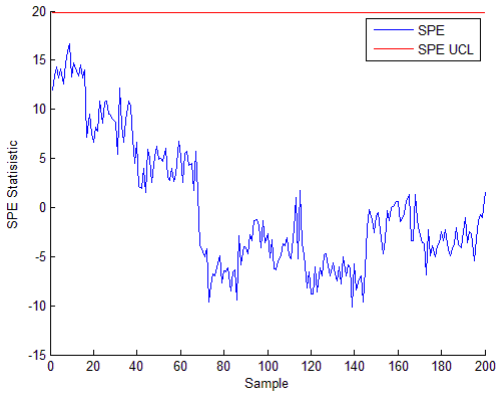


Figure 33 Normal operation region for NLPCA Dataset 2

Models	Legal violations
PCA Dataset 1: $[w_h\ T_{hi}\ T_{ho}T_{ci}\ u]$	11
PCA Dataset 2: $[w_h\ T_{hi}\ T_{ho}T_{ci}\ u\ \hat{w}_c\ \hat{T}_c]$	12
PCA Dataset: 3: $[w_h\ T_{hi}\ T_{ho}T_{ci}\ u\ \hat{w}_c\ \hat{T}_c\ \hat{I}_h\ \hat{I}_c]$	11
NLPCA Dataset 2: $[w_h\ T_{hi}\ T_{ho}T_{ci}\ u\ \hat{w}_c\ \hat{T}_c]$	0

Table 12 NOC violation in SPE plot

Simulation 4, Heat exchanger with Linear MBPCA

FAULT 1

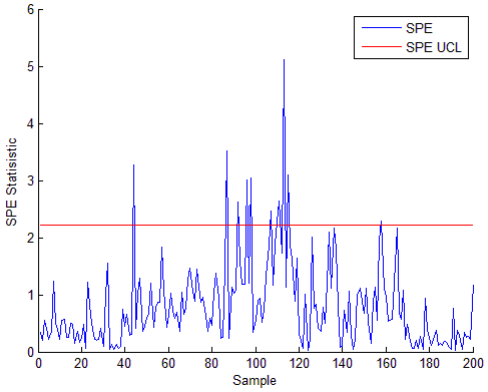


Figure 34 SPE plot PCA Dataset 1

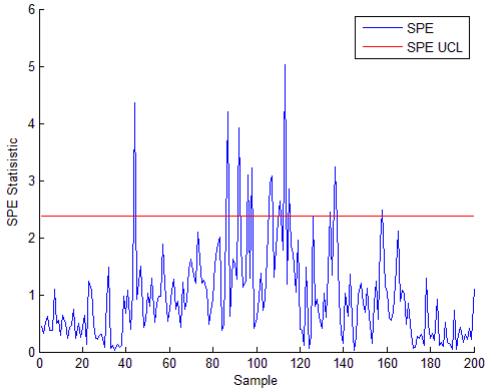


Figure 35 SPE plot PCA Dataset 2

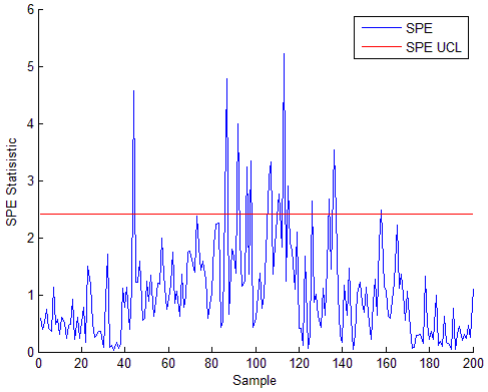


Figure 36 SPE plot PCA Dataset 3

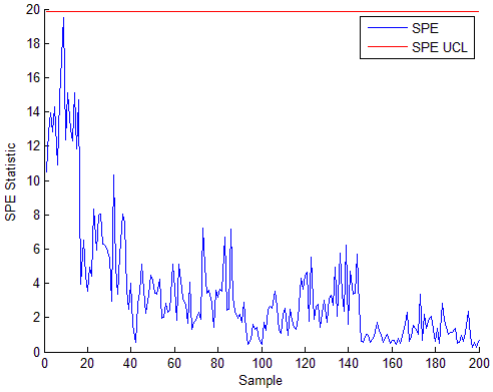


Figure 37 SPE plot NLPCA Dataset 2

In simulation 4, fault in temperature measurement on cold input side is simulated. A bias of 0.4 ℃ is added on the original measurement at sample 50 and is removed at sample 150.

Models		Counted Violations
PCA Dataset	1: $[w_h\ T_{hi}(T_{ho} - \hat{T}_{ho})\ T_{ci}\ u]$	0
PCA Dataset	2: $[w_h\ T_{hi}\ (T_{ho} - \hat{T}_{ho})\ T_{ci}\ u\ \hat{w}_c\ \hat{T}_c]$	1
PCA Dataset	3: $[w_h\ T_{hi}\ (T_{ho} - \hat{T}_{ho})\ T_{ci}\ u\ \hat{w}_c\ \hat{T}_c\ \hat{I}_h\ \hat{I}_c]$	5
NLPCA Dataset	2: $[w_h\ T_{hi}\ (T_{ho} - \hat{T}_{ho})T_{ci}\ u\ \hat{w}_c\ \hat{T}_c]$	0

Table 13 Number of violating samples in 1 day interval

The plots of the SPE statistics together with the upper control limits (found from the normal operating data) are shown in figure 34-37. Table 13 shows how many samples that violate the upper control limits for PCA- and NLPCA models. All the models show poor performance in detecting the fault.

Simulation 5, Heat exchanger with Linear MBPCA

FAULT 2

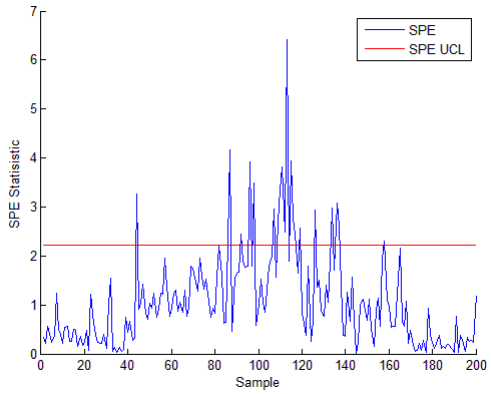


Figure 38 SPE plot PCA Dataset 1

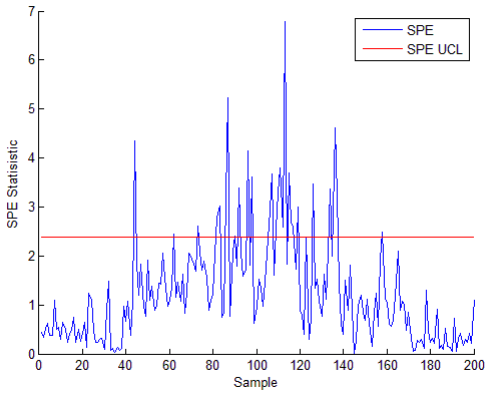


Figure 39 SPE plot PCA Dataset 2

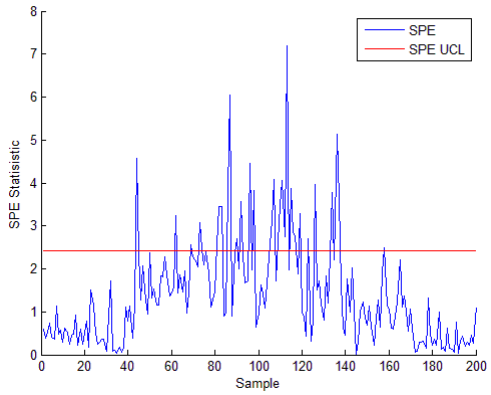


Figure 40 SPE plot PCA Dataset 3

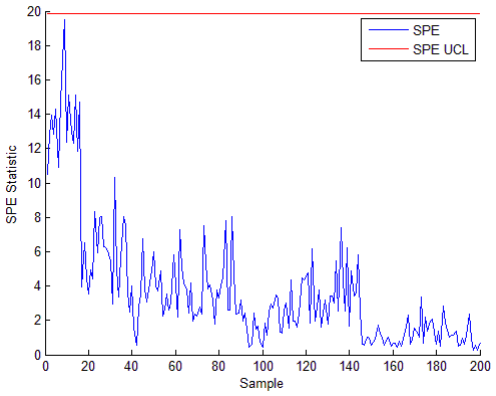


Figure 41 SPE plot NLPCA Dataset 2

In simulation 5, fault in the overall heat transfer coefficient are simulated. A bias of 200 $W/(m^2K)$ is added on the original measurement at sample 50 and is removed at sample 150. The plots of the SPE statistics together with the upper control limits (found from the normal operating data) are shown in figure 38-41. Table 14 shows how many samples that violate the upper control limits for PCA- and NLPCA models. PCA models shows some detection of the fault.

Models		Counted Violations
PCA Dataset	1: $[w_h\ T_{hi}(T_{ho} - \hat{T}_{ho})\ T_{ci}\ u]$	10
PCA Dataset	2: $[w_h\ T_{hi}\ (T_{ho} - \hat{T}_{ho})\ T_{ci}\ u\ \hat{w}_c\ \hat{T}_c]$	15
PCA Dataset	3: $[w_h\ T_{hi}\ (T_{ho} - \hat{T}_{ho})\ T_{ci}\ u\ \hat{w}_c\ \hat{T}_c\ \hat{I}_h\ \hat{I}_c]$	20
NLPCA Dataset	2: $[w_h\ T_{hi}\ (T_{ho} - \hat{T}_{ho})T_{ci}\ u\ \hat{w}_c\ \hat{T}_c]$	0

Table 14 Number of violating samples in 1 day interval

Simulation 6, Heat exchanger with Linear MBPCA

FAULT 3

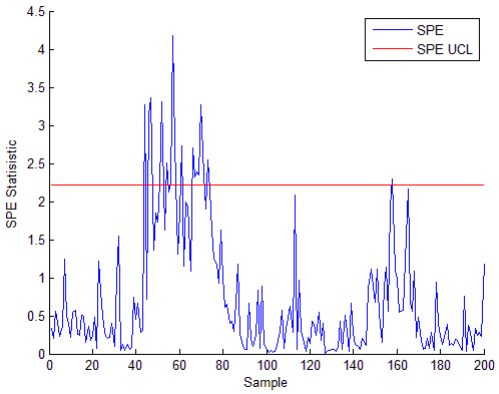


Figure 42 SPE plot PCA Dataset 1

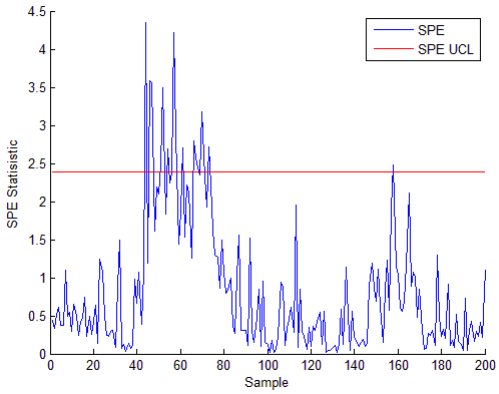


Figure 43 SPE plot PCA Dataset 2

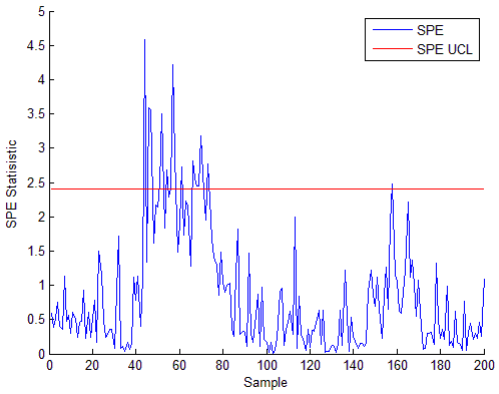


Figure 44 SPE plot PCA Dataset 3

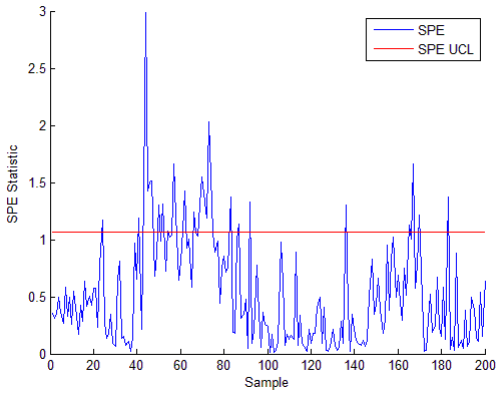


Figure 45 SPE plot NLPCA Dataset 2

Models		Counted Violations
PCA Dataset	1: $[w_h\ T_{hi}(T_{ho} - \hat{T}_{ho})\ T_{ci}\ u]$	7
PCA Dataset	2: $[w_h\ T_{hi}\ (T_{ho} - \hat{T}_{ho})\ T_{ci}\ u\ \hat{w}_c\ \hat{T}_c]$	3
PCA Dataset	3: $[w_h\ T_{hi}\ (T_{ho} - \hat{T}_{ho})\ T_{ci}\ u\ \hat{w}_c\ \hat{T}_c\ \hat{I}_h\ \hat{I}_c]$	6
NLPCA Dataset	2: $[w_h\ T_{hi}\ (T_{ho} - \hat{T}_{ho})T_{ci}\ u\ \hat{w}_c\ \hat{T}_c]$	30

Table 15 Number of violating samples in 1 day interval

In simulation 6, fault in the input flow on hot side is simulated temperature measurement on hot input side is simulated. A bias of 0.54 kg/s is added on the original measurement at sample 50 and is removed at sample 150. The plots of the SPE statistics together with the upper control limits (found from the normal operating data) are shown in figure 42-45. Table 15 shows how many samples violations for the upper control limits for PCA- and NLPCA models. NLPCA is here the best model, but it does not detect the fault longer then sample 80.

Training with not perfect non-linear MBPCA

In this simulation a non-linear estimate is used for MBPCA structure (see chapter 4.1.8). The estimate is based on a dynamic model of the heat exchanger but with the linear equations given I chapter 4.1.8

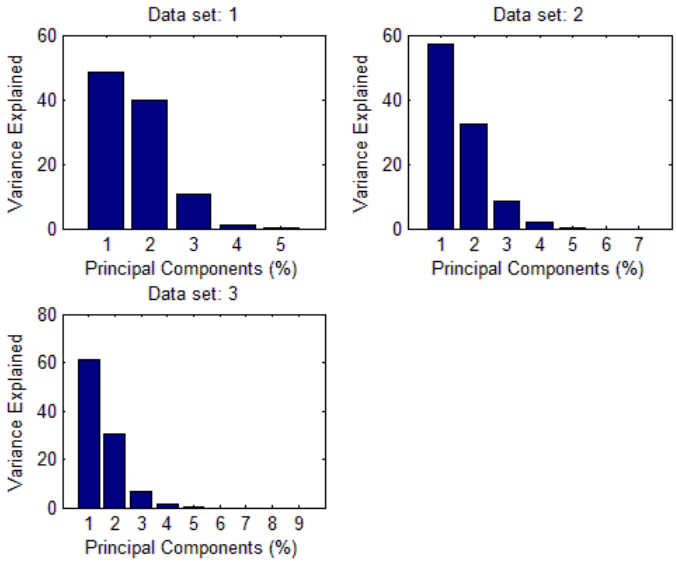


Figure 46 Variance from all variables in the different datasets heat exchanger.

When training with not perfect non-linear MBPCA we got the variance of each PCs given in figure 46. The number of principal components (PCs) was selected based on average eigenvalue method (see chapter 2.3). The average eigenvalue method proposes the use of 2 PCs for all the PCA models. Since the third PCs were far from 1 the propos was followed. In table 16 the overview of variance explained by the selected PCs are listed

PCA models	1.PCs	2. PCs	Variance explained
PCA Dataset:1	48.29%	39.77 %	88.06%
PCA Dataset:2	32.41 %	32.41%	89.55%
PCA Dataset:3	61.28 %	30.34%	91.62%

Table 16 Principal components variance for the different PCA models heat exchanger

After the selection of principal components the SPE plot for normal operation region, upper control limit and number of legal violations was generated as described in chapter 3.2 (see figure 47 -49). The number samples that violated the upper control limits was counted for each of the models and sat as an acceptable number of violations meaning that if no more

samples violate the UCL the process was defined to act normal (without faults). In table 6 the models with there corresponding dataset and number of normal operation region violations (Legal violations) are presented. In NLPCA there is no opportunity to select number of PCs from a proposal. The PCs has to be selected before the training dataset and was here set to 6. The SPE - and UCL SPE values are given in figure 50

SPE charts using trained data with not perfect non-linear MBPCA

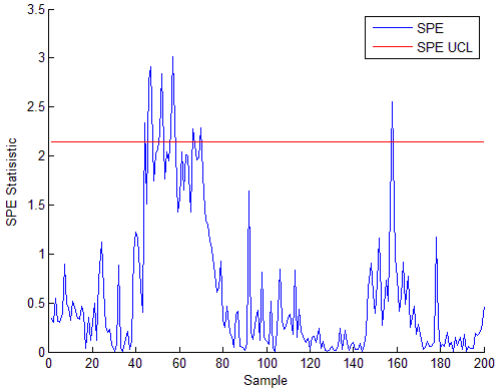


Figure 47 Normal operation region for PCA Dataset 1

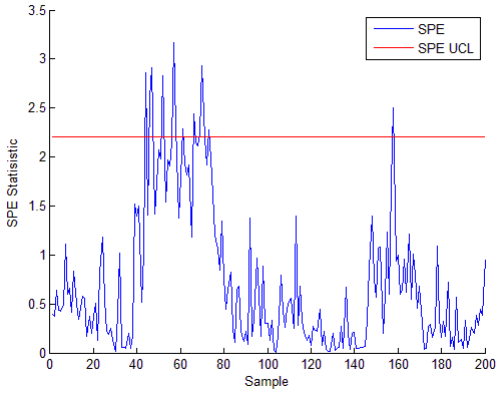


Figure 48 Normal operation region for PCA Dataset 2

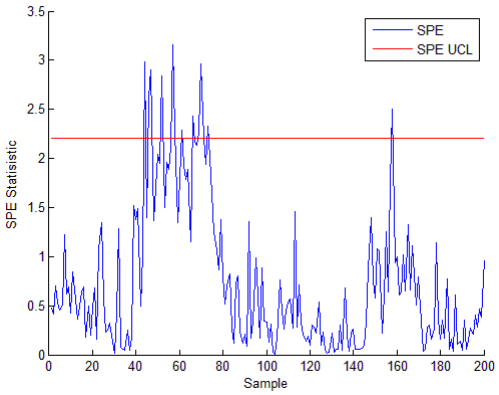


Figure 49 Normal operation region for PCA Dataset 3

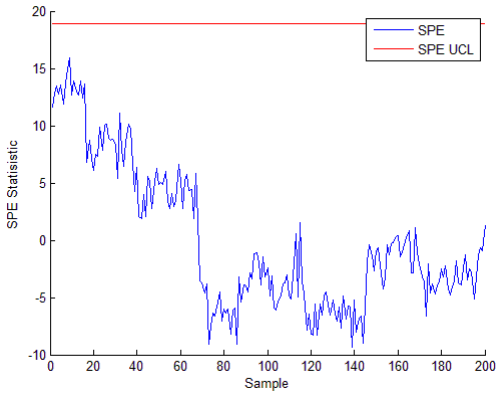


Figure 50 Normal operation region for NLPCA Dataset 2

Models	Legal violations
PCA Dataset 1: $[w_h\ T_{hi}(T_{ho} - \hat{T}_{ho})\ T_{ci}\ u]$	11
PCA Dataset 2: $[w_h\ T_{hi}\ (T_{ho} - \hat{T}_{ho})\ T_{ci}\ u\ \hat{w}_c\ \hat{T}_c]$	12
PCA Dataset 3: $[w_h\ T_{hi}\ (T_{ho} - \hat{T}_{ho})\ T_{ci}\ u\ \hat{w}_c\ \hat{T}_c\ \hat{I}_h\ \hat{I}_c]$	13
NLPCA Dataset 2: $[w_h\ T_{hi}\ (T_{ho} - \hat{T}_{ho})T_{ci}\ u\ \hat{w}_c\ \hat{T}_c]$	0

Table 17 NOC violation in SPE plot

Simulation 7, Non-linear not perfect MBPCA

FAULT 1

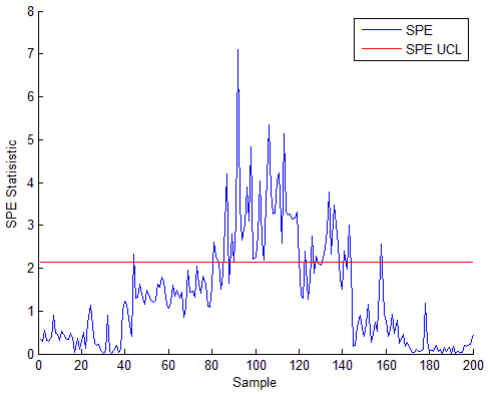


Figure 51 SPE plot PCA Dataset 1

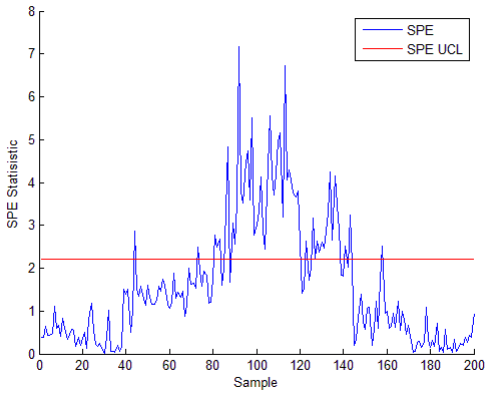


Figure 52 SPE plot PCA Dataset 2

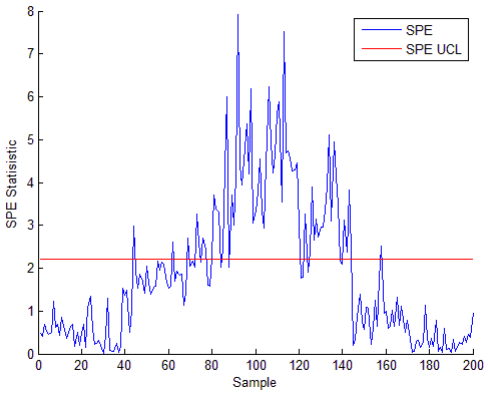


Figure 53 SPE plot PCA Dataset 3

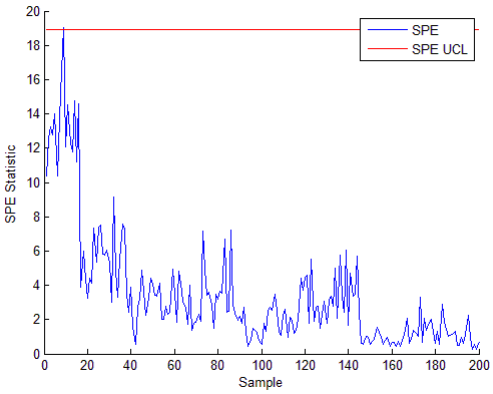


Figure 54 SPE plot NLPCA Dataset 2

In simulation 7, fault in temperature measurement on cold input side is simulated. A bias of 0.4 ℃ is added on the original measurement at sample 50 and is removed at sample 150.

Models		Counted Violations
PCA Dataset	1: $[w_h\ T_{hi}(T_{ho} - \hat{T}_{ho})\ T_{ci}\ u]$	39
PCA Dataset	2: $[w_h\ T_{hi}\ (T_{ho} - \hat{T}_{ho})\ T_{ci}\ u\ \hat{w}_c\ \hat{T}_c]$	44
PCA Dataset	3: $[w_h\ T_{hi}\ (T_{ho} - \hat{T}_{ho})\ T_{ci}\ u\ \hat{w}_c\ \hat{T}_c\ \hat{I}_h\ \hat{I}_c]$	53
NLPCA Dataset	2: $[w_h\ T_{hi}\ (T_{ho} - \hat{T}_{ho})T_{ci}\ u\ \hat{w}_c\ \hat{T}_c]$	1

Table 18 Number of violating samples in 1 day interval

The plots of the SPE statistics together with the upper control limits (found from the normal operating data) are shown in figure 51-54. Table 18 shows how many samples that violate the upper control limits for PCA- and NLPCA models. In this simulation the PCA models detect the fault clearly, but not the whole fault interval.

Simulation 8, Non-linear not perfect MBPCA

FAULT 2

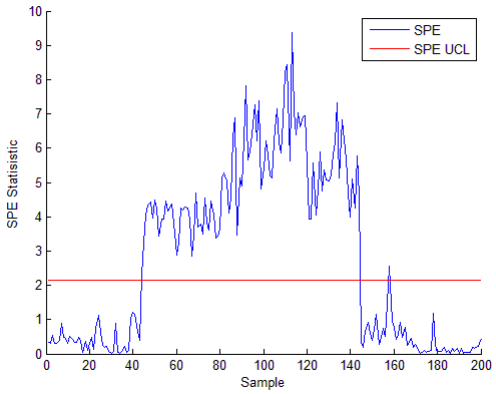


Figure 55 SPE plot PCA Dataset 1

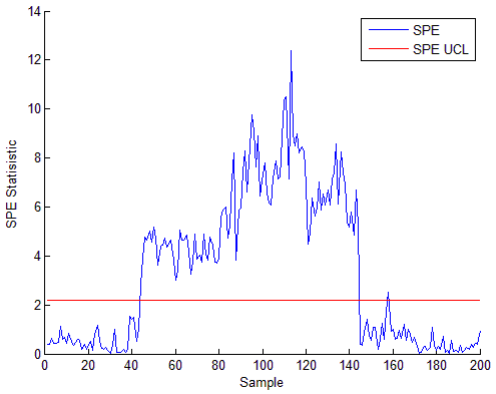


Figure 56 SPE plot PCA Dataset 2

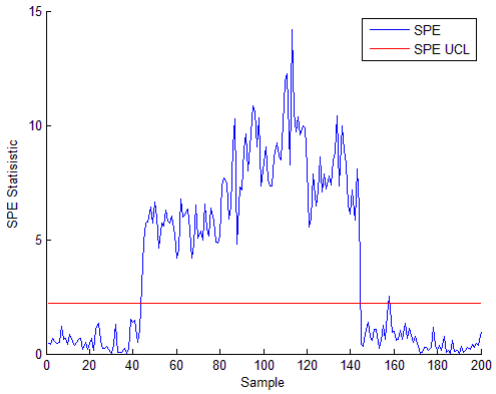


Figure 57 SPE plot PCA Dataset 3

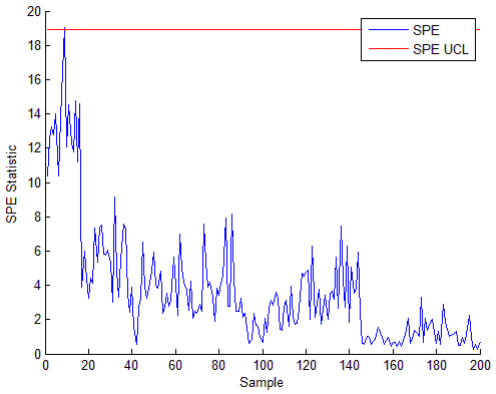


Figure 58 SPE plot NLPCA Dataset 2

In simulation 8, fault in the overall heat transfer coefficient are simulated. A bias of 200 $W/(m^2K)$ is added on the original measurement at sample 50 and is removed at sample 150. The plots of the SPE statistics together with the upper control limits (found from the normal operating data) are shown in figure 55-58. Table 19 shows how many samples that violate the upper control limits for PCA- and NLPCA models. PCA models shows a good performance in detecting the fault.

Models		Counted Violations
PCA Dataset	1: $[w_h\ T_{hi}(T_{ho} - \hat{T}_{ho})\ T_{ci}\ u]$	91
PCA Dataset	2: $[w_h\ T_{hi}\ (T_{ho} - \hat{T}_{ho})\ T_{ci}\ u\ \hat{w}_c\ \hat{T}_c]$	90
PCA Dataset	3: $[w_h\ T_{hi}\ (T_{ho} - \hat{T}_{ho})\ T_{ci}\ u\ \hat{w}_c\ \hat{T}_c\ \hat{I}_h\ \hat{I}_c]$	89
NLPCA Dataset	2: $[w_h\ T_{hi}\ (T_{ho} - \hat{T}_{ho})T_{ci}\ u\ \hat{w}_c\ \hat{T}_c]$	1

Table 19 Number of violating samples in 1 day interval

Simulation 9, Non-linear not perfect MBPCA

FAULT 3

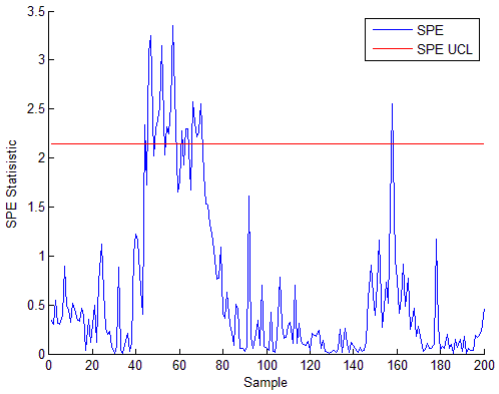


Figure 59 SPE plot PCA Dataset 1

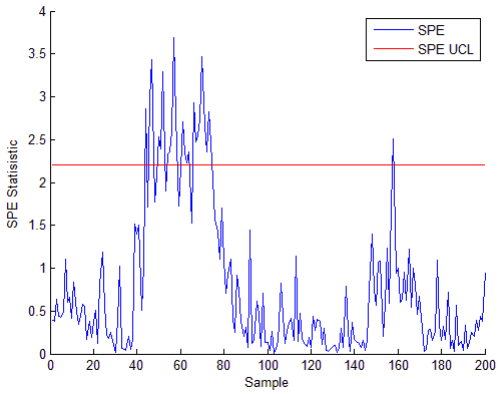


Figure 60 SPE plot PCA Dataset 2

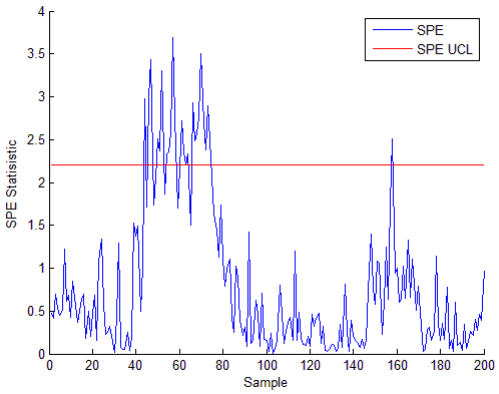


Figure 61 SPE plot PCA Dataset 3

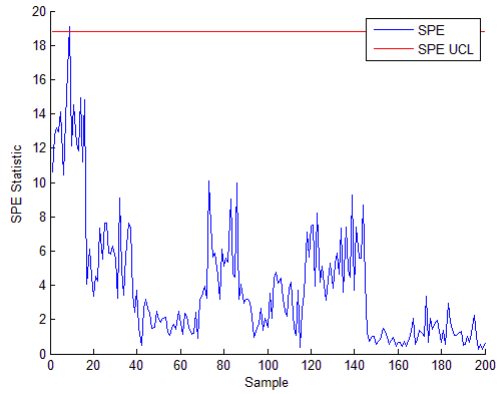


Figure 62 SPE plot NLPCA Dataset 2

In simulation 9, fault in the input flow on hot side is simulated temperature measurement on hot input side is simulated. A bias of 0.54 kg/s is added on the original measurement at sample 50 and is removed at sample 150. The plots of the SPE statistics together with the upper control limits (found from the normal operating data) are shown in figure 59-62.

Models		Counted Violations
PCA Dataset	1: $[w_h\ T_{hi}(T_{ho} - \hat{T}_{ho})\ T_{ci}\ u]$	10
PCA Dataset	2: $[w_h\ T_{hi}\ (T_{ho} - \hat{T}_{ho})\ T_{ci}\ u\ \hat{w}_c\ \hat{T}_c]$	13
PCA Dataset	3: $[w_h\ T_{hi}\ (T_{ho} - \hat{T}_{ho})\ T_{ci}\ u\ \hat{w}_c\ \hat{T}_c\ \hat{I}_h\ \hat{I}_c]$	12
NLPCA Dataset	2: $[w_h\ T_{hi}\ (T_{ho} - \hat{T}_{ho})T_{ci}\ u\ \hat{w}_c\ \hat{T}_c]$	1

Table 20 Number of violating samples in 1 day interval

Table 20 shows how many samples violations for the upper control limits for PCA- and NLPCA models. PCA models do detect the fault for some samples.

Training with perfect non-linear MBPCA

In this simulation a non-linear estimate is used for MBPCA structure (see chapter 4.1.8).

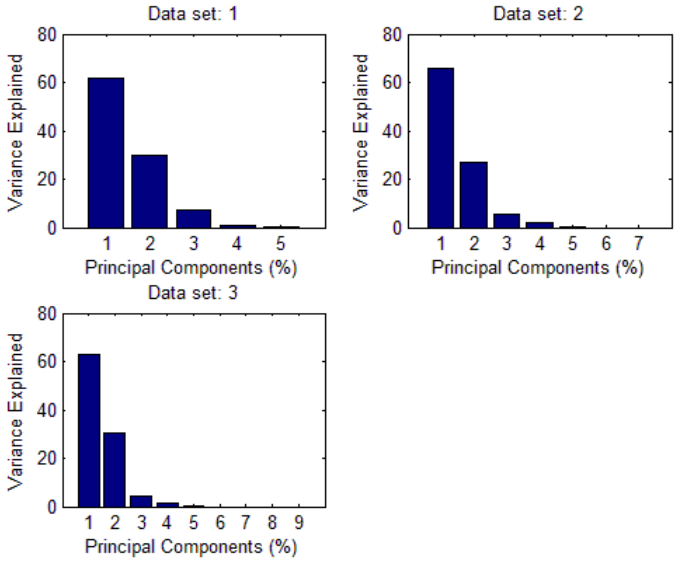


Figure 63 Variance from all variables in the different datasets heat exchanger.

When training with perfect non-linear MBPCA we got the variance of each PCs given in figure 63. The number of principal components (PCs) was selected based on average eigenvalue method (see chapter 2.3). The average eigenvalue method proposes the use of 2 PCs for all the PCA models. Since the third PCs were far from 1 the propos was followed. In table 21 the overview of variance explained by the selected PCs are listed

PCA models	1.PCs	2. PCs	Variance explained
PCA Dataset:1	61.6%	29.96 %	91.56%
PCA Dataset:2	65.73 %	26.85%	92.58%
PCA Dataset:3	63.11 %	30.56%	93.67%

Table 21 Variance from all variables in the different datasets heat exchanger

After the selection of principal components the SPE plot for normal operation region, upper control limit and number of legal violations was generated as described in chapter 3.2 (see figure 64 -66. The number samples that violated the upper control limits was counted for each of the models and sat as an acceptable number of violations meaning that if no more samples violate the UCL the process was defined to act normal (without faults). In table 6 the models with there corresponding dataset and number of normal operation region violations (Legal violations) are presented. In NLPCA there is no opportunity to select number of PCs from a proposal. The PCs has to be selected before the training dataset and was here set to 6. The SPE - and UCL SPE values are given in figure 67

SPE charts using trained data with perfect non-linear MBPCA

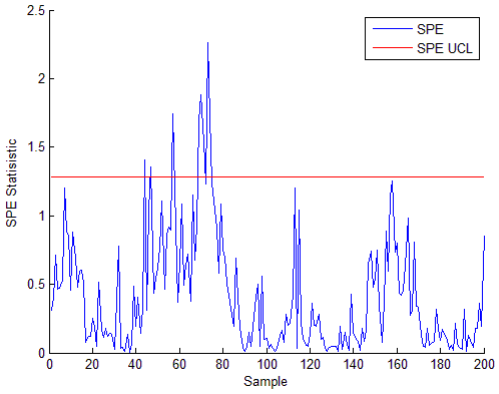


Figure 64 Normal operation region for PCA Dataset 1

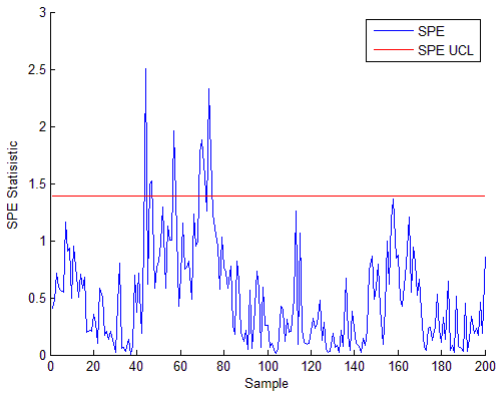


Figure 65 Normal operation region for PCA Dataset 2

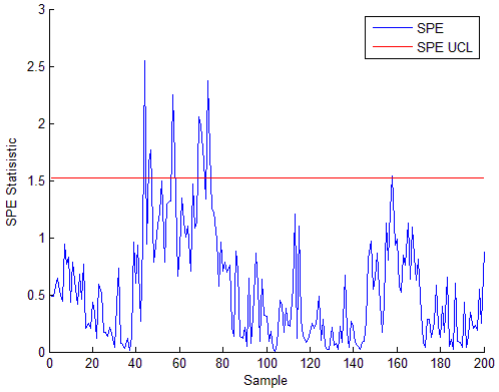


Figure 66 Normal operation region for PCA Dataset 3

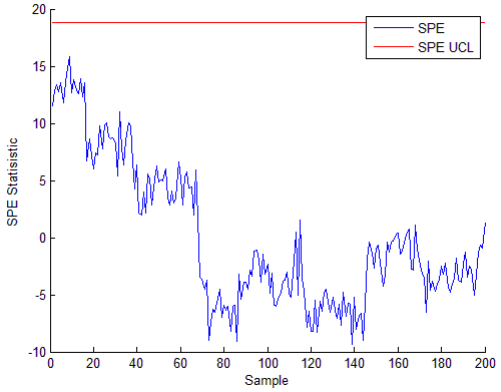


Figure 67 Normal operation region for NLPCA Dataset 2

Models		Legal violations
PCA Dataset	1: $[w_h\ T_{hi}(T_{ho} - \hat{T}_{ho})\ T_{ci}\ u]$	8
PCA Dataset	2: $[w_h\ T_{hi}\ (T_{ho} - \hat{T}_{ho})\ T_{ci}\ u\ \hat{w}_c\ \hat{T}_c]$	9
PCA Dataset	3: $[w_h\ T_{hi}\ (T_{ho} - \hat{T}_{ho})\ T_{ci}\ u\ \hat{w}_c\ \hat{T}_c\ \hat{I}_h\ \hat{I}_c]$	11
NLPCA Dataset	2: $[w_h\ T_{hi}\ (T_{ho} - \hat{T}_{ho})T_{ci}\ u\ \hat{w}_c\ \hat{T}_c]$	0

Table 22 NOC violation in SPE plot

Simulation 10, Non-linear perfect MBPCA

FAULT 1

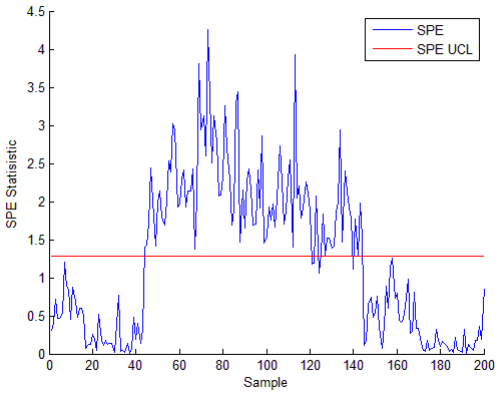


Figure 68 SPE plot PCA Dataset 1

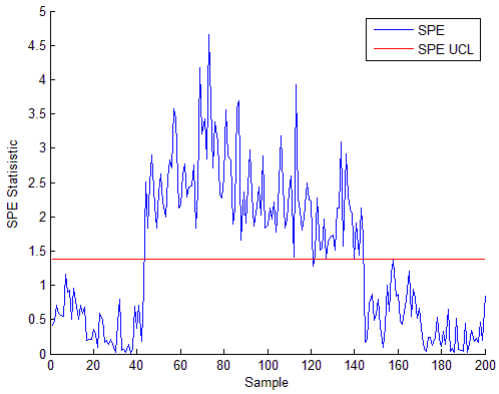


Figure 69 SPE plot PCA Dataset 2

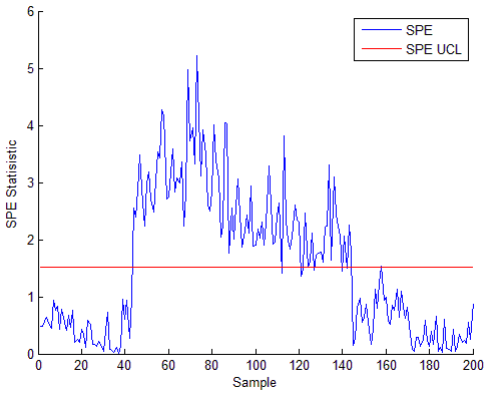


Figure 70 SPE plot PCA Dataset 3

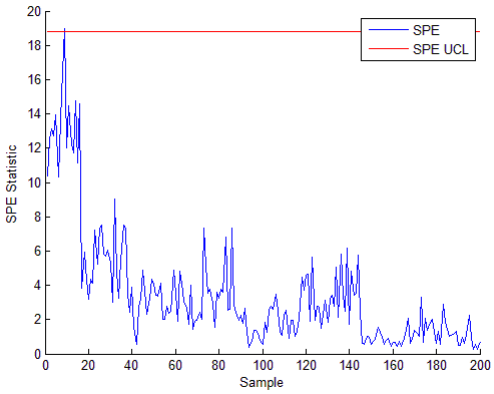


Figure 71 SPE plot NLPCA Dataset 2

In simulation 10, fault in temperature measurement on cold input side is simulated. A bias of 0.4 ℃ is added on the original measurements at sample 50 and is removed at sample 150.

Models		Counted Violations
PCA Dataset	1: $[w_h\ T_{hi}(T_{ho} - \hat{T}_{ho})\ T_{ci}\ u]$	89
PCA Dataset	2: $[w_h\ T_{hi}\ (T_{ho} - \hat{T}_{ho})\ T_{ci}\ u\ \hat{w}_c\ \hat{T}_c]$	91
PCA Dataset	3: $[w_h\ T_{hi}\ (T_{ho} - \hat{T}_{ho})\ T_{ci}\ u\ \hat{w}_c\ \hat{T}_c\ \hat{I}_h\ \hat{I}_c]$	95
NLPCA Dataset	2: $[w_h\ T_{hi}\ (T_{ho} - \hat{T}_{ho})T_{ci}\ u\ \hat{w}_c\ \hat{T}_c]$	1

Table 23 Number of violating samples in 1 day interval

The plots of the SPE statistics together with the upper control limits (found from the normal operating data) are shown in figure 68-71. Table 23 shows how many samples that violate the upper control limits for PCA- and NLPCA models. The PCA models detect the fault well. There are small differences between them.

Simulation 11, Non-linear perfect MBPCA

FAULT 2

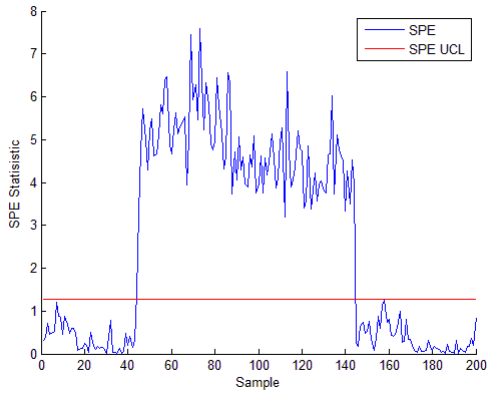


Figure 72 SPE plot PCA Dataset 1

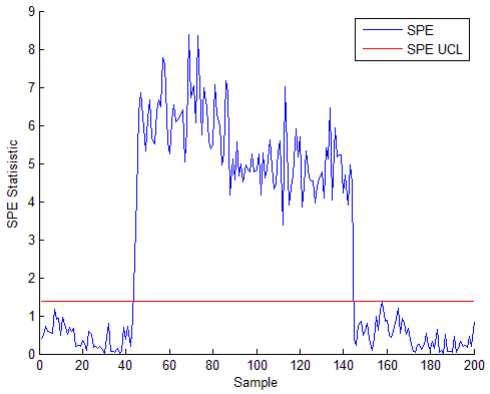


Figure 73 SPE plot PCA Dataset 2

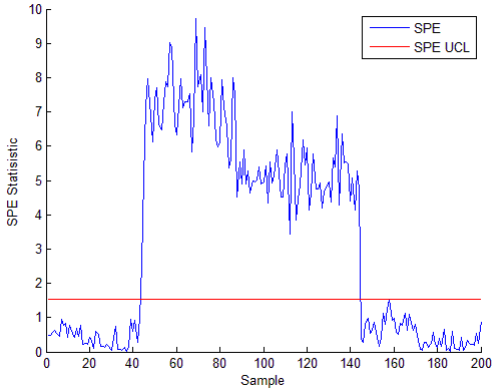


Figure 74 SPE plot PCA Dataset 3

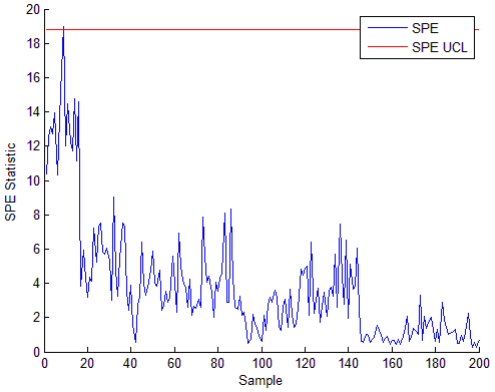


Figure 75 SPE plot NLPCA Dataset 2

In simulation 11, fault in the overall heat transfer coefficient are simulated. A bias of 200 $W/(m^2K)$ is added on the original measurement at sample 50 and is removed at sample 150. The plots of the SPE statistics together with the upper control limits (found from the normal operating data) are shown in figure 72-75. Table 24 shows how many samples that violate the upper control limits for PCA- and NLPCA models. The PCA models detect the fault well. There are small differences between them.

Models		Counted Violations
PCA Dataset	1: $[w_h\ T_{hi}(T_{ho} - \hat{T}_{ho})\ T_{ci}\ u]$	93
PCA Dataset	2: $[w_h\ T_{hi}\ (T_{ho} - \hat{T}_{ho})\ T_{ci}\ u\ \hat{w}_c\ \hat{T}_c]$	92
PCA Dataset	3: $[w_h\ T_{hi}\ (T_{ho} - \hat{T}_{ho})\ T_{ci}\ u\ \hat{w}_c\ \hat{T}_c\ \hat{I}_h\ \hat{I}_c]$	91
NLPCA Dataset	2: $[w_h\ T_{hi}\ (T_{ho} - \hat{T}_{ho})T_{ci}\ u\ \hat{w}_c\ \hat{T}_c]$	1

Table 24 Number of violating samples in 1 day interval

Simulation 12, Non-linear perfect MBPCA

FAULT 3

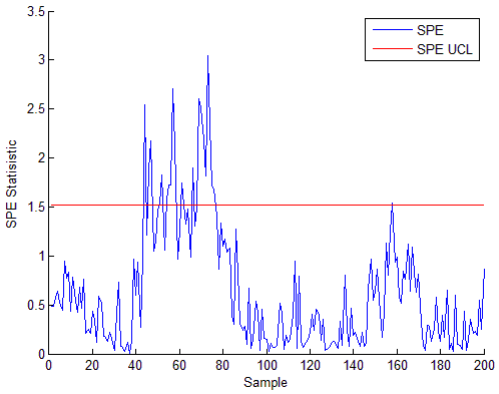


Figure 76 SPE plot PCA Dataset 1

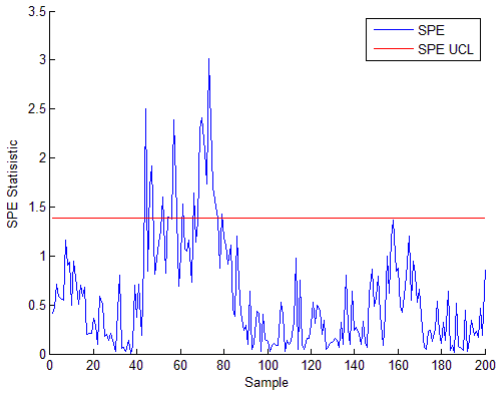


Figure 77 SPE plot PCA Dataset 2

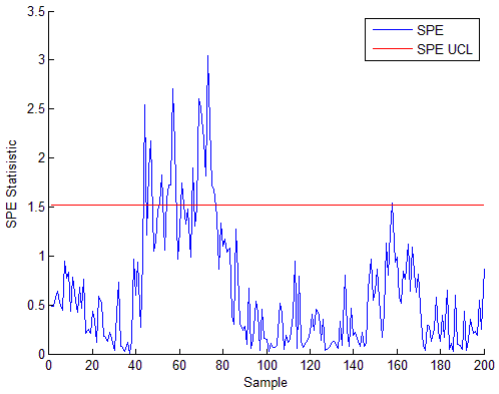


Figure 78 SPE plot PCA Dataset 3

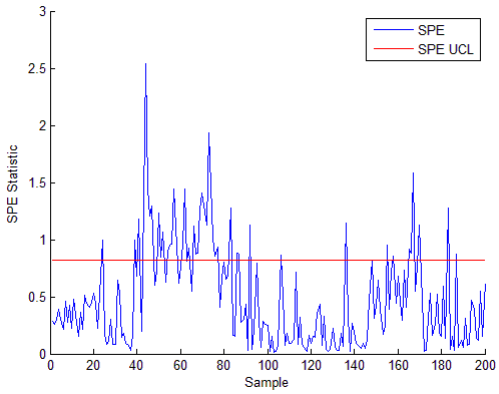


Figure 79 SPE plot NLPCA Dataset 2

In simulation 12 a fault in the input flow on hot side is simulated temperature measurement on hot input side is simulated.

Models		Counted Violations
PCA Dataset	1: $[w_h\ T_{hi}(T_{ho} - \hat{T}_{ho})\ T_{ci}\ u]$	11
PCA Dataset	2: $[w_h\ T_{hi}\ (T_{ho} - \hat{T}_{ho})\ T_{ci}\ u\ \hat{w}_c\ \hat{T}_c]$	10
PCA Dataset	3: $[w_h\ T_{hi}\ (T_{ho} - \hat{T}_{ho})\ T_{ci}\ u\ \hat{w}_c\ \hat{T}_c\ \hat{I}_h\ \hat{I}_c]$	10
NLPCA Dataset	2: $[w_h\ T_{hi}\ (T_{ho} - \hat{T}_{ho}) T_{ci}\ u\ \hat{w}_c\ \hat{T}_c]$	46

Table 25 Number of violating samples in 1 day interval

A bias of 0.54 kg/s is added on the original measurement at sample 50 and is removed at sample 150. The plots of the SPE statistics together with the upper control limits (found from the normal operating data) are shown in figure 76-79. Table 25 shows how many samples violations for the upper control limits for PCA- and NLPCA models. This fault was best detected by NLPCA.

6.2 Centrifugal pump results

For the centrifugal pump process two sections of simulation experiments executed.
The first one contains models build from a NOC with the available measurements and some virtual measurements (measurements based on function from available measurements). And the second contain models build from a NOC where a MBPCA approach is used. When using the MBPCA the same measurements are used except for the flow measurement that is a residual of the flow (flow estimate subtracted from the measurement served as the real flow). Both the two parts containing 3 different PCA models made from different numbers of datasets and one NLPCA model made of one dataset (see table 26). Both the NOC series operates in the same interval.

This section starts with simulations without MBPCA and then continues with MBPCA. But before each part result from training are given.

PCA Dataset 1: $[\Delta p\ w\ q]$
PCA Dataset 2: $[\Delta p\ w\ q\ \mathrm{q}^2]$
PCA Dataset 3:$[\Delta p\ w\ q\ \mathrm{q}^2\ \mathrm{w}^2]$
NLPCA Dataset3: $[[\Delta p\ w\ q\ \mathrm{q}^2\ \mathrm{w}^2]$

Table 26 PCA – and NLPCA models datasets for pump

In table 27 an overview of the simulations executed for the pump process are presented.

Without MBPCA			
Fault 1	Fault 2	Fault 3	Fault 4
Sim.1	Sim.2	Sim.3	Sim.4
MBPCA			
Fault 1	Fault 2	Fault 3	Fault 4
Sim.5	Sim.6	Sim.7	Sim.8

Table 27 Overview of the simulations on Centrifugal pump process

Simulations without MBPCA

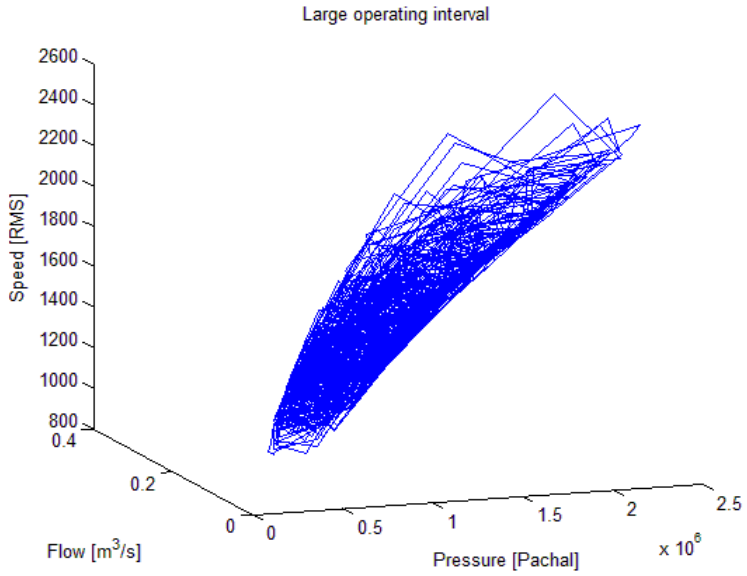


Figure 80 3 dimensional plot of speed, flow and pressure from NOC

In figure 80 a 3 dimensional plot of the data generated for NOC. The data contains 3 measurements (speed, flow and pressure) and we can see the data contains nonlinearities.

The combined data from the experiments to obtain the normal operating region are shown in figure 81. For an explanation of the variables see table 28

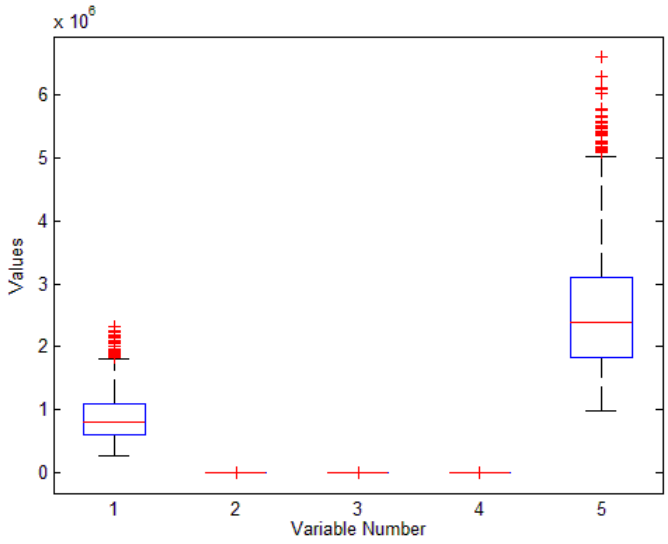


Figure 81 Box plot for NOC on pump

Variable	Range	Noise on measurement	Description
1 Δp	2.7-23.1 [bar]	Variance 0.1	Pressure over pump
2 q	0.022-0.24 $[m^3/s]$	-	Flow through pump
3 n	0.99-2.57 $\times 10^3$ [RMS]	Variance 1	Speed on pump
4 q^2	0.048-4.16 $\times 10^{-2}$ $[m^3/s]$	-	Square Flow through pump
5 n^2	0.98-6.6 $\times 10^6$ [RMS]	Variance 1	Square Speed on pump

Table28 Variable Selected for the Analysis on pump

The different faults simulated on the centrifugal pump are given in table 29.

Fault	Description	Value	Percent of NOC
Fault 1	Sensor fault in pressure measurement	0.2 [bar]	1 %
Fault 1	Fault in output controller	20 [RMS]	1.27 %
Fault 3	Change in parameter c2	3000	-
Fault 4	Novel fault	0.04 $[m^3/s]$	18.35 %

Table 29 Types of faults in centrifugal pump

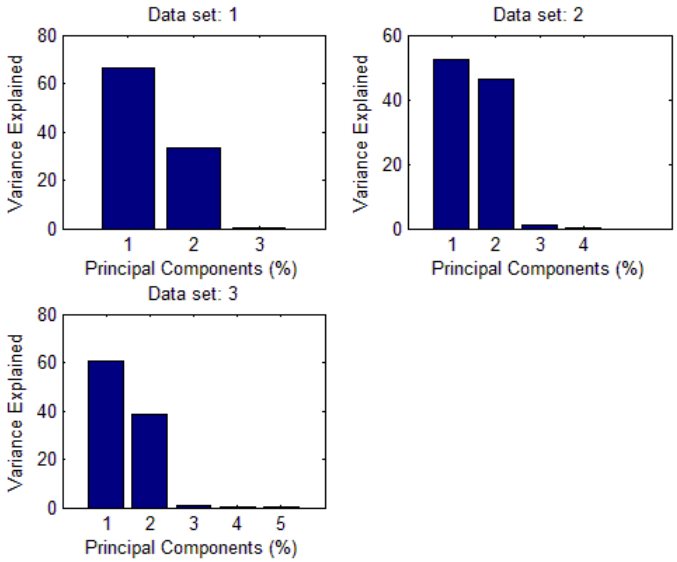


Figure 68 Variance from NOC on the pump

The number of PCs variance is given in figure 68. The PCs was selected based on average eigenvalue method (see chapter 2.4.2). The method suggested two PCs for all three models. Since no PCs other then the one suggested were near 1 the suggestion was followed.

PCA models	1.PCs	2.PCs	Variance explained
PCA Dataset:1	66.43 %	33.44 %	99.87 %
PCA Dataset:2	52.35 %	46.35 %	99.7 %
PCA Dataset:3	60.54 %	38.39 %	99.94 %

Table 30 Principal components variance for the different PCA models from NOC

The PCs variance selected are given in table 30. After the selection of principal components the SPE plot for normal operation region and upper control limit was generated for each of the models (see figure 82 to 84). The number of samples that violated the upper control limits was counted for each of the models and sat as an acceptable number of violations meaning that if no more samples violate the UCL the process was defined to act normal (without faults). In table 31 the models with there corresponding dataset and number of normal operation region violations (Legal violations) are presented. For a discussion of the UCL and the legal violations see chapter 3.2

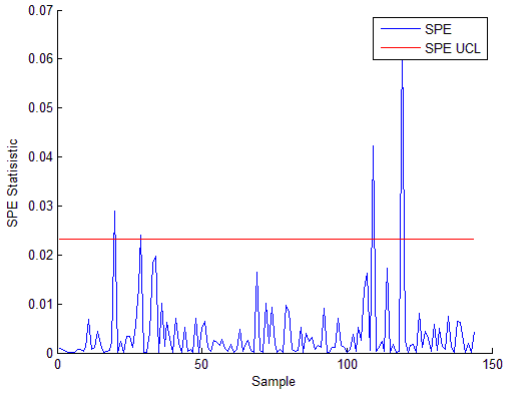


Figure 82 SPE plot PCA dataset 1

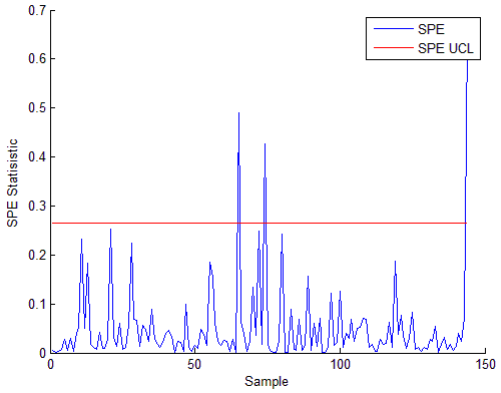


Figure 83 SPE plot PCA dataset 2

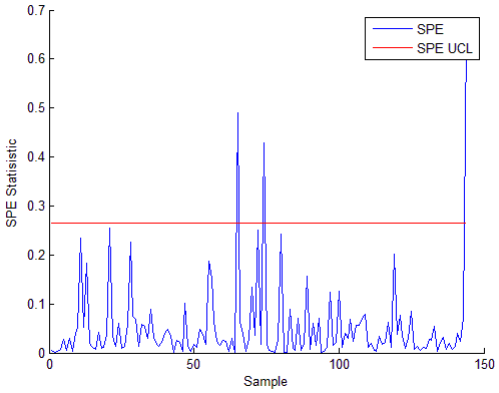


Figure 84 SPE plot PCA dataset 3

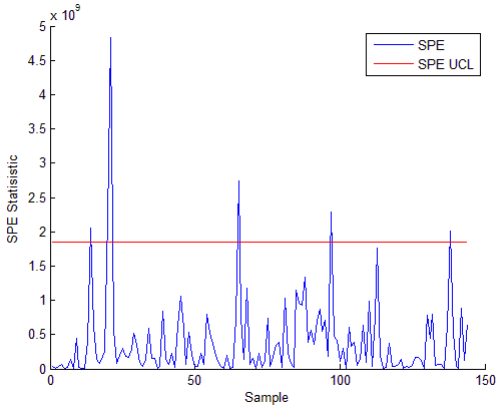


Figure 85 SPE plot NLPCA dataset 3

After the PCA models were generated the NLPCA method based on autoassosiatve network was calculated. This method do not give the opportunity to select PCs based on the magnitude as for the PCA method. The PCs was here selected by a trial and error approach and sat to 1. The data was also here manipulated. After the model was finished the SPE and UCL was plotted (see figure 85).

Models	NOC violations
PCA_{p1} Dataset : $[\Delta p\ w\ q]$	4
PCA_{p2} Dataset : $[\Delta p\ w\ \hat{Q}_p\ \widehat{\mathrm{Q}}_{\mathrm{p}}^2]$	3
PCA_{p3} Dataset: $[\Delta p\ w\ \hat{Q}_p\ \widehat{\mathrm{Q}}_{\mathrm{p}}^2\ \mathrm{w}^2]$	3
$NLPCA_p$ Dataset: $[\Delta p\ w\ \hat{Q}_p\ \widehat{\mathrm{Q}}_{\mathrm{p}}^2\ \mathrm{w}^2]$	6

Table 31 Number of violating samples in 1 day interval

When all the models are made the system is ready for monitor the system. The models are know being tested for different faults mention in chapter 5.2.2 and a overview of the different simulations and faults are given in table 27 and 29 respectively.

In simulations SPE plots would be presented and annotated. In addition a table would be presented containing the different models with the corresponding dataset they are based on. In addition the number of samples that violate the UCL (Violations) when the number of violations from the normal operating region is subtracted is listed.

Simulation 1, without MBPCA

FAULT 1

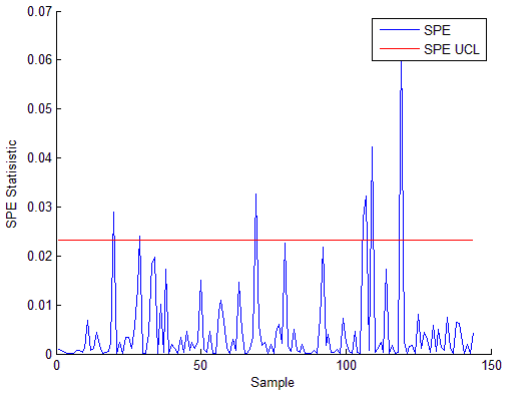


Figure 86 SPE plot PCA dataset 1

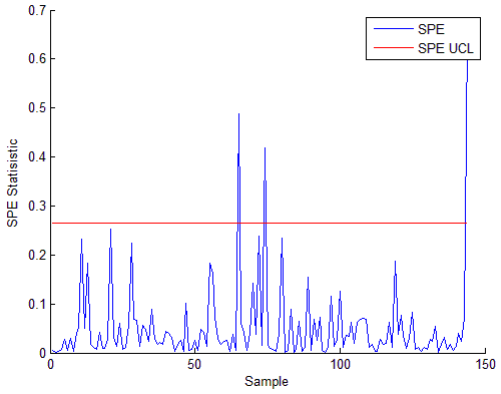


Figure 87 SPE plot PCA dataset 2

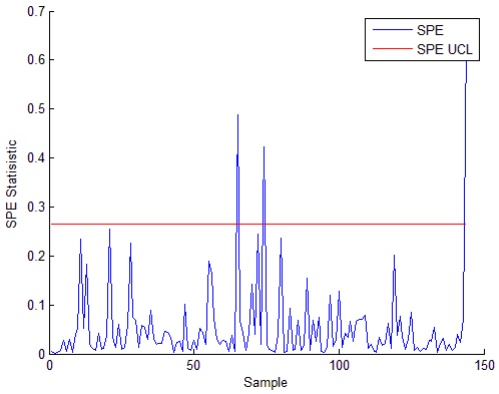


Figure 88 SPE plot PCA dataset 3

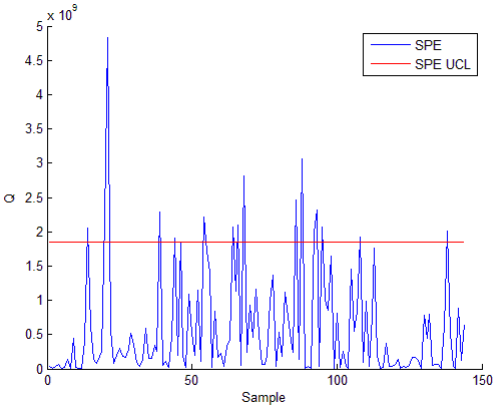


Figure 89 SPE plot NLPCA dataset 3

In simulation 1, fault in pressure measurement over the pump is simulated. A bias of 0.2 bar is added on the original measurement at sample 36 and is removed at sample 108.

Models	Counted Violations
PCA Dataset 1: $[\Delta p\ w\ q]$	3
PCA Dataset 2: $[\Delta p\ w\ q\ \mathrm{q}^2]$	0
PCA Dataset 3: $[\Delta p\ w\ q\ \mathrm{q}^2\ \mathrm{w}^2]$	0
NLPCA Dataset 3: $[\Delta p\ w\ q\ \mathrm{q}^2\ \mathrm{w}^2]$	11

Table 32 Number of violating samples in 1 day interval

The plots of the SPE statistics together with the upper control limits (found from the normal operating data) are shown in figure 86-89. Table 32 shows how many samples that violate the upper control limits for PCA- and NLPCA models. In this simulation the NLPCA is the best for identification of the fault, but the number of violations is not convincing.

Simulation 2, without MBPCA

FAULT 2

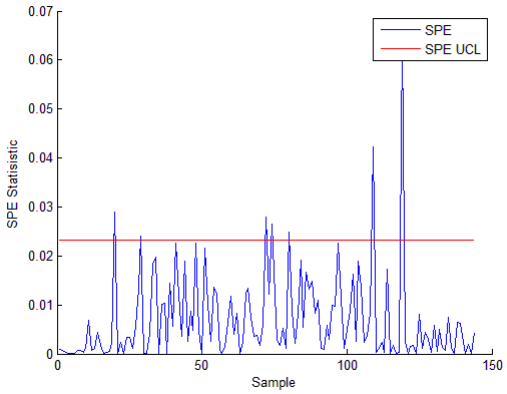


Figure 90 SPE plot PCA dataset 1

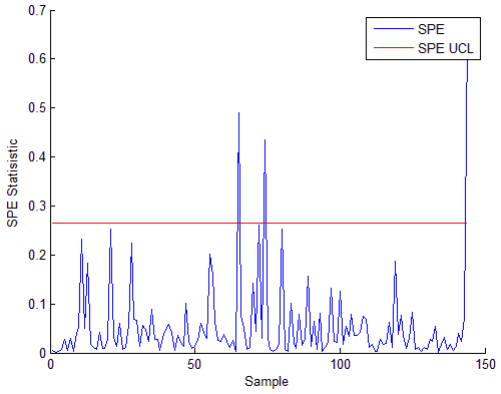


Figure 91 SPE plot PCA dataset 2

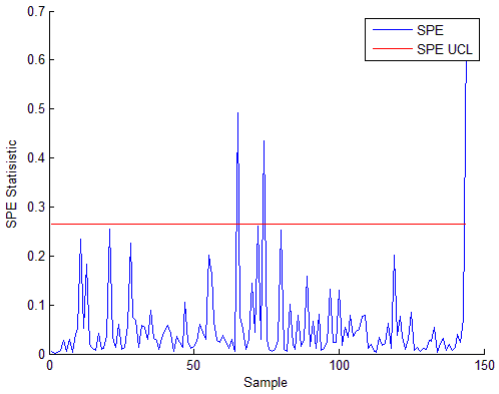


Figure 92 SPE plot PCA dataset 3

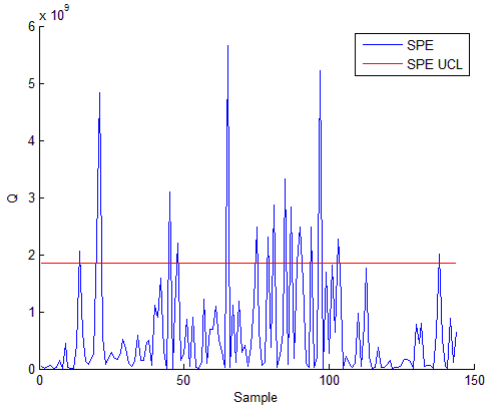


Figure 93 SPE plot NLPCA dataset 3

In simulation 2, fault in speed measurement of pump is simulated. A bias of 20 RMS is added on the original measurement at sample 36 and is removed at sample 108.

Models	Counted Violations
PCA Dataset 1: $[\Delta p\ w\ q]$	3
PCA Dataset 2: $[\Delta p\ w\ q\ \mathrm{q}^2]$	0
PCA Dataset 3: $[\Delta p\ w\ q\ \mathrm{q}^2\ \mathrm{w}^2]$	0
NLPCA Dataset 3: $[\Delta p\ w\ q\ \mathrm{q}^2\ \mathrm{w}^2]$	11

Table 33 Number of violating samples in 1 day interval

The plots of the SPE statistics together with the upper control limits (found from the normal operating data) are shown in figure 90-93. Table 33 shows how many samples that violate the upper control limits for PCA- and NLPCA models. In this simulation the NLPCA is the best for identification of the fault, but the number of violations may not convincing.

Simulation 3, without MBPCA

FAULT 3

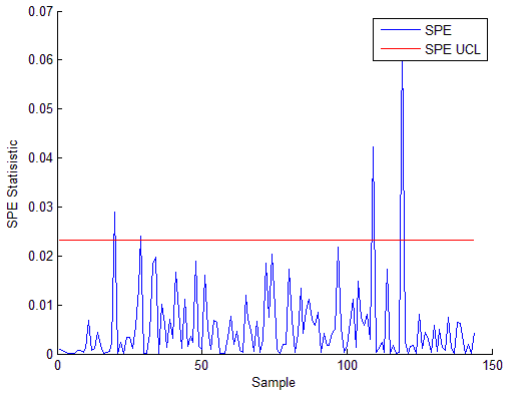


Figure 94 SPE plot PCA dataset

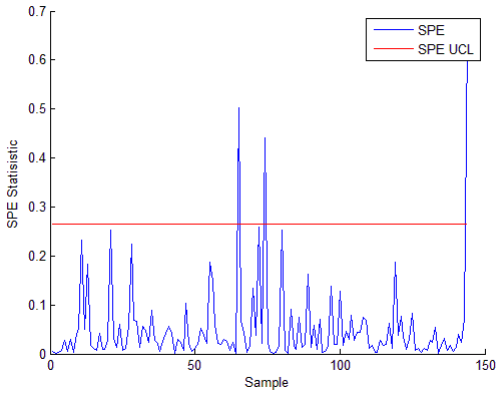


Figure 95 SPE plot PCA dataset 2

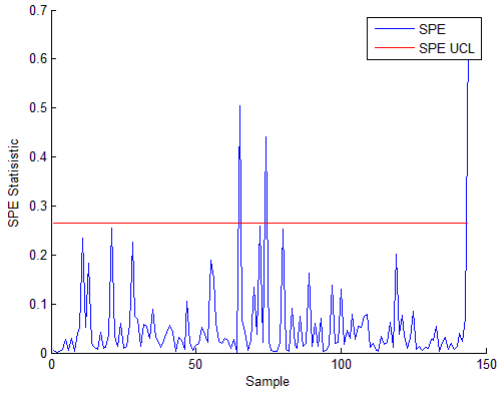


Figure 96 SPE plot PCA dataset 3

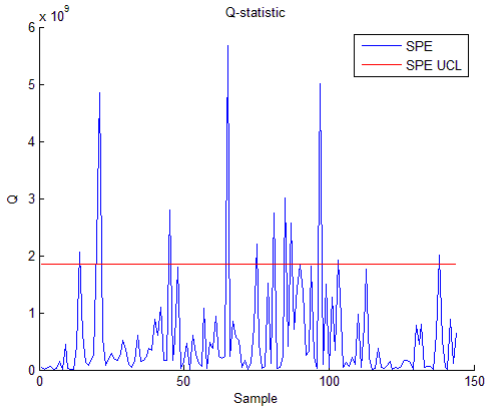


Figure 97 SPE plot NLPCA dataset 3

In simulation 3, fault due to hydraulic losses in the pump passages is simulated. A change of parameter constant is change to 3000 at sample 36 and is removed at sample 108.

Models	Counted Violations
PCA Dataset 1: $[\Delta p\ w\ q]$	0
PCA Dataset 2: $[\Delta p\ w\ q\ \mathrm{q}^2]$	0
PCA Dataset 3: $[\Delta p\ w\ q\ \mathrm{q}^2\ \mathrm{w}^2]$	0
NLPCA Dataset 3: $[\Delta p\ w\ q\ \mathrm{q}^2\ \mathrm{w}^2]$	6

Table 34 Number of violating samples in 1 day interval

The plots of the SPE statistics together with the upper control limits (found from the normal operating data) are shown in figure 94-97. Table 34 shows how many samples that violate the upper control limits for PCA- and NLPCA models. In this simulation the NLPCA is best for identification of the fault, but the number of violations is not convincing.

Simulation 4, without MBPCA

FAULT 4

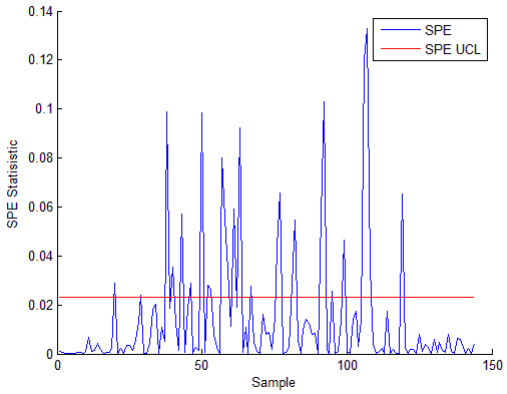


Figure 98 SPE plot PCA dataset 1

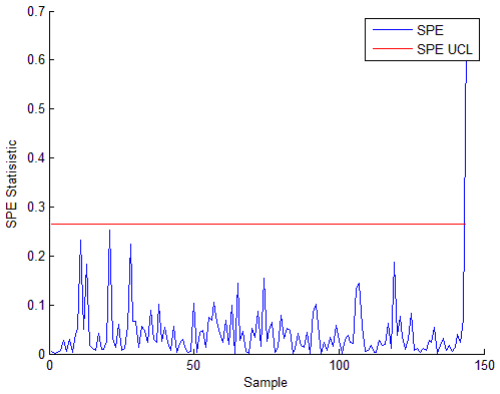


Figure 99 SPE plot PCA dataset 1

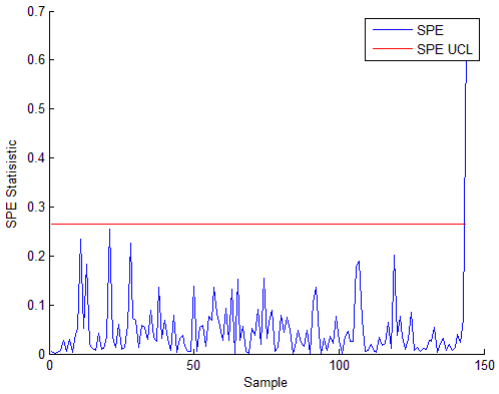


Figure 100 SPE plot PCA dataset 3

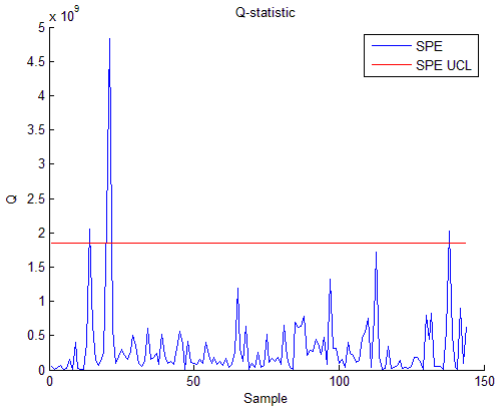


Figure 101 SPE plot NLPCA dataset 3

In simulation 4 a fault a change in operating condition is simulated. In this simulation the flow is increased by 0.04 m^3/s thorough the pump at sample 36 and is removed at sample 108.

Models		Counted Violations
PCA	Dataset 1: $[\Delta p\ w\ q]$	23
PCA	Dataset 2: $[\Delta p\ w\ q\ \mathrm{q}^2]$	-2
PCA	Dataset 3: $[\Delta p\ w\ q\ \mathrm{q}^2\ \mathrm{w}^2]$	-2
NLPCA	Dataset 3: $[\Delta p\ w\ q\ \mathrm{q}^2\ \mathrm{w}^2]$	-2

Table 35 Number of violating samples in 1 day interval

The plots of the SPE statistics together with the upper control limits (found from the normal operating data) are shown in figure 98-101. Table 35 shows how many samples that violate the upper control limits for PCA- and NLPCA models. The fault simulated here is not always considered as a fault. It could be desirable to not detect faults do to small changes in operation conditions to make the models more robust. If this is desirable the PCA model based on dataset 1 would generate a lot of false alarms and therefore the three other models would be preferable.

Simulations with MBPCA

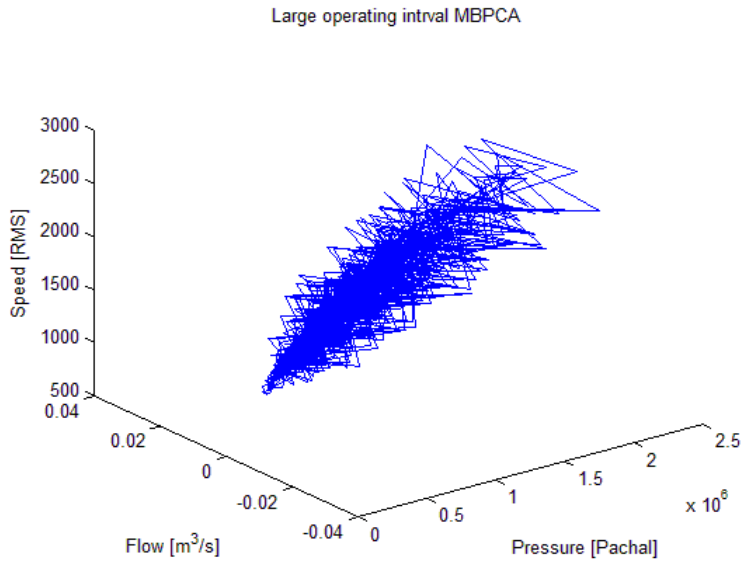


Figure 102 dimensional plot of speed, flow residual and pressure on NOC_{mbpca}

In figure 102 a 3 dimensional plot of the data generated for NOC. The figure contains the same measurements as for NOC without MBPCA except for the flow that is the residual from the estimated flow subtracted from the measurement serve as the real one.

The combined data from the experiments to obtain the NOC are shown in figure 103. For an explanation of the variables see table 28.

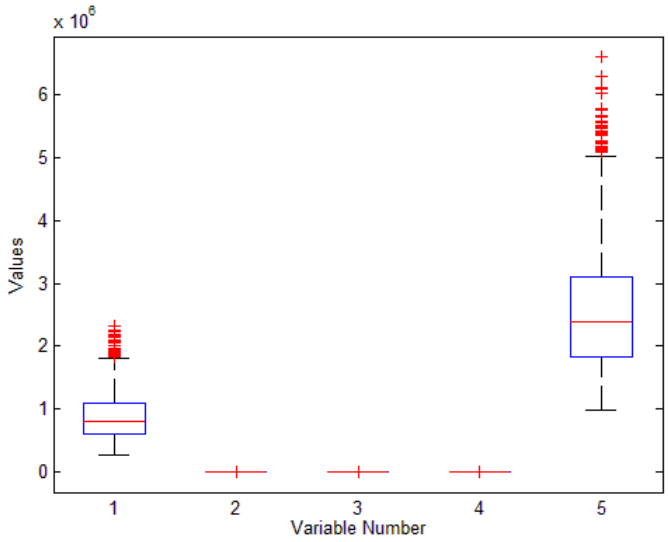


Figure 103 Box plot for NOC using MBPCA

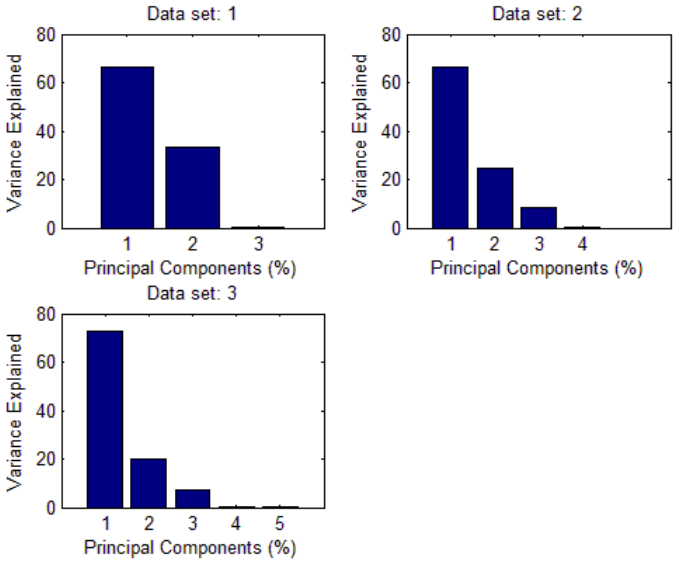


Figure 104 Variance from NOC with MBPCA

After the generation of NOC data from the pump model the data was manipulated (as discussed in the intro of chapter 6.2) before they was used to form PCA - and NLPCA models. The number of PCs variance is given in figure 104. The PCs was selected based on average eigenvalue method (see chapter 2.3). The method suggested two PCs for all three models. Since no PCs other then the one suggested were near 1 the suggestion was followed.

PCA models	**1.PCs**	**2.PCs**	**Variance explained**
PCA Dataset:1	66.6 %	33.15 %	99.76 %
PCA Dataset:2	66.55 %	24.88 %	91.43 %
PCA Dataset:3	72.51 %	19.92 %	92.43 %

Table 36 Principal components variance for the different PCA models on pump large range NOC

The PCs variance selected are given in table 32. After the selection of principal components the SPE plot for normal operation region and upper control limit was generated for each of the models (see figure 105-107). The number of samples that violated the upper control limits was counted for each of the models and sat as an acceptable number of violations meaning that if no more samples violate the UCL the process was defined to act normal (without faults). In table 37 the models with there corresponding dataset and number of normal operation region violations (Legal violations) are presented. For a discussion of the UCL and the legal violations see chapter 3.2

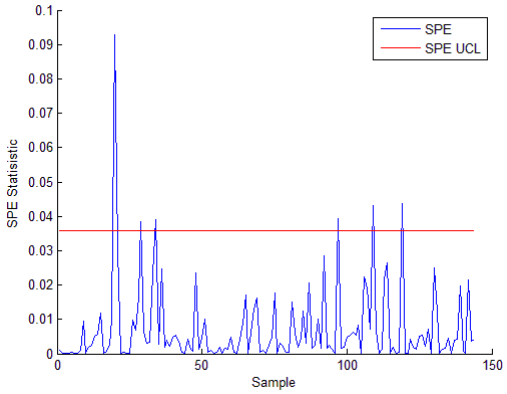


Figure 105 SPE plot PCA dataset 1

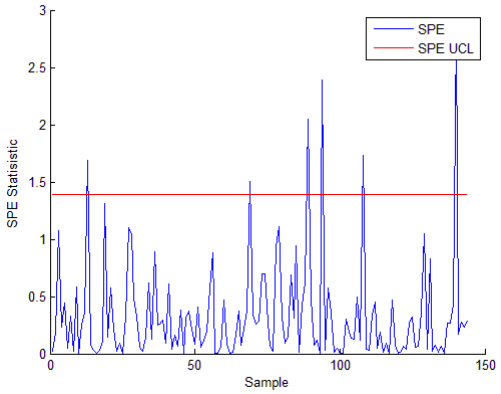


Figure 106 SPE plot PCA dataset 2

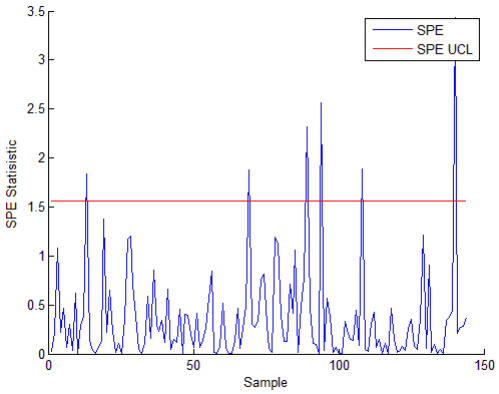


Figure 107 SPE plot PCA dataset 3

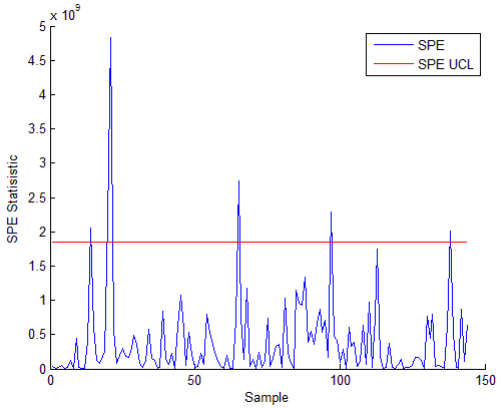


Figure 108 SPE plot NLPCA dataset 3

Models	NOC violations
PCA Dataset 1: $[\Delta p\ w\ (q-\hat{q})]$	6
PCA Dataset 2: $[\Delta p\ w\ (q-\hat{q})\ (q-\hat{q})^2]$	6
PCA Dataset 3: $[\Delta p\ w\ (q-\hat{q})\ (q-\hat{q})^2\ \mathrm{w}^2]$	6
NLPCA Dataset 4: $[\Delta p\ w\ (q-\hat{q})\ (q-\hat{q})^2\ \mathrm{w}^2]$	6

Table 37 Normal operation region violations

After the PCA models were generated the NLPCA method based on autoassosiatve network was calculated. This method does not give the opportunity to select PCs based on the magnitude as for the PCA method. The PCs was here selected by a trial and error approach and sat to 1. The data was also here manipulated. After the model was finished the SPE and UCL was plotted (see figure 108).

Simulation 5, with MBPCA

FAULT 1

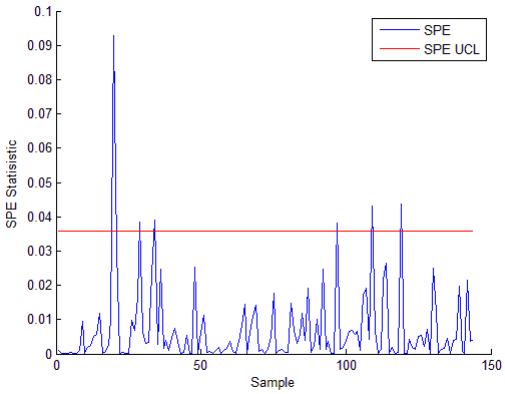


Figure 109 SPE plot PCA dataset 1

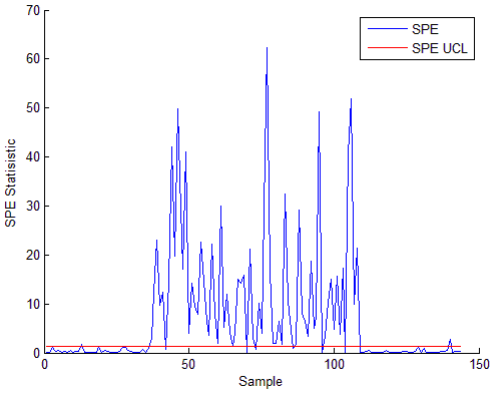


Figure 110 SPE plot PCA dataset 2

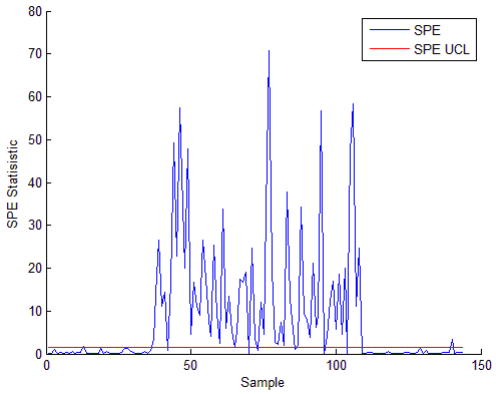


Figure 111 SPE plot PCA dataset 3

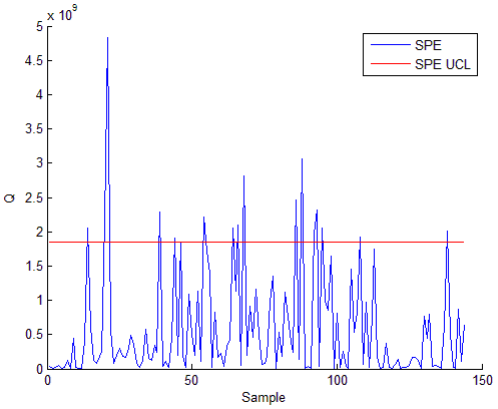


Figure 112 SPE plot NLPCA dataset 3

In simulation 5, fault in pressure measurement over the pump is simulated. A bias of 0.2 bar is added on the original measurement at sample 36 and is removed at sample 108.

Models		Counted Violations
PCA	Dataset 1: $[\Delta p\ w\ (q-\hat{q})]$	0
PCA	Dataset 2: $[\Delta p\ w\ (q-\hat{q})\ (q-\hat{q})^2]$	61
PCA	Dataset 3: $[\Delta p\ w\ (q-\hat{q})\ (q-\hat{q})^2\ \mathrm{w}^2]$	61
NLPCA	Dataset 3: $[\Delta p\ w\ (q-\hat{q})\ (q-\hat{q})^2\ \mathrm{w}^2]$	11

Table 38 Number of violating samples in 1 day interval

The plots of the SPE statistics together with the upper control limits (found from the normal operating data) are shown in figure 109-112. Table 38 shows how many samples that violate the upper control limits for PCA- and NLPCA models. PCA with dataset 2 and 3 are showing superior result in detecting the fault. NLPCA are also indicates some error.

Simulation 6, with MBPCA

FAULT 2

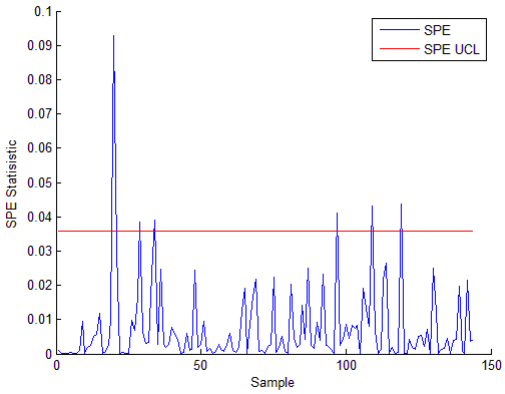


Figure 113 SPE plot PCA dataset 1

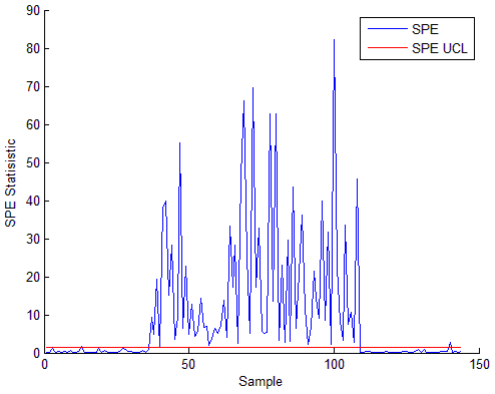


Figure 114 SPE plot PCA dataset 2

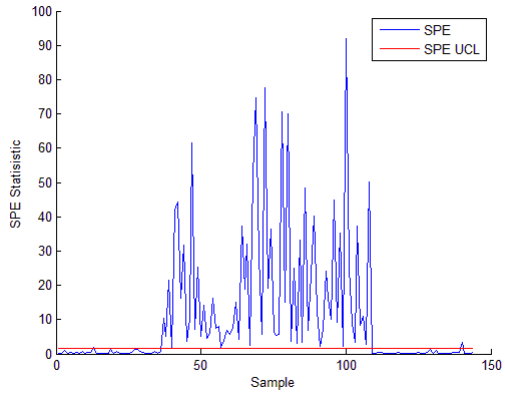


Figure 115 SPE plot PCA dataset 3

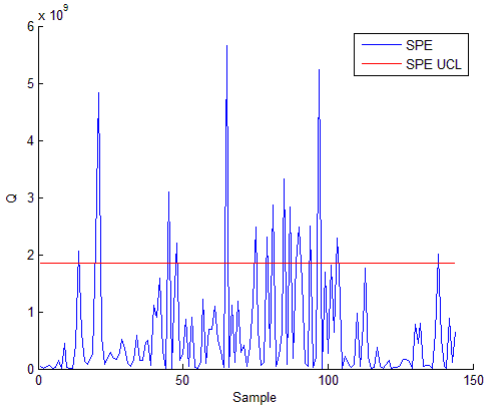


Figure 116 SPE plot NLPCA dataset 3

In simulation 6, fault in speed measurement of pump is simulated. A bias of 20 RMS is added on the original measurement at sample 36 and is removed at sample 108.

Models		Counted Violations
PCA	Dataset 1: $[\Delta p\ w\ (q - \hat{q})]$	0
PCA	Dataset 2: $[\Delta p\ w\ (q - \hat{q})\ (q - \hat{q})^2]$	68
PCA	Dataset 3: $[\Delta p\ w\ (q - \hat{q})\ (q - \hat{q})^2\ \mathrm{w}^2]$	67
NLPCA	Dataset 3: $[\Delta p\ w\ (q - \hat{q})\ (q - \hat{q})^2\ \mathrm{w}^2]$	11

Table 39 Number of violating samples in 1 day interval

The plots of the SPE statistics together with the upper control limits (found from the normal operating data) are shown in figure 113-116. Table 39 shows how many samples that violate the upper control limits for PCA- and NLPCA models. PCA with dataset 2 and 3 are showing superior result in detecting the fault. NLPCA are also indicates some error.

Simulation 7, with MBPCA

FAULT 3

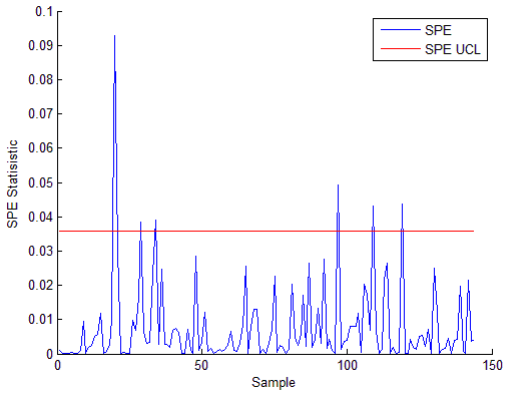


Figure 117 SPE plot PCA dataset 1

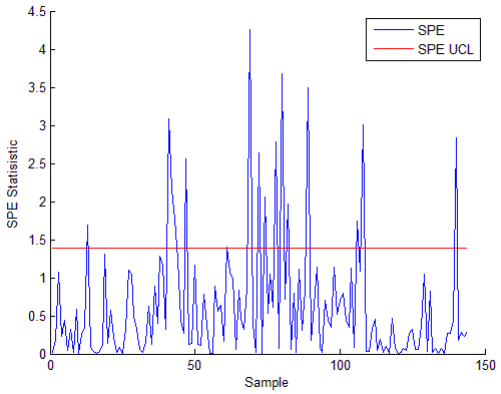


Figure 118 SPE plot PCA dataset 2

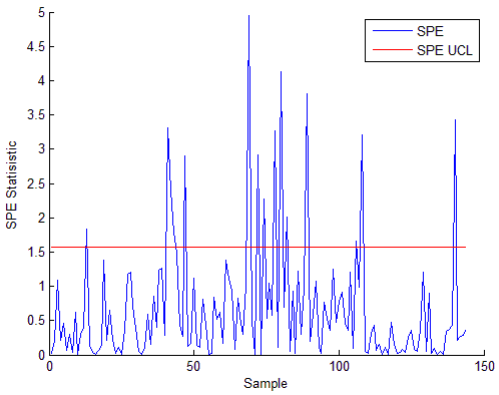


Figure 119 SPE plot PCA dataset 3

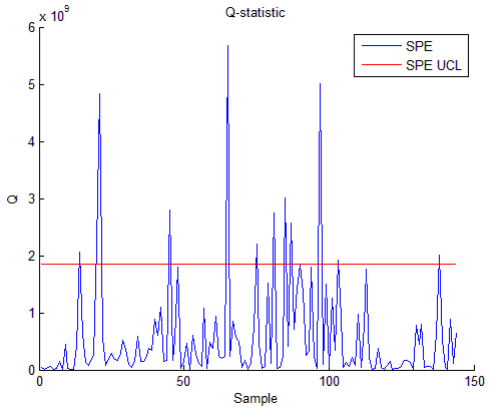


Figure 120 SPE plot NLPCA dataset 3

In simulation 7, fault due to hydraulic losses in the pump passages is simulated. A change of parameter constant is change to 3000 at sample 36 and is removed at sample 108.

Models		Counted Violations
PCA	Dataset 1: $[\Delta p\ w\ (q-\hat{q})]$	0
PCA	Dataset 2: $[\Delta p\ w\ (q-\hat{q})\ (q-\hat{q})^2]$	10
PCA	Dataset 3: $[\Delta p\ w\ (q-\hat{q})\ (q-\hat{q})^2\ \mathrm{w}^2]$	9
NLPCA	Dataset 3: $[\Delta p\ w\ (q-\hat{q})\ (q-\hat{q})^2\ \mathrm{w}^2]$	6

Table 40 Number of violating samples in 1 day interval

The plots of the SPE statistics together with the upper control limits (found from the normal operating data) are shown in figure 117-120. Table 40 shows how many samples that violate the upper control limits for PCA- and NLPCA models. PCA with dataset 2 and 3 and NLPCA indicates some error but not convincing results in detecting the fault.

Simulation 8, with MBPCA

FAULT 4

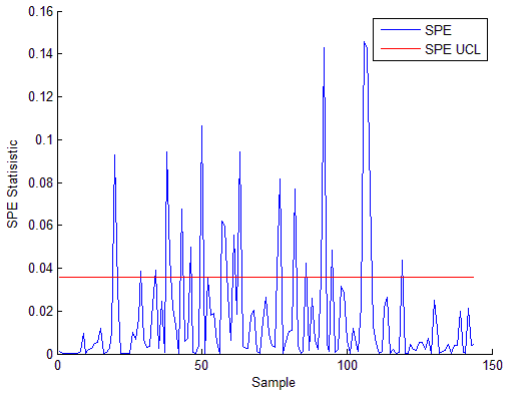


Figure 121 SPE plot PCA dataset 1

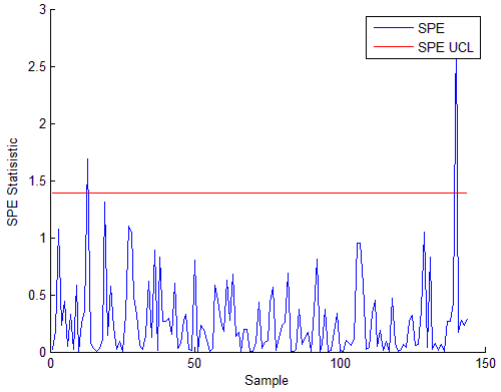


Figure 122 SPE plot PCA dataset 2

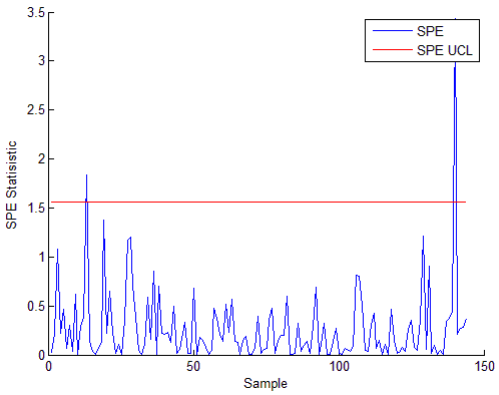


Figure 123 SPE plot PCA dataset 3

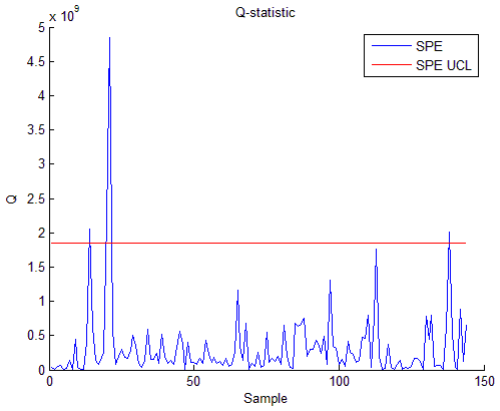


Figure 124 SPE plot NLPCA dataset 3

In simulation 8, change in operating condition is simulated. In this simulation the flow is increased by 0.04 m^3/s thorough the pump at sample 36 and is removed at sample 108.

Models		Counted Violations
PCA	Dataset 1: $[\Delta p\ w\ (q-\hat{q})]$	16
PCA	Dataset 2: $[\Delta p\ w\ (q-\hat{q})\ (q-\hat{q})^2]$	-4
PCA	Dataset 3: $[\Delta p\ w\ (q-\hat{q})\ (q-\hat{q})^2\ \mathrm{w}^2]$	-4
NLPCA	Dataset 3: $[\Delta p\ w\ (q-\hat{q})\ (q-\hat{q})^2\ \mathrm{w}^2]$	-2

Table 41 Number of violating samples in 1 day interval

The plots of the SPE statistics together with the upper control limits (found from the normal operating data) are shown in figure 121-124. Table 41 shows how many samples that violate the upper control limits for PCA- and NLPCA models. The fault simulated here is not always considered as a fault. It could be desirable to not detect faults do to small changes in operation conditions to make the models more robust. If this is desirable the PCA model

Chapter 7

Conclusion

In table 26 a summary of the fault detected from the simulation experiments on heat exchanger process are listed. These samples are the violations from the fault dataset subtracted with the NOC violations. All the PCA models are used with two PCs and all the NLPCA is based on 6 PCs. All the MBPCA methods are based on using the residual from the hot output temperature measurement subtracted with an estimate. There are three different estimates used in the MBPCA approaches are implemented with different degree of nonlinearities.

Based on the simulations it's evident that using MBPCA where the Model is close to the real system concerning to nonlinearities increases the detect ability of the faults dramatically when using the PCA models.

When using a linear model for the MBPCA structure, there is small improvement compared to the PCA models without MBPCA.

When using NLPCA the result is opposite. This method works best without MBPCA. This method is based on selection of PCs without getting a suggestion before making the model. Therefore it would be wrong to brush aside this method for MBPCA. But on the other hand the method is hard to tune and therefore has a drawback from the other methods. In addition if the method takes more time to tune because of optimization and are not as intuitive as the PCA.

In each structure there is used different datasets for making the PCA models but the improvement for using estimates for missing measurements are small for the faults simulated.

Without MBPCA	**Fault1**	**Fault2**	**Fault3**	**Sum**
PCA1	**-4**	**4**	**1**	**1**
PCA2	**-4**	**8**	**0**	**4**
PCA3	**-4**	**5**	**0**	**1**
NLPCA	**48**	**64**	**1**	**113**
Linear MBPCA				
PCA1	**0**	**10**	**7**	**17**
PCA2	**1**	**15**	**3**	**19**
PCA3	**5**	**20**	**6**	**31**
NLPCA	**0**	**0**	**30**	**30**
Non-Linear MBPCA				
PCA1	**39**	**91**	**10**	**140**
PCA2	**44**	**90**	**13**	**147**
PCA3	**53**	**89**	**12**	**154**
NLPCA	**1**	**1**	**1**	**3**
Perfect Non-Linear MBPCA				
PCA1	**89**	**93**	**11**	**193**
PCA2	**91**	**92**	**10**	**193**
PCA3	**95**	**91**	**10**	**196**
NLPCA	**1**	**1**	**46**	**48**

Table 42 Result from number of violations on heat exchanger

In table 43 a summary of the fault detected from the simulation experiments on centrifugal pump process. Based on the result from the simulations in chapter 6.2, the experiments without using MBPCA gave poor performance concerning detection of fault 1-3. Fault 4 is not always considered as a fault, because we sometimes have drifting processes concerning normal changes. If we place emphasis on this the PCA1 model would here trig a lot of false alarms where as the other models do not. This could be to the fact that we are using virtual measurements containing nonlinear functions of the measurements. These nonlinear functions are used because of the pressure – and flow has quadratic relationships in the pump model.

Without MBPCA	**Fault1**	**Fault2**	**Fault3**	**Fault4**	**Sum**
PCA1	**3**	**3**	**0**	**23**	**29**
PCA2	**0**	**0**	**0**	**-2**	**-2**
PCA3	**0**	**0**	**0**	**-2**	**-2**
NLPCA	**11**	**11**	**6**	**-2**	**26**
With MBPCA					
PCA1	**0**	**0**	**0**	**16**	**16**
PCA2	**61**	**68**	**10**	**-4**	**135**
PCA3	**61**	**67**	**9**	**-4**	**133**
NLPCA	**11**	**11**	**6**	**-2**	**26**

Table 43 Result from number of violations on centrifugal pump

For the experiments using MBPCA the increase in performance are tremendous with the respect to PCA2- and PCA3 models. The NLPCA shows no improvement in these simulations.

Based on the result from the two process models, MBPCA and virtual measurements are an efficient method to improve the PCA for fault detections. It looks like the supplement on using MBPCA make the PCA model more suitable for non-linear data even if the model used are not exactly correct. When it comes to using NLPCA here based on Autoassociative Neural Networks the results are less clear.

Chapter 8

Further work

To improve and extend on this work in fault detection, the following avenues are open for exploration:

1. Investigate the use of NLPCA method used in this work with different tuning.

2. Use different datasets for online process

3. Investigate the kernel based PCA method for the process models in this work.

4. Investigate the fault detection and performance with multiple simultaneous faults.

5. Use PCs plots in addition to the SPE statistics on the process models in this work.

6. Quantify process performance improvement with the use of a online fault detection and diagnosis system.

Bibliography

[1] Sourabh Dash and Venkat Venkatasubramanian. Challenges In the Industrial Applications of Fault Diagnostic Systems. Laboratory for intelligent Process Systems, School of Chemical Engineering, Purdue University, W. Lafayette, IN 47907, USA

[2] S.Lachman-shalem, N.Haimovitch, E.N. Shauly and D.R. Lewin. MBPCA Application for Fault Detection in NMOS Fabrication. PSE Research Group, Department of Chemical Engineering, Technoin, Haifa 32000, Israel.

[3] Hyung Dae Jin, Young-Hak Lee, Gibaek Lee, and Chonghun Han. Robust Recursive Principal Component Analysis Modeling for Adaptive Monitoring. Ind.Eng. Chem. Res. 2006, 45 696-703

[4] Jong-Min Lee, ChangKyoo Yoo, Sang Wook Choi, Peter A.Vanrolleghem, In-Beum Lee, Nonlinear process monitoring using kernel principal component analysis. Chemical Engineering Science 59 (2004) 223-234

[5] Xi Sun, Horacio J. Marquez, Tongwen Chen. An improved PCA method with application to boiler leak detection. ISA Transactions 44 (2005) 379–397

[6] Wenfu Ku, Robert H. Storer, Christos Georgakis. Disturbance detection and isolation by dynamic principal component analysis. Chemometric and intelligent laboratory systems 20 (1995) 179-196

[7] Multi-disciplinary, multi-user process monitoring: Cross-discipline development and cross- company collaboration. E. Lunde/Statoil, K. Hovda/ABB and J. Spjøtvold/ABB

[8] Gudmundur Jonsson (1990). Parameter Estimation in Models of Heat Exchangers and Geothermal Reservoirs, Department of Mathematical Statistics, Lund Institute of Technology. Page 3

[9] Y.Rotem, A. Wachs and D.R Lewin. Ethylene Compressor Monitoring using Model-Based PCA. Wolfson Department of Chemical Engineering Technion I.I.T., Haifa 32000, Israel

[10] Mathisen, K.W.; Morari, M.; Skogestad, S. (1993). *Dynamic Models for Heat Exchangers*
and *Heat Exchanger Networks*. ESCAPE 3, Graz, Austria, July.

[11] Rolf Iserman Fault-Diagnosis Systems An Introduction from Fault Detection to Fault Tolerance Springer ISBN-10 3-540-24112-4 Springer Berlin Heidelberg New York. ISBN-13 978-3-540-24112-6 Springer Berlin Heidelberg New York

[12] Jong-Min Lee, ChangKyoo Yoo, Sang Wook Choi, Peter A.Vanrolleghem, In-Beum

Lee, Nonlinear process monitoring using kernel principal component analysis. Chemical Engineering Science 59 (2004) 223-234

[13] S.Joe Qin, Statistical process monitoring: basics and beyond, J. Chemometrics, 2003;17: Chapter 3.6 page 485

[14] Centrifugal pump model MathWorks
Available online at:
http://www.mathworks.com/help/toolbox/physmod/hydro/ref/centrifugalpump.html Date: 02.02.2011

[15] Sergio Valle, Weihua Li, and S. Joe Qin. Selection of the Number of principal Components. Department of Chemical Engineering, The University of Texas at Austin, Austin, Texas 78712

[16] IEC 9906:1999: Rotodynamic pumps. Hydraulic performance acceptance tests. Grades 1 and 2. Brussels 1999.

[17] Nonlinear PCA toolbox for MATLAB by Matthias Scholz
Available online at: http://www.nlpca.de/matlab.html Date: 02.02.2011

[18] Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and*
Psychological Measurement, 20, 141-151.

[19] Huizhi Bao, Faisal Khan, Tariq Iqbal, and Yanjun Chang. Risk-Based Faul Diagnosis and Safety Management for Process Systems. Process Engineering, Faculty of Engineering and Applied Science, Memorial University, St. John's, NL, Canada A1b 3X5 2010.

[20] Uwe Kruger, Junping Zhang, and Lie Xie. Developments and Applications of Nonlinear Principal Component Analysis – a Review. Available online at:
http://pca.narod.ru/1MainGorbanKeglWunschZin.pdf Date: 12.12.2010

[21] Peiling Cui, Junhong Li, Guizeng Wang. Improved kernel principal component analysis for fault detection. Expert Systems with Applications 34 (2008) 1210-1219. Available online at: www.sciencedirect.com Date 07.12.2010.

[22] Jun Liang and Ning Wang. Fault Detection and Isolation Based on PCA: An Industrial Reheating Furnace Case Study. Department of Control Science and Engineering Zhejiang University Hangzhou 310027, China

[23] Philip R.C. Nelson, Paul A. Taylor, John F. MacGregor. Missing data methods in PCA and PLS: Score calculations with incomplete observations. Chemometrics and Intelligent Laboratory Systems 35 (1996) 45-65

Appendixes

A. Calculation of pump coefficients

To calculate the unknown's coefficient in the pump model polynomial below, a pump curve presented in the figure x below are used. This pump curve is from a real pump on a offshore installation. The pump model polynomial is put into matrix and vectors. And head pressure and corresponding flow is taken from one of the pump curves. The pump design flow q_D is also used from the pump curve below.

q_DDifferent values of head pressure and flow is and values of Head pressure and flow is coefficients are calculated with estimation A number of poinhead pressures and flows under the same velocity 2850 RMS is

$$p = \rho_{ref}\left(k\left(c_0 - c_1 q_p\right) - c_2 q_p^2 - c_3\left(q_D - q_p\right)^2\right) \qquad (3.18)$$

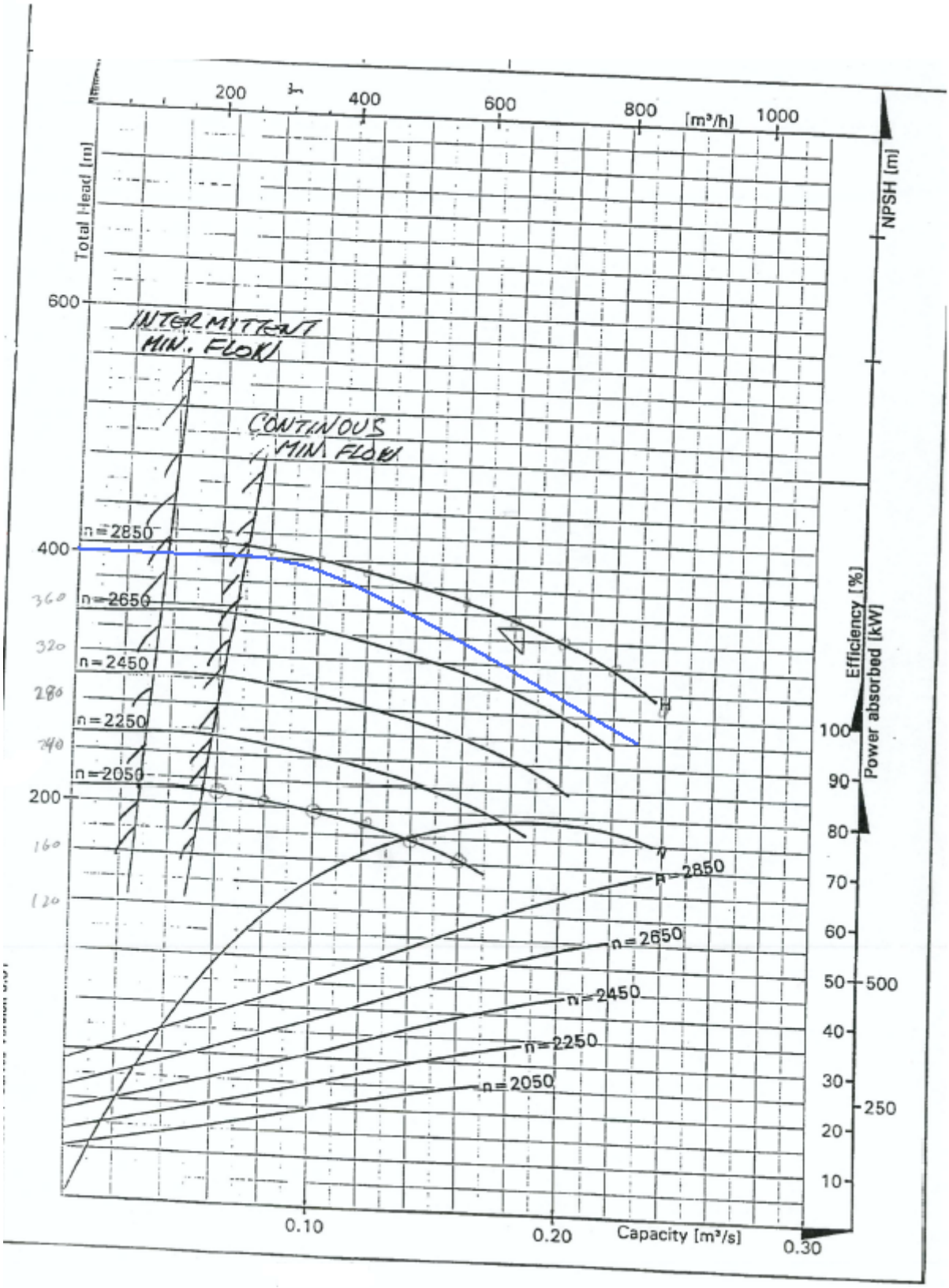
The pump polynomial is set into a matrix with

$$i = [1]^T, \qquad j = -[q_{p1}^2\, q_{p2}^2\, \dots q_{p3}^2\,]^T, \qquad j = -[q_{p1}^2\, q_{p2}^2\, \dots q_{p3}^2\,]^T$$

$$k = -[q_{p1}^2\, q_{p1}^2 \dots q_{p8}^2]^T, l = -[\left(q_d - q_{p1}\right)^2\, \left(q_d - q_{p2}\right)^2 \dots\, \left(q_d - q_{p2}\right)^2]$$

$$A = \frac{1}{g}[i\ j\ k\ l]\,, b = [H_1\ H_2 \dots H_8]^T, x = [c_0\ c_1\ c_2\ c_3]^T$$

Then we solve the equation $Ax = b$ with respect to x. Since there it's more equations then unknowns we use the matlab backlash operator \. This is the solution in the least squares sense to the under- or over determined system of equation $Ax = b$.


200
400
600
800
[m³/h]
1000
Total Head [m]
NPSH [m]
600
INTERMITTENT MIN. FLOW
CONTINOUS MIN. FLOW
n=2850
400
360
n=2650
320
n=2450
280
n=2250
240
n=2050
200
160
120
H
n=2850
n=2650
n=2450
n=2250
n=2050
100
90
80
70
60
50
40
30
20
10
Efficiency [%]
Power absorbed [kW]
500
250
0.10
0.20
Capacity [m³/s]
0.30

B. CD Contents

The enclosed CD contains the following:

1. SIMULINK model of heat exchanger
2. MATLAB script heat exchanger
3. SIMULINK model of centrifugal pump
4. MATLAB script centrifugal pump
5. PCA MATLAB script
6. NLPCA MATLAB script