Robust Probabilistic Slow Feature Analysis for Soft Sensor Development and Model Quality Assessment

by

Cameron Dyson

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

CHEMICAL ENGINEERING

Department of Chemical and Materials Engineering
University of Alberta

Abstract

Model predictive control (MPC) is a popular advanced control technology. Unfortunately, over time the behaviour of the plant may deviate from its initial design conditions resulting in model-plant-mismatch. The detection and diagnosis of such mismatches is an important task to ensure that MPC systems are operating optimally, and any potential model re-identification is targeted to only necessary sub-models. Conventional mismatch detection methods directly use plant operating data for such purposes. The quality of assessment of these methods may suffer in the presence of significant disturbances. In Chapter 2, a linear slow feature analysis (SFA) data reconstruction is proposed to remove fast and typically irrelevant variations, extracting only those slow-varying and important components of the data to detect model-plant-mismatches. This preprocessing approach is shown to improve the performance of a conventional model-plant-mismatch detection method through both simulated and industrial case studies, and thus provide a more targeted selection of sub-models for re-identification.

As SFA does not directly model process noise, and the conventional probabilistic SFA (PSFA) extension treats all noises as Gaussian, these algorithms are susceptible to the presence of outliers in the data. As industrial data often contains outliers there is motivation to remedy this issue. In Chapter 3, a robust PSFA (rPSFA) method with the measurement noises modeled as a scale mixture of Gaussians, switched according to a Bernoulli distribution, is considered for the modelling of systems where data contains outliers. To demonstrate the effectiveness of the proposed method over regular SFA, conventional PSFA and a previously developed Student-t robust PSFA, simulations are conducted through

Tennessee-Eastman benchmark process data. The algorithm is then applied to an industrial zinc roaster process.

The developed rPSFA models the switching between inliers and outliers according to a Bernoulli distribution which is completely random with respect to the previous outlier-inlier state. Many industrial systems exhibit correlated noise behaviour in which an outlier is more likely to occur after another. To account for this, Chapter 4 replaces the Bernoulli distribution with a Hidden Markov Model (HMM) to allow for the previous measurement noise mode to influence the prediction of future noise modes. Further, current literature lacks an outlier robust PSFA based method that is designed to capture the behaviour of multi-modal systems. To this end, the proposed HMM based robust PSFA is implemented in a mixture model fashion, where multiple independent process models are developed simultaneously, and their results are blended according to some weightings. The proposed model is verified in a soft-sensor task for a simulated system with a single operating mode but with outliers generated according to a HMM. Additionally, an industrial system which contains outliers and displays two distinct operating modes is used to demonstrate the development of a soft-sensor and a MPC model-plant-mismatch detection and diagnosis task.

”Slow and steady wins the race.”

- Robert Lloyd

To my family whose continual support cannot be lauded enough.

Acknowledgments

First and foremost, I would like to thank my supervisor Professor Biao Huang. Without him I would not have had the opportunity to be a member of the Computer Process Control group. Further, his help in identifying relevant research and guiding me along my way has been greatly appreciated.

I also greatly appreciate the help that has been given to me by two postdoctoral fellows, Dr. Santhosh Kumar Varanasi, and Dr. Jayaram Valluru. Their expertise and willingness to explain new concepts has ensured that they were both huge aides in my studies. Further their feedback on my work has helped me find mistakes that I could have otherwise missed and has also helped me to improve my writing style.

Next, I would like to acknowledge the support I have had from all members of the Computer Process Control group. The feedback I have received from our weekly presentations has been a blessing towards ensuring that I am able to make what I am presenting understandable. Additionally, those who have sent me relevant papers and tutorials on topics I am interested in have allowed me to spend less time looking for material and more time learning it. The opportunity to be a marker working with Junyao and Krish, two group members, was also a great chance to see another aspect of academia.

I should also show my gratitude to the Chemical and Materials Engineering department at the University of Alberta and the support staff therein for their help in completing my program requirements.

I would also like to thank the Natural Sciences and Engineering Research Council of Canada as well as the groups industrial partners for providing support, both in terms of

finances as well as providing useful and meaningful industrial data.

Lastly, I would like to thank my friends and family for their support through my program, especially through the uncertainty surrounding the COVID-19 pandemic.

Table of Contents

List of Tables

List of Figures

Abbreviations & Acronyms

ARX Autoregressive with Exogenous Variables.

corr. Pearson's Correlation Coefficient.

CV Controlled Variable.

DPCR Dynamic Principal Component Regression.

DPLS Dynamic Partial Least Squares.

EM Expectation Maximization.

FOPDT First Order Plus Dead Time.

HMM Hidden Markov Model.

mhrPSFA Mixtures of Hidden Markov Model Robust Probabilistic Slow Feature Analysis.

MIMO Multiple Input, Multiple Output.

MPC Model Predictive Control.

MQI Model Quality Index.

MQILOO Model Quality Index with Leave One Out approach.

MSE Mean Squared Error.

MV Manipulated Variable.

PCA Principal Component Analysis.

PCR Principal Component Regression.

PLS Partial Least Squares.

PPCR Probabilistic Principal Component Regression.

PPLS Probabilistic Partial Least Squares.

PSFA Probabilistic Slow Feature Analysis.

rPSFA Robust Probabilistic Slow Feature Analysis.

SFA Slow Feature Analysis.

TE Tennessee-Eastman.

tPSFA Student-t Robust PSFA.

Chapter 1

Introduction

1.1 Motivation

Model predictive control (MPC) is a popular advanced control method with many commercial applications in process industries. This popularity is due to the MPC framework providing the ability to automatically handle constraints, achieve economic optimization, recover from upsets, etc., for large scale multiple-input and multiple-output (MIMO) systems. MPC relies on a predictive model to determine the estimated outcome of possible future control actions, and thereby attempts to drive the plant towards a desired trajectory though an optimization solution. As any discrepancy could lead to sub-optimal controller performance, ensuring that the plant model utilized within a predictive controller is accurate with respect to its real-world equivalence is an important task. One approach to ensure model quality would be to perform frequent broad spectrum signal excitation experiments and update the model accordingly. In practice however, this is undesirable as it may require operation away from the economic optimum and have associated safety concerns. Instead, the specific sub-models where mismatch exists should be identified so that a reduced set of experiments can be conducted, and only when necessary. Many methods have been developed to detect and diagnose the presence of model-plant-mismatches at the input-output pair level.

However, real world data often includes quicker varying disturbances that are of little relevance to the plant dynamics and confound the attempts to detect and diagnose model-plant-mismatch. To this end, slow feature analysis (SFA) is an attractive option to eliminate the more quickly varying latent features of a data set while retaining those that vary more slowly and are more likely to be relevant to the model and its quality assessment. Current SFA methods however are limited when compared to other latent variable frameworks, such as PCA and PLS, when it comes to handling issues such as outlier dynamics and multiple process mode operation. While preprocessing, such as the commonly used 3σ outlier replacement algorithm, is possible, this may result in an undesirable loss of information. The motivation of this thesis is to develop a method of SFA based preprocessing to en-

hance model-plant-mismatch detection and diagnosis performance, as well as to develop new probabilistic SFA based methods to better handle common issues found in industrial data sets.

1.2 Literature Overview

In this work, two primary areas are considered, MPC model-plant-mismatch detection and diagnosis, and slow feature analysis. Summaries of the current literature regarding each of these two fields are given below.

MPC Model-Plant-Mismatch Detection and Diagnosis

In most MPC applications a predictive model of the plant is identified from previous operating data, identification experiments, or first principles during the commissioning stage. Over time, changes in feed stock, facility equipment, and operating points can result in the prediction model no longer being accurate. An inaccurate prediction model can result in deteriorated control performance of the MPC itself [1] as well as sub-optimal determination of set points in higher level optimizers that may depend on a steady state version of the prediction model [2, 3]. To this end, regular re-identification of the plant model could be performed. However, to get accurate identification results, greater excitation than that found in typical operating data may be needed [4]. As this additional excitation may require operation outside the economic optimum and have associated safety concerns, such experiments should be conducted sparingly and only target those input-output pairs where re-identification is necessary. To address this, several model-plant-mismatch detection and diagnosis methods have been developed for MPC systems [5–15]. The approaches taken to identify model-plant-mismatch vary, but one common method is to attempt to identify a new linear dynamic model over a short period of time and compare its behaviour to that of the currently used prediction model [16–19]. Such methods are confounded by the presence of high order, nonlinear process dynamics or large disturbances [20, 21] which are

commonly of a relatively quickly varying nature in chemical processes.

Additional detail of the current state of MPC model-plant-mismatch detection and diagnosis for industrial MIMO systems can be found in Chapter 2.

Slow Feature Analysis

Slow feature analysis [22] is a latent variable method that can extract the slowly varying underlying features of data. It has found use in the monitoring and modelling of processes where the features of interest are believed to be slowly varying in nature, such as industrial chemical processes [23–25], and human visual perception and movement [26, 27]. Similar to other latent variable methods it has been extended to a probabilistic framework [28, 29]. This framework has been further extended to address data quality concerns often found in industrial settings that are poorly handled by the original linear approach [30–34].

Chapters 2, 3 and 4 provide more information on the current state of latent variable methods with an emphasis on slow feature analysis for modelling and monitoring of chemical processes.

1.3 Thesis Outline and Contributions

In addition to this introduction, this thesis is structured in a three-paper format with introductions as follows:

In Chapter 2, a linear SFA based encoder-decoder structure is proposed as a data preprocessing step to de-emphasize the fast latent features of historical operating data. The objective of this is to obtain MPC model-plant-mismatch detection and diagnosis that are more targeted to the slow signals which tend to be of greater importance in chemical processes. The proposed method is demonstrated for simulated data with known mismatches as well as industrial data where the location and severity of mismatch within the plant model is unknown.

Chapter 3 discusses the development of a robust PSFA model that can overcome the

issues that outliers in training data cause when using conventional SFA and PSFA. This method is validated with soft-sensor development case studies for simulated and industrial data containing outliers.

In Chapter 4, the developed robust PSFA model is extended to further consider the correlated transition of inlier and outlier modes while also accounting for multiple process operating conditions. This method is validated using soft sensor case studies as well as an industrial MPC model-plant-mismatch detection and diagnosis task.

Finally, in Chapter 5, a summary of conclusions and potential future work related to the previous chapters is presented.

Chapter 2

MPC Model Quality Assessment Through Slow Feature Analysis Preprocessing

A version of this chapter was presented at the AdCONIP 2022 Symposium as: Cameron Dyson, Santhosh Kumar Varanasi, Graham Slot, Primo Majoko, and Biao Huang "MPC Model-Plant-Mismatch Detection Through Slow Feature Analysis Preprocessing with Industrial Application"

2.1 Introduction

Model predictive control (MPC) is a model-based control strategy that has found widespread applications in process industries owing to its ability to handle constraints, achieve economic optimization, etc., for large scale multiple-input and multiple-output (MIMO) systems [35, 36]. MPC computes a sequence of manipulated adjustments that optimize the future behavior of the plant based on a predictive model and operating constraints.

In most applications, a predictive model is identified during the commissioning stage of the MPC from historic operating data, first principles or system identification experiments. Such a model consists of two parts, a plant model which relates measured inputs to the outputs, and a disturbance model which relates unmeasured inputs to the outputs. The plant model, G, often consists of first order plus dead time (FOPDT) transfer functions between relevant input-output pairs, which are selected based on process knowledge or past experience. The disturbance component of the prediction model, H, may be identified in a similar manner, or more commonly, may be assumed simply to be a fixed random walk model. Some commercial MPC software packages include an additional planning module which may use a steady state version of the plant model to determine optimal set points based on some economic, environmental, or safety-based cost function. A simplified block diagram of an MPC structure is shown in Figure 2.1.

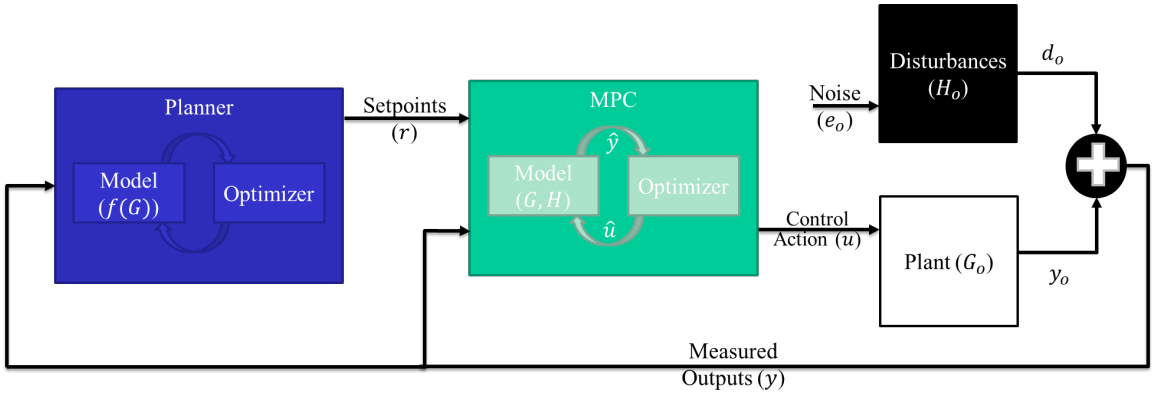

Figure 2.1: Simplified MPC Block Diagram with Planning Module.

While some degree of prediction error is unavoidable due to background noises, inaccuracies in the MPC's prediction model can exacerbate the degree of error. This can lead the MPC to recommend control actions that are not aligned with the plant's desired behaviour. In the case that the prediction model is less responsive to the inputs over the prediction horizon than the real-world plant, this will lead to more severe control actions than are strictly necessary, likely resulting in some overshoot of the set point and increased signal variance. Alternatively, if the prediction model is more responsive than its real-world equivalent, the recommended control actions will result in laggardly tracking of the set point and the system will be slow to reject disturbances. Further, any discrepancies between the predictions model's steady state gain and that of the real plant can result in planning modules recommending set points that are sub-optimal [3].

In industrial settings, changes in operation methods, feedstock, equipment, etc., can gradually lead to increasing mismatch between the plant and the model. As the accuracy of the predictive model used in the MPC is paramount to both its own performance and that of any planning modules, model-plant-mismatches should be detected so that they can be remedied through re-identification. Re-identification of the entire plant model should be avoided whenever possible as it may require perturbations to the plant. This procedure can disturb operation and may lead to safety concerns.

The objective of model quality assessment is then to detect model-plant-mismatch and diagnose which specific sub-models are the most likely source. This allows for a reduced set of experiments that target only the input-output pairs where mismatch exists, thereby reducing the potential loss of profit and safety. Several methods exist in the literature to perform model-plant-mismatch detection and diagnosis utilizing the predictive model and historical data. Some methods claimed to have good performance with industrial data are dynamic partial correlations between the prediction errors and inputs [13], model quality index (MQI) with a leave-one-out approach (MQILOO), the crossover of MQI [17] and correlation analysis between input and disturbance [18], and one-class support vector ma-

chines [37].

In most chemical processes, conditions relating to the plant do not vary quickly and have large inertia compared to the rapid varying nature of disturbances. Therefore, the slow-varying features of most chemical process settings are of main interest while the faster features often relate to the disturbances. As disturbance model changes are of less concern, especially when a fixed disturbance model is used in MPC, the ability to reconstruct the data based upon slower varying features for model-plant-mismatch detection is desirable. Slow feature analysis (SFA) [22] provides a framework to extract slowly varying latent features of data which can then be used to construct a more relevant data set for model quality analysis.

The main objective of the current chapter is to demonstrate the improvements in model-plant-mismatch detection and diagnosis through SFA based data reconstruction. The results are demonstrated utilizing the MQILOO approach for model-plant-mismatch detection and diagnosis in a simulated MIMO system with known mismatches as well as in an industrial data set for which the mismatches are unknown. The remainder of this chapter is organized as follows: in Section 2.2, methods for detecting model-plant-mismatch and concepts of slow feature analysis are provided, improvement though slow feature analysis are demonstrated in Section 2.3, followed by conclusions in Section 2.4.

2.2 Methods

In this section, the methods used in model quality assessment of MPC, and slow feature analysis are provided.

2.2.1 Model Quality Assessment

Consider a MIMO model which has n inputs, u, and m outputs, y. Based on the model, each of the outputs can be calculated with a corresponding one-step ahead prediction error, e. The MQI [17] of an output channel, y_i, is the ratio of the sum of the square of one-step

ahead prediction errors of a high order ARX model, e_{0i}, fitted to the input-output data in the selected window, to that of prediction from the model used by the MPC, e_i,

$$MQI_i = \frac{\sum_{t=1}^{N} e_{0i}^T e_{0i}}{\sum_{t=1}^{N} e_i^T e_i}, \quad i = 1, ..., m \tag{2.1}$$

This allows for model quality assessment on a per output basis. Values less than one indicate that the prediction error of the corresponding output channel could be improved by the re-identification of one or more of its associated sub-models. In order to determine the related inputs that correspond to a mismatched sub-model, a leave-one-out approach (MQILOO) is adopted. Here, the MIMO ARX model is fitted again by fixing the sub-model of one input channel to be the same as the predictive model at a time while all other inputs are used to fit the ARX model. When the MQILOO of an input-output pair increases relative to the MQI of the output, it suggests that the sub-model corresponding to that input-output likely has model-plant-mismatch. The ratio of MQILOO to MQI can then be tracked as an index to monitor the model quality.

Some degree of model mismatch is unavoidable in a nonlinear process modeled by FOPDT plant models with an assumed random walk disturbance model. Therefore, the indices for each sub-model are not monitored directly. Rather, the indices are compared to their averages from a reference period during which performance was considered acceptable. When this ratio exceeds a user determined confidence limit, then model-plant-mismatch can be determined, and re-identification may be performed.

2.2.2 Slow Feature Analysis

Several versions of slow feature analysis (SFA) have been developed since it was first proposed [22]. In general, they share some properties as described in this section.

Given an m-dimensional stochastic and ergodic signal with zero mean, $x(t)$, its temporal variance (referred to as speed) is defined as:

$$\Delta(x(t)) = \langle \dot{x}^2(t) \rangle_t \tag{2.2}$$

where $\langle\rangle_t$ denotes time-based averaging and $\dot{x}(t)$ is the derivative of $x(t)$ with respect to time, typically approximated using the first order finite difference. Note that if SFA is to be performed on a signal without zero mean then a normalization step can be performed beforehand.

The objective of SFA is to find a set of $j = 1,..,q$ functions, $g_j(x(t))$, to extract a set of slowly varying latent variables, $s(t)$, from the input signal, $x(t)$:

$$s_j(t) = g_j(x(t)) \tag{2.3}$$

such that the speed of the extracted features is minimized:

$$\min_{g_j} \Delta(s_j(t)) \tag{2.4}$$

with the following constraints:

$$\langle s_j(t)\rangle_t = 0 \tag{2.5}$$

$$\langle s_j^2(t)\rangle_t = 1 \tag{2.6}$$

$$\forall i \neq j, \langle s_i(t)s_j(t)\rangle_t = 0 \tag{2.7}$$

where (2.5) and (2.6) prevent trivial solutions and (2.7) ensures that slow features are uncorrelated with one another and are found in ascending order of their speed. The extracted slow features can then be used for dimensionality reduction, soft sensor modelling and process monitoring. In systems where the signals of concern are slowly varying and disturbed by some fast-varying noises, such as many chemical processes, SFA provides more relevant latent features than similar methods such as PCA.

In this work the linear version of SFA [22] is considered. Linear SFA assumes that the transform function, $g(x(t))$ is simply a linear mapping of the inputs, i.e.:

$$s(t) = W^T(x(t)) \tag{2.8}$$

where $W = [w_1,...,w_q] \in \mathbb{R}^{m\times q}$ is to be optimized. In order to satisfy (2.7) the number of features obtained must be $q \leq m$. In the case of $q = m$, the global optimum solution can be

found through solving a generalized eigenvalue problem:

$$AW = BW\Omega \tag{2.9}$$

where $A = \langle \dot{x}(t)\dot{x}^T(t)\rangle_t$ is the covariance of the temporal derivatives of $x(t)$, $B = \langle x(t)x^T(t)\rangle_t$ is the covariance of $x(t)$, and $\Omega = diag\{\omega_1, ..., \omega_q\}$ contains the generalized eigenvalues that are the optimal solutions to (2.4):

$$\Delta(s_j(t)) = \omega_j \tag{2.10}$$

Conventionally slow feature analysis is an unsupervised learning method where only the plant's inputs are used to find latent variables. Recently it has been shown that a supervised method that uses both the plant's input and output signals for the slow feature analysis has improved performance for regression tasks [31]. Additionally, dynamic-SFA, where multiple copies of the selected signals are included with introduced lags up to some value d, have been shown to also have improved performance [38, 39]. Therefore, both of these improvements are to be used in this work.

Once the slow features are found, they can be used to reconstruct each of the original input-output signals through a slow feature regression [40]. Here a linear regression is performed using the $q < p$ slowest features as inputs and each of the original plant input and output signals as targets:

$$\hat{x}_i(t) = b^T s_{1:q}(t) + c \tag{2.11}$$

where $b \in \mathbb{R}^q$ are regression coefficients and c is the bias term to be determined. Guidance on selecting the value of q can be found in [30]. In this way the original input-output signals can be encoded as slow features and then a linear regression used to decode the original signal from selected features, thereby reducing the prevalence of the fast components.

2.2.3 Proposed Method for Model-Plant-Mismatch Detection

In the presence of large disturbances and high order, or non-linear process dynamics, model-plant-mismatch methods that rely upon identifying a model, such as MQI, may

falsely attribute the complex disturbances to model-plant-mismatch. An SFA encoder-decoder step is proposed to reduce disturbances that typically vary quickly and retain the signals of interest that typically vary slowly. Based on the temporal variation of the normalized signals, chemical processes tend to be slowly varying but affected by some quickly varying disturbances; therefore, removing the effect of the fastest components should allow for a more targeted analysis of the plant. The proposed method is as follows:

Step 1: In offline training, collect plant inputs, $u(t)$, outputs, $y(t)$ during a period of time when MPC is considered to have adequate performance.

Step 2: Develop a linear SFA model to encode the signals. The input to SFA, $x(t)$, is the normalized versions of the collected plant inputs and outputs augmented with their lagged copies up to lag d.

Step 3: Reconstruct both the input and output data using slow feature regression to decode the measurements from q selected features as in equation (2.11).

Step 4: Use the prediction model and reconstructed data to simulate the 1-step ahead prediction error, $e(t)$.

Step 5: For each input-output pair, determine the ratio of MQILOO to MQI. Divide each of these by its average over the training period.

Step 6: Develop an upper confidence limit for these values. For example, an upper limit of three standard deviations above the mean is considered in this work.

Step 7: During the time period for the model to be monitored, collect the plant data from the same MPC controlled process and apply the standardization along with the lagged augmentation.

Step 8: Apply the SFA model from step 2 to encode, then the slow feature regression from step 3 to decode the newly collected data over a window to be monitored.

Step 9: Again, using the reconstructed data within the moving window, find the MQILOO to MQI ratio for each sub-model and divide these by the averages found in step 4.

Step 10: Determine if there are frequent excursions above the confidence limits. These limits are determined during step 5. The presence of such excursions indicates model-plant-mismatch in the given sub-model.

Step 11: If necessary, design and conduct re-identification experiments for the concerned sub-models to update the prediction model.

The encoding and decoding portion of these steps are represented graphically in Figure 2.2. An example of the signals can be seen in Figure 2.3 where $d = 3$ was used to augment the raw signal ($\because m = 16$), and $q = 4$ slow features were used to decode the signals. Note that only the two slowest and two fastest extracted features are shown.

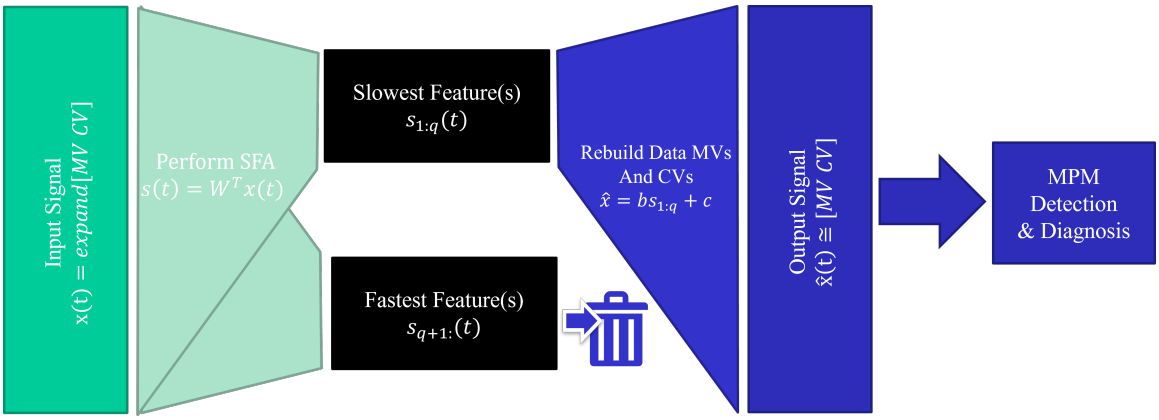

Figure 2.2: SFA Based Encoder-Decoder Structure.

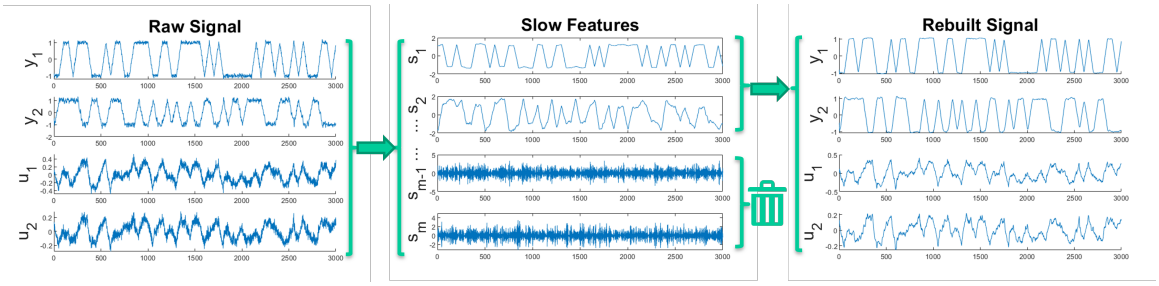

Figure 2.3: SFA Based Encoder-Decoder Prepossessing Sample Results.

2.3 Case Studies

In this section two sets of case studies are performed. A simulated distillation column is used to demonstrate the proposed model-plant-mismatch detection and diagnosis method, and an industrial zinc roaster data set is used to show increased ability in the determination of sub-models that have model-plant-mismatches.

2.3.1 Simulation Case Studies

The first set of case studies were conducted based on the Wood-Berry column model [41]. This is a MIMO FOPDT system with two inputs, two outputs and disturbances driven by two independent Gaussian white noises each with zero mean and a variance of 0.03:

$$\begin{aligned}\begin{bmatrix} Y_1(s) \\ Y_2(s) \end{bmatrix} &= \begin{bmatrix} \frac{12.8e^{-s}}{16.7s+1} & \frac{-18.9e^{-3s}}{21.0s+1} \\ \frac{6.6e^{-7s}}{10.9s+1} & \frac{-19.4e^{-3s}}{14.4s+1} \end{bmatrix} \begin{bmatrix} U_1(s) \\ U_2(s) \end{bmatrix} \\ &+ \begin{bmatrix} \frac{3.3e^{-8s}}{14.9s+1} & \\ & \frac{4.9e^{-3s}}{13.2s+1} \end{bmatrix} \begin{bmatrix} \mathcal{N}(0,0.03) \\ \mathcal{N}(0,0.03) \end{bmatrix}\end{aligned} \tag{2.12}$$

A MPC was designed in MATLAB's MPC Designer App [42] using the true plant model discretized with a one-minute sampling time and an assumed random walk disturbance model. Note that this means that there is always some degree of disturbance-model-mismatch, as is typical whenever a random walk disturbance is assumed in the MPC design. The prediction model is then as follows:

$$\begin{aligned}\begin{bmatrix} \hat{y}_1(t) \\ \hat{y}_2(t) \end{bmatrix} &= \begin{bmatrix} \frac{0.744z^{-1}}{z-0.9419} & \frac{-0.8789z^{-3}}{z-0.9535} \\ \frac{0.5786z^{-7}}{z-0.9123} & \frac{-1.302z^{-3}}{z-0.9329} \end{bmatrix} \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix} \\ &+ \begin{bmatrix} \frac{z}{z-1} & \\ & \frac{z}{z-1} \end{bmatrix} \begin{bmatrix} e_1(t) \\ e_2(t) \end{bmatrix}\end{aligned} \tag{2.13}$$

Simulation Case Study 1

In this study the process is simulated for 10,000 minutes. The gain of the plant sub-model between input-1 and output-1 is doubled (i.e., 100% parameter mismatch) in time instants

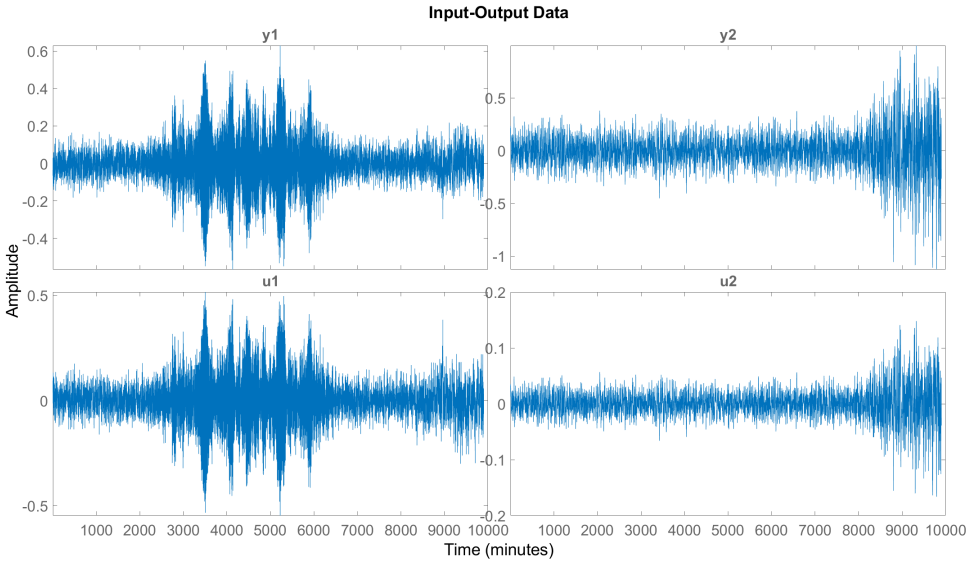

Figure 2.4: Generated Data for Simulation Case Study 1.

2,000-7,000. The gain of the second disturbance is doubled from 8,000 onward. Both changes happen gradually over 1,000 time steps within the bounds of the mismatches. The prediction model used in MPC remains unchanged; therefore, there are both model-plant-mismatch and disturbance-model-mismatch during these time periods. The simulated input and output data are shown in Figure 2.4. Here it is clear that during the periods of mismatch there are increased variances; however, it is difficult to identify what kind of mismatch is occurring, and if it is a model-plant-mismatch, which sub-model(s) are the cause. Therefore, a model-plant-mismatch detection and diagnosis must be performed to identify the source.

Data samples within time 1-2,000 are used to train an SFA reconstruction model with $q = 2$ and $d = 1$. The MQI is found in a data window of 250 minutes. The window moves by 100 minutes at a time. The ARX models identified are of fifth order. A 3-σ threshold is used to determine when the ratio of MQILOO to MQI increases substantially over its average value calculated during the training period. The results can be seen in Figure 2.5. A clear excursion from the limits, and therefore model-plant-mismatch, can be correctly identified in the input-1 and output-1 pair in both the SFA reconstructed and

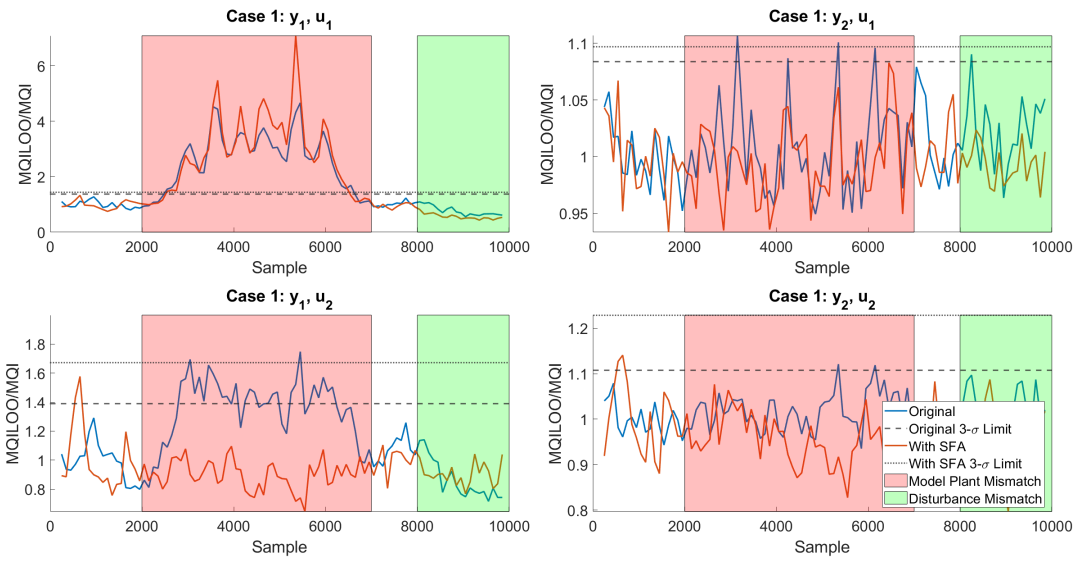

Figure 2.5: Results of MQILOO to MQI Ratio Normalized to Training Period (1-2000) for Simulation Case Study 1.

conventional methods. However, the conventional method also shows an apparent model-plant-mismatch in input-1 and output-2 during the same period, which is a false alarm. Neither method falsely attributes the disturbance mismatch to a specific sub-model.

Simulation Case Study 2

In the second simulation study, the process is simulated in a similar fashion as case 1. However, the gain mismatches are of 20% magnitude rather than 100%. As can be seen in Figure 2.6, it is not immediately apparent that any mismatch has occurred as the MPC is able to maintain control at the fixed set point without an obvious increase in variance. However, a 20% mismatch can create a problem when the process experiences a major disturbance or has a large change of set point instructed by a higher-level economic optimizer. It is therefore of practical interest to detect the model-plant-mismatch even though it may be relatively small.

The detection results are shown in Figure 2.7. While a small increase is seen in the conventional non-SFA based method, it nevertheless lies below the threshold. Therefore, the conventional method fails to clearly indicate that model-plant-mismatch has occurred

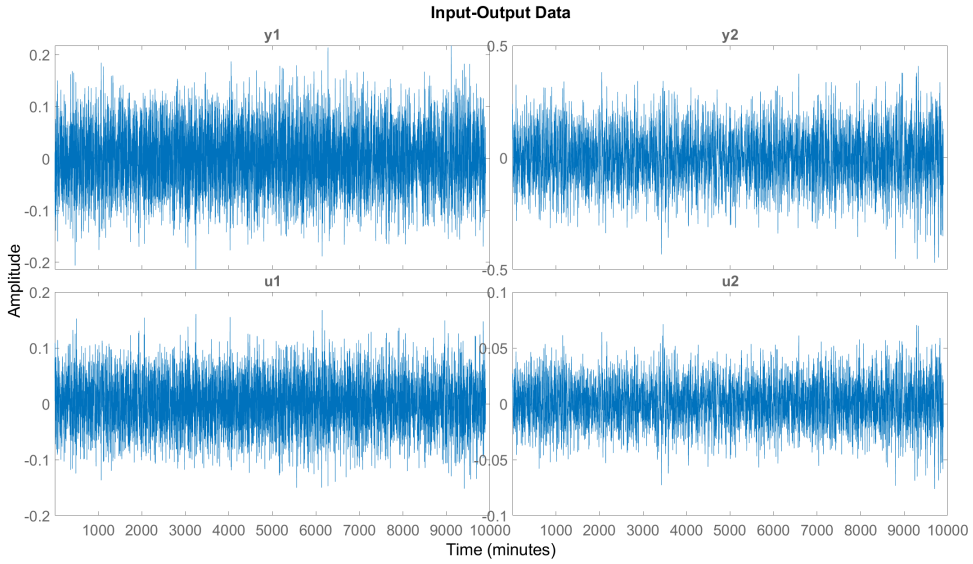

Figure 2.6: Generated Data for Simulation Case Study 2.

during the related time window. The proposed method with SFA reconstruction however reduces the impact of the disturbance and captures slower dynamics and is thus able to correctly identify the model-plant-mismatch. Again, neither method falsely identifies the disturbance mismatch as a plant-model-mismatch.

2.3.2 Industrial Application

A zinc roasting unit with a running MPC is considered in this case study. The main components of the roasting process consist of the furnace, waste heat boiler, cyclone, and electrostatic precipitator as shown in Figure 2.8. In such a process, the zinc concentrates from mines are fed to and roasted in a fluidized bed furnace. To fluidize the bed in a furnace, air with added oxygen is sent from the bottom. Additionally, the oxygen participates in chemical reactions with the concentrate. These reactions are exothermic and the heat that is generated is recovered by the waste-heat boiler. The sulphur dioxide gas from the final step is then cleaned for Hg removal, and H_2SO_4 is produced in an acid plant. A stable bed temperature and good fluidization are essential for effective operation of a fluidized bed furnace.

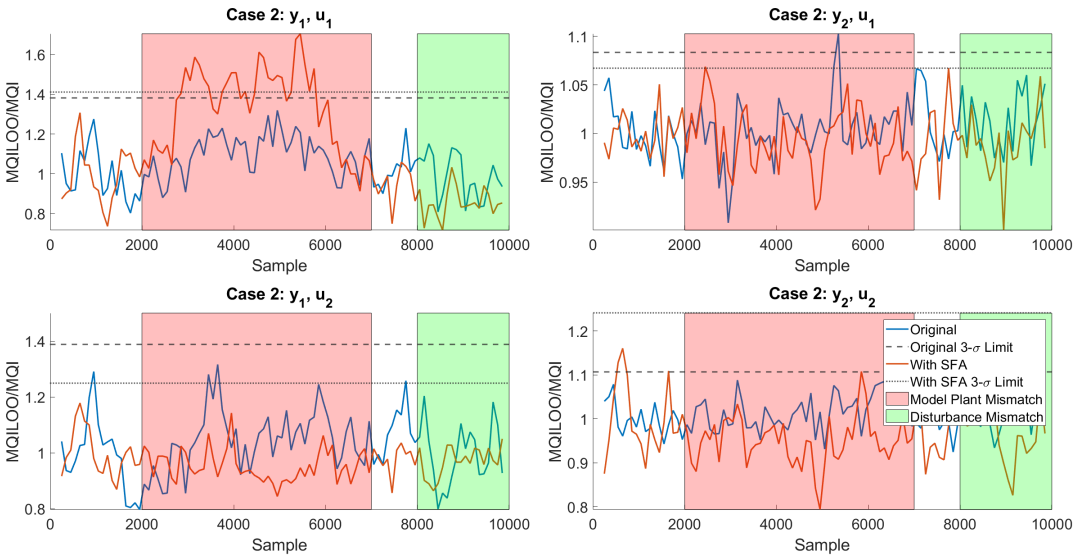

Figure 2.7: Results of MQILOO to MQI Ratio Normalized to Training Period (1-2000) for Simulation Case Study 2.

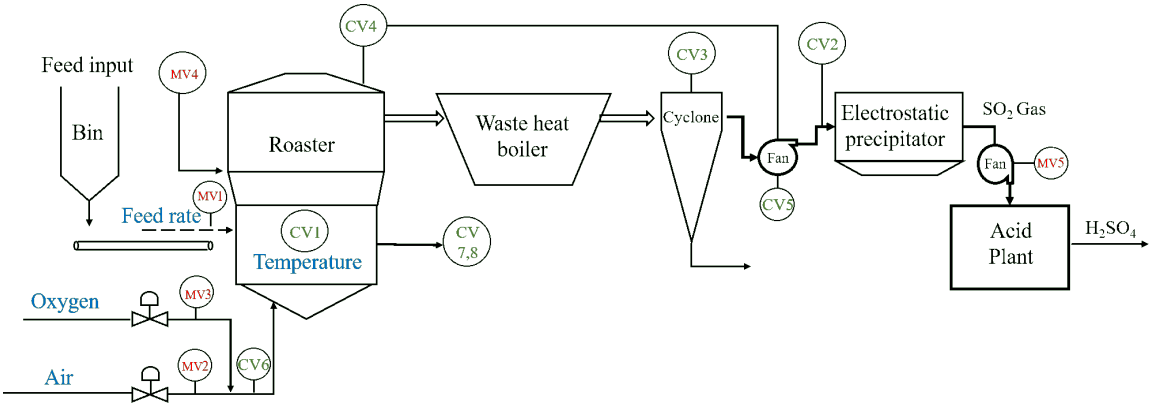

Figure 2.8: Zinc Roaster Process Diagram.

The process consists of 13 variables, which are listed in Table 2.1, and a flow sheet can be seen in Figure 2.8. Ensuring high model quality is essential to maintaining bed temperature stability, which relates strongly to product quality [43, 44], and to the goal of improving the economics of the operation [35]. Other important goals that rely upon an accurate prediction model for an effective MPC include the decrease in environmental impact of the operation and the reduction of safety risks. The predictive model used by the MPC is in the form of several first order plus dead time transfer function sub-models identified during the commissioning stage of the MPC, and a random walk disturbance model is used in

Table 2.1: Description of Zinc Roaster Process Variables

Variable	Description
MV 1	Feed Rate
MV 2	Air Flow Rate
MV 3	Oxygen Flow Rate
MV 4	Bed Spray Water
MV 5	Inlet Pressure
CV 1	Bed Temperature
CV 2	ESP Pressure
CV 3	Cyclone Temperature
CV 4	Pressure Controller Output
CV 5	Fan Speed
CV 6	Oxygen Percentage
CV 7	Required Amount of Oxygen
CV 8	Air:Feed Ratio

the MPC. As can be seen in Table 2.2, sub-models do not exist for every input-output pair since some of them have no apparent dynamics. It should also be noted that this process has two distinct operating modes, a normal mode, and a frequent maintenance mode. Only the normal mode is considered in this Chapter and more detail on the operating modes can be found in Chapters 3 and 4.

A training window was selected from normal operating data through a 2,000 minute period during which the mean squared 1-step ahead prediction errors of the MPC were lowest. This period was selected to ensure that the best model quality could be achieved during the training. The slow feature analysis was performed with all inputs and outputs augmented for lags up to $d = 3$. The reconstruction model was created using $q = 7$. MQILOO and MQI of each sub-model were found using a fifth order ARX model in windows of 500 minutes that are advanced by 100 minutes between each evaluation. The interest for model

Table 2.2: Model-Plant-Mismatch Detection and Diagnosis Results by Sub-Model.

Checkmarks (✓) indicate acceptable performance, crosses (✗) indicate detection of model-plant-mismatch, and dashes (-) indicate the lack of a relevant sub-model. Results without SFA reconstruction are on the left of columns and with the reconstruction on the right.

	MV 1		**MV 2**		**MV 3**		**MV 4**		**MV 5**	
	w/o	with	w/o	with	w/o	with	w/o	with	w/o	with
CV 1	✓	✓	✓	✓	-		✓	✓	-	
CV 2	✓	✓	✓	✓	✓	✓	✗	✗	✓	✓
CV 3	✓	✓	✓	✓	✓	✓	✓	✗	-	
CV 4	✓	✓	✗	✓	✗	✓	✗	✗	-	
CV 5	✓	✓	-		✓	✓	✓	✓	-	
CV 6	-		-		✓	✓	-		-	
CV 7	✗	✓	✓	✓	✗	✓	-		-	
CV 8	✓	✓	✓	✓	✓	✓	-		-	

quality assessment lies in a transitional period which occurred at a later time, shortly after an outage occurred. This period is suspected to have some degree of model-plant-mismatch as several process variables were outside of their normal operating conditions. A summary of the detection and diagnosis results is listed in Table 2.2. Note that without the proposed SFA reconstruction, six sub-models in four different input channels are suspected of having model-plant-mismatch. After the SFA reconstruction step is applied, only three sub-models corresponding to a single input channel (MV 4) are diagnosed as deficient. As a result, only a single input needs to be perturbed for a re-identification task rather than the four suggested by the conventional method. Such an exercise would significantly reduce the effort for model re-identification and MPC tuning.

2.4 Conclusions

In this chapter, linear slow feature analysis was proposed to reconstruct input-output data to improve model-plant-mismatch detection and diagnosis in model predictive control. This was shown to improve the performance of a conventional approach through a simulated case study of a distillation column with known mismatches between the plant and prediction model utilizing the MQILOO metric. An industrial application using data from a zinc roasting unit demonstrates that the proposed method provides more targeted recommendations for sub-models that have model-plant-mismatch and are in need of re-identification. This would allow an operator to reduce costly experimentation on the plant.

As can be seen in Chapter 3 and 4, the linear SFA method used in this chapter does not handle measurement outliers or multiple mode operation well. In order to extend the proposed SFA encoder-decoder step for MPC model-plant-mismatch detection and diagnosis to industrial settings where such data concerns are common, the development of a SFA based method that can address these issues is needed.

Chapter 3

Robust PSFA with Gaussian Scale Mixture Noise Model

To be submitted to journal as: Cameron Dyson, Jayaram Valluru, Biao Huang, and Graham Slot ”Robust PSFA using Switching Gaussian Scale Mixtures”

3.1 Introduction

In industrial processes real-time information regarding critical quality variables is of the utmost importance for accurate process monitoring, control, and real-time optimization of the processes. Soft sensors, derived from either first-principles or data driven approaches, can provide real-time estimation of the quality variables using available measurements. Developing a first principles model for the prediction of quality variables in complex industrial processes requires detailed understanding of the underlying physics of the system. On the other hand, data driven models, are developed based on historical data [4]. As such, the latter have gained popularity in the recent years with multiple implementations using various methods [45–51]. With the increasing availability of historical data, the advantages provided by data-driven methods have led to increased interest in this area.

Industrial data exhibits significant correlations among the process variables. Latent variable (LV) methods, which can extract independent low-dimensional features, have found wide applications for soft-sensing and process monitoring. The most popular ones of these methods are principal component regression (PCR) [52–54], partial least squares (PLS) [55–58], and their probabilistic extensions i.e., probabilistic PCR (PPCR) [59] and probabilistic PLS (PPLS) [60]. One drawback of these methods is that they cannot directly account for dynamic relationships between measured signals. To combat this, one option is to further increase the dimensionality of the original signal by introducing lagged copies of the observations such as in dynamic PCR (DPCR) [61] and dynamic PLS (DPLS). Alternatively, slow feature analysis (SFA) [22], an unsupervised machine learning method that extracts slowly varying latent features from the observed signals, has been successfully used for soft sensor development and process monitoring [23, 38, 40]. Feature extraction through SFA relies on the assumption that the parts of a process relevant to modelling are those that change more slowly; meanwhile those that change more rapidly are often considered as the noise and disturbances, which are of less value for modelling. This assumption often holds in chemical processes, which are typically slow in nature. The probabilistic ex-

tension of SFA (PSFA) [28, 29] models a system in state space form where, conventionally, the slow features are a set of hidden states that drive the process while being influenced by normally distributed noises. In the current work, PSFA is considered to develop soft sensor models for high dimensional data.

The prediction performance of data driven modeling methods heavily relies on the quality of the historical data considered for modelling. Thus, selection of high-quality training data becomes an important task in obtaining accurate data-driven predictive models [30, 45, 62]. Raw industrial data, such as that from the zinc roaster introduced in Chapter 2, often includes potential data quality challenges such as noises, outliers [63, 64], high dimensionality, varied sampling rates across different measurements or missing data [65]. Data-driven latent variable approaches prioritize dimensionality reduction and can be used to address the aforementioned issues by implementing them in probabilistic frameworks. While the Gaussian measurement noise used in PSFA can prove useful, it does not handle outliers in the data by its original formulation. These will typically introduce inaccuracies in the noise parameter estimation and lead to the extraction of less meaningful features [66]. One approach to remedy this issue would be to use an ad hoc method to preprocess the data and replace the outlying samples with values that more closely match the inlying behaviour before training a model. However, outliers may contain some information that is still relevant to the process, therefore discarding them entirely is not recommended. To this end, several different approaches exist in literature to deal with the presence of outliers while retaining them in the training data [62]. A Student-t based approach has been implemented for PSFA for the case of scale outliers [30], and variants utilizing Laplace [67] or skewed [68] distributions could also be developed to handle specific kinds of outliers. However, the possible extensions from these approaches are limited when it comes to different kinds of outliers such as location outliers [69], or those that do not occur completely at random with respect to time and instead follow some correlations of their own. Treating the measurement noise as a Gaussian mixture model would allow for the development of

a framework that could address the scale and location outliers as well as different varieties of outlier contamination [70–74].

The salient contributions of this work are as follows:

- A PSFA algorithm that is robust to measurement outliers is developed through the use of a Gaussian scale mixture noise structure. The switching of the Gaussian components is modeled by a Bernoulli distribution.
- A simulation case study demonstrates the improved performance of the proposed method compared to conventional methods in a soft-sensing task where the training data contains outliers.
- The algorithm is tested utilizing an industrial data set where outliers are present.

Finally, conclusions and future work are presented.

3.1.1 Probabilistic SFA

Classical SFA has been formulated in a probabilistic framework [28–30] which can account for critical data issues such as missing data, outliers, and process drift in a systematic manner while still satisfying the fundamental constraints of SFA introduced in Chapter 2. This enables the use of SFA as a generative model as well as allowing for further extensions to address situations where the regular SFA performs poorly e.g., high signal-to-noise ratio in $x(t)$.

The formulation of PSFA [28, 29] in which $x(t)$ (m-dimensional, N samples) is measured while $s(t)$ (q-dimensional) is a hidden state to be estimated as follows:

$$s(t) = Fs(t-1) + e_s(t), \quad e_s(t) \sim \mathcal{N}(0, \Gamma) \tag{3.1}$$

$$x(t) = Hs(t) + e_x(t), \quad e_x(t) \sim \mathcal{N}(0, \Sigma) \tag{3.2}$$

where $e_s(t)$ and $e_x(t)$ are zero mean white noise signals and follow Gaussian distributions.

$H \in \mathbb{R}^{m \times q}$ is the emission matrix and F, Γ, and Σ are defined as follows:

$$\begin{aligned} F &= diag\{\Gamma_1, ..., \Gamma_q\} \\ \Gamma &= diag\{1-\Gamma_1^2, ..., 1-\Gamma_q^2\} \\ \Sigma &= diag\{\sigma_1^2, ..., \sigma_m^2\} \end{aligned} \tag{3.3}$$

Γ is defined as such to ensure the satisfaction of (2.5)-(2.7). The graphical structure of PSFA is presented in Figure 3.1.

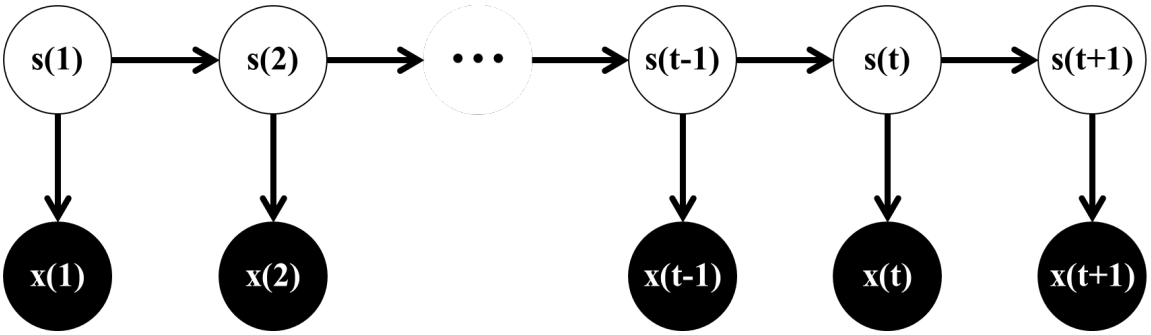

Figure 3.1: Graphical Structure of PSFA.

3.1.2 Slow Feature Regression

Once slow features, $s(t)$, have been extracted, they can be utilized to develop a soft sensor model for $y(t)$ through linear regression:

$$\hat{y}(t) = b^T s_{1:q}(t) + c \tag{3.4}$$

where b and c are regression coefficients to be determined, and q is the selected number of features which are significant.

The performance of such soft sensors may be measured through the mean squared error ($MSE \in [0, \infty)$) and the Pearson's correlation coefficient ($corr. \in [-1, 1]$).

$$MSE = \frac{1}{N} \sum_{t=1}^{N} \left(y(t) - \hat{y}(t) \right)^2 \tag{3.5}$$

$$corr. = \frac{cov(\hat{y}, y)}{\sigma_{\hat{y}} \sigma_y} \tag{3.6}$$

where *cov* represents the covariance between two signals and σ is the standard deviation of a signal.

3.2 Robust PSFA using Gaussian Mixture Models

In the presence of outliers, the assumption that the noise in the input measurement follows a Gaussian distribution can lead to inaccurate parameter estimates. In this work, to account for measurement outliers, the noise in the input measurements is modelled using a scaled Gaussian mixture model with two Gaussian components, one representing regular noise (inliers) and the other representing outliers [62]. The noise component of (3.2) is replaced with (3.7) in which inliers are represented by the first term and outliers by the second term with the additional parameters of a mixing factor δ and a scaling factor $\rho \in (0,1]$:

$$e_x(t) \sim (1-\delta)\mathcal{N}(0,\Sigma) + \delta\mathcal{N}(0,\rho^{-1}\Sigma) \tag{3.7}$$

Figure 3.2 shows the difference in probability density between a normal Gaussian distribution with $\mu = 0$, $\Sigma = 1$ and a Gaussian scale mixture with the additional parameters $\rho = 0.1$ and $\delta = 0.25$. The scale mixture displays longer tails which enable it to better handle outliers without inflating the estimated variance of the inliers. In comparison with the existing robust PSFA approach based on a t-distribution, the mixture Gaussian approach allows for the occurrence of outliers to follow a certain probability distribution.

To do this, define $Q_x = [q_x(1),...q_x(N)]$. This is a binary vector indicating which of the two mixed Gaussians a specific sample of $e_x(t)$ is drawn from. It takes on a value of $q_{x_t} = 1$ when $e_x(t) \sim \mathcal{N}(0,\Sigma)$, or a value of $q_{x_t} = \rho$ when $e_x(t) \sim \mathcal{N}(0,\rho^{-1}\Sigma)$. In this work it is represented as a Bernoulli distribution [72, 75]:

$$P(q_x(t)|\delta,\rho) = \delta^{1-\frac{q_x(t)-\rho}{1-q_x(t)\rho}}(1-\delta)^{\frac{q_x(t)-\rho}{1-q_x(t)\rho}} \tag{3.8}$$

Thus, for measurement data, $X = [x_1t,...,x_mt]$, $X \in \mathbb{R}^{m\times N}$, by considering the Gaussian mixture model the unknown parameters of the robust PSFA (rPSFA) model to be estimated

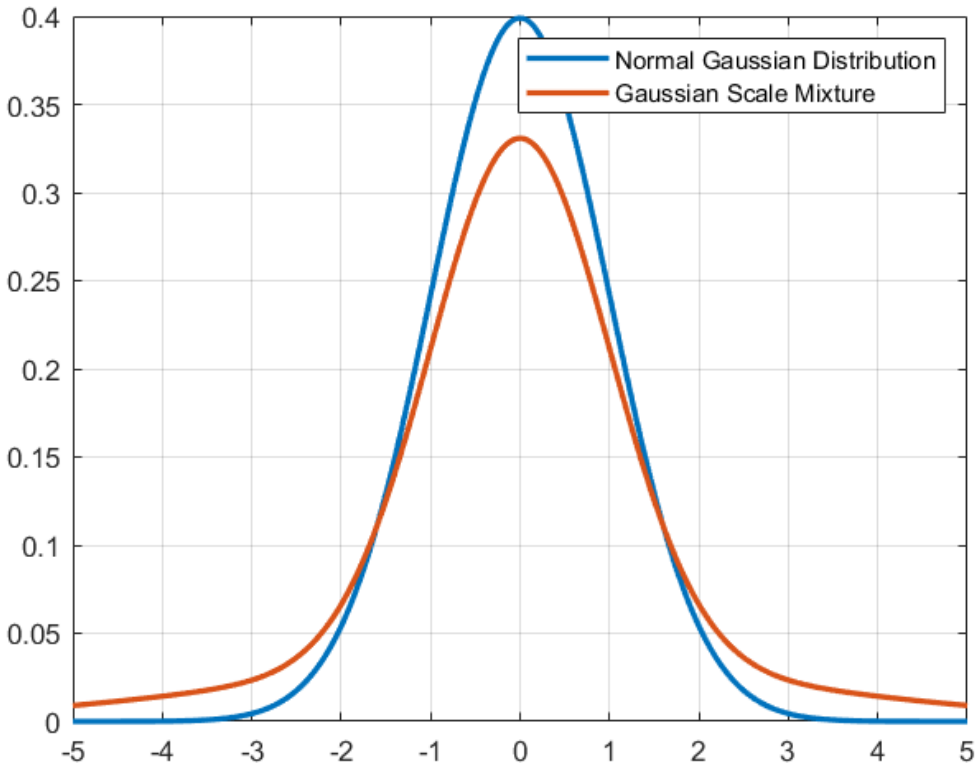

Figure 3.2: Probability Density Function of Normal Gaussian Distribution and Scale Mixture of Two Gaussians with $\rho = 0.1$ and $\delta = 0.25$.

are $\theta = \{H, \Gamma_1, ..., \Gamma_q, \sigma_1^2, ..., \sigma_m^2, \delta, \rho\}$ and the hidden variables are $S = [s_1(t), ..., s_q(t)]$, $S \in \mathbb{R}^{q \times N}$ and Q_x.

3.2.1 Solution

In the presence of hidden variables, direct maximization of the log-likelihood function (i.e., $P(X|\theta)$) w.r.t. unknown parameters is intractable. To address this kind of problem the Expectation Maximization algorithm [76] is widely used in literature, where instead of maximizing the log-likelihood function, the expectation of the joint log-likelihood function is maximized w.r.t. to the unknown parameters. The EM algorithm involves two steps, the E-step (expectation step) and M-step (maximization step), which are discussed in detail as follows.

In the E-Step, given the parameters θ for iteration count r, the expectation of the joint log-likelihood function (Q-function) is evaluated w.r.t. the hidden variables (S, Q_x) i.e.:

$$\mathbb{Q} = \mathop{\mathbb{E}}_{S,Q_x|X,\theta_r}(logP(\overbrace{X}^{Observed}, \overbrace{S,Q_x}^{Hidden}|\theta)) \tag{3.9}$$

where:

$$\theta = \{H, \Gamma_1, ..., \Gamma_q, \sigma_1^2, ..., \sigma_m^2, \delta, \rho\} \tag{3.10}$$

Since the noises acting on input variables and latent variables are independent, the complete likelihood function can be written as follows:

$$\begin{aligned} P(X,S,Q_x|\theta) &= P(X|S,Q_x,\theta)P(S,Q_x|\theta) \\ &= P(X|S,Q_x,\theta)P(S|\theta)P(Q_x|\theta) \\ &= \prod_{t=1}^{N} P(x(t)|s(t),q_x(t),\theta)P(s(t)|s(t-1),q_x(t),\theta)P(q_x(t)|\theta) \end{aligned} \tag{3.11}$$

Substituting (3.11) into (3.9) and expanding the Q-function yields:

$$\begin{aligned} \mathbb{Q} &= \mathop{\mathbb{E}}_{S,Q_x|X,\theta_r}(logP(X,S,Q_x|\theta)) \\ &= \mathop{\mathbb{E}}_{S,Q_x|X,\theta_r}\Bigg(\underbrace{logP(s(1)|\theta)}_{\mathbb{Q}_A} + \underbrace{\sum_{t=2}^{N} logP(s(t)|s(t-1),q_x,\theta)}_{\mathbb{Q}_B} \\ &+ \underbrace{\sum_{t=1}^{N} logP(x(t)|s(t),q_x(t),\theta)}_{\mathbb{Q}_C} + \underbrace{\sum_{t=1}^{N} logP(q_x(t)|\theta)}_{\mathbb{Q}_D}\Bigg) \end{aligned} \tag{3.12}$$

Generally, initial state of slow features are assumed to follow $\mathcal{N}(0,I)$ [29], $\mathbb{Q}_A$ can therefore be expressed as follows:

$$\mathbb{Q}_A = -\frac{q}{2}log(2\pi) - \frac{1}{2}\mathbb{E}\big(s(1)^T s(1)\big) \tag{3.13}$$

$\mathbb{Q}_B$ can be found as follows:

$$\begin{aligned} \mathbb{Q}_B = \mathop{\mathbb{E}}_{S,Q_x|X,\theta_r}\Bigg[&-\frac{q(N-1)}{2}log(2\pi) - \frac{N-1}{2}\sum_{j=1}^{q} log(1-\Gamma_j^2) \\ &-\frac{1}{2}\sum_{t=2}^{N}\sum_{j=1}^{q}\frac{1}{1-\Gamma_j^2}\big(s_j(t)-\Gamma_j s_j(t-1)\big)^2\Bigg] \end{aligned} \tag{3.14}$$

$\mathbb{Q}_C$ and $\mathbb{Q}_D$ can be split into two parts corresponding to when $q_x(t) = 1$ and $q_x(t) = \rho$ respectively. For $\mathbb{Q}_C$ this is derived as follows:

$$\begin{aligned}
\mathbb{Q}_C &= \mathop{\mathbb{E}}_{S,Q_x|X,\theta_r}\big(\sum_{t=1}^{N} logP(x(t)|s(t),q_x(t),\theta)\big) \\
&= \mathop{\mathbb{E}}_{S|X,Q_x,\theta_r}\Bigg(\mathop{\mathbb{E}}_{Q_x|S,X,\theta_r}\big(\sum_{t=1}^{N} logP(x(t)|s(t),q_x(t),\theta)\big)\Bigg) \\
&= \mathop{\mathbb{E}}_{S|X,Q_x,\theta_r}\Bigg(\big(\sum_{t=1}^{N}\sum_{i=\rho,1} P_i(t) logP(x(t)|s(t),q_x(t)=i,\theta)\big)\Bigg) \\
&= \mathop{\mathbb{E}}_{S|X,Q_x,\theta_r}\Bigg(\sum_{t=1}^{N}\big(P_1(t) logP(x(t)|s(t)q_x(t)=1,\theta) \\
&\quad + P_\rho(t) logP(x(t)|s(t)q_x(t)=\rho,\theta)\big)\Bigg) \\
&= \mathop{\mathbb{E}}_{S|X,Q_x,\theta_r}\Bigg(\sum_{t=1}^{N}\big(P_1(t) logP(x(t)|s(t)q_x(t)=1,\theta)\big)\Bigg) \\
&\quad + \mathop{\mathbb{E}}_{S|X,Q_x,\theta_r}\Bigg(\sum_{t=1}^{N}\big(P_\rho(t) logP(x(t)|s(t)q_x(t)=\rho,\theta)\big)\Bigg) \\
&= \mathbb{Q}_C^{first} + \mathbb{Q}_C^{second}
\end{aligned} \tag{3.15}$$

The two terms of $\mathbb{Q}_C$ can then be written as:

$$\begin{aligned}
\mathbb{Q}_C^{first} &= \mathop{\mathbb{E}}_{S|X,Q_x,\theta_r}\Bigg(\sum_{t=1}^{N} P_1(t) \times logP(x(t)|s(t)q_x(t)=1,\theta)\Bigg) \\
&= \sum_{t=1}^{N} P_1(t) \times \Bigg[-\frac{1}{2} log((2\pi)^m|\Sigma|) \\
&\quad -\frac{1}{2} E_1\big(x(t)-Hs(t)\big)^T \Sigma^{-1}\big(x(t)-Hs(t)\big)\Bigg]
\end{aligned} \tag{3.16}$$

$$\begin{aligned}
\mathbb{Q}_C^{second} &= \mathop{\mathbb{E}}_{S|X,Q_x,\theta_r}\Bigg(\sum_{t=1}^{N} P_\rho(t) \times logP(x(t)|s(t)q_x(t)=\rho,\theta)\Bigg) \\
&= \sum_{t=1}^{N} P_\rho(t) \times \Bigg[-\frac{1}{2} log((2\pi)^m|\rho^{-1}\Sigma|) \\
&\quad -\frac{1}{2} E_\rho\big(x(t)-Hs(t)\big)^T \rho\Sigma^{-1}\big(x(t)-Hs(t)\big)\Bigg]
\end{aligned} \tag{3.17}$$

where $P_1(t) = P(q_x(t)=1|x(t),\theta)$, $E_1(\cdot) = \mathbb{E}(\cdot|x,q_x(t)=1,\theta)$, $P_\rho = P(q_x(t)=\rho|x(t),\theta)$, and $E_\rho(\cdot) = \mathbb{E}(\cdot|x,q_x(t)=\rho,\theta)$.

The $\mathbb{Q}_D$ term can also be split into two parts similar to $\mathbb{Q}_C$:

$$\begin{aligned}
\mathbb{Q}_D &= \mathop{\mathbb{E}}_{S,Q_x|X,\theta_r}\big(\sum_{t=1}^{N} logP(q_x(t)|\theta)\big) \\
&= \sum_{t=1}^{N} \mathop{\mathbb{E}}_{S,Q_x|X,\theta_r}\big(logP(q_x(t)|\theta)\big) \\
&= \sum_{t=1}^{N} \sum_{i=1,\rho} P_i(t) logP(q_x(t)=i|\theta) \\
&= \sum_{t=1}^{N} \big(P_1(t) logP(q_x(t)=1|\theta) + P_\rho(t) logP(q_x(t)=\rho|\theta)\big) \\
&= \sum_{t=1}^{N} \big(P_1(t) logP(q_x(t)=1|\theta)\big) + \sum_{t=1}^{N} \big(P_\rho(t) logP(q_x(t)=\rho|\theta)\big) \\
&= \mathbb{Q}_D^{first} + \mathbb{Q}_D^{second}
\end{aligned} \tag{3.18}$$

Each of the two terms of $\mathbb{Q}_D$ can then be found as:

$$\begin{aligned}
\mathbb{Q}_D^{first} &= \sum_{t=1}^{N} \big(P_1(t) logP(q_x(t)=1|\theta)\big) \\
&= \sum_{t=1}^{N} P_1(t) \times log(1-\delta)
\end{aligned} \tag{3.19}$$

$$\begin{aligned}
\mathbb{Q}_D^{second} &= \sum_{t=1}^{N} \big(P_\rho(t) logP(q_x(t)=\rho|\theta)\big) \\
&= \sum_{t=1}^{N} P_\rho(t) \times log(\delta)
\end{aligned} \tag{3.20}$$

In the M-step, namely the parameter estimation step, the Q-function is maximized w.r.t. θ:

$$\theta^{new} = arg\max_{\theta} \mathbb{Q}(\theta, \theta^{old}) \tag{3.21}$$

As can be seen in (3.3) the F and Γ matrices share the same subset of parameters, $\Gamma_j s$, which only appear in $\mathbb{Q}_B$. Each Γ_j can be updated as follows:

$$\begin{aligned}
\frac{\partial \mathbb{Q}}{\partial \Gamma_j} &= \frac{\partial \mathbb{Q}_B}{\partial \Gamma_j} = 0 \\
&= \frac{N-1}{2} \frac{2\Gamma_j}{1-\Gamma_j^2} - \frac{1}{2} \frac{(1-\Gamma_j^2) J'(\Gamma_j) + 2\Gamma_j J(\Gamma_j)}{(1-\Gamma_j^2)^2}
\end{aligned} \tag{3.22}$$

where:

$$
\begin{aligned}
J(\Gamma_j) &= \sum_{t=2}^{N} \mathbb{E}_{X,\theta^{old}}\Big((s_j(t) - \Gamma_j s_j(t-1))^2 \Big), \quad 1 \le j \le q \\
J'(\Gamma_j) &= 2\Gamma_j \sum_{t=2}^{N} \mathbb{E}_{X,\theta^{old}}\Big(s_j^2(t-1) \Big) - 2 \sum_{t=2}^{N} \mathbb{E}_{X,\theta^{old}}\Big(s_j(t) s_j(t-1) \Big)
\end{aligned}
\tag{3.23}
$$

Each Γ_j^{new} can be found as the root in $[0,1)$ of a cubic polynomial:

$$
a_3 \Gamma_j^{3,new} + a_2 \Gamma_j^{2,new} + a_1 \Gamma_j^{new} + a_0 = 0 \tag{3.24}
$$

where:

$$
\begin{aligned}
a_3 &= N - 1 \\
a_2 &= -\sum_{t=2}^{N} E\Big(s_j(t) s_j(t-1) \Big) \\
a_1 &= \sum_{t=2}^{N} \Big(\big(E(s_j(t) s_j(t)^T) + E(s_j(t-1) s_j(t-1)^T) \big) - 1 \Big) \\
a_0 &= a_2
\end{aligned}
\tag{3.25}
$$

The updated H^{new} can be found as follows:

$$
\frac{\partial \mathbb{Q}}{\partial H} = \frac{\partial \mathbb{Q}_C}{\partial H} = 0 \tag{3.26}
$$

$$
\begin{aligned}
H^{new} = &\Bigg(\sum_{t=1}^{N} P_1(t) x(t) E_1(s(t))^T + \rho P_\rho(t) x(t) E_\rho(s(t))^T \Bigg) \times \\
&\Bigg(\sum_{t=1}^{N} P_1(t) E_1(s(t) s(t)^T) + \rho P_\rho(t) E_\rho(s(t) s(t)^T) \Bigg)^{-1}
\end{aligned}
\tag{3.27}
$$

Each element of the updated $\Sigma^{new} = diag\{\sigma_1^{2new}, ..., \sigma_m^{2new}\}$ can be derived as follows:

$$
\frac{\partial \mathbb{Q}}{\partial \sigma_i^2} = \frac{\partial \mathbb{Q}_C}{\partial \sigma_i^2} = 0, 1 \le i \le m \tag{3.28}
$$

$$
\begin{aligned}
\sigma_i^{2new} = \frac{1}{N} \sum_{t=1}^{N} \Bigg[& P_1(t) \Bigg(\big(x_i^2(t) - 2 h_i^{T,new} E_1(s(t)) x_i(t) \big) \\
& + h_i^{T,new} E_1(s(t) s(t)^T) h_i^{new} \Bigg) \\
& + \rho P_\rho(t) \Bigg(\big(x_i^2(t) - 2 h_i^{T,new} E_\rho(s(t)) x_i(t) \big) \\
& + h_i^{T,new} E_\rho(s(t) s(t)^T) h_i^{new} \Bigg) \Bigg]
\end{aligned}
\tag{3.29}
$$

The updated ρ^{new} can be found as follows:

$$\frac{\partial \mathbb{Q}}{\partial \rho} = \frac{\partial \mathbb{Q}_C}{\partial \rho} = 0 \tag{3.30}$$

$$\begin{aligned} \rho^{new} = m\left[\sum_{t=1}^{N} P_\rho(t)\right] \\ \times \left[\sum_{t=1}^{N} P_\rho(t) E_\rho\left(\left(x(t) - Hs(t)\right)^T \Sigma^{-1,new}\left(x(t) - Hs(t)\right)\right)\right]^{-1} \end{aligned} \tag{3.31}$$

The updated mixing factor, δ^{new}, can be solved similarly:

$$\frac{\partial \mathbb{Q}}{\partial \delta} = \frac{\partial \mathbb{Q}_D}{\partial \delta} = 0 \tag{3.32}$$

$$\delta^{new} = \frac{\sum_{t=1}^{N} P_\rho(t)}{N} \tag{3.33}$$

3.2.2 Posterior Distribution of Hidden Variables

In order to obtain the necessary values to update the parameter estimation, the expectations related to the hidden variables must be found. Further, to develop a soft sensor model, the latent slow feature S is needed. The required expectations are as follows:

$$E_*(s(t)) = \mathbb{E}(s(t)|x, q_x(t) = *, \theta) \tag{3.34}$$

$$E_*(s(t)s(t)^T) = \mathbb{E}(s(t)s(t)^T|x, q_x(t) = *, \theta) \tag{3.35}$$

$$E_*(s(t)s(t-1)^T) = \mathbb{E}(s(t)s(t-1)^T|x, q_x(t) = *, \theta) \tag{3.36}$$

$$P_1(t|x(t), \theta) + P_\rho(t|x(t), \theta)\rho = \mathbb{E}(q_x(t)|x, \theta) \tag{3.37}$$

$$E(s(t)) = \mathbb{E}(s(t)|x, q_x, \theta) \tag{3.38}$$

$$E(s(t)s(t)^T) = \mathbb{E}(s(t)s(t)^T|x, q_x, \theta) \tag{3.39}$$

$$E(s(t)s(t-1)^T) = \mathbb{E}(s(t)s(t-1)^T|x, q_x, \theta) \tag{3.40}$$

Equations (3.34)-(3.36) represent the estimate of the slow features given that all samples are either inliers or outliers. Each E_* is to be found twice, once for the inlier case where $* = 1$, and again for the outlier case where $* = \rho$.

$$\begin{aligned} E_*(s(t)) &= \hat{\mu_{t*}} \\ E_*(s(t)s(t)^T) &= \hat{V_{t*}} + \hat{\mu_{t*}}\hat{\mu_{t*}}^T \\ E_*(s(t)s(t-1)^T) &= J_{t-1}\hat{V_{t*}} + \hat{\mu_*}t\hat{\mu_{t-1*}}^T \end{aligned} \tag{3.41}$$

E_*s that correspond to the inliers and outliers can be found utilizing the Kalman smoothing for a Linear dynamical system [29, 77, 78]. First, a forward recursion is performed to estimate the posterior distribution up to time t, $P(s(t)|x(t),...,x(t), q_x(1) = *, ..., q_x(t) = *, \theta^{old})$:

$$\begin{aligned} P_{t-1*} &= FV_{t-1*}F^T + \Gamma \\ \mu_t &= F\mu_{t-1*} + K_{t*}[x(t) - HF\mu_{t-1*}] \\ V_{t*} &= (I - K_{t*}H)P_{t-1*} \\ K_{t*} &= P_{t-1*}H^T(HP_{t-1*}H^T + \Sigma_*)^{-1} \end{aligned} \tag{3.42}$$

with the initialization:

$$\begin{aligned} \mu_{1*} &= K_{1*}x(1) \\ V_{1*} &= I - K_{1*}H \\ K_{1*} &= H^T(HH^T + \Sigma_*)^{-1} \end{aligned} \tag{3.43}$$

The backwards recursion can then be done to find the overall posterior distribution $P(S|X, Q_x = *, \theta^{old})$:

$$\begin{aligned} \hat{\mu_{t*}} &= \mu_{t*} + J_{t*}(\hat{\mu_{t+1*}} - F\mu_{t*}) \\ \hat{V_{t*}} &= V_{t*} + J_{t*}(\hat{V_{t+1*}} - P_{t*})J_{t*}^T \\ J_{t*} &= V_{t*}F^TP_{t*}^{-1} \end{aligned} \tag{3.44}$$

which is initialized with:

$$\begin{aligned} \hat{\mu_{t*}} &= \mu_{t*} \\ \hat{V_{t*}} &= V_{t*} \end{aligned} \tag{3.45}$$

Σ_*s are defined using the Gaussian scale mixture parameters as follows:

$$\begin{aligned} \Sigma_1 &= \Sigma \\ \Sigma_\rho &= \rho^{-1}\Sigma \end{aligned} \tag{3.46}$$

$$P(q_x(t)|x(t),\theta^{old}) = \frac{P(x(t)|q_x(t),\theta^{old})P^*(q_x(t)|\theta^{old})}{P(x(t)|\theta^{old})} \tag{3.47}$$

In which P* is the prior of Q_x and the first factor of the numerator is easily found using Gaussian properties. The overall expectations (3.38)-(3.40) can then be found as the sum of the inlier and outlier cases ($* = 1$ and ρ respectively) weighted by their respective posterior at each sample. For example, (3.38) can be found as follows:

$$\begin{aligned} E(s(t)) &= \sum_{*=1,\rho} P(q_x(t)|x(t),\theta)E(s(t)|x,q_x(t)=*,\theta) \\ &= P_1(t)E_1(s(t)|x,q_x(t)=1,\theta) + P_\rho(t)E_\rho(s(t)|x,q_x(t)=\rho,\theta) \end{aligned} \tag{3.48}$$

3.3 Verification and Application

In this section, the efficacy of the proposed robust PSFA for soft sensor modelling is demonstrated on a benchmark simulation and on an industrial data set. Performance of the rPSFA is compared with the conventional SFA and PSFA, where identical data sets are used for comparison.

3.3.1 Simulated Study

For the simulation case study, benchmark Tennessee Eastman (TE) [79] simulation data [80] is considered to verify the developed algorithm. The TE process consists of five main units: a reactor, stripper, condenser, compressor, and separator. The feed consists of four reactants (A, C, D and E) and three products (F, G and H). In Fan et al. [30], a slow feature regression soft sensor is used to build a model for y which is the concentration (mole %) of component A in the reactor feed. Accordingly, five highly correlated variables were

considered as inputs:

$$
\begin{aligned}
&x_1\text{: Normalized Reactor Pressure (kPag)}\\
&x_2\text{: Normalized Stripper Temperature } (^oC)\\
&x_3\text{: Normalized Stripper Steam Flow (kg/hr)}\\
&x_4\text{: Normalized Compressor Work (kW)}\\
&x_5\text{: Normalized Component C in Purge Gas (Mole \%)}
\end{aligned}
\tag{3.49}
$$

In order to show the effect of outliers, in the parameter learning step (or model training step), 10% of the training data is corrupted by adding outliers, which are generated from a Gaussian distribution with $\mathcal{N}(0,4I)$. The training and testing data for model building and evaluation are reported in Figures 3.3 and 3.4 respectively.

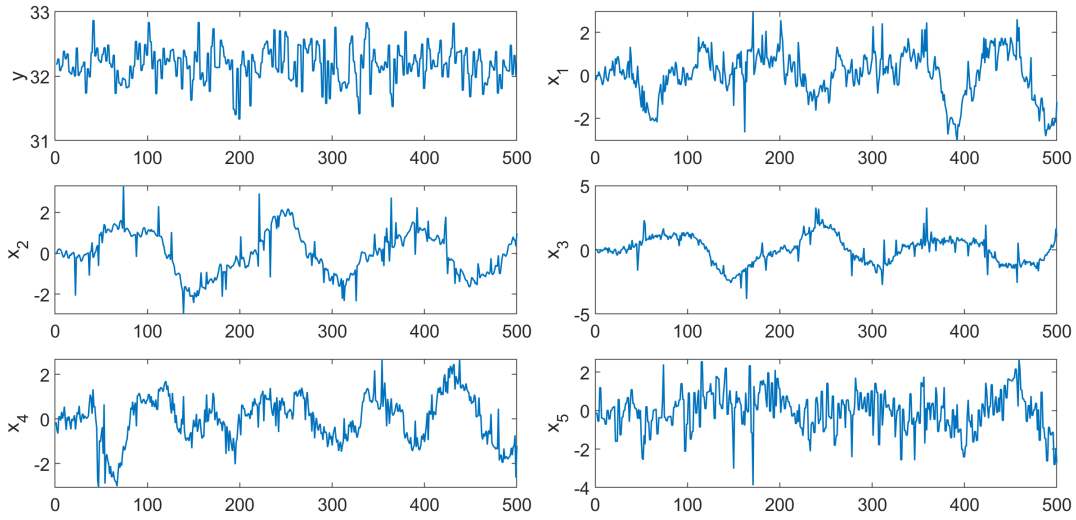

Figure 3.3: Simulated Tennessee Eastman Training Data with 10% Measurement Outliers.

SFA, PSFA, Student-t PSFA (tPSFA) and the proposed robust PSFA are applied to $x = [x_1,..,x_5]$ with the SFA solution as the initialization for the PSFA, tPSFA and rPSFA methods. The extracted features along with their lag-1 autocorrelation coefficients can be seen in Figure 3.5 and Table 3.1. Here it can be seen that the features extracted by each method have similar overall shapes, with the proposed rPSFA and existing tPSFA yielding the smoothest results.

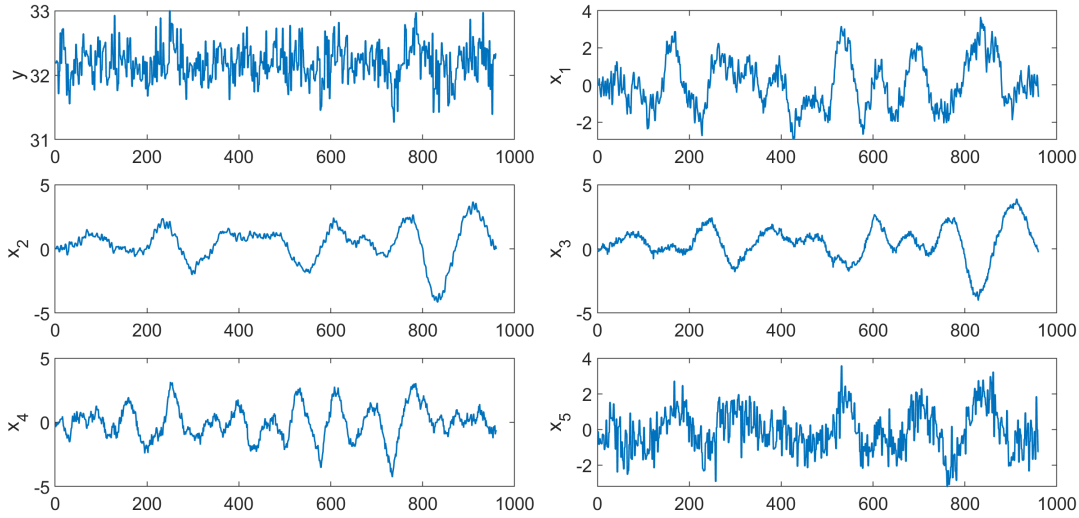

Figure 3.4: Simulated Tennessee Eastman Testing Data.

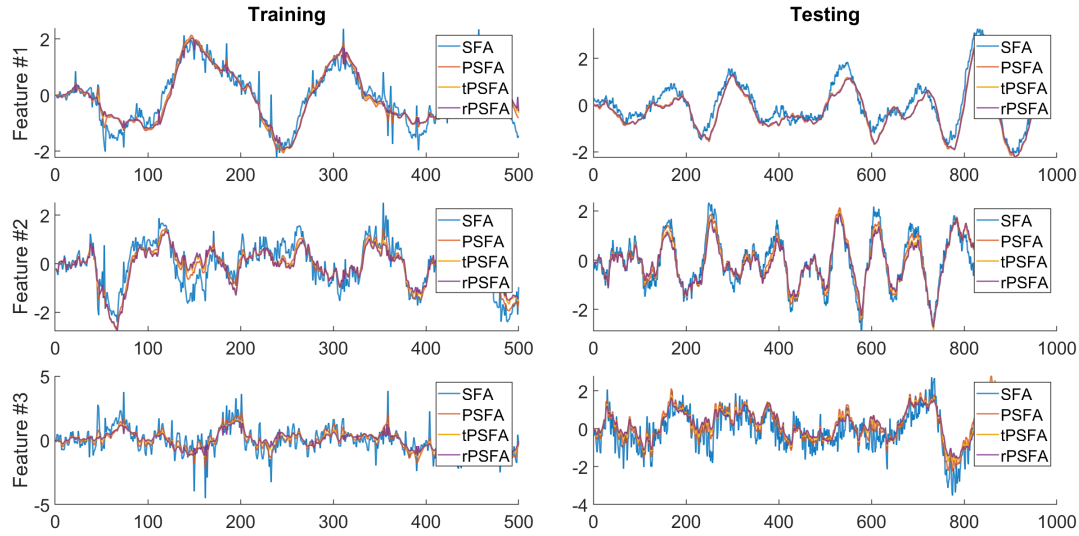

Figure 3.5: Slow Features Extracted from Tennessee Eastman Training and Testing Data with 10% Outliers in Training.

As the lag-1 autocorrelation directly relates to the slowness of a signal [28], it is clear that the slow features extracted by SFA and PSFA are negatively impacted by the presence of outliers in the training data while tPSFA and the proposed rPSFA are robust to them. In the soft sensor task the objective is to build a linear regression model for y utilizing the slow features, s, extracted from x. The obtained slow features are then used to develop a

Table 3.1: Slow Features Lag-1 Autocorrelation Coefficients for Tennessee Eastman Training and Testing Data with 10% Outliers in Training.

	Training			**Testing**		
Feature	#1	#2	#3	#1	#2	#3
SFA	0.935	0.891	0.592	0.991	0.873	0.681
PSFA	0.992	**0.985**	0.941	0.997	0.922	0.978
tPSFA	0.994	0.979	**0.948**	0.997	**0.990**	0.986
rPSFA	**0.995**	0.975	0.932	**0.998**	0.989	**0.990**

Table 3.2: Slow Features Regression Performance for Tennessee Eastman Training and Testing Data with 10% Outliers in Training.

	Training		**Testing**	
	Corr.	MSE	Corr.	MSE
SFA	0.288	0.0723	0.465	0.0671
PSFA	0.308	0.0713	0.481	0.0641
tPSFA	0.320	0.0707	0.485	0.0639
rPSFA	**0.322**	**0.0707**	**0.487**	**0.0638**

regression model for prediction of the quality variable, y. The predictions and performance of the slow feature based regression models in the training and testing phases are reported in Figures 3.6 and Table 3.2, respectively. Here it is evident that in the presence of outliers in the training data, rPSFA has better correlation with the actual data and lower MSE values compared to the SFA and PSFA approaches in the testing data. rPSFA performs similar to tPSFA with a minor improvement. A zoomed-in subset of the testing results can also be found in Figure 3.7. From the figure it can be seen that the regression results for SFA and PSFA are relatively more affected by measurement noise carried over from the input variables.

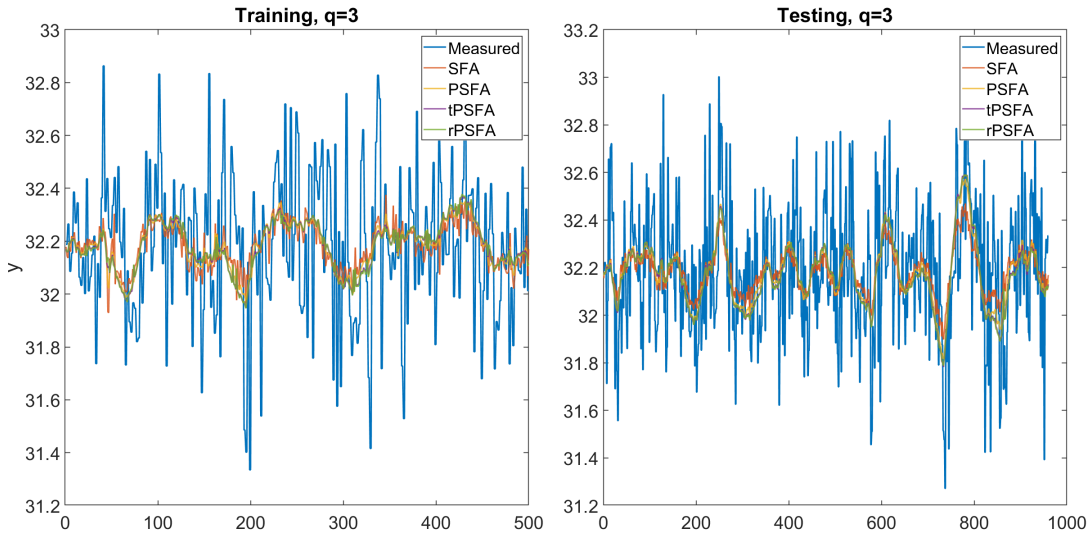

Figure 3.6: Slow Features Regression for Tennessee Eastman Training and Testing Data with 10% Outliers in Training.

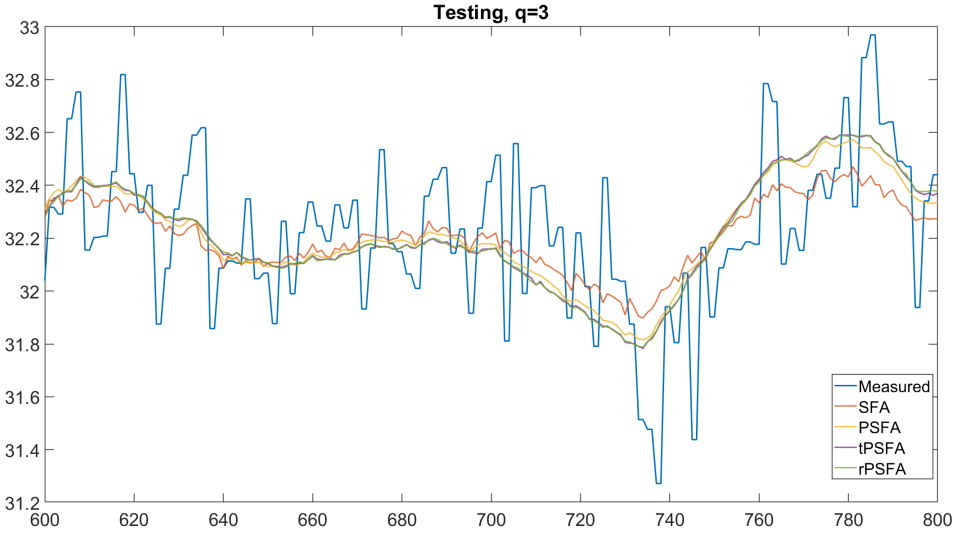

Figure 3.7: Subset of Slow Features Regression for Tennessee Eastman Testing Data.

Further, the efficacy of the developed robust soft sensor is tested for 50 Monte-Carlo simulations with outliers added according to $\mathcal{N}(0,4I)$ for each of the following outlier percentages: 1%, 3%, 5%, 10%, 20%, 30%, 40% and 50%. The average correlation coefficient and average mean squared error over each set of trials can be seen in Figure 3.8. Here it can be observed that all the methods show deteriorating performance with increased

outlier occurrence frequency. However, the proposed robust PSFA outperforms SFA and PSFA for all outlier fractions with slight improvement over tPSFA.

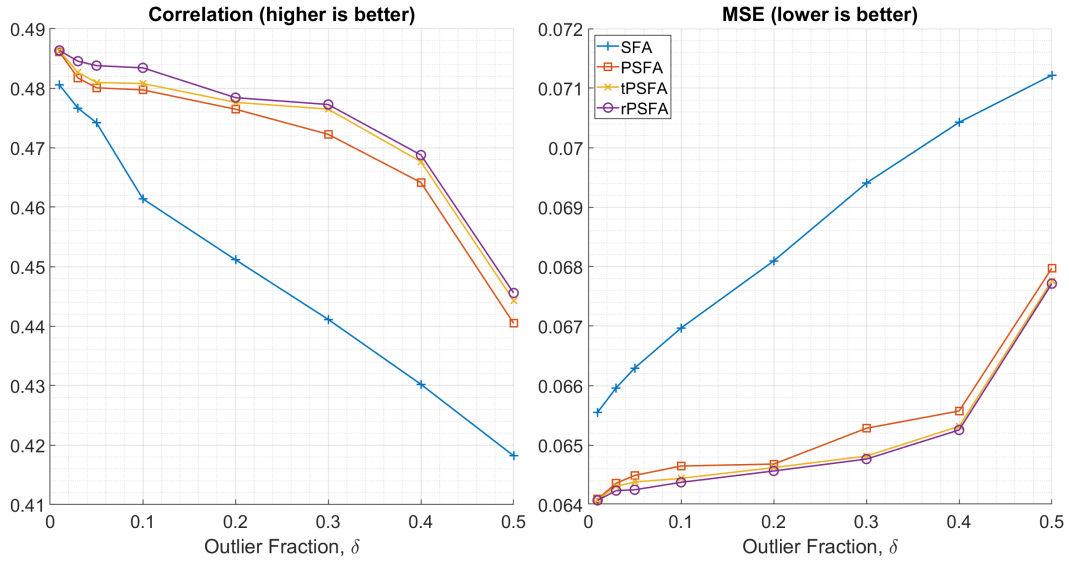

Figure 3.8: Slow Features Regression Correlation and MSE for Tennessee Eastman Testing Data with Various Gaussian Mixture Outlier Percentages in Training.

An additional set of Monte-Carlo simulations with 50 trials was performed by adding 10% outliers selected from $\mathcal{N}(0,\Sigma_{out}I)$ where Σ_{out} varied from one to ten. The average correlation coefficient and average mean squared error over each set of trials can be seen in Figure 3.9. Again, tPSFA outperforms PSFA and SFA, with rPSFA being slightly better still.

As the outliers added in the previous two Monte-Carlo simulations followed a Gaussian mixture, it is expected that modelling the noise as such would result in an improvement when compared to modelling it as a Student-t distribution. For the sake of fairness, then the simulations were repeated, and the added noise was selected from a Student-t distribution. First, 50 Monte-Carlo simulations were performed with outliers added according to $\mathcal{S}_t(0,2I,4)$ (which has a variance of $4I$) for each of the following outlier percentages: 1%, 3%, 5%, 10%, 20%, 30%, 40% and 50%. Next the outlier fraction was fixed as 10% and simulations were performed with outliers selected from $\mathcal{S}_t(0,2I,\nu)$ where ν is the degrees

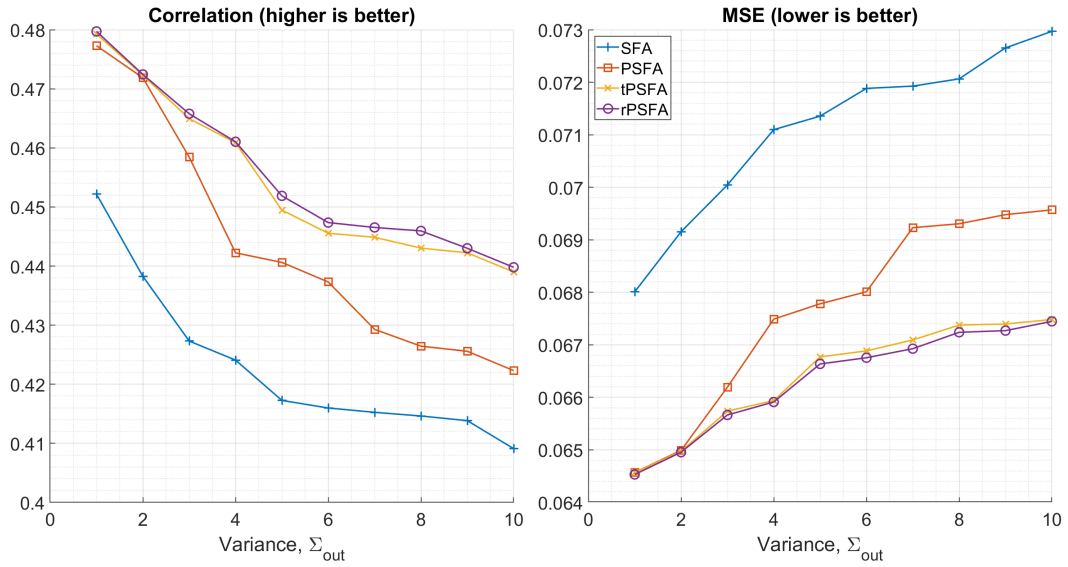

Figure 3.9: Slow Features Regression Correlation and MSE for Tennessee Eastman Testing Data with Various Gaussian Mixture Outlier Severities in Training.

of freedom of the Student-t distribution and lower values indicate more severe outliers. ν from one to ten were used with 50 trials at each value. The average correlation coefficient and average mean squared error over each set of trials can be seen in Figures 3.10 and 3.11 respectively.

Here it can again be seen that the rPSFA and tPSFA methods outperform the SFA and PSFA methods in the presence of outliers. The tPSFA method performs slightly better than the proposed rPSFA in these cases.

From these four sets of Monte-Carlo simulations it can be seen that the rPSFA method outperforms the tPSFA method when the outliers follow a Gaussian mixture. Conversely, when the outliers follow a Student-t distribution, tPSFA outperforms the rPSFA method as expected. In either case, the difference is small and both methods show clear improvement over SFA and PSFA. Thus, these simulations demonstrate both the proposed rPSFA and the existing tPSFA serve their own purposes respectively. However, the effectiveness depends on the actual outlier distributions. Since the actual outlier distribution is unknown, the proposed rPSFA provides an effective option and complements the tPSFA for practical

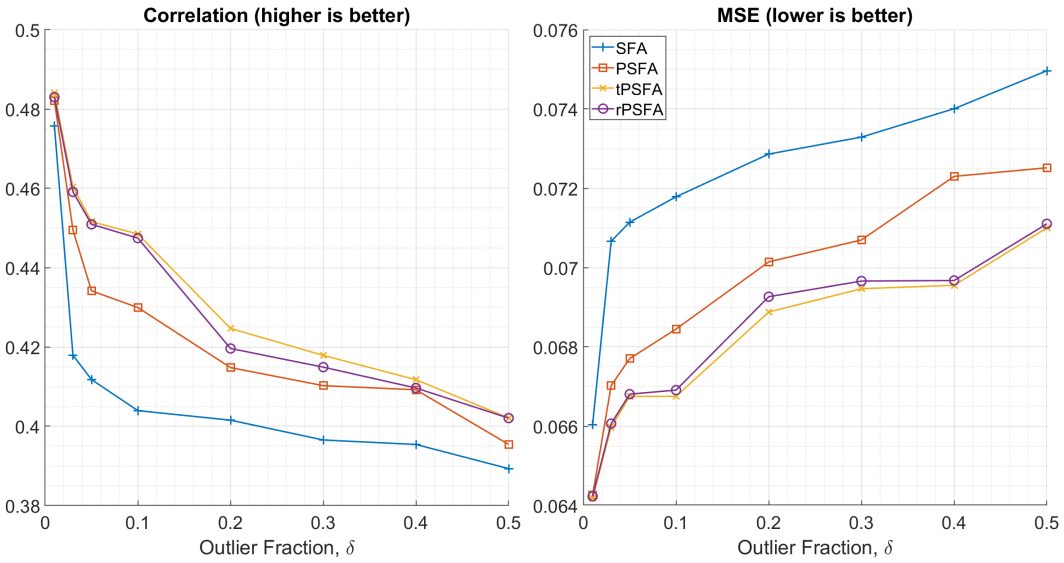

Figure 3.10: Slow Features Regression Correlation and MSE for Tennessee Eastman Testing Data with Various Student-t Outlier Percentages in Training.

applications. Further, the proposed rPSFA distinguishes between the outlier and the regular noise so that it provides a flexibility to model possible correlation between the outliers themselves as well as correlation with the regular noise as will be discussed in the next Chapter 3. In the next example with an industrial application where the outliers do not follow a known distribution, the proposed rPSFA shows better performance than tPSFA.

3.3.2 Industrial Case Study

In this case study the development of a soft sensor for a zinc roasting unit is considered. The main components of the roasting process consist of the furnace, waste heat boiler, cyclone, and electrostatic precipitator (ESP) as shown in Figure 3.12. In this process, the mined zinc concentrates are fed to and roasted in a fluidized bed furnace. To fluidize the bed, air supplemented with additional oxygen is sent in from the bottom of the furnace. The oxygen participates in exothermic chemical reactions with the concentrate and the generated heat is recovered by the waste-heat boiler. The sulphur dioxide gas from the final step is then cleaned for Hg removal, and H_2SO_4 is produced in an acid sub-plant. In total, the process consists of 13 variables of interest, which are listed in Table 3.3. Several of these variables

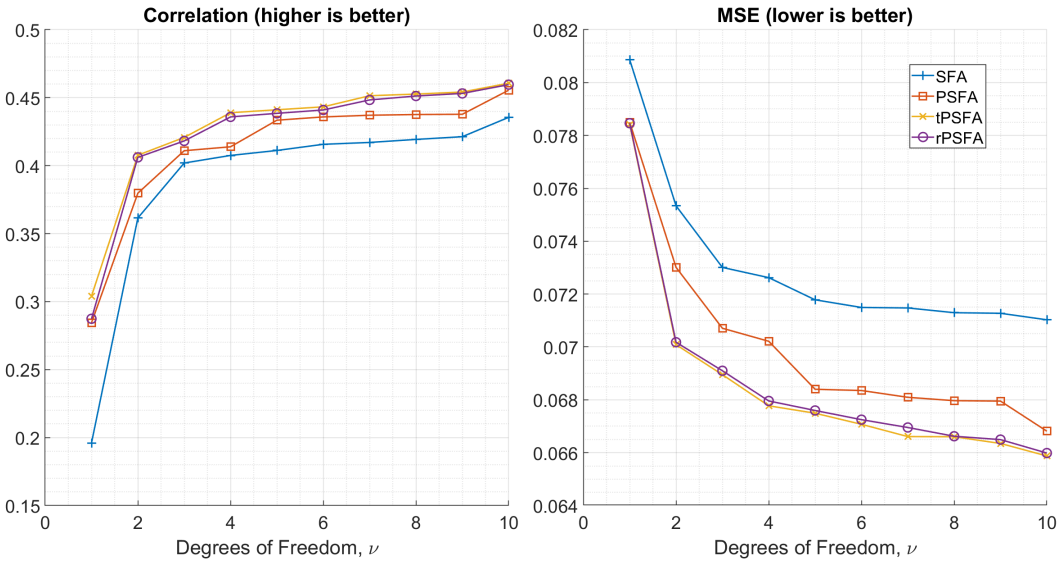

Figure 3.11: Slow Features Regression Correlation and MSE for Tennessee Eastman Testing Data with Various Student-t Outlier Severities in Training.

contain occasional outliers during the training period, which are believed to come from sensor faults and should therefore be considered in the soft sensor development.

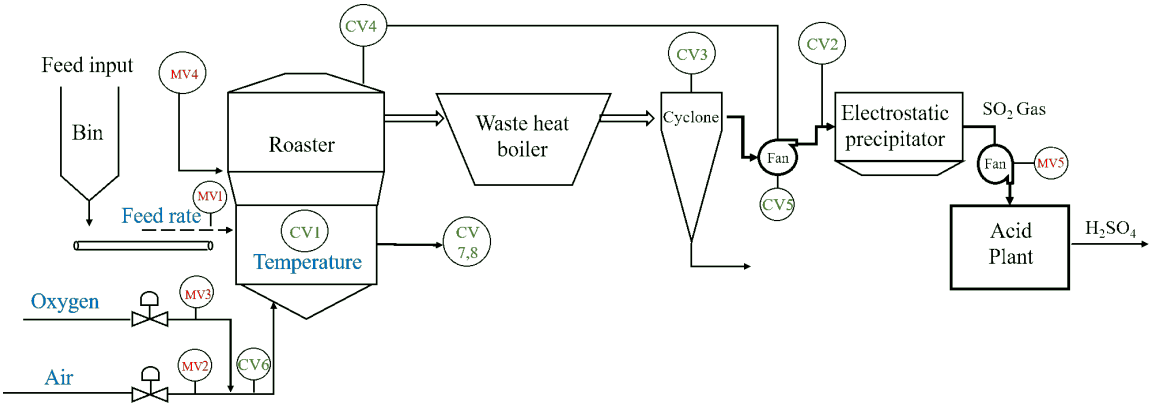

Figure 3.12: Zinc Roaster Process Diagram.

In this case study the goal is to build a slow feature regression soft sensor for the ESP pressure using the other variables. A set of slow features $q = 5$ are to be used and the parameters are estimated using the proposed algorithm. The lag-1 autocorrelations of the slow features can be seen in Figure 3.13. The conventional SFA and PSFA methods both show significantly reduced lag-1 autocorrelations and therefore increased speed during the

Table 3.3: Description of Zinc Roaster Process Variables.

Variable Description
Feed Rate
Air Flow Rate
Oxygen Flow Rate
Bed Spray Water
Inlet Pressure
Bed Temperature
ESP Pressure
Cyclone Temperature
Pressure Controller Output
Fan Speed
Oxygen Percentage
Required Amount of Oxygen
Air:Feed Ratio

testing phase while the proposed algorithm is still able to extract features that are more slowly varying.

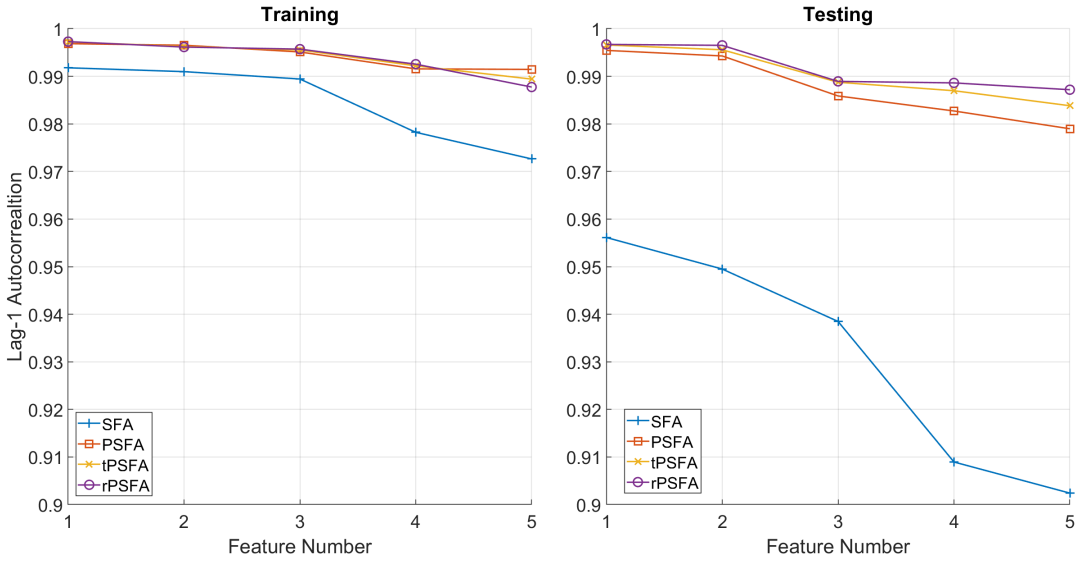

Figure 3.13: Slow Feature Lag-1 Autocorrelations for Zinc Roaster Training and Testing Data.

The results of the slow feature regression soft sensors can be seen in Figure 3.14 and Table 3.4 where it can be observed that during the testing phase the proposed robust model outperforms the conventional SFA and PSFA and shows improved or similar performance to tPSFA depending on the distribution of outliers. Additionally, all models have declined in performance when compared to the training data while the proposed method has the smallest decline.

3.4 Conclusions

This work proposed a PSFA algorithm that is robust to outliers, which occur with certain statistical distributions. This was done by treating the measurement disturbance as scale mixture of two Gaussian distributions and using a Bernoulli distribution to describe the statistics of outlier occurrence. The efficacy of the proposed approach was demonstrated both on a benchmark simulation and an industrial data set. From the results of both the case

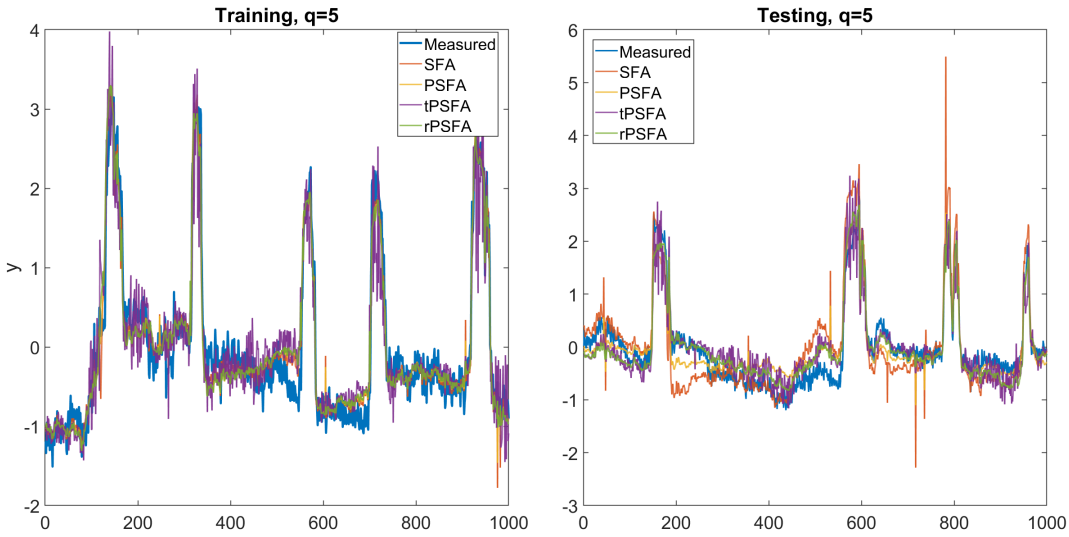

Figure 3.14: Slow Features Regression for Zinc Roaster Training and Testing Data.

studies it was observed that the proposed robust PSFA soft sensor is more robust against outliers when compared to the conventional SFA and PSFA and shows a slight improvement over tPSFA when the outliers follow a Gaussian mixture. It also shows better performance for an industrial data set where the outlier distribution is unknown.

Table 3.4: Slow Features Regression Performance for Zinc Roaster Training and Testing Data.

	Training		**Testing**	
	Corr.	MSE	Corr.	MSE
SFA	0.924	0.145	0.879	0.209
PSFA	0.968	0.0620	0.903	0.113
tPSFA	**0.974**	**0.0515**	0.922	0.0943
rPSFA	0.967	0.0967	**0.945**	**0.0654**

Chapter 4

Mixtures of Hidden Markov Model Robust PSFA

In preparation to submit as: Cameron Dyson, Jayaram Valluru, Biao Huang, and Graham Slot ”Hidden Markov Approach to Robust PSFA with Dynamic Switching”

4.1 Introduction

In Chapter 3, a measurement outlier robust PSFA (rPSFA) model was developed by assuming that the measurement noise followed a scale mixture of Gaussians switched according to a Bernoulli distribution. When compared to the existing Student-t based robust PSFA [30], the Gaussian mixture framework has similar performance but greater opportunity to be extended to handle alternate outlier related issues since it treats each noise mode separately.

One such extension would be to consider the case where outliers do not occur completely at random with respect to time, and instead follow some correlation of their own. In practice, sensor faults and samples with increased noise variance often occur in quick succession of one another before a return to the typical disturbance distribution. To capture this dynamic outlier behaviour the noise could be assumed to still consist of two scaled Gaussians, however that they are switched according to a Hidden Markov Model (HMM) rather than a Bernoulli distribution. The former considers dynamic switching with autocorrelation between the modes while the latter assumes static switching without consideration of noise mode dynamics.

Further, current PSFA literature lacks a solution that is robust to measurement outliers while also handling multiple operating point processes. As there are many industrial processes with multiple operating points, such as in the full data set of the model-plant-mismatch detection and diagnosis task presented in Chapter 2, where slow feature analysis could otherwise prove useful, there is motivation to resolve this deficiency.

In this chapter, to address both of these concerns simultaneously, i.e., correlated switching between inlier and outlier noise modes and multiple operating point processes, a mixture modelling approach to a Hidden Markov Model robust PSFA algorithm (mhrPSFA) is developed. In Section 4.2, the relevant methods are presented along with a summary of their related literature. Next in Section 4.3, the methodology under which the parameters of the mhrPSFA model may be estimated is presented. To demonstrate and verify the proposed

method, Section 4.4 presents a soft sensor case study for a simulated system with correlated inlier-outlier noise mode switching, as well as soft sensor and model-plant-mismatch detection and diagnosis case studies for an industrial system with outliers and multiple process operating conditions. Lastly, Section 4.5 provides conclusions and possible future directions.

4.2 Fundamentals

This section develops the approach used in this work. First a summary of the rPSFA algorithm proposed in Chapter 3 is presented. Next an overview of Hidden Markov Models and mixture modelling is provided. Last, the mixture slow feature regression, and the Model Quality Index (MQI) used in the case studies are discussed.

4.2.1 Robust Probabilistic SFA

Conventional [22] and probabilistic SFA [28, 29] methods have been well studied and greater detail pertaining to them can be found in Chapter 2 and Chapter 3 respectively.

Chapter 3 proposed the rPSFA model that is robust to outliers in the measurement space using a scaled Gaussian mixture noise switched according to a Bernoulli distribution. This retains the benefits of PSFA while enabling the use of training data that contains outliers.

The formulation of rPSFA is similar to conventional PSFA and is defined as follows:

$$s(t) = Fs(t-1) + e_s(t), \quad e_s(t) \sim \mathcal{N}(0, \Gamma) \tag{4.1}$$

$$x(t) = Hs(t) + e_x(t), \quad e_x(t) \sim (1-\delta)\mathcal{N}(0, \Sigma) + \delta\mathcal{N}(0, \rho^{-1}\Sigma) \tag{4.2}$$

where the two additional parameters when compared to conventional PSFA are $\delta \in [0, 1)$, which represents the fraction of samples that are outliers, and $\rho \in (0, 1]$, which is the outlier scaling factor with lower values corresponding to increased outlier severity. This approach introduces an additional hidden variable, $q_x(t)$, which acts as an indicator to determine whether the measurement noise in each time sample came from the inlier distribution,

where $q_x(t) = 1$, or outlier distribution, where $q_x(t) = \rho$. The switching of this hidden variable was modelled with a Bernoulli distribution which is indiscriminate with respect to its own state in the previous time sample.

4.2.2 Hidden Markov Models

In the presence of outliers in the process data, the common assumption that the noise follows a single Gaussian distribution can lead to inaccurate parameter estimates in the training step. In this work, to account for outliers in the measured variables, the noise in the measurements is modelled using scale mixture of Gaussians which are switched following a Hidden Markov Model with two states [69, 81] representing a dynamic relation between noise modes. HMMs have shown to be useful in various applications with the possibility for several extensions including handling missing data [82] and nonlinearity [83, 84]. Similar to the previously developed rPSFA that uses a Bernoulli distribution to switch between inliers and outliers, an additional hidden variable is introduced to the conventional PSFA, $Q_x = [q_x(1), ...q_x(N)]$. This is a binary vector indicating which of the two Gaussians a particular sample of $e_x(t)$ is drawn from. This takes on a value of $q_x(t) = 1$ in the presence of regular noise, or a value of $q_x(t) = \rho$ for samples that correspond to outliers. Rather than the completely at random approach associated with a Bernoulli distribution [72, 75], here the indicator variable is assumed to be governed by a transition probability. This enables the consideration of correlation between the previous state of the indicator variable, $q_x(t-1)$, and the current value of $q_x(t)$. When compared to conventional PSFA the additional parameters of the model are a transition matrix, α, and a scaling factor, $\rho \in (0, 1]$:

$$s(t) = Fs(t-1) + e_s(t), \quad e_s(t) \sim \mathcal{N}(0, \Gamma) \tag{4.3}$$

$$x(t) = Hs(t) + e_x(t), \quad e_x(t) \sim \begin{cases} \mathcal{N}(0, \Sigma), \ for\ q_x(t) = 1 \\ \mathcal{N}(0, \rho^{-1}\Sigma), \ for\ q_x(t) = \rho \end{cases} \tag{4.4}$$

The transition matrix α represents the probability of $q_x(t)$ switching between time steps:

$$\begin{aligned}\alpha &= \begin{bmatrix} \alpha_{1,1} & \alpha_{\rho,1} \\ \alpha_{1,\rho} & \alpha_{\rho,\rho} \end{bmatrix} \\ &= \begin{bmatrix} \alpha_{1,1} & 1-\alpha_{1,1} \\ 1-\alpha_{\rho,\rho} & \alpha_{\rho,\rho} \end{bmatrix}\end{aligned} \tag{4.5}$$

where $\alpha_{i,j} = P(q_x(t) = i | q_x(t-1) = j)$. Additionally, to estimate the initial condition of the HMM, the prior of each mode, π_i, is needed. In the presently considered case of two measurement noise modes, this can be simplified to a single parameter $\pi = \pi_1 = 1 - \pi_\rho$. Thus, for measurement data, $X = [x_1(t), ..., x_m(t)]$, $X \in \mathbb{R}^{m \times N}$, by considering the scaled Gaussian mixture with HMM dynamics, the unknown parameters of the PSFA model to be estimated are $\theta = \{H, \Gamma_1, ..., \Gamma_q, \sigma_1^2, ..., \sigma_m^2, \alpha, \pi, \rho\}$ and the hidden variables are $S = [s_1(t), ..., s_q(t)]$, $S \in \mathbb{R}^{q \times N}$ and Q_x.

As an illustrative example of HMM behaviour, when $\alpha_{1,1} = 0.95$ and $\alpha_{\rho,\rho} = 0.8$, the probability that an outlier occurs given that an outlier occurred in the previous sample is 80% while there is a 20% chance to transition back to the inliers. If instead in the previous sample an inlier had occurred then there would be a 5% chance to transition to an outlier, meanwhile there would be a 95% chance for another inlier to occur. This is graphically represented in Figure 4.1, where starting from the circle corresponding to the state at $t-1$, a state for t can be generated by selecting one of the two arrows out of the current circle at the probabilities labelled on the arrows. The result will either be to return to the original circle or to transition to the other, with the value in the current circle being the noise mode indicator $q_x(t)$. This process repeats for each new time step. Note that the Bernoulli approach can be considered as a special case of the HMM approach when the latter is constrained to $\alpha_{1,1} = 1 - \delta$, $\alpha_{\rho,\rho} = \delta$ and $\pi = 1 - \delta$.

4.2.3 Mixture Models

Many industrial processes include multiple operating conditions, each with their own set of dynamic behaviours. Several latent variable methods including linear and conventional

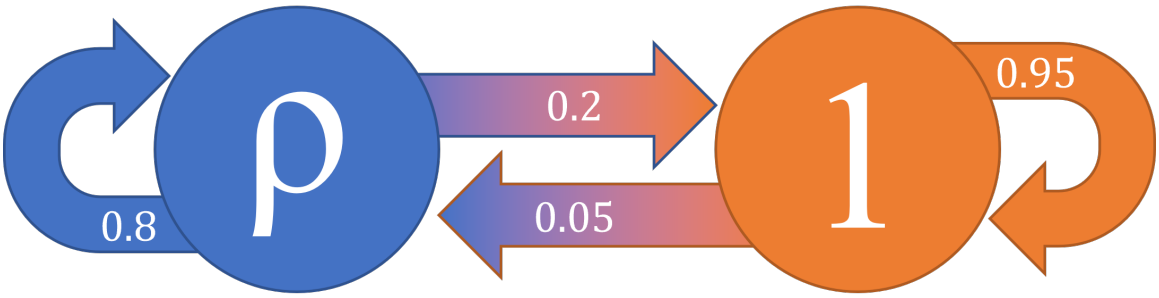

Figure 4.1: Diagram of HMM Transition Structure.

probabilistic SFA, have been extended to handle such cases and have shown improved performance [33, 74]. Similar methods exist for other non-SFA based latent variable approaches and have been used to perform soft sensor modelling and fault detection tasks in systems which contain multiple operating conditions [52, 53, 85, 86]. Further, some latent variable methods have extensions that consider both robustness to outliers, and multiple operating point processes simultaneously [87–90]. However, a PSFA based method which is robust to the presence of measurement outliers while also addressing multiple operating point behaviour has not been well studied in the literature. To resolve this deficiency, an extension of the robust PSFA method presented in Chapter 3 with the additional consideration of HMM based noise mode dynamics is proposed. A mixture of K independent HMM robust PSFA models are to be developed. Each of the $k = 1,...,K$ models should correspond to an operating point. The value of K is assumed to be known in advanced, relying on process knowledge. Independent slow feature regression models can then be developed for each operating point and the resulting predictions blended using a mixing factor, $P(k(t))$ which is the posterior of a given sample coming from model k and is subject to the following constraint:

$$\sum_{k=1}^{K} P(k(t)) = 1, \quad \forall t = 1,...,N \tag{4.6}$$

Further, SFA assumes that all measured signals have zero mean, which in the conventional method is easily obtained by subtracting the means from each signal or by performing a normalization step before conducting the analysis. However, when there are multiple operating points with different means, as is often the case in industrial data, then this step must be performed separately for each operating point. Combined, these two concerns then require the introduction of an additional hidden variable that indicates the process operating point from which a given sample is originated and a parameter for each of the models which represents the mean of the measured signals across that operating point, $\mu_{x,k}$.

4.2.4 Mixture Slow Feature Regression

In order to develop a mixture slow feature model, a weighted linear regression is to be performed for each operating point with the sample weight vector of model k being $P(k(t))$:

$$\underset{b_k,c_k}{arg\ min} \sum_{t=1}^{N} P(k(t)) \left| y(t) - c_k - \sum_{j=1}^{q} s_{j,k}(t) b_{j,k} \right|^2 \tag{4.7}$$

$$\hat{y}_k(t) = b_k^T s_{1:q,k}(t) + c_k \tag{4.8}$$

The overall soft sensor prediction is then the weighted mixture across each operating point:

$$\hat{y} = \sum_{k=1}^{K} P(k(t)) \hat{y}_k \tag{4.9}$$

The performance of such soft sensors can be measured through the mean squared error ($MSE \in [0, \infty)$) and the Pearson's correlation coefficient ($corr. \in [-1, 1]$).

$$MSE = \frac{1}{N} \sum_{t=1}^{N} \left(y(t) - \hat{y}(t) \right)^2 \tag{4.10}$$

$$corr. = \frac{cov(\hat{y}, y)}{\sigma_{\hat{y}} \sigma_y} \tag{4.11}$$

where cov represents the covariance between two signals and σ is the standard deviation of a signal.

4.2.5 Model Quality Assessment

The selected model quality assessment method used in the model-plant-mismatch detection and diagnosis case study is the Model Quality Index with a leave one out approach [17]. This metric compares the sum of the squared one-step ahead prediction error of the model used in the MPC to the prediction error of a fitted high-order dynamic model. Further detail can be found in Chapter 2.

4.3 Solution

In this section the development of a mixture modelling approach to HMM robust PSFA (mhrPSFA) is presented. In the presence of hidden variables, direct maximization of the log-likelihood function (i.e., $P(X|\theta)$ where X is the set of observations and θ the parameter set) w.r.t. unknown parameters is intractable. To address these kinds of problems, the Expectation Maximization algorithm [76] is widely used in literature, where instead of maximizing the log-likelihood function, the expectation of the joint log-likelihood function is maximized w.r.t. to the unknown parameters. The EM algorithm involves two steps, the E-step (expectation step) and M-step (maximization step), which are discussed in detail as follows.

In the E-Step, given the parameters θ for iteration count r, the expectation of the joint log-likelihood function (Q-function) is evaluated w.r.t. the hidden variables (S, Q_x) i.e.:

$$\mathbb{Q} = \mathop{\mathbb{E}}_{S,Q_x,k|X,\theta_r}(logP(\overbrace{X}^{Observed},\overbrace{S,Q_x,k}^{Hidden}|\theta)) \tag{4.12}$$

where:

$$\theta = \theta_1,...,\theta_K \tag{4.13}$$

$$\theta_k = \{H_k,\Gamma_{1,k},...,\Gamma_{q,k},\sigma^2_{1,k},...,\sigma^2_{m,k},\alpha_k,\pi_k,\rho_k,\mu_{x,k}\} \tag{4.14}$$

As the noises corresponding to input variables and latent variables are independent, the complete likelihood function can be written as follows:

$$\begin{aligned} P(X,S,Q_x,k|\theta) &= P(X|S,Q_x,k,\theta)P(S,Q_x,k|\theta) \\ &= P(X|S,Q_x,k,\theta)P(S|\theta)P(Q_x|\theta)P(k) \\ &= \prod_{t=1}^{N} P(x(t)|s(t),q_x(t),\theta)P(s(t)|s(t-1),q_x(t),\theta) \\ &\times P(q_x(t)|\theta)P(k(t)) \end{aligned} \tag{4.15}$$

Substituting (4.15) into (4.12) and expanding the Q-function yields:

$$\begin{aligned} \mathbb{Q} &= \mathop{\mathbb{E}}_{S,Q_x,k|X,\theta_r}(logP(X,S,Q_x,|\theta)) \\ &= \underbrace{\mathop{\mathbb{E}}_{S,Q_x,k|X,\theta_r}\Big(logP(s(1)|k(t),\theta)\Big)}_{\mathbb{Q}_A} \\ &+ \underbrace{\mathop{\mathbb{E}}_{S,Q_x,k|X,\theta_r}\Big(\sum_{t=2}^{N} logP(s(t)|s(t-1),q_x(t),k(t),\theta)\Big)}_{\mathbb{Q}_B} \\ &+ \underbrace{\mathop{\mathbb{E}}_{S,Q_x,k|X,\theta_r}\Big(\sum_{t=1}^{N} logP(x(t)|s(t),q_x(t),k(t),\theta)\Big)}_{\mathbb{Q}_C} \\ &+ \underbrace{\mathop{\mathbb{E}}_{S,Q_x,k|X,\theta_r}\Big(\sum_{t=2}^{N} logP(q_x(t)|q_x(t-1),k(t),\theta)\Big)}_{\mathbb{Q}_D} \\ &+ \underbrace{\mathop{\mathbb{E}}_{S,Q_x,k|X,\theta_r}\Big(logP(q_x(1)|k(t),\theta)\Big)}_{\mathbb{Q}_E} \\ &+ \underbrace{\mathop{\mathbb{E}}_{S,Q_x,k|X,\theta_r}\Big(\sum_{t=1}^{N} logP(k(t))\Big)}_{\mathbb{Q}_F} \end{aligned} \tag{4.16}$$

4.3.1 Parameter Estimation

In the M-step the Q-function is to be maximized w.r.t. θ. Note that this process is similar to that presented in Chapter 3 with the addition of each sample being weighted by $P(k(t))$,

which indicates from which operating point a sample is drawn.

$$\theta^{new} = arg\max_{\theta} \mathbb{Q}(\theta, \theta^{old}) \tag{4.17}$$

F_k and Γ_k share the same parameters ($\Gamma_{j,k}s$) which only appear in $\mathbb{Q}_B$, and by setting the respective partial derivative to zero, each Γ_j^{new} can be found as the root in $[0,1)$ of a cubic polynomial:

$$\frac{\partial \mathbb{Q}}{\partial \Gamma_{j,k}} = \frac{\partial \mathbb{Q}_B}{\partial \Gamma_{j,k}} = 0 \tag{4.18}$$

$$a_{3,k}\Gamma_{j,k}^{3,new} + a_{2,k}\Gamma_{j,k}^{2,new} + a_{1,k}\Gamma_{j,k}^{new} + a_{0,k} = 0 \tag{4.19}$$

where:

$$\begin{aligned}
a_{3,k} &= \sum_{t=2}^{N} P(k(t)|x(t), \theta^{old}) \\
a_{2,k} &= -\sum_{t=2}^{N} P(k(t)|x(t), \theta^{old}) E\Big(s_{j,k}(t)s_{j,k}(t-1)\Big) \\
a_{1,k} &= \sum_{t=2}^{N} P(k(t)|x(t), \theta^{old}) \Big(\big(E(s_{j,k}(t)s_{j,k}(t)^T) + E(s_{j,k}(t-1)s_{j,k}(t-1)^T)\big) - 1\Big) \\
a_{0,k} &= a_{2,k}
\end{aligned} \tag{4.20}$$

For each $k \in 1, ..., K$ the updated H_k^{new} can be found as follows:

$$\frac{\partial \mathbb{Q}}{\partial H_k} = \frac{\partial \mathbb{Q}_C}{\partial H_k} = 0 \tag{4.21}$$

$$\begin{aligned}
H_k^{new} = &\Bigg(\sum_{t=1}^{N} P(k(t)|x(t), \theta^{old}) \big(P_{1,k}(t)x_k(t)E_1(s_k(t))^T \\
&+ \rho_k P_{\rho,k}(t)x_k(t)E_\rho(s_k(t))^T\big)\Bigg) \\
&\times \Bigg(\sum_{t=1}^{N} P(k(t)|x(t), \theta^{old}) \big(P_{1,k}(t)E_1(s_k(t)s_k(t)^T) \\
&+ \rho_k P_{\rho,k}(t)E_\rho(s_k(t)s_k(t)^T)\big)\Bigg)^{-1}
\end{aligned} \tag{4.22}$$

where each measured state is centered w.r.t each operating point as follows:

$$x_k(t) = x(t) - \mu_{x,k} \tag{4.23}$$

and the subscript of 1 indicates an inlier while ρ indicates an outlier, i.e $P_{1,k}(t) = P(q_{x,k}(t) = 1|x_k, \theta_k)$, $E_{1,k}(\cdot) = \mathbb{E}(\cdot|x_k, q_{x,k} = 1, \theta_k)$, $P_{\rho,k} = P(q_{x,k}(t) = \rho_k|x_k(t), \theta_k)$, and $E_{\rho,k}(\cdot) = \mathbb{E}(\cdot|x_k, q_{x,k}(t) = \rho_k, \theta_k)$. Each element of the updated $\Sigma_k^{new} = diag\{\sigma_{1,k}^{2new}, ..., \sigma_{m,k}^{2new}\}$ can be found as follows:

$$\frac{\partial \mathbb{Q}}{\partial \sigma_{i,k}^2} = \frac{\partial \mathbb{Q}_C}{\partial \sigma_{i,k}^2} = 0, 1 \leq i \leq m \tag{4.24}$$

$$\begin{aligned}\sigma_{i,k}^{2new} = \frac{1}{N}\sum_{t=1}^{N} P(k(t)|x(t), \theta^{old}) \Bigg[& P_{1,k}(t)\Bigg(\big(x_{i,k}^2(t) - 2h_{i,k}^{T,new} E_1(s_k(t))x_{i,k}(t)\big) \\ & + h_{i,k}^{T,new} E_1(s_k(t)s_k(t)^T) h_{i,k}^{new}\Bigg) \\ & + \rho_k P_{\rho,k}(t)\Bigg(\big(x_{i,k}^2(t) - 2h_{i,k}^{T,new} E_\rho(s_k(t))x_{i,k}(t)\big) \\ & + h_{i,k}^{T,new} E_\rho(s_k(t)s_k(t)^T) h_{i,k}^{new}\Bigg)\Bigg]\end{aligned} \tag{4.25}$$

The updated $\rho_k^{new}s$ can be found as follows:

$$\frac{\partial \mathbb{Q}}{\partial \rho_k} = \frac{\partial \mathbb{Q}_C}{\partial \rho_k} = 0 \tag{4.26}$$

$$\begin{aligned}\rho_k^{new} = & \, m\Bigg[\sum_{t=1}^{N} P(k(t)|x(t), \theta^{old}) P_{\rho,k}(t)\Bigg] \\ & \times \Bigg[\sum_{t=1}^{N} P(k(t)) P_{\rho,k}(t) \\ & \times E_\rho\Bigg(\big(x_k(t) - H_k s_k(t)\big)^T \Sigma_k^{-1,new}\big(x_k(t) - H_k s_k(t)\big)\Bigg)\Bigg]^{-1}\end{aligned} \tag{4.27}$$

Each element of each α_k^{new}, $\alpha_{j,*,k}^{new}$, can be found as follows:

$$\frac{\partial \mathbb{Q}}{\partial \alpha_k} = \frac{\partial \mathbb{Q}_D}{\partial \alpha_k} = 0 \tag{4.28}$$

$$\alpha_{j,*,k}^{new} = \Bigg[\sum_{t=2}^{N} \xi_{j,*,k}\Bigg] \times \Bigg[\sum_{*=1,\rho}\sum_{t=2}^{N} \xi_{j,*,k}\Bigg] \tag{4.29}$$

where:

$$\xi_{j,*,k} = P(k(t)|x(t),\theta^{old})P(q_{x,k}(t-1)=j, q_{x,k}(t)=*) \tag{4.30}$$

The updated π_k^{new}s can be found as follows:

$$\frac{\partial \mathbb{Q}}{\partial \pi_k} = \frac{\partial \mathbb{Q}_E}{\partial \pi_k} = 0 \tag{4.31}$$

$$\pi_k^{new} = \frac{\gamma_{1,1,k}}{\sum_{*=1,\rho} \gamma_{1,*,k}} \tag{4.32}$$

where:

$$\gamma_{t,*,k} = P(k(t)|x(t),\theta^{old})P(q_{x,k}(t)=*) \tag{4.33}$$

Last the updated mean of $x(t)$:

$$\mu_{x,k}^{new} = \frac{\sum_{t=1}^{N} P(k(t)|x(t),\theta^{old})x(t)}{\sum_{t=1}^{N} P(k(t)|x(t),\theta^{old})} \tag{4.34}$$

Further information regarding the motivation for these calculations can be found in Chapter 3.

4.3.2 Posterior Distribution of Hidden Variables

In order to obtain the necessary values to update the parameter estimation in each iteration step, the expectations related to the hidden variables must be found. Further, to develop a soft sensor model, the overall S_k of each sub-model is needed. The required expectations are derived as follows:

$$E_{*,k}(s_k(t)) = \mathbb{E}(s_k(t)|x_k, q_{x,k}(t)=*, k, \theta_k) \tag{4.35}$$

$$E_{*,k}(s_k(t)s_k(t)^T) = \mathbb{E}(s_k(t)s_k(t)^T|x_k, q_{x,k}(t)=*, k, \theta_k) \tag{4.36}$$

$$E_{*,k}(s_k(t)s_k(t-1)^T) = \mathbb{E}(s_k(t)s_k(t-1)^T|x_k, q_{x,k}(t)=*, k, \theta_k) \tag{4.37}$$

$$P_{1,k}(t|x_k,\theta_k)+P_{\rho,k}(t|x_k,\theta_k)\rho_k=\mathbb{E}(q_{x,k}(t)|x_k,\theta_k) \tag{4.38}$$

$$E(s_k(t))=\mathbb{E}(s_k(t)|x_k,q_{x,k},\theta_k) \tag{4.39}$$

$$E(s_k(t)s_k(t)^T)=\mathbb{E}(s_k(t)s_k(t)^T|x_k,q_{x,k},\theta_k) \tag{4.40}$$

$$E(s_k(t)s_k(t-1)^T)=\mathbb{E}(s_k(t)s_k(t-1)^T|x_k,q_{x,k},\theta_k) \tag{4.41}$$

Equations (4.35)-(4.37) represent the estimate of the slow features given that all samples are either inliers or outliers. Each $E_{*,k}$ is to be found twice, one for the inlier case where $*=1$, and again for the outlier case where $*=\rho_k$.

$$\begin{aligned}
&E_{*,k,k}(s_k(t))=\hat{\mu_{t*,k}}\\
&E_{*,k,k}(s_k(t)s_k(t)^T)=\hat{V_{t*,k}}+\hat{\mu_{t*,k}}\hat{\mu_{t*,k}}^T\\
&E_{*,k,k}(s_k(t)s_k(t-1)^T)=J_{t-1*,k}\hat{V_{t*,k}}+\hat{\mu_{*,k}}t\hat{\mu_{t-1*,k}}^T
\end{aligned} \tag{4.42}$$

With the forward recursions of a Linear Dynamical System [29, 77, 78]:

$$\begin{aligned}
&P_{t-1*,k}=F_kV_{t-1*,k}F_k^T+\Gamma_k\\
&\mu_{t,k}=F_k\mu_{t-1*,k}+K_{t*,k}[x_k(t)-H_kF_k\mu_{t-1*,k}]\\
&V_{t*,k}=(I-K_{t*,k}H_k)P_{t-1*,k}\\
&K_{t*,k}=P_{t-1*,k}H_k^T(H_kP_{t-1*,k}H_k^T+\Sigma_{*,k})^{-1}
\end{aligned} \tag{4.43}$$

Initialized with:

$$\begin{aligned}
&\mu_{1*,k}=K_{1*,k}x_k(1)\\
&V_{1*,k}=I-K_{1*,k}H_k\\
&K_{1*,k}=H_k^T(H_kH_k^T+\Sigma_{*,k})^{-1}
\end{aligned} \tag{4.44}$$

The backwards recursion can then be done:

$$\begin{aligned}
&\hat{\mu_{t*,k}}=\mu_{t*,k}+J_{t*,k}(\hat{\mu_{t+1*,k}}-F_k\mu_{t*,k})\\
&\hat{V_{t*,k}}=V_{t*,k}+J_{t*,k}(\hat{V_{t+1*,k}}-P_{t*,k})J_{t*,k}^T\\
&J_{t*,k}=V_{t*,k}F_k^TP_{t*,k}^{-1}
\end{aligned} \tag{4.45}$$

which is initialized with:

$$\begin{aligned}
&\hat{\mu_{t*,k}}=\mu_{t*,k}\\
&\hat{V_{t*,k}}=V_{t*,k}
\end{aligned} \tag{4.46}$$

The $\Sigma_{*,k}$s are defined using the Gaussian scale mixture parameters as follows:

$$\begin{aligned}\Sigma_{1,k} &= \Sigma_k \\ \Sigma_{\rho,k} &= \rho_k^{-1}\Sigma_k\end{aligned} \tag{4.47}$$

Then the expectation (4.38) can be found using the posterior of the hidden variable $q_x(t)$ at each sample.

The overall expectations (4.39)-(4.41) can then be found as the weighted sum of the two noise mode cases. The weighs are to be determined according to the properties of HMMs [91]. For example (4.39) can be found as follows:

$$\begin{aligned}E(s_k(t)) &= \sum_{*=1,\rho} P(q_x(t)|x(t),k,\theta)E(s_k(t)|x,q_x(t)=*,\theta) \\ &= P_{1,k}(t)E_1(s(t)|x,q_x(t)=1,\theta)+P_{\rho,k}(t)E_\rho(s(t)|x,q_x(t)=\rho,\theta)\end{aligned} \tag{4.48}$$

The posterior of the mixture model weights can be updated using one of several methodologies. Lacking any process knowledge regarding when the operating point switching occurs, $P(k(t))$ can be updated during each iteration using the current parameter set, θ, according to Bayes rule [74] and Gaussian properties [91]:

$$P(k(t)|x(t),\theta^{old}) = \frac{P(x(t)|k,\theta^{old})P(k|\theta^{old})}{P(x(t)|\theta^{old})} \tag{4.49}$$

$$P(x(t)|k(t),\theta^{old}) \sim \mathcal{N}(\mu_{x,k}, H_k H_k^T + \Sigma_k) \tag{4.50}$$

In practice the direct determination of the operating point indicator according to the above method is difficult. Alternatively, some process operators may record the current operating point directly in the data historian. In such a case this information can be used to obtain $P(k(t))$ by assigning the recorded operating point to have $P(k(t)) = 1$ and all others to zero. In other situations, the operating point may not be recorded directly, however certain scheduling variables important to the operating point can be identified through process knowledge. The values of the scheduling variables during training can be used to develop a soft clustering model using methods such as a Gaussian mixture model or a HMM [92–94]. The posterior of each operating point k at each time sample in the clustering model can then be taken as the corresponding $P(k(t))$.

4.4 Verification and Application

In this section two sets of case studies are presented. First, a soft sensor is developed for a simulated process with a single operating point where inlier and outlier noises are switched between according to a transition probability matrix. In the second set of case studies an industrial zinc roaster process with two operating points and occasional outliers is considered for the development of a soft sensor and a model-plant-mismatch detection and diagnosis task.

4.4.1 Simulated Study

This case study is performed using Tennessee Eastman (TE) [79] simulation data [80] to verify the developed mhrPSFA model in the case of a single operating point ($K = 1 \therefore P(k(t)) = 1,\ \forall t = 1,...,N$) where outliers occur according to a transition probability matrix. The TE process consists of five main units: a reactor, stripper, condenser, compressor, and separator. The feed consists of four reactants (A, C, D and E) and three products (F, G and H). Motivated by Fan et al. [30], a slow feature regression soft sensor is used to build a model for y which is the concentration (mole %) of component A in the reactor feed. As per Fan et al. [30] five highly correlated variables were selected as inputs:

$$\begin{aligned} &x_1\text{: Normalized Reactor Pressure (kPag)} \\ &x_2\text{: Normalized Stripper Temperature } (^oC) \\ &x_3\text{: Normalized Stripper Steam Flow (kg/hr)} \\ &x_4\text{: Normalized Compressor Work (kW)} \\ &x_5\text{: Normalized Component C in Purge Gas (Mole \%)} \end{aligned} \tag{4.51}$$

In order to show the effect of correlated outliers, they were added to the input data according to a HMM with $\alpha_{1,1} = 0.95$ and $\alpha_{\rho,\rho} = 0.8$. The outliers were selected from $\mathcal{N}(0,4I)$. The training and testing data can be seen in Figures 4.2 and 4.3 respectively, and the generated Q_x can be seen in Figure 4.4. The difference between this simulation and the one presented

in Chapter 3 is that here, rather than occurring randomly at each time step, the outliers are more likely to occur after they have previously occurred.

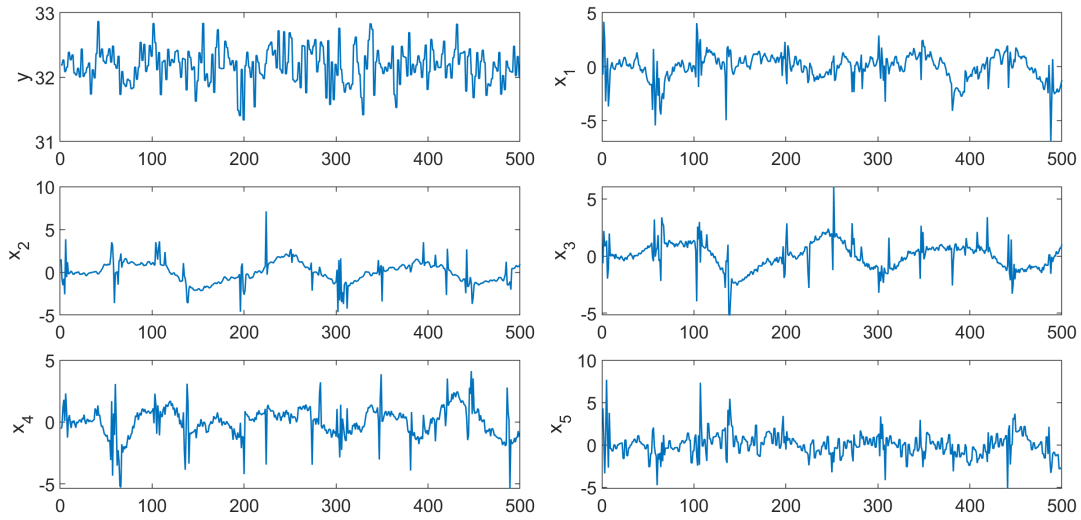

Figure 4.2: Simulated Tennessee Eastman Training Data with $\alpha_{1,1} = 0.95$ and $\alpha_{\rho,\rho} = 0.8$.

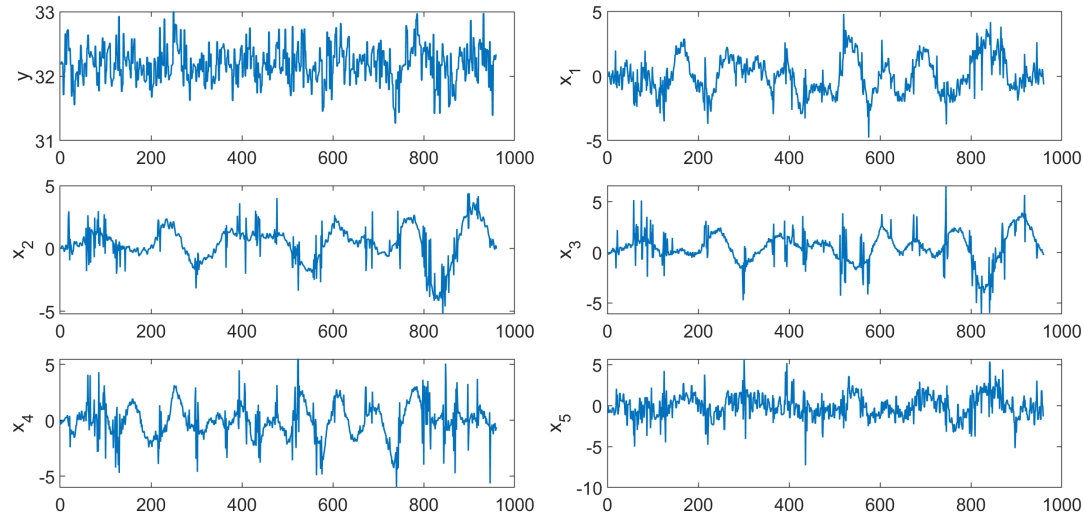

Figure 4.3: Simulated Tennessee Eastman Testing Data with $\alpha_{1,1} = 0.95$ and $\alpha_{\rho,\rho} = 0.8$.

The extracted slow features along with their lag-1 autocorrelations can be seen in Figure 4.5 and Table 4.1. Here it is seen that while all methods were able to obtain meaningful features, the HMM version extracted slower features, i.e., it extracted features with higher lag-

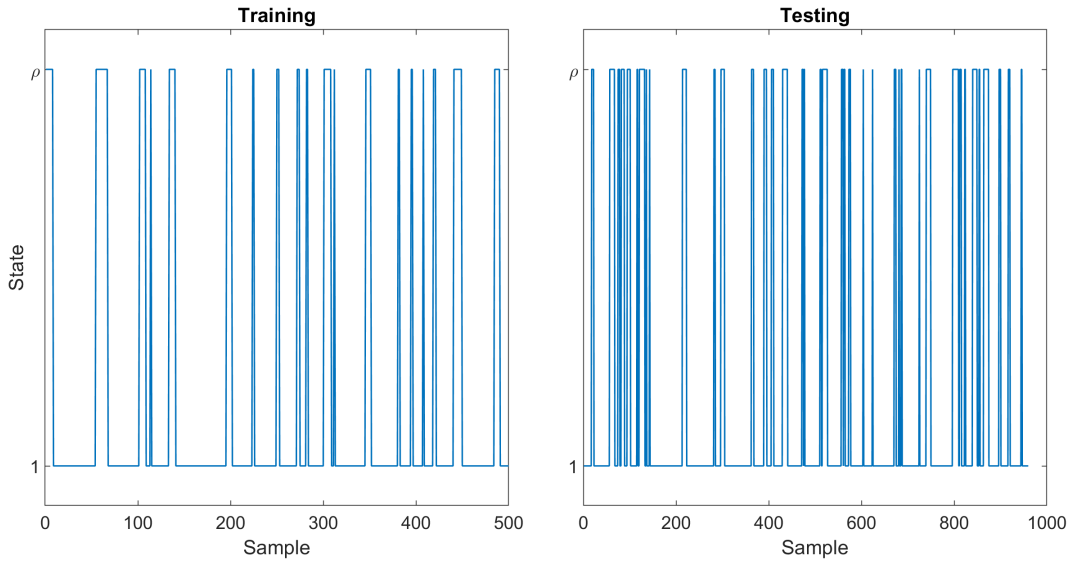

Figure 4.4: Generated Inlier-Outlier HMM States, $q_x(t)$, with $\alpha_{1,1} = 0.95$ and $\alpha_{\rho,\rho} = 0.8$.

1 autocorrelations when compared to the Bernoulli based Gaussian mixture and Student-t robust PSFA (tPSFA) [30] approaches, which implies slower features are extracted by the proposed method. The rPSFA and tPSFA performed similarly to one another. The HMM version was able to capture slow trends better and was less affected by outliers, particularly those that are of moderate severity.

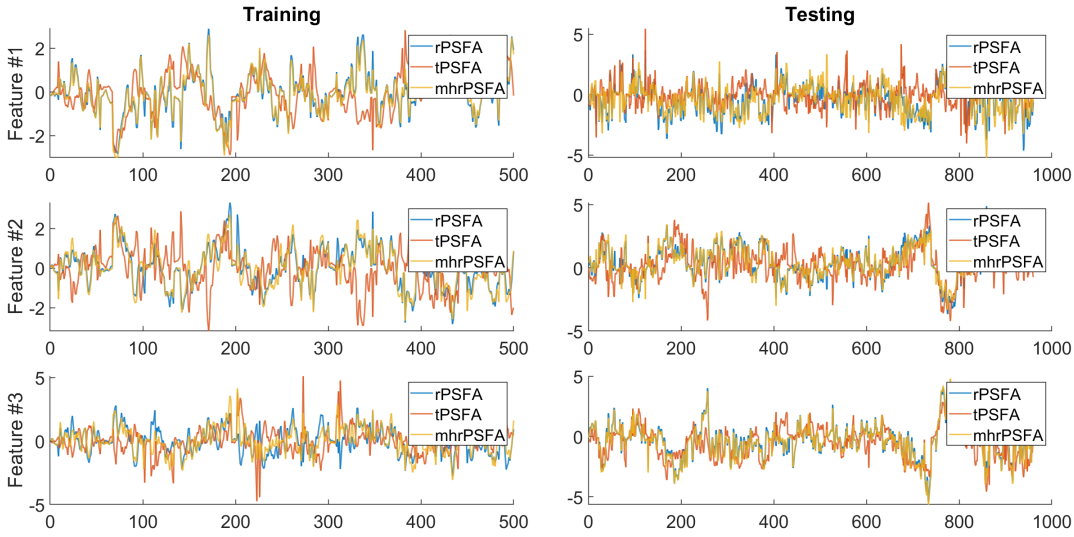

Figure 4.5: Extracted Slow Features for Tennessee Eastman Training and Testing Data with $\alpha_{1,1} = 0.95$ and $\alpha_{\rho,\rho} = 0.8$.

Table 4.1: Slow Features Lag-1 Autocorrelation Coefficients for Tennessee Eastman with $\alpha_{1,1} = 0.95$ and $\alpha_{\rho,\rho} = 0.8$.

	Testing			**Training**		
Feature	#1	#2	#3	#1	#2	#3
rPSFA	0.961	0.825	**0.789**	0.889	0.885	0.787
tPSFA	**0.998**	0.822	0.636	0.900	0.883	0.635
mhrPSFA	0.944	**0.871**	0.736	**0.910**	**0.886**	**0.790**

In the soft sensor task, the objective is to build a linear regression model for $y(t)$ utilizing the slow features, $s(t)$, extracted from $x(t)$. The results are presented in Figure 4.6 and Table 4.2 where during the testing phase, the proposed mhrPSFA method had a higher correlation coefficient and a lower mean squared error than rPSFA and tPSFA. From these results it is apparent that when the inlier-outlier state at time t is correlated with its previous state, modelling this behaviour with a HMM can yield improved performance when compared to methods that treat the occurrence of outliers as completely random.

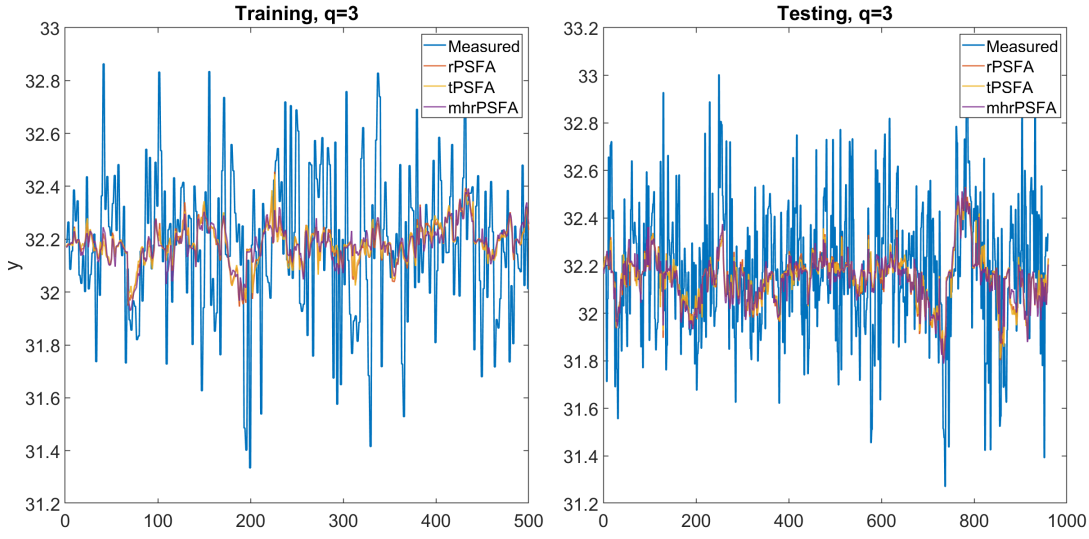

Figure 4.6: Slow Features Regression for Tennessee Eastman Training and Testing Data with $\alpha_{1,1} = 0.95$ and $\alpha_{\rho,\rho} = 0.8$.

Table 4.2: Slow Features Regression Performance for Tennessee Eastman Training and Testing Data with $\alpha_{1,1} = 0.95$ and $\alpha_{\rho,\rho} = 0.8$.

	Training		**Testing**	
	Corr.	MSE	Corr.	MSE
rPSFA	0.270	**0.0731**	0.313	0.0772
tPSFA	0.257	0.0736	0.312	0.0768
mhrPSFA	**0.253**	0.0738	**0.338**	**0.0757**

4.4.2 Industrial Case Studies

In these case studies a zinc roasting unit with a running model predictive controller is considered. The process consists of 5 manipulated variables (MVs) and 8 controlled variables (CVs), which are listed in Table 4.3 and greater detail is provided in Chapter 3. Several of these variables contain occasional outliers during the training and testing periods which are believed to come from sensor faults and measurement noise, this should therefore be taken into account for in the soft sensor development. Additionally, the process operates with two distinct operating points. The first is the normal operation and the second is a maintenance operation point where different behaviour can be observed, particularly in $y_{2:5}$.

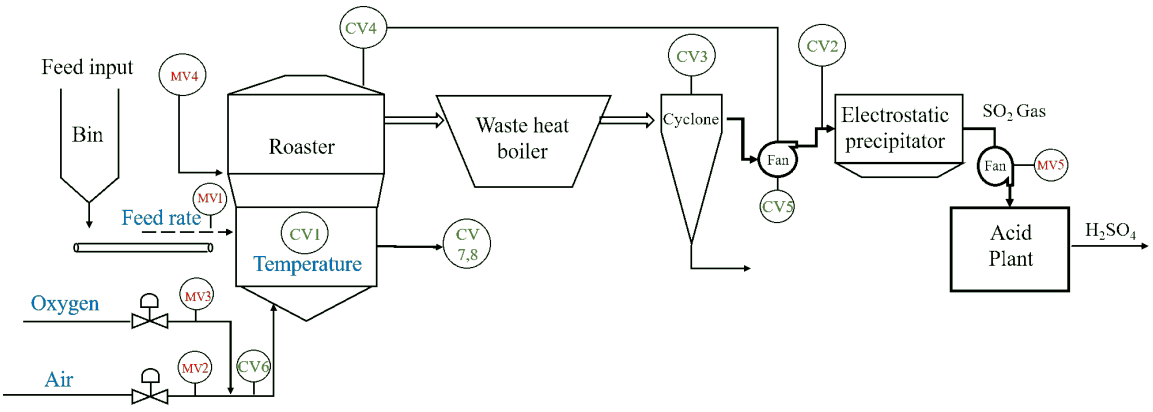

Figure 4.7: Zinc Roaster Process Diagram.

Two case studies are considered utilizing the proposed mhrPSFA algorithm. First, a soft

Table 4.3: Description of Zinc Roaster Process Variables.

Variable	**Description**
u_1	Feed Rate
u_2	Air Flow Rate
u_3	Oxygen Flow Rate
u_4	Bed Spray Water
u_5	Inlet Pressure
y_1	Bed Temperature
y_2	ESP Pressure
y_3	Cyclone Temperature
y_4	Pressure Controller Output
y_5	Fan Speed
y_6	Oxygen Percentage
y_7	Required Amount of Oxygen
y_8	Air:Feed Ratio

sensor is developed. This serves as an extension of the industrial case study presented in Chapter 3 and shows that utilizing a mixture model approach allows for multiple process operating points to be more accurately described than with the previously developed rPSFA. Second, an encoder-decoder is developed for use in MPC model-plant mismatch detection and diagnosis. This serves as an extension to Chapter 2 where a similar task was performed for the normal operating point alone utilizing conventional SFA.

Industrial Soft Sensor Case Study

In this case study, the goal is to build a slow feature regression soft sensor for the normalized ESP pressure (y_2) using the normalized versions of the other variables. A set of $q = 5$ slow features are to be used and the parameters are estimated using the proposed mhrPSFA model. The number of operating points is specified as $K = 2$ based on process knowledge and the mixture weightings are determined using the model parameters as described in (4.50). The training prediction, testing prediction, testing parity plot, and results are presented in Figure 4.8, Figure 4.9, Figure 4.10, and Table 4.4, respectively.

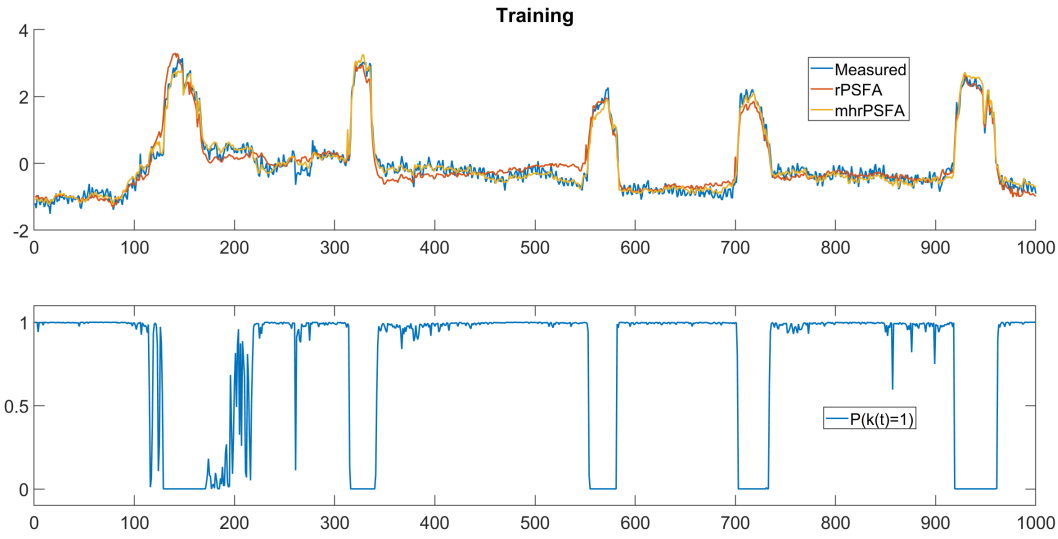

Figure 4.8: Slow Features Regression for Zinc Roaster Training Data with Mode-1 Indicator Variable.

From these figures it is evident that the previously developed rPSFA does not perform

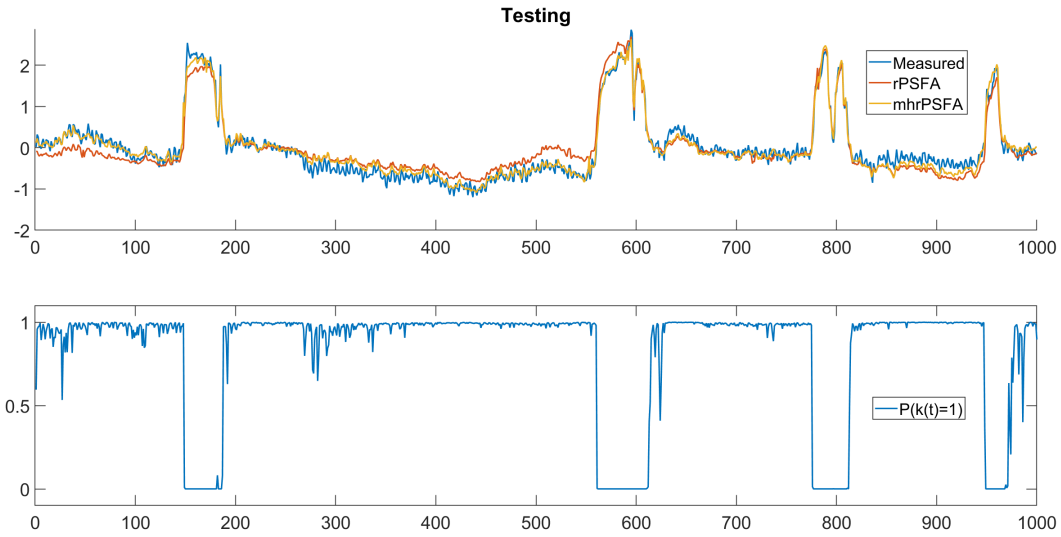

Figure 4.9: Slow Features Regression for Zinc Roaster Testing Data with Mode-1 Indicator Variable.

Table 4.4: Slow Features Regression Performance for Zinc Roaster Training and Testing Data.

	Training		**Testing**	
	Corr.	MSE	Corr.	MSE
rPSFA	0.967	0.0658	0.945	0.0654
mhrPSFA	**0.987**	**0.0253**	**0.983**	**0.0209**

well in processes with multiple operating points when compared to the proposed mhrPSFA. In Figure 4.10 an apparent bias in the rPSFA predictions can be seen in each of the identified operating modes, which approximately correspond to when the measured signal for y_2 is above or below one. The predictions for which $P(k(t) = 1) > P(k(t) = 2)$ (approximately corresponding to the normal operation point where $y(t) < 1$) show an apparent bias and tend to overestimate the measured value. The predictions for those samples where $P(k(t) = 2) > P(k(t) = 1)$ (approximately corresponding to the maintenance operation point where $y(t) > 1$) also show a bias and tend to underestimate the measured value. Further, increased

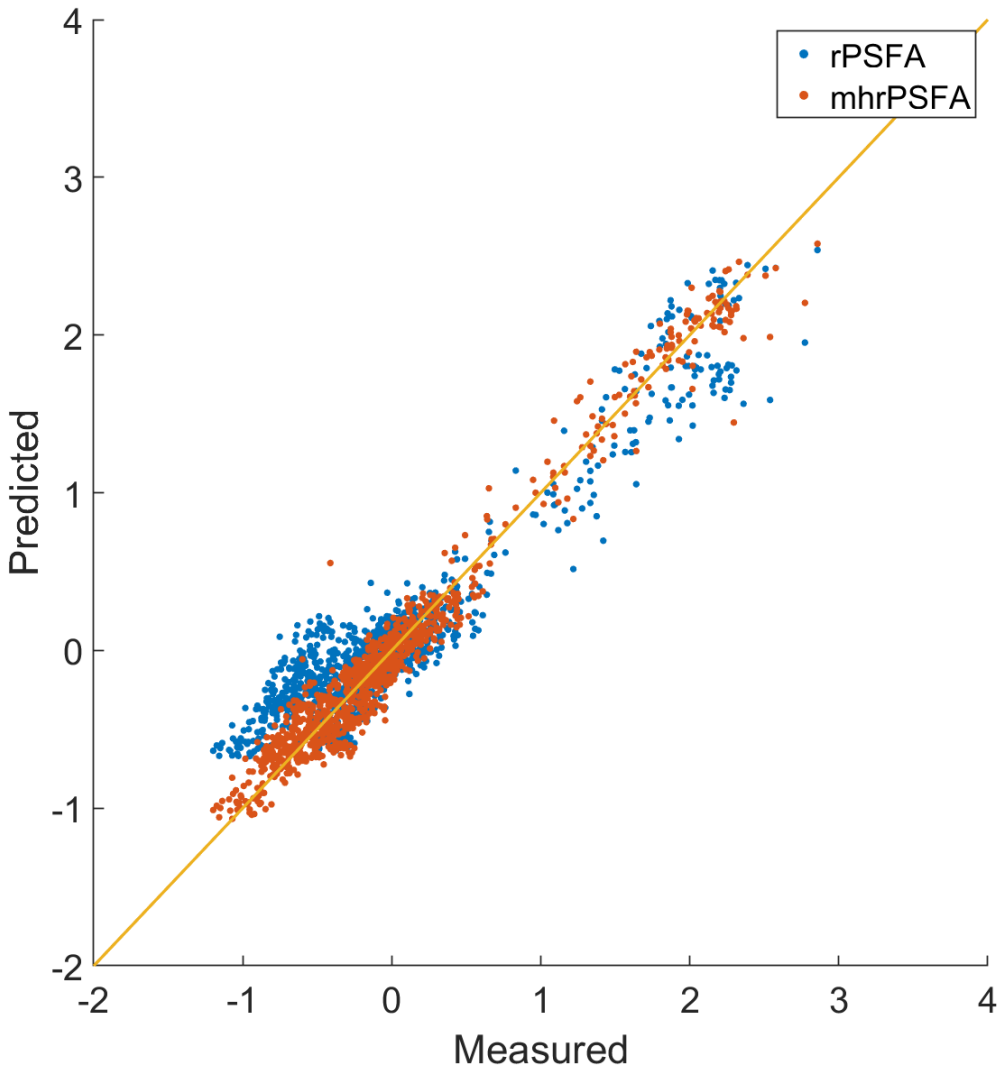

Figure 4.10: Parity Plot for Slow Features Regression for Zinc Roaster Testing Data.

error variance within each operating point when compared to the mhrPSFA can be seen. These issues are largely eliminated in the mhrPSFA predictions which follow the line of parity closely.

Figure 4.11 shows the soft sensor testing performance with alternate numbers of features with $q = 1, ..., 10$. From this it can be seen that the mhrPSFA algorithm is able to capture the plant behaviour more efficiently with a lower number of features when compared to rPSFA. In this system the disparity is particularly prevalent before the fourth feature.

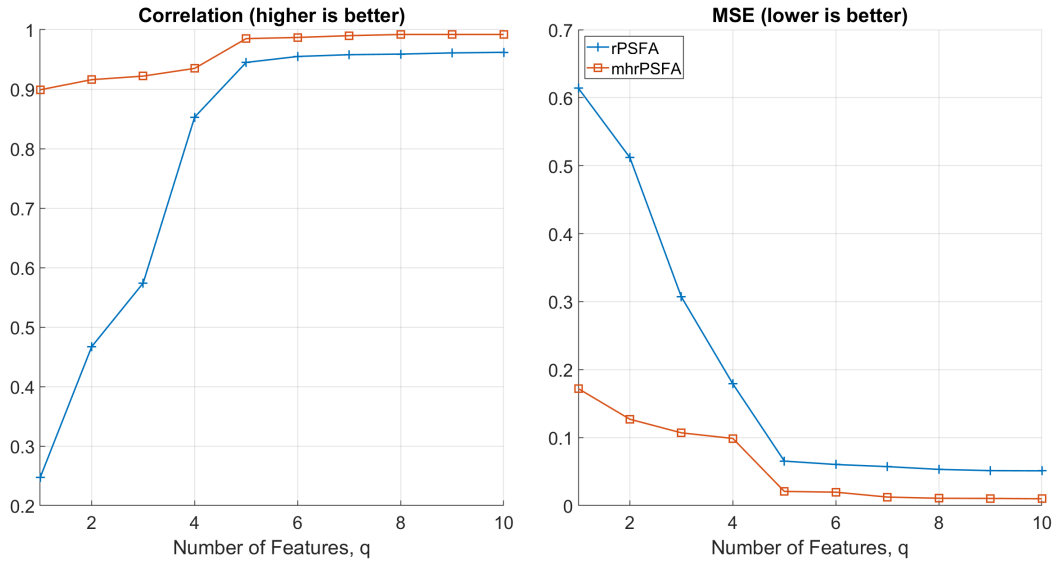

Figure 4.11: Slow Feature Regression Testing Performance with Varied Number of Features.

Industrial Model-Plant-Mismatch Detection and Diagnosis Case Study

In this case study the proposed mhrPSFA model is used to develop an encoder-decoder to preprocess data for use in a model-plant-mismatch detection task according to the methodology proposed in Section 2.2.3. This serves as an extension to the industrial case study presented in Section 2.3.2. However, without the limitations of a single SFA model, a larger data set can be considered containing both the normal and maintenance operation points at once. Here a period of 5,000 data points is considered for each of the testing and training and the MQILOO [17] metric is again selected as the model quality assessment method. The training period was selected from normal operating data as the period with the lowest mean squared 1-step ahead prediction error. The testing period of concern was selected as it occurred during a transitional period, after a brief shutdown of the plant where potential model-plant-mismatch was suspected to have been introduced. Utilizing the training data, a mhrPSFA model is developed to encode the signals using $q = 7$ components where all MVs and CVs are treated as inputs. A mixture slow feature regression model is then developed using the training data to decode the signals by reconstructing all MVs and CVs

from the encoded signal. The developed encoder-decoder is then applied to the training and testing data before conducting model-plant-mismatch detection and diagnosis. In this way the fast-varying signals that are suspected of being less relevant to the plant model are de-emphasized. Note that no lagged copy augmentation was introduced ($d = 0$) but all other parameters remain the same as the case study presented in Chapter 2, i.e., MQI is assessed using fifth order ARX models that use a window of 500 samples moved by 100 samples at a time, with sub-model deterioration being diagnosed by a three standard deviation increase from the mean of the ratio of MQILOO to MQI during the training period.

A summary of the results can be seen in Table 4.5. Compared to the case study presented in Section 2.3.2, similar results were obtained before and after preprocessing. Without using the mhrPSFA to preprocess the data, several potential model quality issues are detected, suggesting the requirement of extensive signal excitation to re-identify them accordingly. When utilizing a mhrPSFA preprocessing step however, a reduced set of potential model-plant-mismatch sources are detected only pertaining to u_4, suggesting less extensive re-identification is required. The mhrPSFA method required significantly less work in identifying relevant data to consider in this analysis when compared to the conventional SFA used in Chapter 2 which does not handle outliers or multiple operation conditions, and thereby limits its applications or requires the use of additional preprocessing methods that result in a loss of information.

4.5 Conclusions

This work proposed an extension to the previously developed scale Gaussian mixture robust PSFA to replace the random switching of the inlier and outlier mode occurrence with a correlated switching. This allows for the capture of dynamic noise mode switching and can better account for situations in which the occurrence of outliers is not completely random. Further, this extension was implemented in a mixture model fashion allowing for the proposed method to better capture the behaviour of multiple operating point processes

Table 4.5: Model-Plant-Mismatch Detection and Diagnosis Results by Sub-Model.

Checkmarks (✓) indicate acceptable performance, crosses (✗) indicate detection of model-plant-mismatch, and dashes (-) indicate the lack of a relevant sub-model. Results without mhrPSFA reconstruction are on the left of columns and with the reconstruction on the right.

	u_1		u_2		u_3		u_4		u_5	
	w/o	with	w/o	with	w/o	with	w/o	with	w/o	with
y_1	✓	✓	✓	✓	-		✓	✓	-	
y_2	✓	✓	✗	✓	✓	✓	✗	✗	✓	✓
y_3	✓	✓	✓	✓	✓	✓	✓	✓	-	
y_4	✓	✓	✗	✓	✗	✓	✗	✗	-	
y_5	✓	✓	-		✓	✓	✓	✓	-	
y_6	-		-		✗	✓	-		-	
y_7	✗	✓	✓	✓	✗	✓	-		-	
y_8	✓	✓	✓	✓	✓	✓	-		-	

when compared to the previously developed rPSFA. Further extensions of this work could include higher order HMM structures that consider multiple historical time steps [95–97], or alternate Gaussian mixtures to address outliers not-well described by a scale mixture of two components.

Chapter 5

Conclusions and Future Work

5.1 Conclusions

The main objectives of this thesis were to propose a SFA based framework for preprocessing industrial data for MPC model-plant-mismatch detection and diagnosis in chemical processes, and to develop robust PSFA models to address data quality concerns commonly found in industrial processes.

Industrial MPC systems commonly display some degree of change in behaviour over time which can introduce or worsen existing discrepancies between the prediction model and the plant. Such mismatches can result in sub-optimal set point tracking and determination. Frequent broad model re-identification experiments to resolve these issues is unfavourable due to potential economic and safety concerns. Instead, several methods exist in the literature to detect errors in specific sub-models that could benefit from model re-identification, thereby narrowing the scope of the potential experiments. However, the quality of assessment of these methods is diminished by the presence of significant disturbances. In chemical processes the plant is typically slowly varying in nature and disturbed by some quickly varying noises. To this end, a conventional SFA encoder-decoder structure to preprocess data and eliminate quickly varying disturbances that are typically of less value when detecting model-plant-mismatch was proposed in Chapter 2. Simulated and industrial case studies showed the proposed method's improved ability to more accurately determine the sub-models whose mismatch was present and provide a reduced set of candidates for re-identification.

Industrial processes often have data quality issues including outliers and multiple operating points that result in deteriorated performance of the conventional SFA algorithm. To this end, Chapters 3 and 4 introduced improvements to the conventional probabilistic SFA method to address the aforementioned concerns. First, in Chapter 3 a framework for a robust PSFA method was proposed by considering a Gaussian scale mixture measurement noise structure. The noise modes were assumed to be switched according to a Bernoulli distribution as indicated by the introduction of an additional hidden variable. The param-

eter and online state estimations for this model were performed according to the iterative EM algorithm. Soft-sensor applications showed that the proposed method was able to extract meaningful features better when outliers are present in the training data compared to SFA and PSFA while performing similarly well as the existing Student-t robust PSFA. Last, in Chapter 4, to simultaneously resolve the issues of correlated outlier occurrence and multiple process operating points, the Bernoulli distribution was replaced with a Hidden Markov Model and a multi-model approach was adopted. The necessary changes to the parameter and state estimation process from Chapter 3 were discussed. Then the soft-sensor studies from Chapter 3 and the industrial model-plant-mismatch detection and diagnosis task from Chapter 2 were revisited with this model structure, and improved performance was observed.

5.2 Future Considerations

In order to further improve upon the proposed methods, there are several directions one could take. In this work all measurement outliers were modeled as coming from a Gaussian scaled distribution with two components. Adaptation for alternate Gaussian mixtures, such as location outliers or skewed distributions, could be considered for systems where the outliers are not well described by a scale mixture.

Further, another common issue that can be found in industrial data that was not addressed in this work is missing samples, or differing sampling rates across measured signals. In industrial data, sensor faults may not always result in a measurement outlier and instead may result in no value, or some placeholder value, being recorded to the data historian. Such values contain no information and should not be treated as outliers. Additionally, in many systems there are key variables that require manual or lengthy analysis to evaluate. Operators may wish to have an estimate of the values at the same high frequency that simpler variables are measured. The proposed PSFA methods could be extended to take these concerns relating to missing data into consideration.

Finally, in this work only MPC systems where the slowest features are assumed to be the most important were considered. While this is often the case for chemical systems, the assumption does not apply to all cases. For example, in the field of robotics the modelling may be more concerned with the fastest latent features and an alternate encoder-decoder structure that keeps these instead of the slowest features could be considered.

Thesis Related Publication List (as of August 2022)

This section lists the academic contributions made during the pursuit of this thesis.

Papers

1. To be submitted to journal as: Cameron Dyson, Jayaram Valluru, Biao Huang, and Graham Slot ”Robust PSFA using Switching Gaussian Scale Mixtures”

2. In preparation to submit as: Cameron Dyson, Jayaram Valluru, Biao Huang, and Graham Slot ”Hidden Markov Approach to Robust PSFA with Dynamic Switching”

Conferences

1. Cameron Dyson, Santhosh Kumar Varanasi, Graham Slot, Primo Majoko, and Biao Huang ”MPC Model-Plant-Mismatch Detection Through Slow Feature Analysis Preprocessing with Industrial Application” in *2022 7th International Symposium on Advanced Control of Industrial Processes (AdCONIP)*, Vancouver, British Columbia, Canada, August 2022.

Bibliography

[1] A. S. Badwe, R. S. Patwardhan, S. L. Shah, S. C. Patwardhan, and R. D. Gudi, "Quantifying the impact of model-plant mismatch on controller performance," *Journal of Process Control*, vol. 20, no. 4, pp. 408–425, 2010.

[2] F. Xu, B. Huang, and E. C. Tamayo, "Assessment of economic performance of model predictive control through variance/constraint tuning," *IFAC Proceedings Volumes*, vol. 39, no. 2, pp. 899–904, 2006.

[3] B. Huang, F. Xu, K. Lee, and Y. Shardt, *Lmipa toolbox*, Computer Process Control Group, 2008.

[4] L. Ljung, *System Identification: Theory for the User*. Prentice Hall PTR, 1999.

[5] V. Botelho, J. O. Trierweiler, M. Farenzena, and R. Duraiski, "Methodology for detecting model–plant mismatches affecting model predictive control performance," *Industrial & Engineering Chemistry Research*, vol. 54, no. 48, pp. 12 072–12 085, 2015.

[6] J. M. Simkoff, S. Wang, M. Baldea, L. H. Chiang, I. Castillo, R. Bindlish, and D. B. Stanley, "Plant–model mismatch estimation from closed-loop data for state-space model predictive control," *Industrial & Engineering Chemistry Research*, vol. 57, no. 10, pp. 3732–3741, 2018.

[7] S. Yerramilli and A. K. Tangirala, "Detection and diagnosis of model-plant mismatch in multivariable model-based control schemes," *Journal of Process Control*, vol. 66, pp. 84–97, 2018.

[8] X. Tian, G. Chen, and S. Chen, "A data-based approach for multivariate model predictive control performance monitoring," *Neurocomputing*, vol. 74, no. 4, pp. 588–597, 2011.

[9] L. Shang, Y. Wang, X. Deng, Y. Cao, P. Wang, and Y. Wang, "A model predictive control performance monitoring and grading strategy based on improved slow feature analysis," *IEEE Access*, vol. 7, pp. 50 897–50 911, 2019.

[10] F. Loquasto and D. E. Seborg, "Monitoring model predictive control systems using pattern classification and neural networks," *Industrial & engineering chemistry research*, vol. 42, no. 20, pp. 4689–4701, 2003.

[11] C. A. Harrison and S. J. Qin, "Discriminating between disturbance and process model mismatch in model predictive control," *Journal of Process Control*, vol. 19, no. 10, pp. 1610–1616, 2009.

[12] L. E. Olivier and I. K. Craig, "Model-plant mismatch detection and model update for a run-of-mine ore milling circuit under model predictive control," *Journal of Process Control*, vol. 23, no. 2, pp. 100–107, 2013.

[13] A. S. Badwe, R. D. Gudi, R. S. Patwardhan, S. L. Shah, and S. C. Patwardhan, "Detection of model-plant mismatch in mpc applications," *Journal of Process Control*, vol. 19, no. 8, pp. 1305–1313, 2009.

[14] H. Jiang, S. L. Shah, B. Huang, B. Wilson, R. Patwardhan, and F. Szeto, "Model analysis and performance analysis of two industrial mpcs," *Control engineering practice*, vol. 20, no. 3, pp. 219–235, 2012.

[15] S. Wang, J. M. Simkoff, M. Baldea, L. H. Chiang, I. Castillo, R. Bindlish, and D. B. Stanley, "Autocovariance-based plant-model mismatch estimation for linear model predictive control," *Systems & Control Letters*, vol. 104, pp. 5–14, 2017.

[16] S. Yerramilli and A. K. Tangirala, "Detection and diagnosis of model-plant mismatch in mimo systems using plant-model ratio," *IFAC-PapersOnLine*, vol. 49, no. 1, pp. 266–271, 2016.

[17] L. Li, J. Song, X. Zhang, J. Ye, and S. Yang, "Model deficiency diagnosis and improvement via model residual assessment in model predictive control," *Industrial & Engineering Chemistry Research*, vol. 56, no. 42, pp. 12 151–12 162, 2017.

[18] L. Li, L. Lu, Z. Huang, X. Chen, and S. Yang, "A model mismatch assessment method of mpc by decussation," *ISA Transactions*, vol. 106, pp. 51–60, 2020.

[19] Q. Lu, M. G. Forbes, P. D. Loewen, J. U. Backström, G. A. Dumont, and R. B. Gopaluni, "Support vector machine approach for model-plant mismatch detection," *Computers & Chemical Engineering*, vol. 133, p. 106 660, 2020.

[20] S. S. Wilson and C. L. Carnal, "System identification with disturbances," in *Proceedings of 26th Southeastern Symposium on System Theory*, IEEE, 1994, pp. 502–506.

[21] Y. Zhu, *Multivariable system identification for process control*. Elsevier, 2001.

[22] L. Wiskott and T. Sejnowski, "Slow feature analysis: Unsupervised learning of invariances.," *Neural Computation*, vol. 14, no. 4, pp. 715–770, 2002.

[23] F. Guo, C. Shang, B. Huang, K. Wang, F. Yang, and D. Huang, "Monitoring of operating point and process dynamics via probabilistic slow feature analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 151, pp. 115–125, 2016.

[24] C. Shang, F. Yang, B. Huang, and D. Huang, "Recursive slow feature analysis for adaptive monitoring of industrial processes," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 11, pp. 8895–8905, 2018.

[25] C. Shang, B. Huang, F. Yang, and D. Huang, "Slow feature analysis for monitoring and diagnosis of control performance," *Journal of Process Control*, vol. 39, pp. 21–34, 2016.

[26] P. Berkes and L. Wiskott, "Slow feature analysis yields a rich repertoire of complex cell properties," *Journal of vision*, vol. 5, no. 6, pp. 9–9, 2005.

[27] Z. Zhang and D. Tao, "Slow feature analysis for human action recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 3, pp. 436–450, 2012.

[28] R. Turner and M. Sahani, "A Maximum-Likelihood Interpretation for Slow Feature Analysis," *Neural Computation*, vol. 19, no. 4, pp. 1022–1038, Apr. 2007. eprint: https://direct.mit.edu/neco/article-pdf/19/4/1022/816872/neco.2007.19.4.1022.pdf.

[29] C. Shang, B. Huang, F. Yang, and D. Huang, "Probabilistic slow feature analysis-based representation learning from massive process data for soft sensor modeling," *AIChE Journal*, vol. 61, no. 12, pp. 4126–4139, 2015. eprint: https://aiche.onlinelibrary.wiley.com/doi/pdf/10.1002/aic.14937.

[30] L. Fan, H. Kodamana, and B. Huang, "Identification of robust probabilistic slow feature regression model for process data contaminated with outliers," *Chemometrics and Intelligent Laboratory Systems*, vol. 173, pp. 1–13, 2018.

[31] L. Fan, H. Kodamana, and B. Huang, "Semi-supervised dynamic latent variable modeling: I/o probabilistic slow feature analysis approach," *AIChE Journal*, vol. 65, no. 3, pp. 964–979, 2019.

[32] L. Fan, "Robust latent variable modeling using probabilistic slow feature analysis," 2020.

[33] K. Tsujimoto and T. Omori, "Switching probabilistic slow feature analysis for time series data," *International Journal of Machine Learning and Computing*, vol. 10, no. 6, 2020.

[34] J. Zhang, D. Zhou, M. Chen, and X. Hong, "Continual learning-based probabilistic slow feature analysis for multimode dynamic process monitoring," *arXiv preprint arXiv:2202.11295*, 2022.

[35] K. H. Lee, B. Huang, and E. C. Tamayo, "Sensitivity analysis for selective constraint and variability tuning in performance assessment of industrial mpc," *Control engineering practice*, vol. 16, no. 10, pp. 1195–1215, 2008.

[36] E. F. Camacho and C. B. Alba, *Model predictive control*. Springer science & business media, 2013.

[37] Q. Lu, M. G. Forbes, P. D. Loewen, J. U. Backström, G. A. Dumont, and R. B. Gopaluni, "Support vector machine approach for model-plant mismatch detection," *Computers & Chemical Engineering*, vol. 133, p. 106 660, 2020.

[38] C. Shang, F. Yang, X. Gao, X. Huang, J. A. Suykens, and D. Huang, "Concurrent monitoring of operating condition deviations and process dynamics anomalies with slow feature analysis," *AIChE Journal*, vol. 61, no. 11, pp. 3666–3682, 2015.

[39] X. Ma, Y. Si, Z. Yuan, Y. Qin, and Y. Wang, "Multistep dynamic slow feature analysis for industrial process monitoring," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 12, pp. 9535–9548, 2020.

[40] C. Shang, F. Yang, X. Gao, and D. Huang, "Extracting latent dynamics from process data for quality prediction and performance assessment via slow feature regression," in *2015 American Control Conference (ACC)*, 2015, pp. 912–917.

[41] R. Wood and M. Berry, "Terminal composition control of a binary distillation column," *Chemical Engineering Science*, vol. 28, no. 9, pp. 1707–1717, 1973.

[42] MathWorks, *Mpc designer app*, Natick, Massachusetts, United State, 2020.

[43] J. Nyberg, "Characterisation and control of the zinc roasting process," *University of Oulu*, 2005.

[44] Z. Feng, Y. Li, B. Sun, C. Yang, H. Zhu, and Z. Chen, "A trend-based event-triggering fuzzy controller for the stabilizing control of a large-scale zinc roaster," *Journal of Process Control*, vol. 97, pp. 59–71, 2021.

[45] P. Kadlec, B. Gabrys, and S. Strandt, "Data-driven soft sensors in the process industry," *Computers & chemical engineering*, vol. 33, no. 4, pp. 795–814, 2009.

[46] W. Shao, Z. Ge, Z. Song, and K. Wang, "Nonlinear industrial soft sensor development based on semi-supervised probabilistic mixture of extreme learning machines," *Control Engineering Practice*, vol. 91, p. 104 098, 2019.

[47] A. Khosbayar, J. Valluru, and B. Huang, "Multi-rate gaussian bayesian network soft sensor development with noisy input and missing data," *Journal of Process Control*, vol. 105, pp. 48–61, 2021.

[48] F. Guo, W. Bai, and B. Huang, "Output-relevant variational autoencoder for just-in-time soft sensor modeling with missing data," *Journal of Process Control*, vol. 92, pp. 90–97, 2020.

[49] R. Xie, N. M. Jan, K. Hao, L. Chen, and B. Huang, "Supervised variational autoencoders for soft sensor modeling with missing data," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 4, pp. 2820–2828, 2019.

[50] R. B. Gopaluni, A. Tulsyan, B. Chachuat, B. Huang, J. M. Lee, F. Amjad, S. K. Damarla, J. W. Kim, and N. P. Lawrence, "Modern machine learning tools for monitoring and control of industrial processes: A survey," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 218–229, 2020.

[51] B. Shen, L. Yao, and Z. Ge, "Nonlinear probabilistic latent variable regression models for soft sensor application: From shallow to deep structure," *Control Engineering Practice*, vol. 94, p. 104 198, 2020.

[52] Z. Ge, F. Gao, and Z. Song, "Mixture probabilistic pcr model for soft sensing of multimode processes," *Chemometrics and Intelligent Laboratory Systems*, vol. 105, no. 1, pp. 91–105, 2011.

[53] Z. Ge, B. Huang, and Z. Song, "Mixture semisupervised principal component regression model and soft sensor application," *AIChE Journal*, vol. 60, no. 2, pp. 533–545, 2014.

[54] I. T. Jolliffe, "Principal component analysis," *Technometrics*, vol. 45, no. 3, p. 276, 2003.

[55] P. Geladi and B. R. Kowalski, "Partial least-squares regression: A tutorial," *Analytica chimica acta*, vol. 185, pp. 1–17, 1986.

[56] B. Kowalski, R. Gerlach, and H. Wold, "Chemical systems under indirect observation," *Systems under indirect observation*, pp. 191–209, 1982.

[57] R. Rosipal and N. Krämer, "Overview and recent advances in partial least squares," in *International Statistical and Optimization Perspectives Workshop" Subspace, Latent Structure and Feature Selection"*, Springer, 2005, pp. 34–51.

[58] S. Wold, M. Sjöström, and L. Eriksson, "Pls-regression: A basic tool of chemometrics," *Chemometrics and intelligent laboratory systems*, vol. 58, no. 2, pp. 109–130, 2001.

[59] M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 61, no. 3, pp. 611–622, 1999. eprint: https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/1467-9868.00196.

[60] J. Zheng, Z. Song, and Z. Ge, "Probabilistic learning of partial least squares regression model: Theory and industrial applications," *Chemometrics and Intelligent Laboratory Systems*, vol. 158, pp. 80–90, 2016.

[61] W. Ku, R. H. Storer, and C. Georgakis, "Disturbance detection and isolation by dynamic principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 30, no. 1, pp. 179–196, 1995, InCINC '94 Selected papers from the First International Chemometrics Internet Conference.

[62] H. Kodamana, B. Huang, R. Ranjan, Y. Zhao, R. Tan, and N. Sammaknejad, "Approaches to robust process identification: A review and tutorial of probabilistic methods," *Journal of Process Control*, vol. 66, pp. 68–83, 2018.

[63] R. K. Pearson, "Outliers in process modeling and identification," *IEEE Transactions on control systems technology*, vol. 10, no. 1, pp. 55–63, 2002.

[64] H. Liu, S. Shah, and W. Jiang, "On-line outlier detection and data cleaning," *Computers & chemical engineering*, vol. 28, no. 9, pp. 1635–1647, 2004.

[65] D. B. Rubin, "Inference and missing data," *Biometrika*, vol. 63, no. 3, pp. 581–592, 1976.

[66] G. A. Gottwald, L. Mitchell, and S. Reich, "Controlling overestimation of error covariance in ensemble kalman filters with sparse observations: A variance-limiting kalman filter," *Monthly weather review*, vol. 139, no. 8, pp. 2650–2667, 2011.

[67] S. Kotz, T. Kozubowski, and K. Podgórski, *The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance*, 183. Springer Science & Business Media, 2001.

[68] C. Zhao and J. Yang, "A robust skewed boxplot for detecting outliers in rainfall observations in real-time flood forecasting," *Advances in Meteorology*, vol. 2019, 2019.

[69] A. Sadeghian, N. M. Jan, O. Wu, and B. Huang, "Robust probabilistic principal component regression with switching mixture gaussian noise for soft sensing," *Chemometrics and Intelligent Laboratory Systems*, p. 104 491, 2022.

[70] S. Khatibisepehr and B. Huang, "A bayesian approach to robust process identification with arx models," *AIChE Journal*, vol. 59, no. 3, pp. 845–859, 2013.

[71] G. J. McLachlan and D. Peel, *Finite mixture models*. John Wiley & Sons, 2004.

[72] A. Sadeghian and B. Huang, "Robust probabilistic principal component analysis for process modeling subject to scaled mixture gaussian noise," *Computers & Chemical Engineering*, vol. 90, pp. 62–78, 2016.

[73] A. Sadeghian, O. Wu, and B. Huang, "Robust probabilistic principal component analysis based process modeling: Dealing with simultaneous contamination of both input and output data," *Journal of Process Control*, vol. 67, pp. 94–111, 2018.

[74] A. Memarian, S. K. Varanasi, and B. Huang, "Mixture robust semi-supervised probabilistic principal component regression with missing input data," *Chemometrics and Intelligent Laboratory Systems*, vol. 214, p. 104 315, 2021.

[75] S. Khatibisepehr and B. Huang, "A bayesian approach to robust process identification with arx models," *AIChE Journal*, vol. 59, no. 3, pp. 845–859, 2013. eprint: https://aiche.onlinelibrary.wiley.com/doi/pdf/10.1002/aic.13887.

[76] G. J. McLachlan and T. Krishnan, *The EM algorithm and extensions*. John Wiley & Sons, 2007, vol. 382.

[77] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*, 4. Springer, 2006, vol. 4.

[78] R. E. Kalman, "A new approach to linear filtering and prediction problems," 1960.

[79] J. Downs and E. Vogel, "A plant-wide industrial process control problem," *Computers & Chemical Engineering*, vol. 17, no. 3, pp. 245–255, 1993, Industrial challenge problems in process control.

[80] R. Braatz, *Tennessee eastman problem simulation data*, http://web.mit.edu/braatzgroup/links.html, 2002.

[81] P. Smyth, "Hidden markov models for fault detection in dynamic systems," *Pattern recognition*, vol. 27, no. 1, pp. 149–164, 1994.

[82] F. Koushanfar and M. Potkonjak, "Markov chain-based models for missing and faulty data in mica2 sensor motes," in *SENSORS, 2005 IEEE*, IEEE, 2005, 4–pp.

[83] N. Sammaknejad, B. Huang, W. Xiong, A. Fatehi, F. Xu, and A. Espejo, "Operating condition diagnosis based on hmm with adaptive transition probabilities in presence of missing observations," *AIChE Journal*, vol. 61, no. 2, pp. 477–493, 2015.

[84] N. Sammaknejad, B. Huang, and Y. Lu, "Robust diagnosis of operating mode based on time-varying hidden markov models," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 2, pp. 1142–1152, 2015.

[85] R. Sharifi and R. Langari, "Nonlinear sensor fault diagnosis using mixture of probabilistic pca models," *Mechanical Systems and Signal Processing*, vol. 85, pp. 638–650, 2017.

[86] S. Sedghi, A. Sadeghian, and B. Huang, "Mixture semisupervised probabilistic principal component regression model with missing inputs," *Computers & Chemical Engineering*, vol. 103, pp. 176–187, 2017.

[87] J. Zhu, Z. Ge, and Z. Song, "Robust modeling of mixture probabilistic principal component analysis and process monitoring application," *AIChE journal*, vol. 60, no. 6, pp. 2143–2157, 2014.

[88] J. Wang, W. Shao, and Z. Song, "Semi-supervised variational bayesian student'st mixture regression and robust inferential sensor application," *Control Engineering Practice*, vol. 92, p. 104 155, 2019.

[89] W. Song, W. Yao, and Y. Xing, "Robust mixture regression model fitting by laplace distribution," *Computational Statistics & Data Analysis*, vol. 71, pp. 128–137, 2014.

[90] P. Zhu, X. Yang, and H. Zhang, "Mixture robust l1 probabilistic principal component regression and soft sensor application," *The Canadian Journal of Chemical Engineering*, vol. 98, no. 8, pp. 1741–1756, 2020.

[91] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.

[92] X. Jin, S. Wang, B. Huang, and F. Forbes, "Multiple model based lpv soft sensor development with irregular/missing process output measurement," *Control Engineering Practice*, vol. 20, no. 2, pp. 165–172, 2012.

[93] Y. Lu and B. Huang, "Robust multiple-model lpv approach to nonlinear process identification using mixture t distributions," *Journal of Process Control*, vol. 24, no. 9, pp. 1472–1488, 2014.

[94] Y. Zhao, B. Huang, H. Su, and J. Chu, "Prediction error method for identification of lpv models," *Journal of process control*, vol. 22, no. 1, pp. 180–193, 2012.

[95] Z. Li, Y. He, F. Chu, J. Han, and W. Hao, "Fault recognition method for speed-up and speed-down process of rotating machinery based on independent component analysis and factorial hidden markov model," *Journal of Sound and Vibration*, vol. 291, no. 1-2, pp. 60–71, 2006.

[96] W. K. Ching, E. S. Fung, and M. K. Ng, "Higher-order hidden markov models with applications to dna sequences," in *International Conference on Intelligent Data Engineering and Automated Learning*, Springer, 2003, pp. 535–539.

[97] I. Stanculescu, C. K. Williams, and Y. Freer, "Autoregressive hidden markov models for the early detection of neonatal sepsis," *IEEE journal of biomedical and health informatics*, vol. 18, no. 5, pp. 1560–1570, 2013.