Accepted Manuscript

Data-driven Soft Sensors in the Process Industry

Petr Kadlec [a], Bogdan Gabrys [a], Sibylle Strandt [b]

[a]*Smart Technology Research Centre, Computational Intelligence Research Group, Bournemouth University, Poole BH12 5BB, United Kingdom*

[b]*Evonik Degussa AG, 45128 Essen, Germany*

Abstract

In the last two decades Soft Sensors established themselves as a valuable alternative to the traditional means for the acquisition of critical process variables, process monitoring and other tasks which are related to process control. This paper discusses characteristics of the process industry data which are critical for the development of data-driven Soft Sensors. These characteristics are common to a large number of process industry fields, like the chemical industry, bioprocess industry, steel industry, etc. The focus of this work is put on the data-driven Soft Sensors because of their growing popularity, already demonstrated usefulness and huge, though yet not completely realised, potential. A comprehensive selection of case studies covering the three most important Soft Sensor application fields, a general introduction to the most popular Soft Sensor modelling techniques as well as a discussion of some open issues in the Soft Sensor development and maintenance and their possible solutions are the main contributions of this work.

Key words: Soft Sensors; Process industry; Data-driven models; PCA; ANN;

1 Introduction

Industrial processing plants are usually heavily instrumented with a large number of sensors. The primary purpose of the sensors is to deliver data for process monitoring and control. But approximately two decades ago researchers started to make use of the large amounts of data being measured and stored in the process industry by building predictive models based on this data. In the context of process industry, these predictive models are called *Soft Sensors*. This term is a combination of the words "software", because the models are usually computer programs, and "sensors", because the models are delivering similar information as their hardware counterparts. Other common terms for predictive sensors in the process industry are *inferential sensors* (see e.g. Jordaan et al., 2004; Qin et al., 1997), *virtual online analyser* as they are called in the Six-Sigma context (Han and Lee, 2002) and *observer-based* sensors (Goodwin, 2000).

At a very general level one can distinguish two different classes of Soft Sensors, namely model-driven and data-driven. The model-driven family of Soft Sensors is most commonly based on First Principle Models (FPM) but model-driven Soft Sensors based on extended Kalman Filter (Welch and Bishop, 2001) or adaptive observer (Bastin and Dochain, 1990) have also been published (e.g. Chruy, 1997; Jos de Assis and Maciel Filho, 2000). First Principle Models describe the physical and chemical background of the process. These models are developed primarily for the planning and design of the processing plants, and therefore usually focus on the description of the ideal steady-states of the processes which is only one of their drawbacks which makes it difficult to base Soft Sensors on them. As a solution the data-driven Soft Sensors gained increasing popularity in the process industry.

2

Because data-driven models are based on the data measured within the processing plants, and thus describe the real process conditions, they are, compared to the model-driven Soft Sensors, more reality related and describe the true conditions of the process in a better way. Nevertheless there is a lot of different issues which have to be dealt with while developing data-driven Soft Sensors. These issues will be discussed later on in this paper. The most popular modelling techniques applied to data-driven Soft Sensors are the Principle Component Analysis (Jolliffe, 2002) in a combination with a regression model, Partial Least Squares (Wold et al., 2001), Artificial Neural Networks (Bishop, 1995; Principe et al., 2000; Hastie et al., 2001), Neuro-Fuzzy Systems (Jang et al., 1997; Lin and Lee, 1996) and Support Vector Machines (Vapnik, 1998).

The range of tasks fulfilled by Soft Sensors is broad. The original and still most dominant application area of Soft Sensors is the prediction of process variables which can be determined either at low sampling rates or through off-line analysis only. Because these variables are often related to the process output quality, they are very important for the process control and management. For these reasons it is of great interest to deliver additional information about these variables at higher sampling rate and/or at lower financial burden, which is exactly the role of the Soft Sensors. The modelling methods applied to this kind of applications are either statistical or soft computing supervised learning approaches. This Soft Sensor application field is further on referred to as *on-line prediction*. Other important application fields of Soft Sensors are those of *process monitoring and process fault detection*. These tasks refer to detection of the state of the process and in the case of a deviation from the normal conditions to identification of the deviation source. Traditionally, the process state is monitored by process operators in the control rooms of the

processing plants. The observation and interpretation of the process state is often based on univariate statistics and it is up to the experience of the process operator to put the particular variables into relations and to make decisions about the process state. The role of process monitoring Soft Sensors is, based on the historical data, to build multivariate features which are relevant for the description of the process state. By presenting the predicted process state or the multivariate features the Soft Sensor can support the process operators and allow them to make faster, better and more objective decisions. Process monitoring Soft Sensors are usually based on the Principle Component Analysis and Self Organizing Maps (Kohonen, 1997). It was already mentioned that processing plants embody large number of various sensors, therefore there is a certain probability that a sensor can occasionally fail. Detection of this failure is the next application area of Soft Sensors. In more general terms this application field can be described as *sensor fault detection and reconstruction.* Once a faulty sensor is detected and identified, it can be either reconstructed or the hardware sensor can be replaced by another Soft Sensor, which is trained to act as a *back-up Soft Sensor* of the hardware measuring device. If the back-up sensor proves to be an adequate replacement of the physical sensor, this idea can be driven even further and the Soft Sensor can replace the measuring device also in normal working conditions. The software tool can be easier maintained and is not subject to mechanical failures and therefore such a substitution can provide a financial advantage for the process owner.

Despite all the previously listed Soft Sensor application fields and the high number of publications dealing with Soft Sensor applications, there are still some unaddressed issues of the Soft Sensor development and maintenance. A lot of the origins of these issues are in the process data which is used for the Soft Sensor building.

4

Common effects present in the data are measurement noise, missing values, data outliers, co-linear features and varying sampling rates. To solve these problems, there is typically a large amount of manual work needed. Another problem is that the processing plants are rather dynamic environments. Often they *develop* gradually during the operation time but there can be also sudden abrupt changes of the process, for example, if the quality of the process input changes. It is very difficult for the Soft Sensors to react to these changes which usually results in prediction accuracy deterioration. At present time, these issues are solved in a rather ad-hoc manner, which leads to unnecessary high costs of the Soft Sensor development and maintenance. Further on in this work, all the aspects, which have been briefly outlined in this section, are going to be reviewed in a more comprehensive way. The rest of the paper is organized as follows. Section 2 gives an overview of different process types and deals with their aspects from the Soft Sensor modelling point of view. Section 3 focuses on data-driven Soft Sensors, namely on their development methodology, on the methods which are commonly applied to soft sensing and on open issues of the Soft Sensor modelling. A review of publications dealing with Soft Sensor application to diverse processes is also given in Section 3. Section 4 provides a brief description of most popular data-driven pre-processing and modelling techniques to soft sensing. Section 5 contains a discussion of the most important open issues of Soft Sensor development and maintenance as well as an outline of the future research directions in the Soft Sensors field. Finally, the work is concluded in Section 6.

2 Industrial Processes

This section deals with the process industry environment. First, the two different types of industrial processes and their distinguishing characteristics are discussed in Section 2.1. This is followed by a detailed discussions of the data produced in the process industry in Section 2.2.

2.1 Industrial process types

2.1.1 Continuous Processes

Continuous processing plants are, as their name suggests, running in uninterrupted, continuous way. After the start-up phase of the plant they are operated in more or less constant and hopefully optimal state. As the process should stay most of the time in this optimal state, the Soft Sensors applied to continuous processes focus usually on the description of this steady-state and are not able to deal with any transient states like the process start-up and shut-down phase. Nonetheless, even the steady state is progressively changing with time, which has a negative effect on the prediction quality of the Soft Sensor. The most common causes of the process operating point changes are the changes of the process product demand, the change of the catalyst activity, clogging of heat exchangers, etc.

As continuous processes generate the majority of revenue for most of the most process industry companies this review is biased towards this type of processes. Majority of the application examples listed in Section 3.3 deal with continuous processes and therefore Section 4 presents the most popular techniques for Soft Sensor development mainly from the continuous process modelling point of view.

Batch, semi-batch or discontinuous processes (further on referred to as *batch* processes only) are processes with a definite duration. Very often these processes are started *on demand* for the production of required product amount. Many processes in food and biochemistry industry, like fermentation processes, are of this type. Another field where batch processes are very common is the speciality chemistry. Here, the special chemicals have to be produced infrequently and often in very small amounts and thus it would not be economical to run the plants in continuous mode.

Bonne and Jorgensen (2004) commented that: *Batch processes are experiencing a renaissance as products-on-demand and first-to-market strategies impel the need for flexible and specialised production methods.* This statement is clearly demonstrating the increased demand for modelling tools based on batch process data.

In terms of data-driven modelling there is a difference between continuous and batch processes. Batch process modelling has to deal with an additional discrete dimension of the data, namely the batch-to-batch variation (Nomikos and MacGregor, 1995b). While modelling these processes, one has to take into account the finite and varying duration of the processes, the time variance of the particular batches described by the batch trajectory, the often high batch-to-batch variance and the starting conditions of the batches (Champagne et al., 2002). The techniques applied for modelling and monitoring of batch processes are most commonly multivariate statistical techniques. In the case of batch process monitoring, the most common applied method is the principle component analysis. There are several batch processes monitoring applications reviewed in Section 3.3.2 and a discussion of batch process modelling tools is given in Section 4.6.

7

2.2 Characteristics of process industry data

This section presents the most critical characteristics of the process industry data as they are identified from the Soft Sensor development and maintenance point of view. Another general view on process data was published in Pearson (2001), where the focus is put on the discussion of the process data distribution and the information which can be extracted from it.

2.2.1 Missing values

Missing data are single samples or consequent sets of samples, where one or more variables (i.e. measurements) have a value which does not reflect the real state of the physical measured quantity. The affected variables usually have values like $\pm\infty$, 0 or any other constant value.

Missing values in the context of process industry have various causes. The most common causes are the failure of a hardware sensor, its maintenance or removal. As it was already mentioned, processing plants are heavily instrumented for the purpose of control of the processes, therefore also the recorded process data consists of large number of diverse variables. In such a scenario, there is a certain probability that some of the sensors will occasionally fail. One should keep in mind that some of the sensor types are mechanical devices (e.g. flow rate sensors) and thus suffer from abrasion effects. Another possible causes of missing data are related to the transmission of the data between the sensors and the database, errors in the database, problems in accessing the database, etc.

Since most of the techniques applied to data-driven soft sensing cannot deal with

8

missing data, a strategy for their replacement have to be usually implemented. There are different strategies to replace missing values. An approach, which is very primitive and not recommended but still commonly applied in practical scenarios, is to replace the missing values with the mean values of the affected variable. Another non-optimal approach is to skip the data samples consisting of variable or variables with the missing values, i.e. *case deletion* (Scheffer, 2002). More efficient approach to missing values handling takes into account the multivariate statistics of the data and thus makes the reconstruction of the missing values dependent on the other available variables of the affected samples (see e.g. Walczak and Massart (2001b) for maximum-likelihood multivariate approach to missing values replacement). This kind of approaches are related to "sensor fault detection and reconstruction" (for some practical algorithms see Section 3.3.3). From another point of view, one can distinguish two different approaches for dealing with missing values Scheffer (2002). These are: (i) single imputation where the missing values are replaced in a single step (using e.g. mean/median values); and (ii) multiple imputation which are iterative techniques where several imputation steps are performed.

A study dealing with missing data was presented in Schafer and Graham (2002). In this study the authors also propose two general approaches to handle missing data based on maximum-likelihood and Bayesian multiple imputation.

In Chen and Chen (2000) an algorithm based on iteratively reweighed least squares is applied to deal with missing and noisy data. This algorithm is limited to the estimation of dynamic linear system parameters only. The authors show, that the algorithm can deal with situation where the probability of missing data is less then 50% provided that a high number of samples is available.

Walczak and Massart (2001a) and Walczak and Massart (2001b) is a two-part publication dealing with multiple imputation techniques for missing values handling. The first part (Walczak and Massart, 2001a) focuses on the influence of the missing data handling techniques on methods typically applied in chemometrics, i.e. PCR/PLS, etc. Whereas the second part (Walczak and Massart, 2001b) proposes a maximum-likelihood based algorithm for dealing with missing data.

An altrnative approach to dealing with missing data in a probabilistic framework was published in Gabrys (2002). This work particularly focuses on missing data treatment in the context of decision making and diagnostic analysis.

2.2.2   Data outliers

Outliers are sensor values which deviate from the typical or sometimes also meaningful, ranges of the measured values. One can distinguish between two types of outliers, namely *obvious outliers* and *non-obvious outliers* (Qin, 1997). Obvious outliers are those values which violate the physical or technological limitations. For example the absolute pressure may not reach negative values or flow sensor may not deliver values which exceed the technological limitations of the sensor. To be able to detect this type of outliers efficiently the system has to be provided with the limiting values in the form of a-priori information. In contrast to this, non-obvious outliers are even harder to identify because they do not violate any limitations but still lay out of the typical ranges and do not reflect the correct variable states.

Outlier detection as part of the data pre-processing remains very critical for the Soft Sensor development because not detected outliers have negative effect on the performance of the Soft Sensor models. For example, the influence of a single outlier

can be critical for the PCA (Walczak and Massart, 1995; Stanimirova et al., 2007; Serneels and Verdonck, 2008). Another problem of outlier detection is that even when applying automatic outlier handling pre-processing steps, usually the results have to be validated manually by the model developer. The goal of the manual inspection is to detect any possible outlier *maskings* (i.e. false negative detections-not detected outliers) and outlier *swamping* (i.e. false positive detections - correct values labelled as outliers).

Typical approaches to outlier detection are based on the statistics of the historical data. The most simple approach is the 3σ outlier detection algorithm (e.g. Lin et al., 2007; Pearson, 2002), which is based on univariate observations of the variable distributions. This method labels all data samples out of the range $\mu(x) \pm 3\sigma(x)$, where $\mu(x)$ is the mean value and $\sigma(x)$ the standard deviation of the variable x, as outliers. More robust version of this approach is the Hampel identifier (Davies and Gather, 1993) which is in contrast to the 3σ method uses more outlier resistant median and median absolute deviation from median (MAD) values (Pearson, 2001, 2002) to calculate the limits.

In Pearson (2001), the author discusses the outliers problem. He focuses on the influence of outliers on the identification of linear and non-linear models. For the handled models the Hampel identifier, which is based on a robust estimation of the variables' statistics, is found to be an effective approach for dealing with outliers. In Menold et al. (1999) a moving window filter is combined with the Hampel identifier to obtain an outlier detection and removal system. In contrast to the univariate approaches the multivariate methods use combinations of more features to detect the outliers. An example from this group based on the PCA is the Jolliffe parameter (Jolliffe, 2002; Warne et al., 2004a). Gonzalez et al. (2003) is using a two-stage

11

outlier detection approach. The first stage is the application of the PCA, after this the T^2 measure can be used to detect outlier candidates which are located outside of the 99% confidance ellipse. These candidates are then further analysed in the second step, where Scheffé's test (Gomez et al., 1996) is applied to these points.

Another, rather general review of the outlier detection problem and several outlier detection algorithms is presented in Hodge and Austin (2004).

2.2.3   Drifting data

There are two types of drifting data and dependent on the cause of the drifts one can distinguish between process and sensor drifts. The causes of the process drift are the changes of the process or of some external process conditions. The processing plants consist of a large number of mechanical elements which undergo steady abrasion during the operation of the plant. This may have an effect on the process itself, e.g. the flow between two parts of the process can decrease due to the abrasion of mechanical pumps. Another cause of the drifting data can also be external influences like changing environmental conditions (e.g. weather influence), the purity of the input materials, catalyst deactivation, etc. These factors have not only an influence on the data but affect the process state as well. Therefore the drifts should be recognised, reported and appropriate actions have to be taken to remove their cause. This is different in the case of sensor drifts which are caused by changes in the measuring devices and not by the process itself. The critical point is that this type of drifts, while still observed in the measured data, does not reflect any changes in the process. Therefore in the case of sensor drifts, the action to be taken should be the re-calibration of the measurement devices or the adaptation of the Soft Sensor without performing any corrective actions to the process.

12

In terms of the effects on the process data, one can observe changes in the means and variances of the single variables as well as changes of the correlation structure of the data Li et al. (2000).

Distinguishing between the two discussed different drift causes is challenging and once again a lot of expert knowledge is needed in order to take appropriate action. Another challenging aspect of dealing with drifting data is the fact that the changes may progress very slowly and may influence each other, and thus have non-linear form, which makes them difficult to detect and compensate.

The most common approach to deal with dynamics in the data is to apply the *moving window* techniques. In this case the model is updated on periodical basis using only a defined number of the most recent samples. Some examples of the application of this technique in the context of Soft Sensor modelling are: Wang et al. (2005); Zhao and Chai (2004); Qin (1998); Dayal and MacGregor (1997). Further approaches for Soft sensor adaptation are discussed in Section 3.2.5.

The problems with drifting data are not unique to the process industry data and they can be found in several other fields dealing with changing environments. In the machine learning terminology these problems are summarised under the term *concept drift*. For detailed treatment and some solutions see Widmer and Kubat (1996); Gama et al. (2004).

2.2.4   Data co-linearity

Another challenging issue for soft sensing, apart from those stated above, is related to the structure of the data. Typically, the data measured in the process industry are strongly co-linear. This results from the partial redundancy in the sensor

arrangement, e.g. two neighbouring temperature sensors will deliver strongly correlated measurements. At this place it should be recalled that the primary purpose of the data collected within the processing plants is for the process control. For this purpose it is necessary to have detailed information about all process components which results in a large number of measurements. Such environments are often called *data rich but information poor* (Dong and McAvoy, 1996) but for soft sensing the requirements are different, in this case only informative variables are required. Anything else is unnecessarily increasing the model complexity, which has often negative effect on the model training and performance.

There are two ways to deal with the co-linearity problem. One way is by transforming the input variables into a new reduced space with less co-linearity as it is done in the case of the PCA (Jolliffe, 2002) and PLS (Wold et al., 2001; Abdi, 2003). These two approaches are the most popular ones to deal with data co-linearity in the process industry. Examples of applications where PCA is used are: Lin et al. (2007); Amazouz and Pantea (2006); Wang and Cui (2005); Zhao and Chai (2004) and for the PLS Marjanovic et al. (2006); Zhang and Lennox (2004); Zamprogna et al. (2004a). Another way to handle co-linearity is to select a subset of the input variables which is less co-linear. These approaches are summarised under the umbrella of *variable (or feature) selection* methods in the computational learning research. A general review of these methods is presented in Guyon and Elisseeff (2003). Some feature selection methods in the context of soft sensing are also discussed in Warne et al. (2004a). Among the discussed approaches in their work are the correlation- and partial correlation-based feature selection as well as Mallows' C_p statistics.

2.2.5 Sampling rates and measurement delays

Various sensors usually work at different sampling rates and thus one has to take care to synchronize them. The synchronization of the data is usually handled by the Process Information Management System (PIMS) which records new data samples only if one of the observed variables changes more than a pre-defined threshold value. The definition of such threshold is another critical point, which influences the quality of the historical data. This is because too low values would cause the recording of unnecessarily large number of samples, whereas too high threshold can lead to missing of important process changes. Soft sensing is often applied in multi-rate systems with several operating sampling rates. Such a scenario occurs in a system where some of the variables, usually critical for the process control, are evaluated in laboratories at much lower sampling rate than the rest of the automatically measured data. This fact causes problems for the modelling and control of the processes. A summary of the last fifty years of multi-rate research is provided in Ding and Chen (2005).

Additional issue of the process data are the process related delays of the measurements. The materials in the processes have usually a given run-time through the process (e.g. the dwell period within a reactor or distillation column) and thus it is not reasonable to relate two different measurements taken at the same time at different locations within the process. Instead of this, the delays of the particular measurements should be compensated by synchronizing the variables. In order to perform the synchronisation there is an extensive knowledge about the process required.

In the case of batch processes a particular problem is that the different runs of batch

processes can have different run times. To be able to apply data-driven methods to batch process historical data the data must have the same length (i.e. the same number of samples) and thus also require synchronisation. Detailed discussion of various synchronisation approaches is given in Section 4.6.

3 Soft Sensors in the process industry

This section deals with Soft Sensors in a detailed way. After distinguishing two types of them in Section 3.1 a discussion of a state-of-the-art Soft Sensor development methodology is given in Section 3.2. Section 3.3 provides a comprehensive overview of published Soft Sensor application case studies.

3.1 Model-driven and data-driven Soft Sensors

At a very general level one can distinguish two types of Soft Sensors, namely *Model-Driven* and *Data-Driven* Soft Sensors. Model-driven models are also called white-box models because they have full phenomenological knowledge about the process background. In contrast to this purely, data-driven models are called black-box techniques because the model itself has no knowledge about the process and is based on empirical observations of the process. In between the two extremes there are many combinations of these two major types of models possible. A typical example of such a combination is a model-driven Soft Sensor making use of data-driven method for the modelling of fractions which can not be modelled easily in terms of phenomenological models. These models are sometimes called *hybrid* models but in order to avoid any confusion with the hybrid combinations of two or

16

more computational learning methods (e.g. neuro-fuzzy systems) we refer to them as grey-box models in the remaining of this paper.

Model-driven Models (MDM), or more specifically First Principles Models (FPM), are primarily developed for the purpose of planning and development of the process plants. These models are based on equations describing the chemical and physical principles underlying the process. A typical example is using mass-preservation principles, exothermal equation, energy balances, reaction kinetics in the form of reaction rate equations for this purpose. The drawback of this type of models is that their development requires a lot of process expert knowledge. This knowledge is not always available. For example, for biochemical process there is often not enough phenomenological knowledge for accurate description of the processes at hand. Another problem is that the models often describe a simplified theoretical background of the process rather than the real-life conditions of the process which is influenced by many factors out of the scope of the MDM. Additionally, the model-driven models usually focus on the description of the optimal steady-state of the process and are thus not suitable for the description of any transient states. Nonetheless, model-driven Soft Sensors are popular as a support for inferential control. Examples of inferential control applications of first principle Soft Sensor are (De Wolf et al., 1996) and (Doyle, 1998) where the first example is based on Kalman filter and the latter one on non-linear observer method. Another example of model-driven Soft Sensor is (Prasad et al., 2002) where a multi-rate Kalman filter is applied to the control of a polymerisation process.

The focus of this review and of soft sensing in general is therefore put on the Data-Driven Models (DDM) which have emerged as very attractive modelling approaches enhancing the toolbox of diagnostic, prognostic and decision support methods avail-

17

able for plant operators and embedded in automated control systems. These models are based on the real-life measurements which are recorded, stored and provided as historical data by the Process Information Management Systems (PIMS). The models themselves are empirical predictive methods like Principle Component Regression (PCR), Multi-layer Perceptron (MLP), etc.

3.2 Soft Sensor development methodology

This section describes the typical steps and issues of the common practice of Soft Sensor development. The presented procedure is rather general and can thus be applied for both continuous and batch processes as well as to any of the application areas discussed in Section 3.3. An overview of the methodology is presented in Figure 1.
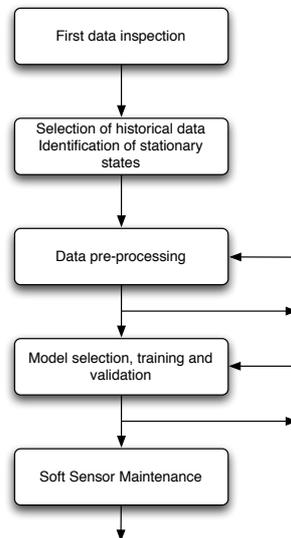
Fig. 1. Methodology for Soft Sensor development

3.2.1 First data inspection

During this initial step, the first inspection of the data is performed. The aim of this step is to gain an overview of the data structure and identify any obvious problems which may be handled at this initial stage (e.g. locked variables having constant value, etc.). The next aim of this stage is to assess the requirements for the model complexity. An experienced Soft Sensor developer can, already at this stage, make a reasonable decision whether, in the case of an on-line prediction Soft Sensor, to use a simple regression model, a rather more complex and powerful PCA regression model or a non-linear neural network to build the Soft Sensor. In some cases, the model family decision at this stage may not be correct, therefore the models and their performance should be always evaluated and compared to alternative models at the later development stages.

A particular attention is paid to the assessment of the target variable. It has to be checked, if there is enough variation in the output variable and if this can be modelled at all.

3.2.2 Selection of historical data and identification of stationary states

Here, data to be used for the training and evaluation of the model are selected. Next, the stationary parts of the data have to be identified and selected. In vast majority of the cases further modelling will only deal with the stationary states of the process. The identification of the stationary process states is usually performed by manual annotation of the data.

In Jiang et al. (2003) the steady state detection of continuous processes is discussed and a wavelet transform based approach is applied to perform this task.

In the case of batch processes there are usually no steady states and thus the model developer focuses on the selection of representative batch runs rather than on the identification of steady states.

3.2.3 Data pre-processing

The aim of this step is to transform the data in such a way, that it can be more effectively processed by the actual model. An example of a typical pre-processing step is the normalisation of the data to the zero-mean and unit variance (as it is required by the PCA). In the case of the data which are produced in the process industry there are several pre-processing steps necessary which is indicated by the loop around the "Data pre-processing" box in Figure 1. The usual steps are the handling of missing data, outliers detection and replacement, selection of relevant variables (i.e. feature selection), handling of drifting data and detection of delays between the particular variables. A lot of the listed steps are at the moment handled manually or need at least a supervised inspection of the results. The data pre-processing is usually done in an iterative way, e.g. after the standardisation and missing values treatment which are usually performed only once, an outlier removal and feature selection are repeatedly applied until the model developer considers the data as being ready to be used for the training and evaluation of the actual model. Due to the characteristics of the data discussed in Section 2.2 the importance of the pre-processing is critical. At the moment, the pre-processing of the data is the step which requires a large amount of manual work and expert knowledge about the underlying process.

3.2.4 Model selection, training and validation

This phase is critical for the final Soft Sensor. As the model is the engine of the Soft Sensor, selection of the optimal type is crucial for the Soft Sensors performance. So far, there is no unified theoretical approach for this task and thus the model type and its parameters are often selected in an ad-hoc manner for each Soft Sensor. Model selection is also often subject to developer's past experience and personal preference which can be of disadvantage for the final Soft Sensor. This can be observed in the domain of published Soft Sensor applications where many of the authors strongly focus on one model type (e.g. PLS) which is in their field of expertise.

Nevertheless, despite the lack of a common theoretically superior approach to model selection there are few techniques which can be adopted to this task. A possible approach is to start with a simple model type or structure (e.g. linear regression model) and gradually increase model complexity as long as significant improvement in the model's performance can be observed (using e.g. the Student's t-test (Gosset, 1908)). While performing this task it is important to asses the performance of the model on independent data (Weiss and Kulikowski, 1991; Hastie et al., 2001). The same approach can also be applied to the parameters selection of the pre-processing methods like for instance variable selection.

Additionally, for some industrial processes it can be difficult to obtain sufficient amount of historical data for the model development. In such cases it is of advantage to resort to statistical error-estimation techniques like *K-fold cross-validation* (Kohavi, 1995). This method makes an optimal uses of the available data by partitioning it in such a way that all of the samples are used for the model performance validation. Another alternative in these circumstances is to apply statistical re-

sampling methods like for example *bagging* (Breiman, 1996) and *boosting* (Freund and Schapire, 1997). In the case of the first method, a set of training data sets is generated by randomly drawing samples (with replacement) from the available data and training one model for each of the random sets. The final model is obtained by averaging over the particular models' predictions. In contrast to this, in the case of boosting, the probability of each sample to be drawn is not random but related to the prediction error of the model given the data sample. Additionally in case of boosting, the weights of the contributions of the particular models are calculated based on the models performance on a validation data set.

The generalisation performance of the developed Soft Sensor can be also increased by applying ensemble methods. Comprehensive reviews of ensemble building techniques were published in Kuncheva (2004); Valentini and Masulli (2002). Ensemble building have been proved theoretically Wolpert (1992); Krogh and Vedelsby (1995); Kittler et al. (1998); Freund and Schapire (1997) and practically Opitz and Maclin (1999); Bauer and Kohavi (1999); Ruta and Gabrys (2000); Gabrys and Ruta (2006) to improve the model's prediction performance. The underlying idea is to train a set of base models and to make a combination of their responses in order to obtain the final prediction. Different strategies for building of the combinations were discussed in Gabrys (2004). The idea of ensemble methods was brought further in Ruta and Gabrys (2005) where approach to the selection of single predictors, ensembles and multi-level structres were studied.

After finding the optimal model structure and training the model, the trained Soft Sensor has to be evaluated on independent data once again (Weiss and Kulikowski, 1991). There are several tools for the evaluation of the model performance. In the case of numerical performance evaluation the most popular is the Mean Squared

Error (MSE), which measures the average square distance between the predicted and the correct value. Another way of performance judgement is using visual representation of the predictions. In these, the *four-plot analysis* is a useful tool since it provides useful information about the relation between the predictions and the correct values together with the analysis of the prediction residuals (Fortuna, 2007). A disadvantage of the visual methods is that they require an assistance of the model developer and the final decision if the model performs adequately, is up to the subjective judgement of the model developer.

A more detailed discussion on model selection and validation is provided in Fortuna (2007) where apart from the discussion of several techniques for model selection and validation the authors of the book stress the necessity for the application of process knowledge during the Soft Sensor development phase.

3.2.5 Soft Sensor maintenance

After developing and deploying the Soft Sensor, it has to be maintained and tuned on a regular basis. The maintenance is necessary due to the drifts and other changes of the data (see Section 2.2.3) which cause the performance of the Soft Sensor to deteriorate and have to be compensated for by adapting or re-developing the model.

Currently most of the Soft Sensors do not provide any automated mechanisms for their maintenance. This fact together with the previously discussed evidence of changing data results in the requirement for manual quality control and maintenance of the Soft Sensors which is a significant cost factor for the application of Soft Sensors. Even worse, there is often no objective measure for assesing the Soft Sensor quality level and the judgement if a model works well or not is dependent on the

model operator subjective perception based on visual interpretation of the deviation between the correct target value and its prediction.

Nevertheless, there are several adaptive approaches in the literature related to the Soft Sensors. The majority of these approaches are based on adaptive versions of the PCA or PLS, like *Moving Window PCA* (Wang et al., 2005) or the *Recursive PCA* (Li et al., 2005) (see Section 4.1 for the PCA and Section 4.2 for the PLS). All of these methods rely on periodical or continuous adaptation of the principle component base. Neuro-fuzzy based Soft Sensors (see section 4.4 for an overview), such as (Macias and Zhou, 2006), often intrinsically provide mechanisms for automatic adaptation. These mechanisms are based on the deployment of new units in the neural structure of the model once a new state of the data is found. An approach related to the neuro-fuzzy methods also providing adaptation possibilities and is local learning (Atkeson et al., 1997). An adaptive Soft Sensor developed in this framework was published in Kadlec and Gabrys (2008a).

Despite the methods for the automated Soft Sensor adaptation the model operator still plays an important role as it is his judgment and knowledge of the underlying process which decides about the way the parameters of the individual adaptation methods are selected (e.g. the length of the window in case of the moving window technique, or a threshold for the deployment of a new receptive field in case of the neuro-fuzzy methods).

3.2.6   Related methodologies

The discussed methodology, though it is the one most commonly used, is not the only possible way for developing a Soft Sensor. For example in Warne et al. (2004a)

an alternative methodology for Soft Sensor, or inferential sensor in Warne's terminology, development has been presented. It is less detailed but still consistent with the methodology presented here. It focuses on three different steps, namely:(i) Data collection and conditioning, (ii) Influential variable selection and (iii) Correlation building. These three steps correspond to the "Selection of historical data", "Data pre-processing" and "Model selection, training and evaulation" steps in Figure 1.

Another work mentioning Soft Sensor development methodology is Fortuna (2007). Again, there is no significant difference to the methodology presented in this section.

Han and Lee (2002) presents a rather general methodology for Soft Sensor development in the light of the Six Sigma process management methodology (see Smith and Fingar (2003) for details on Six Sigma).

In Park and Han (2000), in addition to a general 3-step Soft Sensor methodology consisting of the (i) process understanding, (ii) data preprocessing and (iii) model determination steps, there is a more specialised methodology for the development of models based on multivariate smoothing procedure discussed.

3.3 Soft Sensor Applications

The applications of Soft Sensors can be found across many fields of the process industry. The most typical examples are the chemical industry, paper/pulp industry and steel industry. The following sections list examples of the previously introduced three most common application types of Soft Sensors across these different fields of the process industry.

3.3.1   On-line prediction

The most common application of Soft Sensors is the prediction of values which cannot be measured on-line using automated measurements. This may be for technological reasons (e.g. there is no equipment available for the required measurement), economical reasons (e.g. the necessary equipment is too expensive), etc. This often applies to critical values which are related to the final product quality. Soft Sensors can in such scenarios provide useful information about the values of interest and in the case when the Soft Sensor prediction fulfils given standards, it can be also incorporated into the automated control loops of the process. Soft Sensors have been widely used in fermentation, polymerisation and refinery processes. The common denominator of these processes is their dynamics which can not be easily described in terms of rigorous models and that there is often no way of collecting the necessary information on-line. From the computational learning point of view these problems are equivalent to supervised regression. The data-driven models are based on historical data of the process. This data consists of the past plant measurements which form the input data space of the Soft Sensor. The target values are the lab measurements, infrequent observations, etc., of the values of interest.

Linear regression models are the most straightforward way of modelling the target values. In this case, the modelled variable is a linear combination of the input variables.

A Soft Sensor for the modelling of the particle size in a grinding plant was published in Casali et al. (1998). The developed Soft Sensor is an ARMAX-type stepwise regression model. The input for the model are systematically selected based on the correlation between the analysed input feature and the output including delayed

26

versions of the input variables. The authors present a set of models using different types of input including combined inputs based on the a-priori (phenomenological) knowledge about the process. The best performance is achieved by a model combining historical data and physically significant combinations of the input variables, i.e. a grey-box model.

Locally Weighted Regression (LWR) together with non-linearity handling pre-processing is applied in Park and Han (2000). As the process data are non-linear, the authors propose to use models with limited field of influence (local models). The advantage of this kind of models is that one can use less complex linear models to deal with the problem. The performance of the proposed Soft Sensor is compared to another common modelling approaches like ANN in terms of two industrial data sets (toluene composition in a splitter column and diesel temperature estimation in a crude oil column). The results show that the LWR based method provides comparable or better results when compared to the other modelling techniques.

Another Soft Sensor based on local learning was published in Kadlec and Gabrys (2008a). This Soft Sensor is based on a combination of set of locally valid models. These local models are combinations of ten Multiple Linear Regression (MLR) models. The receptive fields are modelled using the Parzen window technique. Based on an application of the Soft Sensor to an industrial drier process the model shows much better performance than a traditional MLP based Soft Sensor. Furthermore, the presented approach provides several possibilities for adaptation of the Soft Sensor which leads to further performance improvement.

Another typical modelling approach used for these problems is the application of Multi-Layer Perceptron (MLP) which is one of the most popular Artificial Neural

Network (ANN) models used for function approximation. An introduction to ANNs is given in Section 4.3.

Thorough analysis of the application of MLPs for Soft Sensor building has been presented in Qin (1997). This work discusses a lot of practical issues of the application of neural networks for Soft Sensor modelling. A particular focus is put on the necessary pre-processing steps like the handling of missing values and outliers. Focusing on the identified issues, there is also a modification of the error measure of the back-propagation algorithm (i.e using of Manhattan distance instead of mean squared error) proposed. Furthermore the MLP based Soft Sensor is compared to an NNPLS model. Based on the case study dealing with batch refinery process, it is shown that the NNPLS outperforms the MLP due to better generalisation performance and more effective dealing with data co-linearity.

In Jos de Assis and Maciel Filho (2000) an MLP is compared to model-driven approaches based on First Principle Model (FPM), adaptive observer technique and extended Kalman Filter (eKF) models, which are common approaches to model-driven Soft Sensor building. The disadvantages of FPM and eKF are the complexity of the development and amount of a-priori knowledge which has to be available for the model development. On the other hand, the applicability of the MLP for solving on-line estimation of fermentation batch processes is limited due to the changing dynamics of the particular batch runs. The authors therefore suggest a hybrid solution where the process dynamics is described by a model-driven model and the MLP black-box approach is used to model only parts of the model, like the growth rate of bioprocesses.

Meleiro and Finho (2000) are presenting a grey-box Soft Sensor which delivers

necessary control information for self-tuning adaptive controller of a fermentation process. The Soft Sensor is an MLP which is trained using simulated data based on a phenomenological model of an ethanol production plant. After training the model is validated using industrial process data. The Soft Sensor is successfully implemented into the control loop of the process controller.

Radhakrishnan and Mohamed (2000) publishes an extensive discussion of application aspects of MLP to steel industry data modelling. They provide a detailed procedure, including data preprocessing, model selection, etc., for the application of MLP to the modelling of metal quality in a blast furnace. There is also an expert system for the control of the silica content, which is based on the developed Soft Sensor, presented. In a real-life application, the installation of the Soft Sensor and the expert system leads to significant improvement of the steel production.

An application of MLP for sugar quality estimation was published in Devogelaere et al. (2002). The approached problem in this work is the modelling of the massecuite electrical conductivity which is an important value for the control loop controlling the sugar production process. The eight input features of the model were selected manually using a-priori knowledge about the process. The results achieved by the MLP were good enough to take the Soft Sensor into real-life operation.

Fortuna et al. (2005) developed and published a complex Soft Sensor based on MLP. The Soft Sensor models the butane and stabilised gasoline concentrations of a distillation column. The model is a cascaded 3-level neural network. Apart from the input variables which are measurements within the column the model uses delayed versions of the input variables. The model gives satisfactory results for the on-line prediction of the concentrations.

29

The performance of two ANN variants, namely the Multi-Layer Perceptron (MLP) and the Radial Basis Function Network (RBFN), are compared to a Support Vector Regression (SVR) model in Desai et al. (2006). The data sets for the comparison are two simulated batch bioprocesses. It is clearly shown, that the performance of the SVR Soft Sensor is superior in comparison to the other two methods. The authors also provide a theoretical explanation of the performance benefits. The ability to locate global minima of the presented problems and the interpretability of the learnt knowledge in terms of the training data (support vectors) are stated as advantages of the proposed SVR Soft Sensor.

Another performance comparison between an MLP and an RBFN was published in James et al. (2002). In this work, these two models are also compared to a grey-box model based on a first principle model and either an MLP or an RBFN. The performance was tested in terms of a biomass concentration prediction in a biochemical batch process. They describe the hybrid model as the best performing one. However, the performance gain comes at the cost of a-priori knowledge which have to be input into the model.

In Wang et al. (2006) an RBFN-based Soft Sensor for the modelling of a membrane separation process was developed. The Multiple Input Multiple Output (MIMO) Soft Sensor predicts some critical process performance values (like gas concentrations). The aim of the Soft Sensor is to deliver additional on-line information for the process control.

An ensemble approach for Soft Sensor development based on Multi-Layer Perceptrons was published in Kadlec and Gabrys (2008b). In this work the problem of optimal network complexity selection was approached in the context of ensemble

methods. The optimal MLP topology was established by training several models with different complexities and assessing their relative performance. In such a way performance distributions across the different parameter values were calculated. The final ensemble is built by weighting the contributions of ensemble members by their estimated generalisation performance. This Soft Sensor was applied to an industrial drier process.

Su et al. (1998) published an application of Recurrent Neural Network (RNN) to the modelling of the degree-of-cure, which is an important quality indicator in an epoxy/graphite fiber composites production process. The Soft Sensor is a grey-box model, making partial use of a-priori information about the process. The Soft Sensor was parametrised, trained and evaluated using simulated process data and after some minor tuning tested using real process data and target values obtained from off-line laboratory measurements. The authors were satisfied with the performance of the Soft Sensor and deployed it in the real-life process environment.

Also an RNN was applied to the prediction of biomass concentration in Chen et al. (2004a). RNN was applied in this work due to its theoretical ability to capture dynamic effects underlying the data. Although the RNN model performance is not compared to any other model type, the authors conclude that recurrent artificial neural networks are capable of achieving a satisfactory prediction performance.

Another RNN application to the prediction of the melt-flow-length for filling of molds in injection molding process was presented in Chen et al. (2004b). The authors decided to use the recursive version of ANN because of its capability to store temporal patterns which is of advantage in the modelled process. The developed Soft Sensor provides accurate results of the melt-flow-length prediction.

Yang and Chai (1997) focus on soft sensing in a dynamic environment. The authors discuss the application of a multi-step predictor and decide to use an RNN for its implementation. They are using an Inner Recurrent Neural Network, where only the hidden layer has recursive connections. The usefulness of the algorithm is demonstrated based on three dynamic simulated processes.

The authors of Fellner et al. (2003) propose a grey-box technique for the implementation of a-priori knowledge in a data-driven model. They focus on ANN, which provides the possibility to deploy nodes (neurons) which represent the process knowledge, e.g. single differential equations, etc. The nodes are abstract signal processing units transforming the input information to their output using arbitrary, but differentiable, equations. The authors apply the proposed ANN to the estimation of diacetyl in a biochemical process.

Another method commonly applied to soft sensing is the PCA/PLS-based regression (see Section 4.1).

A self-validating Soft Sensor is presented in Qin et al. (1997). The input data is validated using a PCA-based approach for fault detection published in Dunia et al. (1996). In the case of a detected failure, the sensor can be reconstructed using the correlation structure of the affected input measurement to the other input space variables, which is one of the valuable capabilities of the PCA. After this pre-processing step, which on one hand removes the co-linearity of the input data and on the other hand provides the ability for the reconstruction of sensor faults, a Soft Sensor using traditional modelling techniques is built. This Soft Sensor is successfully evaluated on a real-life problem dealing with air emission monitoring process data.

Dayal and MacGregor (Dayal and MacGregor, 1997) proposed a novel recursive version of the least squares algorithm based on the Exponentially Weighted PLS (EWPLS). The authors use an adaptive approach for the time window length calculation. Within the time window the samples are exponentially weighted dependent on their age. The model is successfully applied to two processes: a simulated continuous stirred tank reactor and an industrial flotation circuit.

Another recursive version of the PLS algorithm is devised in Qin (1998). In this work the recursive PLS algorithm is extended to a version which works block-wise and is thus suitable for adaptive modelling. The algorithm is combined with the two common techniques for adaptive modelling, namely with the moving window and the forgetting factor approaches. The performance of the proposed algorithms is demonstrated by applying it to octane number modelling in a refinery process.

Zamprogna et al. (2004b) is dealing with application aspects of the PCA and PLS to the modelling of batch processes. Furthermore, there is a set of PLS regression models using different regressors developed and evaluated. The data set used for the evaluation is a simulated distillation column. The PCA algorithm is used for the identification and discarding of erroneous process states. The best prediction results are, due to the non-linearity of the process, achieved using the Multi-way PLS.

In Lin et al. (2007) in addition to a systematic procedure for PCA-based Soft Sensor development, two case studies applying the proposed method to process industry problems, namely a free lime prediction and NO_x prediction in a cement kiln, are presented. Within the proposed development procedure firstly missing values are handled using an heuristic approach. This is followed by outlier detection using

33

univariate Hampel identifier and multivariate robust statistics, like the Q-Statistics and the Hotelling's T^2. After the data pre-processing, a PLS-based regression model performing a one-step-ahead prediction is derived.

In accordance to increasing popularity of Support Vector Machines (SVM) in the machine learning community, there are also some recent applications of this technique to soft sensing. Support Vector Machines are in more detail described in Section 4.5.

Yan et al. (2004) presents a Soft Sensor based on SVR, or more accurately on Least Squares Support Vector Machines (LS-SVM). The authors define an iterative procedure which, apart from involving the LS-SVM model, uses Bayesian evidence framework for the optimal selection of the LS-SVM model parameters. The model is successfully applied to the estimation of the freezing point of light diesel oil in a Fluid Catalytic Cracking unit.

In Feng et al. (2003), there is also an LS-SVM model applied to a process industry problem. The LS-SVM is chosen due to an evidence for better generalisation properties when compared to an RBFN. Indeed the LS-SVM outperforms an RBFN on the case study dealing with the prediction of gasoline absorption rate in an Fluid Catalytic Cracking unit. The LS-SVM model is also described as being less dependent on the size of the training data set, providing stronger learning ability.

Another very popular and successful family of approaches applied to soft sensing (see Section 4.4) are neuro-fuzzy models combining the advantages of ANNs, most commonly the multi-layer perceptrons, and Fuzzy Inference Systems (FIS).

A Neuro-Fuzzy System (NFS) model was developed and published by Wang and

Rong (Wang and Rong, 1997). The presented NFS is trained using a two-step approach consisting of a clustering and a back-propagation algorithm. One of its advantages is that the connectionist structure is determined automatically. The proposed approach is applied to the modelling of a distillation column, more specifically, to the propylene purity modelling at the output of the column.

An example of this type of Soft Sensor is an ANFIS-based Soft Sensor applied to rubber viscosity prediction in Merikoski et al. (2001). Because there is no automated way to measure rubber viscosity, which is an important quality indicator, a Soft Sensor is necessary to deliver the data. In the publication, it is claimed that the accuracy of the Soft Sensor meets the requirements for implementation in the process control loop.

Another ANFIS-based Soft Sensor was presented in Warne et al. (2004b). In this work the data is pre-processed using PCA transformation which on one hand helps to deal with the co-linearity of the data and on the other hand limits the size of the input space of the ANFIS model which in turn reduces the complexity of the model significantly. The presented methodology is applied to the prediction of polymeric-coated substrate anchorage which is an important quality measure of the process product.

Neuro-fuzzy Soft Sensor based on rough set theory and optimized by a genetic algorithm is discussed in Luo and Shao (2006). The rough set theory is used to obtain a reduced set of rules which are then implemented in the form of an MLP. The genetic algorithm is used to get an optimal discretisation of the input variables. The performance of the algorithm is demonstrated on a refinery case study, namely on the prediction of freezing point of the light diesel fuel in a Fluid Catalytic

Cracking unit.

Neuro-fuzzy FasArt and FasBack were applied in Arazo-Bravo et al. (2004) for the modelling and control of a penicillin production batch process. A Soft Sensor for the prediction of the biomass, viscosity and penicillin production delivers the necessary information for the control mechanisms of the FasBack adaptive controller. The holistic control model is trained and evaluated using simulated process data. The trained model is then able to deliver satisfactory results for the real process control.

In Macias and Zhou (2006) an extended Takagi-Sugeno (exTS) model has been applied to the prediction of the quality of crude oil distillation in a refinery process. The advantages of applying an evolving neuro-fuzzy model to this problem is reported to be the ability of the model to deal with non-linear problems and dealing with a large number of features. The presented model has the ability to evolve its rule base together with the dynamics of the process, which is an advantage of evolving neuro-fuzzy methods, distinguishing the NFS from other models.

Apart from the combination of ANN and FIS there is a large number of other hybrid models, which are combination of two or more computational learning techniques.

The work of Qin (1997) has been already mentioned, and one of the contributions of this work is the definition of Neural Network Partial Least Squares (NNPLS) algorithm which is a hybrid system combining the PLS algorithm with an MLP. This algorithm makes use of the capabilities of the MLP to map the input variables non-linearly onto the latent variables of the PLS. The discussed hybrid algorithm is also applied to a refinery process.

Another application of NNPLS to soft sensing was presented in Dong et al. (1995),

36

where the NNPLS and the Non-Linear Principle Component Analysis (NLPCA) algorithms were applied to the prediction of emissions of NO_x gas in exhaust streams. In this case, the input data is pre-processed by mapping it on principle components space using the NLPCA algorithm. After this pre-processing the actual model, an NNPLS technique, predicts the target values. The application shows that the model outperforms a linear model and also demonstrates an immunity with regards to missing values.

A hybrid system consisting of Particle Swarm Optimisation which is used for the training of an MLP was presented in Li et al. (2005). In this work the PSO algorithm is combined with the Alopex algorithm (see Tzanakou et al. (1979)) to avoid local minima to which the PSO is prone. The proposed algorithm is applied to an ethylene distillation column data set.

Another hybrid approach to Soft Sensor modelling has been developed by Kordon et al. (Kordon et al., 2002; Kordon, 2004, 2005; Kordon et al., 2005). In this case, the hybridisation is done on a lower level. The involved methods perform pre-processing of the data for the succeeding modelling steps. The methodology for the inferential sensor building consists of three different steps. The first step is the analysis of the data by an analytical neural network (Kordon, 2004). The aim of this step is to perform feature selection on the input data and to deal with time delays between the selected features. In the next step the data is processed using SVM. During this step the outlier detection is done. In the third step the actual Soft Sensor is built. This is performed by applying the Genetic Programming (GP) algorithm. The GP algorithm selects a function from a pool of available functions and trains it to model the output variable using the pre-processed input data. The Soft Sensor is a set of analytical functions which maps the input space to the target variable space. The

proposed approach was applied to several real-life problems, e.g. the interface level estimation in an organic process in Kalos et al. (2003).

The work of Chen et al. (Chen et al., 2000) was already briefly mentioned. The Soft Sensor presented in this work is a grey-box model of a model-driven first principle model and a data-driven artificial neural network. The ANN, which is Radial Basis Function Network (RBFN), is used to model the non-linear reaction rates. This model is then incorporated into the mass-balance model of a stirred-tank bioreactor. The performance of the proposed hybrid Soft Sensor is illustrated on an experimental case-study dealing with single microbial population.

A non-traditional approach to soft sensing is presented in Rao et al. (1993). There is an "Intelligent Soft Sensor" presented in this publication. It is a large system consisting of a symbolic rule-based part, numerical part and a graphical part. This allows to integrate quantitative as well as qualitative knowledge into the model. The three parts are merged by a meta-system. The system is developed for a batch digester quality control support of a sulphite pulping system.

Gonzalez et al. are discussing the performance of an ARMAX stepwise regression, Takagi and Sugeno, fuzzy combinational, PLS, wavelet-based and MLP models in Gonzalez et al. (2003). All these models are applied to a rougher flotation bank modelling. The model input are both the process measurements and the combined features. The combined features are built using a-priori process knowledge and represent meaningful process descriptors. Apart from this contribution a novel 2-level approach for outlier detection combining PCA capabilities and Scheffé's test is provided. After the application to the modelling of copper concentration grade the authors conclude that the dynamic PLS as well as the MLP and wavelet-based

models are providing best performance.

3.3.2 Process monitoring and process fault detection

Another application area of Soft Sensors is the process monitoring. Process monitoring can be either an unsupervised learning or binary classification task. The systems can be either trained to describe/analyse the normal operating state or to recognize possible process faults. Commonly, process monitoring techniques are based on multivariate statistical techniques like PCA, or more precisely on Hotelling's T^2 (Hotelling, 1931) and Q-statistics (Jackson and Mudholkar, 1979). These measures have on one hand the advantage of considering all input features, i.e. using multivariate statistics, and on the other hand providing information about the contribution of the particular features to a possible violation of the monitoring statistics (Choi et al., 2006). Another popular method for process monitoring are the Self Organizing Maps (see Section 4.3).

Nomikos and MacGregor published a pioneering work on the application of PCA-based techniques for batch and semi-batch process monitoring in Nomikos and Mac-Gregor (1995b). In this work they provide a thorough analysis of the applicability of Statistical Process Control (SPC) charts to the batch process on-line monitoring. The monitoring of new batches is based on the comparison of their PCA-space representation to reference curves. The reference curves are based on a set of past "good" processes. Based on the reference batches there is also a possibility to calculate the control limits. In the case of the violation of these limits an alarm is raised and an analysis of the process fault can be done. The presented technique is evaluated on an industrial polymerisation batch process.

Li et al. (2000) is dealing with the application aspects of the PCA and related methods to the process industry problems. The focus is put on the development of a Recursive PCA (RPCA) approach targeting adaptive process monitoring. Within this framework it has also been shown that the method can deal with outliers, missing values and delayed measurements. The authors presented an effective approach for the update of the correlation matrices as well as two algorithms for the incremental update of the PCA base using the old PCA structure. Additionally a review of the most common techniques for the selection of the number of principle components, which is an important question while developing PCA models, is also presented. Based on the review a new technique for recursive selection of the number of principle components is shown. For the purpose of the adaptive process monitoring, it is necessary to update the confidence limits of the model with the new incoming data, therefore the authors define also a monitoring scheme, which detects and handles data outliers, missing values and process faults before updating the model. Finally, the proposed monitoring scheme is applied to a rapid thermal annealing process monitoring.

Rotem et al. (2000) applied the model-based PCA (MBPCA) method to the fault detection of an ethylene compressor. The detection system is based on the first principle model of the process, which makes the method applicable only to this specific process.

A process monitoring Soft Sensor using an adaptive version of the PCA (Fast Moving Window PCA - FMWPCA) was published in Wang et al. (2005). The adaptivity of the model is achieved by updating the data structures necessary for the PCA calculation using a novel moving window technique. This technique updates the PCA base (i.e. removes the oldest data sample and adds the new, current, one)

in a single step which makes this technique computationally efficient. Additionally, an N-step-ahead process monitoring approach is presented which increases the immunity towards the faulty data. The effectiveness of the described algorithm is demonstrated using a simulated Fluid Catalytic Cracking unit process.

In Amazouz and Pantea (2006) an application of PCA and PLS to batch process monitoring is presented. The proposed procedure is split into two steps, the first is applying the PCA to manually explore the data space and to identify reference or "good" batches which are in the second stage used to develop the PLS model. Having a PLS model of this reference batch one can compare the new incoming process data (test data) to this model. If there is a deviation between the new data and the reference model data an analysis of the PLS scores provides information about the variable(s) causing the deviation. The authors are also planing to develop a database of typical process faults and to use an expert system for automatic process fault identification.

The applicability of the PLS, namely the Multi-way PLS algorithm to modelling of batch process quality variables as well as process monitoring and control was presented in Zhang and Lennox (2004). The studied process is a simulated penicillin production fermentation batch process. The quality variable prediction is done using standard PLS regression model. The process monitoring is carried out using the SPE and T^2-statistics of the model.

He et al. (2005) are discussing an alternative approach to process monitoring and process fault detection. The presented method is a three-step approach to process monitoring. The first step is called "Pre-analysis" and at this stage a number of clusters in the process data is manually estimated using 2D and 3D PCA-scores

plots. Using the estimated number of clusters, the data is partitioned by the k-means algorithm. In the second step, the data are visualised after transforming them using the Fisher Discriminant Analysis (FDA). The authors tend to use the FDA instead of the PCA due to the discrimination abilities of the FDA. Within this step the normal and faulty process states are annotated. The final step is then the calculation of the fault directions for the separate fault classes using the pairwise FDA. The calculated fault direction provides information about the source of the particular process fault. The algorithm is applied to a simulated as well as to an industrial process.

Marjanovic et al. (2006) deals with the identification of batch process end points which can improve the process effectiveness. The applied technique is the Multi-way PLS (MPLS). The devised technique proves to be very effective and can thus be implemented for the real-time batch process monitoring.

A set of practical applications of process monitoring and quality prediction using Self Organizing Map (SOM) was published in Alhoniemi (1999). In this work SOMs have been found useful for the monitoring of a continuous pulp digester. Before feeding the data into the SOM model they have been manually pre-processed using a-priori knowledge of the process. Another application presented in the work is the quality prediction of steel production based on the concentration of the input elements and some process parameters. The last application of SOMs presented in the work is the analysis of the data from paper and pulp industry.

A complex Soft Sensor for process fault detection and identification has been presented in Yang et al. (2000). The Soft Sensor is based on an MLP and is applied to the detection of three typical faults in a Fluid Catalytic Cracking (FCC) reactor.

The MLP is fed with input from different sources. One source of input is a model-driven Soft Sensor. This sensor predicts the catalyst circulation rate based on the energy balance equation within the FCC reactor. The output of the Soft Sensor is then mapped to trends of the catalyst circulation rate, e.g. stable, increasing, etc. The trends are then provided to the MLP. The other inputs to the MLP are trends of directly measurable process variable like the reactor temperature, reactor feed flow rate, etc, which are determined using the wavelet transformation. The developed approach works well for the given process but because of the involvement of the process specific FPM, it is not applicable to any other processes.

In a recent publication (Kampjarvi et al., 2008) a complex Soft Sensor for the detection and isolation of process faults is devised which is based on PCA, RBFN and SOM. The Soft Sensor is developed in the framework of an ethylene cracking process. The authors demonstrate improved accuracy of the system after including calculated variables, which are built using process knowledge. The final Soft Sensor achieves high performance and is included into the model predictive control of the process.

3.3.3 Sensor fault detection and reconstruction

The vast majority of modelling techniques applied within the process industry as Soft Sensors are not able to handle data from faulty sensors as a matter of their normal operation, therefore there is a need to identify and replace sensor and process faults before the actual model building and application.

Process and sensor faults are detected and handled using the PCA in Dunia and Qin (1998a) and Dunia and Qin (1998b). The faults are detected in the PCA resid-

ual space. This has the advantage that one can, on one hand, identify the sensor or process faults effectively and on the other hand, by projecting the fault state to the original space one can also find which particular sensor or set of sensors are responsible for the fault. By manipulating the PCA residual space one can also achieve a reconstruction of the fault. The work also defines conditions of the fault detectability, identifiability and reconstructability. For the task of process fault detection there is a need for the description of the "fault direction" which requires the input of process knowledge to the Soft Sensor. For the sensor fault detection there is no need for such a knowledge. The proposed approach is again evaluated in terms of an industrial boiler continuous process.

In Lee et al. (2004) the previous approach was extended to dynamic processes. The extension to dynamic processes is achieved by using the Time-Lagged PCA (TLPCA) instead of the traditional static PCA method. Although there is a need to remove low auto- and cross-correlated variables from the data set, the presented method is claimed to be suitable for highly dynamic processes, which is demonstrated on one simulated and one industrial data set.

Another PCA-based sensor fault detection and diagnosis Soft Sensor was published in Wang and Cui (2005). The Soft Sensor uses the Q-statistics to detect faults and the sensors responsible for them. The underlining process is a centrifugal chiller system. The same authors published another fault detection Soft Sensor (Wang and Xiao, 2004), this time monitoring an Air Handling Unit (AHU). In order to deal with the non-linearity of the process the model is split into two separate models. Additionally, the model is extended using a simple expert system which handles the signals from the two PCA sub-models.

3.3.4 Soft Sensor applications summary

Table 1 provides a list of the Soft Sensor applications discussed in this review and summarizes the most important properties of the Soft Sensors.

The list of Soft Sensor application examples presented in this work is not exhaustive because the amount of published Soft Sensor applications is too large to be fully covered. Instead of this, this work focuses on one hand on recent publications and on the other hand on non-traditional approaches.

Assuming the presented examples are a representative sample of the recent Soft Sensors, the distribution of current soft sensing methods is presented in Figure 2. The figure shows clearly the current trend in soft sensing. The most popular methods for Soft Sensor building are the multivariate statistical techniques, i.e. the PCA and the PLS, which together cover 38% of the applications presented in this review. Another technique commonly applied in soft sensing are the neural networks based methods like MLP, RNN, etc. But some of the most recent applications rely on methods which have been recently finding their way into much broader application areas. These are for example the neuro-fuzzy methods, which have the advantage of providing intrinsic mechanism for adaptation/evolution as well as SVM which have their justification in the theory of machine learning and additionally proved to have very good generalisation ability accross a number of different application areas.

A common point of most of the presented Soft Sensors is the need for the involvement of process related a-priori knowledge. This can be done in several ways. If we ignore the purely model-driven Soft Sensors, which are out of the scope of this review, one can distinguish different levels of a-priori information influence. One type of a-priori information involvement is the construction of additional features
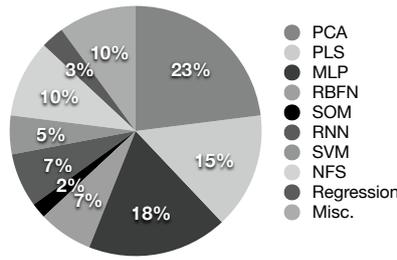
Fig. 2. Distribution of computational learning methods in soft sensing

which describe some process related properties. The hope is that these features will be correlated with the modelled target variable and thus have a positive effect on its modelling. Another way of applying process knowledge to data-driven soft sensing is during the initial modelling steps (see Section 3.2). Especially the pre-processing steps require a lot of attention of the model developer, who has often to interview the process experts in order to be able to carry out manual variable selection, to evaluate the results of the particular pre-processing steps, etc.

Section 4 continues the above discussion and provides a critical review of the most common Soft Sensor techniques as they were identified in this section.

3.3.5 Further reviews of Soft Sensor applications

There are several other publications providing reviews of Soft Sensor applications, among them Gonzalez (1999) giving a review of regression based models, ANN, PCA, Kalman Filter and Expert Systems applications in process industry. Gonzalez focuses on the applications aspects of the before mentioned methods and provides a list of application examples.

In Fortuna (2007) apart from extensive handling of Soft Sensors and their application to process monitoring and control, an overview of applications of mainly

ANN-based Soft Sensors is given.

Dote and Ovaska also provide a list and a discussion of applications of soft computing techniques in the process industry in their general review of industrial application of soft computing methods (Dote and Ovaska, 2001).

Focusing on process fault detection and diagnosis Venkatasubramanian published an extensive three part review. The first part Venkatasubramanian et al. (2003c) provides an introduction to process fault detection and abnormal event management. Apart from a taxonomy of the different approaches, this part presents quantitative model-based methods for process fault detection and criteria which are used to evaluate and compare the different approaches. The second part of the series (Venkatasubramanian et al., 2003a) deals with qualitative model representations and search strategies for process fault detection. These methods are usually based on first principle descriptions of the processes. Finally the third part (Venkatasubramanian et al., 2003b) focuses on process data based techniques. These models can be both, qualitative, e.g. enhanced Kalman Filter models, as well as quantitative which can be based on any data-driven method. From the data driven approaches the authors describe PCA/PLS, Statistical Classifier and ANN based techniques. Additionally, a comparative study of the various techniques presented in the three-part review is given.

4 Data-driven methods for soft sensing

This section outlines and provides further references to the most popular techniques for Soft Sensor development as they were identified in Section 3.3.4. These are the

47

multivariate Principle Component Analysis (Section 4.1), Partial Least Squares (Section 4.2), Artificial Neural Networks (Section 4.3), Neuro-Fuzzy Systems (Section 4.4) and Support Vector Machines (Section 4.5). Finally, the last part of the section deals with the modelling of batch processes (Section 4.6).

4.1   Principle Component Analysis

The PCA algorithm reduces the number of variables by building linear combinations of them. This is done in such a way that these combinations cover the highest possible variance in the input space and are additionally orthogonal to each other. In the context of the process industry data this is a very useful feature because the data there are often co-linear (see Section 2.2). In this way the co-linearity can be handled and the dimensionality of the input space can be decreased at the same time. The PCA is usually applied as pre-processing step followed by the actual computational learning method.

An extensive and general derivation, interpretation and application aspects of the PCA is provided in Jolliffe (2002). Application possibilities of the PCA in the process industry are reviewed in Warne et al. (2004a) and Gonzalez (1999).

There are several extensions of the original PCA algorithm which target some of their published drawbacks. Among them are the Model-Based PCA Rotem et al. (2000) and the Non-Linear PCA (Dong and McAvoy, 1996). Another extensions of the PCA focused on the derivation of an adaptive version of this transformation. Such extensions are the Recursive PCA in (Li et al., 2000), the Moving Window PCA (Wang et al., 2005) and the Time-Lagged PCA in Lee et al. (2004). An overview of different PCA versions is provided in Figure 3.
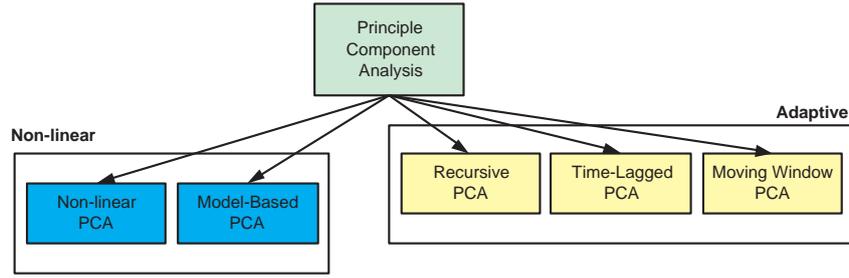
Fig. 3. PCA and its derivations

Although the PCA is a well established and powerful algorithm it has several drawbacks and limitations. One of the limitations is that the pure PCA can only effectively handle linear relationships (correlations) of the data and thus can not deal with non-linearity of the data. This limitation has been solved by extending the original PCA algorithm as discussed in the previous paragraph. Another issue is the selection of optimal number of principal components. This problem is most commonly approached by using cross validation techniques. Another problem is that the principal components describe very well the input space but do not reflect the relation between the input and the output data space which actually have to be modelled. A solution to this problem is given by the Partial Least Squares method discussed in the next Section.

4.2   Partial Least Squares

This algorithm, instead of focusing on the covering of the input space variance, pays attention to the covariance matrix that brings together the input and the output data space. The algorithm decomposes the input and output space simultaneously while keeping the orthogonality constraint. In this way it is assured that the model focuses on the relation between the input and output variables.

49

A general description of the PLS technique is provided in Geladi and Esbensen (1991) and Abdi (2003). As PLS is a very popular technique in chemical engineering and in chemometrics, there are several publications dealing with the application aspects of PLS to this domains (Frank and Friedman, 1993; Kourti, 2002).

The original PLS algorithm suffers from similar problems as its PCA counterpart. It is also modelling only linear relations between the data. Therefore there have been also some advanced versions of the PLS proposed. Making the PLS applicable to non-linear problem is the target of the Multi-way PLS (MPLS) (Bro, 1996) and of the Neural Network PLS (NNPLS) (Qin and McAvoy, 1992). An adaptive version of the PLS called Recursive PLS (RPLS) is proposed in Qin (1998). Another adaptive version of the PLS based on the moving window technique is the Exponentially Weighted PLS (EWPLS) (Dayal and MacGregor, 1997). The different versions of the PLS algorithm are reviewed in Figure 4.
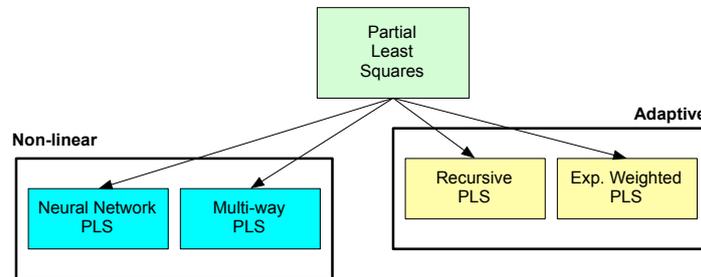
Fig. 4. PLS and its derivations

4.3 Artificial Neural Networks

The original intention of Artificial Neural Networks (ANN) was to build computational models motivated by the operation of biological *neurons* which are the basic

information processing units in nervous systems. The task of both the biological and the artificial neuron is to collect information at the inputs, to process this information and to output it. There is a large variety of computational intelligence models which are more or less biologically plausible and can be summarized under the term *Artificial Neural Network*.

A general introduction to the theory of ANN is given in Bishop (1995). This book describes a large number of different ANN variants, learning algorithms, application areas, etc. Another theoretical considerations of ANN is presented in Hastie et al. (2001), in this work the ANN are presented in a general statistical context. The application aspects of large number of ANN variants, including dynamic and adaptive ANN, are discussed in Principe et al. (2000). This book is especially recommended for Soft Sensor modelling as there is a large number of mutual topics between the book and the process industry applications of ANN. Detailed discussion of some problems of process industry data and of the suitability of ANN to solve this problem is provided in Qin (1997). Apart from the discussion of the ANN issues in the process industry the work proposes also some possible solution to them.

Among the large number of ANN variants mentioned above, the most common, in the process industry as well as in general, are the feed-forward networks, like the Multi-Layer Perceptron (MLP) (Bishop, 1995) and the Radial-Basis Function Network (RBFN) (Poggio and Girosi, 1990). Both of these models are universal function approximators (Funahashi, 1989).

The structure of Recursive Neural Networks (RNN) is similar to the one of the feed-forward networks, the only significant difference is a feed-back connection (Mandic and Chambers, 2001). This gives the network the capability to extract and learn

temporal sequences from the data which can be of advantage in the context of process industry data as these often show re-occurring temporal patterns.

Self-Organizing Map (SOM) or Kohonen Map (Kohonen, 1997) is a type of ANN which is able to deal with unsupervised problems and is thus applicable to process monitoring tasks. SOMs consist of an usually high dimensional input layer and an output layer (also called competitive layer) which is arranged in a two- or three-dimensional grid. During the learning the grid is arranged in such a way that the low dimensional representation of the data preserves its high dimensional topology. This makes them also useful for the visualisation of high-dimensional data. Figure 5 shows an overview of the ANN variants which are interesting from the process industry point of view.
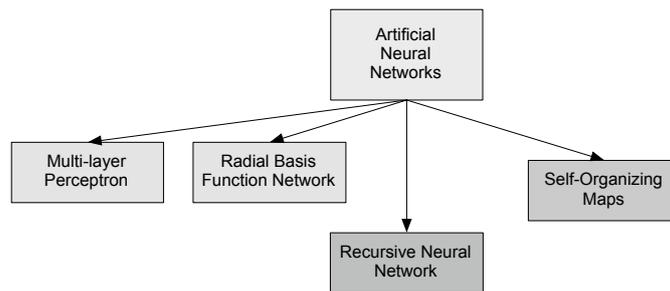
Fig. 5. ANN versions commonly used in process industry

The drawback of ANNs is that during their learning that they are prone to get stuck in local minima, which can result in sub-optimal performance. Another problem are the difficulties with the estimation of optimal topology of the networks. The topology of the ANN is critical for their performance because their generalisation power is to a large extent dependent on the complexity of the networks. There is also an issue with the interpretability of the learnt knowledge. The learnt knowledge is distributed in the weights between the particular neurons and is not available

in terms of human understandable representation. The generalisation performance of the ANN is dependent on the model parameters. This dependency cannot be described in clear analytical terms and is very much dependent on the underlying data.

4.4 Neuro-Fuzzy Systems

Neuro-Fuzzy System (NFS) is a hybrid intelligent model which combines the learning and universal approximation abilities of the ANN with the human-like reasoning of the Fuzzy Inference System (FIS) (Zadeh, 1996). It is a realisation of the fuzzy system by a connectionist structure of an ANN. The aim of the fusion of the two methods is to provide a learning system which provides the advantages of both of the involved techniques while at the same time dealing with their drawbacks. Another appealing property for the process industry application of the NFS models is that the technique is based on receptive fields and thus intrinsically provides means for the building of local models. An introduction to NFS is provided in Jang et al. (1997) and Nauck et al. (1997).

The evolving variants of NFS are very well suited to dealing with dynamic environment. These systems are called *evolving* because they adapt automatically together with the changing environment represented by the data. An evolving system is thought to be able to change its structure, to grow and shrink and to update its parameters (Angelov and Kasabov, 2005). In this way the model is able to deploy new local models related to new states of the input data if necessary.

An early example of such a method, which was able to adapt its structure according to the complexity of the underlying problem, was published in Gabrys and Bargiela

(1999). Examples of several other evolving neuro-fuzzy methods are Angelov and Buswell (2002); Angelov and Filev (2004); Rong et al. (2006).

4.5 Support Vector Machines

Due to their theoretical background in the statistical learning theory Support Vector Machines (SVM) gained attention in the computational learning community. Their derivation and theoretical justification can be found in Vapnik (1998). Application aspects of the SVM are discussed in Yan et al. (2004); Feng et al. (2003); Li and Li (2005).

While grounded in the theory, SVMs have been demonstrated to work very well for a wide spectrum of applications so it is not surprising that they have also been successfully applied as Soft Sensors. While some successes have been reported there is still a lot of work needed especially when dealing with very large data sets for which the computational complexity of SVM training process can be prohibitive.

4.6 Techniques for batch process modelling

When measuring K variables at J time points one batch leads to $J \times K$ data tables. Therefore a set of N batches leads to a three-way matrix of the size $N \times J \times K$.

To apply the so far discussed methods it is necessary to unfold the three dimensional matrix into a two dimensional table. This task has been approached by Nomikos and MacGregor (1994, 1995b), where the Batch-Wise-Unfolding (BWU) method for unfolding the data has been discussed for graphical presentation of this approach. This unfolding technique is based on the Multi-way PCA or PLS (Wold et al.,

1987; Nomikos and MacGregor, 1995a) and nowadays widely used in batch process industry.

Another approach to unfolding the three-way matrix was developed by Wold et al. (1998). This method is called Observation-Wise-Unfolding (OWU) and it was described in details in (Eriksson et al., 2001). While BWU mainly deals with the batch-to-batch variation, OWU describes the dynamic behaviour of the batches and summarizes the main trajectory of all variables. Time or maturity is regarded as a y-variable and the progress of the batch is modelled using the PLS method. The two discussed unfolding methods (BWU and OWU) have the conditions for selecting the batch database in common. In the case of monitoring batch processes, it is often of interest to compare the current batch to a set of well performing batches. These batches are selected according to quality or any other performance indicator and often referred to as the *golden batch*. The model is built based on the data set which fulfils the quality requirements. The tested batch is then monitored with respect to the model and therefore compared to the *golden batch*. However, if on-line prediction of the output quality is of interest, it is advisable to build a model of all batches, those with good and those with poor quality. This will introduce enough variability into the model and lead to a better performing prediction model.

In van Sprang et al. (2002) a good comparison of both unfolding approaches is given.

A major drawback of the BWU method is the necessity of having batches with equal length which is not always the case due to the variability in the production and some other influences which are out of control of the process controllers. To overcome this problem a maturity variable (e.g. conversion) can be introduced as it

was presented in Nomikos and MacGregor (1995b) and in Neogi and Schlags (1998). In the case of OWU, the varying length of the data tables is not a big issue as long as the variation of the batch length is kept in certain limits.

Applying the PCA or PLS also requires mean centering and scaling to unit variance of the unfolded data. In this way the mean trajectory is removed from the data and the remaining data explains the variation of the batch around the average batch which in turn allows an effective batch-to-batch comparison of the underlying process. A major drawback of the discussed approach is the assumption that the data of the whole batch is available. To solve this issue Rnnar et al. (1998) proposed a hierarchical method, where the batch is divided into several stages and separate models are built for each of the stages. The final model is a hierarchical combination of the component models.

Another issue of batch process modelling lies in the fixed model. For example in case of process monitoring, when the process shows slow changes from batch to batch it will lead to a lot of false alarms. To overcome this Lee and Vanrolleghem (2003) and Lee et al. (2003), suggest a moving window techniques for updating the model after a new successful batch was finished. It has been shown that through the continuous update of the model batch-to-batch changes can be effectively handled. Lee and Vanrolleghem (2003) applied the technique to a sequencing batch reactor for biological waste water treatment and Lee et al. (2003) to a simulated fed-batch penicillin production process.

An advanced to batch process monitoring and fault detection was presented in Kourti et al. (1995). In this work, in addition to the three-way data table, a two-dimensional data matrix with initial conditions like, product quality measurements

is incorporated into the model.

Apart from the before mentioned work, there are several extensions and further developments of the described methods. For example, Nomikos and MacGregor (1995b) and Chen and Liu (2002) discuss SPC charts for batch control and the latter additionally proposes a dynamic approach by using time-lagged variables in the model. To overcome the drawbacks of linear models Lee and Dorsey (2004) use state space models in combination with batch modelling. Van sprang et al. (2005) is improving the model performance by combining grey models having process related expert knowledge with batch modelling. In the field of the PAT-initiative (Process Analytical Tools) of the FDA (Food and Drug Association) Gabrielsson et al. (2006) combines spectroscopic data and *design of experiments* together with batch monitoring.

5 Open Issues and future steps of Soft Sensor development

There are two main issues in the Soft Sensor development and maintenance respectively. At the development phase, there has to be a lot of effort spent on the manual pre-processing of the data as well as on the model selection and validation steps. To be able to deal with issues like missing values, data outliers, etc., discussed in Section 2.2, the model developer has to manually try different pre-processing approaches and select the one giving the best performance as estimated on the training or validation data. Furthermore, the Soft Sensor development process is usually iterative, which means that after optimizing one part of the model development process the developer has to check the influence of this action on the other parts of the model and possibly tune the affected parts of the model. The solution to this problem

can be approached from different directions. One of them is obtaining as much of process knowledge as possible and to incorporate this knowledge into the model. A problem of this approach is that the process knowledge differs from one process to another and has to be manually incorporated into the models each time a new Soft Sensor is developed. As such this approach is not a real solution to the previously discussed issue. Another way to approach the mentioned problem is to equip the model with the ability to select the most appropriate approach from a pool of available methods. This of course increases the complexity of the model but at the same time, if implemented effectively, moves at least a part of the manual development burden to the model. For this purpose techniques from the active research field called *Meta Learning* (see e.g. Vilalta and Drissi (2002) for a review of meta learning techniques) can be applied.

Another major issue is related to the model maintenance. After successful launch of the Soft Sensor in the most cases one can observe a gradual deterioration of the performance of the Soft Sensor. The decrease of the prediction quality is caused by the gradual changes in the process. Usually after some time the performance of the model reaches unacceptable level and the model has to be retrained or in the worst case rebuilt from scratch. This problem has already been realized by the Soft Sensors developers and some approaches to solve the problem have been reported as discussed in Section 3.2.5.

6 Summary

Figure 6 provides a summary of this review. We focus on two main aspects of the Soft Sensor development: (i) on the process industry and (ii) on the most common

computational learning techniques applied for the Soft Sensor modelling.
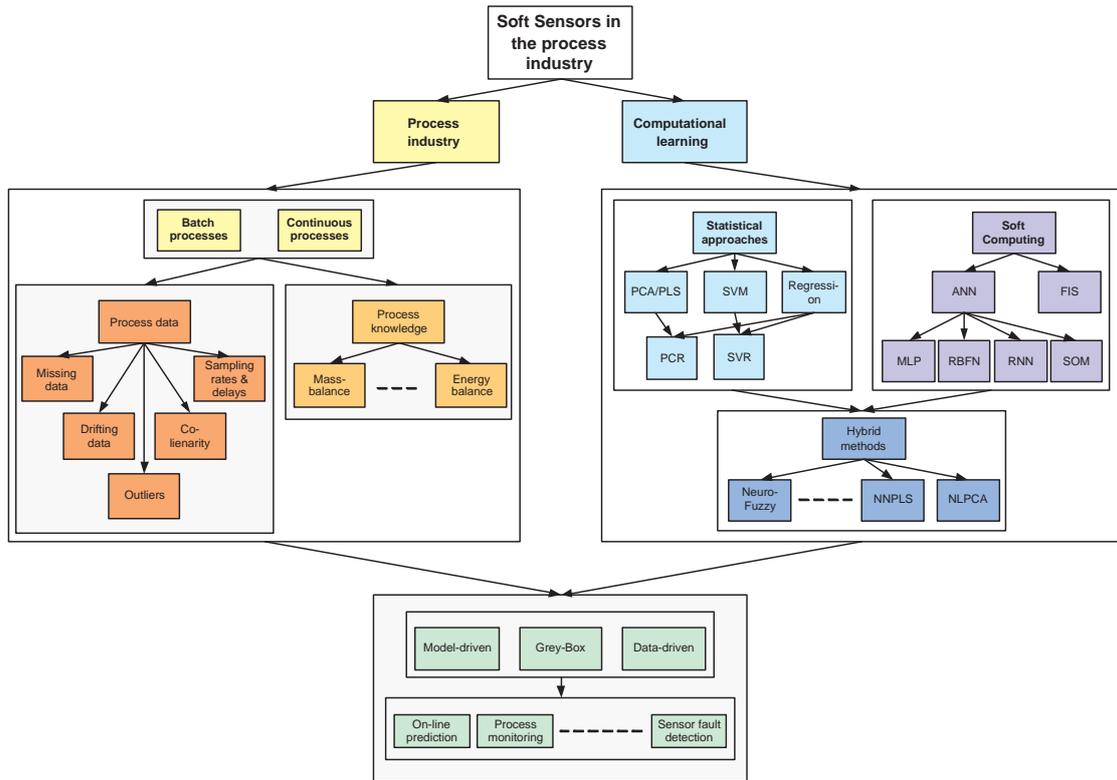
Fig. 6. Overview of this review. For the list of abbreviations see 2

This review mainly focused on data-driven and grey-box Soft Sensors. The data for the training, evaluation and testing of the models is delivered by the process industry. The industrial data has some common properties like missing values and outliers, which are listed on the left hand side of Figure 6. Currently, in order to deal with these issues there is a lot of manual pre-processing and process knowledge necessary. But there are already first Soft Sensors publications dealing with those critical issues in a semi-automated way.

The methods currently applied to process industry problems are coming on one hand from the statistical part and on the other hand from the soft computing part of computational learning. This review outlines the most common techniques for

Soft Sensor modelling from both fields. Apart from the traditional methods for soft sensing like the PCA and the ANN, hybrid methods are currently becoming popular. Another method which recently caught the attention of Soft Sensor developers is SVM based regression. The applications presented in this review outline the advantages and drawbacks of these approaches.

This review provides also an extensive list of applications of Soft Sensors across many fields of the process industry. The presented examples focus on the application of Soft Sensor as: on-line predictors, process monitoring and sensor fault and reconstruction tools.

Based on the reviewed applications we have identified a set of the most important data-driven techniques, which are applied to Soft Sensor modelling, and provide a discussion of these methods which focuses on a brief introduction into the method followed by a review of publications dealing with theoretical and practical aspects of the methods.

References

Abdi, H., 2003. Partial least squares (pls) regression. Encyclopedia of Social Sciences, Research Methods. Thousand Oaks (CA): Sage (2003)Encyclopedia of Social Sciences, Research Methods. Thousand Oaks (CA): Sage (2003).

Alhoniemi, E. S. A., 1999. Process monitoring and modeling using the self-organizing map. Integrated Computer-Aided Engineering 6 (1), 3–14, 6 (1).

Amazouz, M., Pantea, R., 2006. Use of multivariate data analysis for lumber drying process monitoring and fault detection. In: Crone, S. F., S., L., Stahlbock, R. (Eds.), International Conference on Data Mining. pp. 329–332, pp. 329–332.

Angelov, P., Buswell, R., 2002. Identification of evolving fuzzy rule-based models. IEEE TRANSACTIONS ON FUZZY SYSTEMS 10 (5), 667, 10 (5).

Angelov, P., Kasabov, N., 2005. Evolving computational intelligence systems. In: IEEE Workshop on Genetic and Fuzzy Systems GFS2005. Grenada, Spain, Grenada, Spain.

Angelov, P. P., Filev, D. P., 2004. Flexible models with evolving structure. International Journal of Intelligent Systems 19 (4), 327–340, 19 (4).

Arazo-Bravo, M. J., Cano-Izquierdo, J. M., Gmez-Snchez, E., Lpez-Nieto, M. J., Dimitriadis, Y. A., Lpez-Coronado, J., 2004. Automatization of a penicillin production process with soft sensors and an adaptive controller based on neuro fuzzy systems. Control Engineering Practice 12 (9), 1073–1090, 12 (9).

Atkeson, C. G., Moore, A. W., Schaal, S., 1997. Locally weighted learning. Artificial Intelligence Review 11 (1), 11–73, 11 (1).

Bastin, G., Dochain, D., 1990. On-line estimation and adaptive control of bioreactors. Elsevier New York, Amsterdam, Elsevier New York, Amsterdam.

Bauer, E., Kohavi, R. O. N., 1999. An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. Machine Learning 36, 105–139, 36.

Bishop, C. M., 1995. Neural Networks for Pattern Recognition. Oxford University Press, USA, Oxford University Press, USA.

Bonne, D., Jorgensen, S. B., 2004. Data-driven modeling of batch processes. In: Proc. of 7th International Symposium on Advanced Control of Chemical Processes, ADCHEM. In: Proc. of 7th International Symposium on Advanced Control of Chemical Processes, ADCHEM.

Breiman, L., 1996. Bagging predictors. Machine Learning 24 (2), 123–140, 24 (2).

Bro, R., 1996. Multiway calibration. multilinear pls. Journal of Chemometrics 10 (1), 47–61, 10 (1).

Casali, A., Gonzalez, G., Torres, F., Vallebuona, G., Castelli, L., Gimenez, P., 1998. Particle size distribution soft-sensor for a grinding circuit. Powder Technology 99 (1), 15–21, 99 (1).

Champagne, M., Dudzic, M., Inc, T., Temiscaming, Q., 2002. Industrial use of multivariate statistical analysis for process monitoring and control. American Control Conference, 2002. Proceedings of the 2002 1, 1.

Chen, J., Liu, K. C., 2002. On-line batch process monitoring using dynamic pca and dynamic pls models. Chemical Engineering Science 57 (1), 63–75, 57 (1).

Chen, J. M., Chen, B. S., 2000. System parameter estimation with input/output noisy data andmissing measurements. Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on] 48 (6), 1548–1558, 48 (6).

Chen, L., Bernard, O., Bastin, G., Angelov, P., 2000. Hybrid modelling of biotechnological processes using neural networks. Control Engineering Practice 8 (7), 821–827, 8 (7).

Chen, L. Z., Nguang, S. K., Li, X. M., Chen, X. D., 2004a. Soft sensors for on-line biomass measurements. Bioprocess and Biosystems Engineering 26 (3), 191–195, 26 (3).

Chen, X., Gao, F., Chen, G., 2004b. A soft-sensor development for melt-flow-length measurement during injection mold filling. Materials Science & Engineering A 384 (1-2), 245–254, 384 (1-2).

Choi, S. W., Martin, E. B., Morris, A. J., Lee, I. B., 2006. Adaptive multivariate statistical process control for monitoring time-varying processes. Industrial & Engineering Chemistry Research 45, 3108–3118, 45.

Chruy, A., 1997. Software sensors in bioprocess engineering. Journal of Biotechnology 52 (3), 193–199, 52 (3).

Davies, L., Gather, U., 1993. The identification of multiple outliers. Journal of the American Statistical Association 88 (423), 782–792, 88 (423).

Dayal, B. S., MacGregor, J. F., 1997. Recursive exponentially weighted pls and its applications to adaptive control and prediction. Journal of Process Control 7 (3), 169–179, 7 (3).

De Wolf, S., Cuypers, R. L. E., Zullo, L. C., Vos, B. J., Bax, B. J., 1996. Model predictive control of a slurry polymerisation reactor. Computers and Chemical Engineering 20, 955–961, 20.

Desai, K., Badhe, Y., Tambe, S. S., Kulkarni, B. D., 2006. Soft-sensor development for fed-batch bioreactors using support vector regression. Biochemical Engineering Journal 27 (3), 225–239, 27 (3).

Devogelaere, D., Rijckaert, M., Leon, O. G., Lemus, G. C., 2002. Application of feedforward neural networks for soft sensors in the sugar industry. In: Neural Networks, 2002. SBRN 2002. Proceedings. VII Brazilian Symposium on. pp. 2–6, pp. 2–6.

Ding, F., Chen, T., 2005. Modeling and identification for multirate systems. Acta Automatica Sinica 31 (1), 105–122, 31 (1).

Dong, D., McAvoy, T. J., 1996. Nonlinear principal component analysis–based on principal curves and neural networks. Computers and Chemical Engineering 20 (1), 65–78, 20 (1).

Dong, D., McAvoy, T. J., Chang, L. J., 1995. Emission monitoring using multivariate soft sensors. In: American Control Conference, 1995. Proceedings of the. Vol. 1. Vol. 1.

Dote, Y., Ovaska, S. J., 2001. Industrial applications of soft computing: A review. In: Proceedings of the IEEE. Vol. 89. pp. 1243–1265, pp. 1243–1265.

Doyle, F. J., 1998. Nonlinear inferential control for process applications. J. Process

Contr 8 (5-6), 339–353, 8 (5-6).

Dunia, R., Qin, J., Edgar, T. F., McAvoy, T. J., 1996. Sensor fault identification and reconstruction using principal component analysis. In: Proceedings of the 13th Triennial IFAC World Congress. pp. 259–264, pp. 259–264.

Dunia, R., Qin, S. J., 1998a. Joint diagnosis of process and sensor faults using principal component analysis. Control Engineering Practice 6 (4), 457–469, 6 (4).

Dunia, R., Qin, S. J., 1998b. Subspace approach to multidimensional identification and reconstruction. AIChE Journal 44 (8), 1813, 44 (8).

Eriksson, L., Johansson, E., Kettaneh-Wold, N., Wold, S., 2001. Multi-and Megavariate Data Analysis: Principles and Applications. Umetrics, Umetrics.

Fellner, M., Delgado, A., Becker, T., 2003. Functional nodes in dynamic neural networks for bioprocess modelling. Bioprocess and Biosystems Engineering 25 (5), 263–270, 25 (5).

Feng, R., Shen, W., Shao, H., 2003. A soft sensor modeling approach using support vector machines. American Control Conference, 2003. Proceedings of the 2003 5, 5.

Fortuna, L., 2007. Soft Sensors for Monitoring and Control of Industrial Processes. Springer, Springer.

Fortuna, L., Graziani, S., Xibilia, M. G., 2005. Soft sensors for product quality monitoring in debutanizer distillation columns. Control Engineering Practice 13 (4), 499–508, 13 (4).

Frank, I. E., Friedman, J. H., 1993. A statistical view of some chemometrics regression tools. Technometrics 35 (2), 109–135, 35 (2).

Freund, Y., Schapire, R. E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences 55 (1), 119–139, 55 (1).

Funahashi, K., 1989. On the approximate realization of continuous mappings by neural networks. Neural Networks 2 (3), 183–192, 2 (3).

Gabrielsson, J., Jonsson, H., Trygg, J., Airiau, C., Schmidt, B., Escott, R., 2006. Combining process and spectroscopic data to improve batch modeling. AICHE JOURNAL 52 (9), 3164, 52 (9).

Gabrys, B., 2002. Neuro-fuzzy approach to processing inputs with missing values in pattern recognition problems. International Journal of Approximate Reasoning 30 (3), 149–179, 30 (3).

Gabrys, B., 2004. Learning hybrid neuro-fuzzy classifier models from data: To combine or not to combine? Fuzzy Sets and Systems 147, 39–56, 147.

Gabrys, B., Bargiela, A., 1999. Neural networks based decision support in presence of uncertainties. Journal of Water Resources Planning and Management 125 (5), 272–280, 125 (5).

Gabrys, B., Ruta, D., 2006. Genetic algorithms in classifier fusion. Applied Soft Computing 6 (4), 337–347, 6 (4).

Gama, J., Medas, P., Castillo, G., Rodrigues, P., 2004. Learning with drift detection. In: Advances in Artificial Intelligence SBIA 2004: 17th Brazilian Symposium on Artificial Intelligence. Vol. 3171. p. 286295, p. 286295.

Geladi, P., Esbensen, K., 1991. Regression on multivariate images: principal component regression for modeling, prediction and visual diagnostic tools. J. Chemometr 5 (97), 111, 5 (97).

Gomez, E., Unbehauen, H., Kortmann, P., Peters, S., 1996. Fault detection and diagnosis with the help of fuzzy-logic and with application to a laboratory turbogenerator. Proc. of the 13th IFAC World Congress, volume N, 175–180Proc. of the 13th IFAC World Congress, volume N.

Gonzalez, G. D., 1999. Soft sensors for processing plants. Intelligent Processing and

Manufacturing of Materials, 1999. IPMM'99. Proceedings of the Second International Conference on 1, 1.

Gonzalez, G. D., Orchard, M., Cerda, J. L., Casali, A., Vallebuona, G., 2003. Local models for soft-sensors in a rougher flotation bank. Minerals Engineering 16 (5), 441–453, 16 (5).

Goodwin, G. C., 2000. Predicting the performance of soft sensors as a route to low cost automation. Annual Reviews in Control 24, 55–66, 24.

Gosset, W. S., 1908. The probable error of a mean. Biometrika 6 (1), 1–25, 6 (1).

Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. Journal of Machine Learning Research 3 (7-8), 1157–1182, 3 (7-8).

Han, C., Lee, Y. H., 2002. Intelligent integrated plant operation system for six sigma. Annual Reviews in Control 26 (1), 27–43, 26 (1).

Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, Springer.

He, Q. P., Qin, S. J., Wang, J., 2005. A new fault diagnosis method using fault directions in fisher discriminant analysis. AIChE Journal 51 (2), 555–571, 51 (2).

Hodge, V., Austin, J., 2004. A survey of outlier detection methodologies. Artificial Intelligence Review 22 (2), 85–126, 22 (2).

Hotelling, H., 1931. The generalization of student's ratio. The Annals of Mathematical Statistics 2 (3), 360–378, 2 (3).

Jackson, J. E., Mudholkar, G. S., 1979. Control procedures for residuals associated with principal component analysis. Technometrics 21 (3), 341–349, 21 (3).

James, S., Legge, R., Budman, H., 2002. Comparative study of black-box and hybrid estimation methods in fed-batch fermentation. Journal of Process Control 12 (1), 113–121, 12 (1).

Jang, J. S. R., Sun, C. T., Mizutani, E., 1997. Neuro-fuzzy and soft computing.

Prentice Hall Upper Saddle River, NJ, Prentice Hall Upper Saddle River, NJ.

Jiang, T., Chen, B., He, X., Stuart, P., 2003. Application of steady-state detection method based on wavelet transform. Computers and Chemical Engineering 27 (4), 569–578, 27 (4).

Jolliffe, I. T., 2002. Principal Component Analysis. Springer, Springer.

Jordaan, E., Kordon, A., Chiang, L., Smits, G., 2004. Robust inferential sensors based on ensemble of predictors generated by genetic programming. In: Proceedings of PPSN2004, Birmingham, UK. pp. 522–531, pp. 522–531.

Jos de Assis, A., Maciel Filho, R., 2000. Soft sensors development for on-line bioreactor state estimation. Computers and Chemical Engineering 24 (2), 1099–1103, 24 (2).

Kadlec, P., Gabrys, B., 2008a. Adaptive local learning soft sensor for inferential control support. In: CIMCA 2008. IEEE, Vienna, IEEE, Vienna.

Kadlec, P., Gabrys, B., 2008b. Learnt topology gating artificial neural network. In: 2008 IEEE International Joint Conference on Neural Networks (IJCNN2008). IEEE, Hong Kong, pp. 2605–2612, pp. 2605–2612.

Kalos, A., Kordon, A., Smits, G., Werkmeister, S., 2003. Hybrid model development methodology for industrial soft sensors. In: 2003 American Control Conference. pp. 5417–5422, pp. 5417–5422.

Kampjarvi, P., Sourander, M., Komulainen, T., Vatanski, N., Nikus, M., Jms-Jounela, S. L., 2008. Fault detection and isolation of an on-line analyzer for an ethylene cracking process. Control Engineering Practice 16 (1), 1–13, 16 (1).

Kittler, J., Hatef, M., Duin, R. P. W., Matas, J., 1998. On combining classifiers. Pattern Analysis and Machine Intelligence, IEEE Transactions on 20 (3), 226–239, 20 (3).

Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation

and model selection. Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence 2, 11371145, 2.

Kohonen, T., 1997. Self-organizing maps. Springer-Verlag New York, Inc. Secaucus, NJ, USA, Springer-Verlag New York, Inc. Secaucus, NJ, USA.

Kordon, A., Smits, G., Jordaan, E., Rightor, E., Co, D. C., Freeport, T. X., 2002. Robust soft sensors based on integration of genetic programming, analytical neural networks, and support vector machines. In: Evolutionary Computation, 2002. CEC'02. Proceedings of the 2002 Congress on. Vol. 1. Vol. 1.

Kordon, A. K., 2004. Hybrid intelligent systems for industrial data analysis. International Journal of Intelligent Systems 19 (4), 367–383, 19 (4).

Kordon, A. K., 2005. Application issues of industrial soft computing systems. Fuzzy Information Processing Society, 2005. NAFIPS 2005. Annual Meeting of the North American, 110–115Fuzzy Information Processing Society, 2005. NAFIPS 2005. Annual Meeting of the North American.

Kordon, A. K., Kalos, A. N., Castillo, F. A., Kotanchek, M. E., Jordaan, E. M., Smits, G. F., 2005. Competitive advantages of evolutionary computation for industrial applications. In: Evolutionary Computation, 2005. The 2005 IEEE Congress on. Vol. 1. Vol. 1.

Kourti, T., 2002. Process analysis and abnormal situation detection: from theory to practice. Control Systems Magazine, IEEE 22 (5), 10–25, 22 (5).

Kourti, T., Nomikos, P., MacGregor, J. F., 1995. Analysis, monitoring and fault diagnosis of batch processes using multiblock and multiway pls. JOURNAL OF PROCESS CONTROL 5, 277–277, 5.

Krogh, A., Vedelsby, J., 1995. Neural network ensembles, cross validation and active learning. Advances in Neural Information Processing Systems-7, 231238Advances in Neural Information Processing Systems-7.

Kuncheva, L. I., 2004. Combining Pattern Classifiers: Methods and Algorithms. Wiley-IEEE, Wiley-IEEE.

Lee, C., Choi, S. W., Lee, I.-B., 2004. Sensor fault identification based on time-lagged pca in dynamic processes. Chemometrics and Intelligent Laboratory Systems 70 (2), 165–178, 70 (2).

Lee, D. S., Vanrolleghem, P. A., 2003. Monitoring of a sequencing batch reactor using adaptive multiblock principal component analysis. Biotechnology and Bioengineering 82 (4), 489–497, 82 (4).

Lee, J. H., Dorsey, A. W., 2004. Monitoring of batch processes through state-space models. AIChE Journal 50 (6), 1198–1210, 50 (6).

Lee, J. M., Yoo, C. K., Lee, I. B., 2003. On-line batch process monitoring using a consecutively updated multiway principal component analysis model. Computers and Chemical Engineering 27 (12), 1903–1912, 27 (12).

Li, M. Z. Z., Li, W., 2005. Study on least squares support vector machines algorithm and its application. Proceedings of the 17th IEEE International Conference on Tools with Artificial Intelligence, 1082–3409Proceedings of the 17th IEEE International Conference on Tools with Artificial Intelligence.

Li, S. J., Zhang, X. J., Qian, F., 2005. Soft sensing modeling via artificial neural network based on pso-alopex. Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on 7, 7.

Li, W., Yue, H. H., Valle-Cervantes, S., Qin, S. J., 2000. Recursive pca for adaptive process monitoring. Journal of Process Control 10 (5), 471–486, 10 (5).

Lin, B., Recke, B., Knudsen, J., Jrgensen, S. B., 2007. A systematic approach for soft sensor development. Computers and Chemical Engineering 31 (5), 419–425, 31 (5).

Lin, C., Lee, C., 1996. Neural fuzzy systems: a neuro-fuzzy synergism to intelligent

systems. Prentice-Hall, Inc. Upper Saddle River, NJ, USA, Prentice-Hall, Inc. Upper Saddle River, NJ, USA.

Luo, J. X., Shao, H. H., 2006. Developing soft sensors using hybrid soft computing methodology: a neurofuzzy system based on rough set theory and genetic algorithms. Soft Computing-A Fusion of Foundations, Methodologies and Applications 10 (1), 54–60, 10 (1).

Macias, J. J., Zhou, P. X., 2006. A method for predicting quality of the crude oil distillation. In: Evolving Fuzzy Systems, 2006 International Symposium on. pp. 214–220, pp. 214–220.

Mandic, D. P., Chambers, J. A., 2001. Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability. Wiley, Wiley.

Marjanovic, O., Lennox, B., Sandoz, D., Smith, K., Crofts, M., 2006. Real-time monitoring of an industrial batch process. Computers and Chemical Engineering 30 (10-12), 1476–1481, 30 (10-12).

Meleiro, L. A. C., Finho, R. M., 2000. A self-tuning adaptive control applied to an industrial large scale ethanol production. Computers and Chemical Engineering 24 (2-7), 925–930, 24 (2-7).

Menold, P. H., Pearson, R. K., Allgower, F., 1999. Online outlier detection and removal. Proceedings of the 7th Mediterranean on Control and Automation (MED99), Haifa, Israel, 1110–1133Proceedings of the 7th Mediterranean on Control and Automation (MED99), Haifa, Israel.

Merikoski, S., Laurikkala, M., Koivisto, H., 2001. An adaptive neuro-fuzzy inference system as a soft sensor for viscosity in rubber mixing process. Tech. rep., WSEAS NNA-FSFS-EC 2001, Puerto de la Cruz, Tenerife, Spain, 11-15 February 2001, paper 446; http://www.ad.tut.fi, WSEAS NNA-FSFS-EC 2001, Puerto de la Cruz, Tenerife, Spain, 11-15 February 2001, paper 446; http://www.ad.tut.fi.

Nauck, D., Klawonn, F., Kruse, R., Klawonn, F., 1997. Foundations of Neuro-Fuzzy Systems. John Wiley & Sons, Inc. New York, NY, USA, John Wiley & Sons, Inc. New York, NY, USA.

Neogi, D., Schlags, C. E., 1998. Multivariate statistical analysis of an emulsion batch process. INDUSTRIAL AND ENGINEERING CHEMISTRY RESEARCH 37, 3971–3979, 37.

Nomikos, P., MacGregor, J. F., 1994. Monitoring batch processes using multiway principal component analysis. AIChE Journal 40 (8), 1361–1375, 40 (8).

Nomikos, P., MacGregor, J. F., 1995a. Multi-way partial least squares in monitoring batch processes. Chemometrics and Intelligent Laboratory Systems 30 (1), 97–108, 30 (1).

Nomikos, P., MacGregor, J. F., 1995b. Multivariate spc charts for monitoring batch processes. Technometrics 37 (1), 41–59, 37 (1).

Opitz, D., Maclin, R., 1999. Popular ensemble methods: An empirical study. Journal of Artificial Intelligence Research 11 (169-198), 12, 11 (169-198).

Park, S., Han, C., 2000. A nonlinear soft sensor based on multivariate smoothing procedure for quality estimation in distillation columns. Computers and Chemical Engineering 24 (2-7), 871–877, 24 (2-7).

Pearson, R. K., 2001. Exploring process data. Journal of Process Control 11 (2), 179–194, 11 (2).

Pearson, R. K., 2002. Outliers in process modeling and identification. Control Systems Technology, IEEE Transactions on 10 (1), 55–63, 10 (1).

Poggio, T., Girosi, F., 1990. Regularization algorithms for learning that are equivalent to multilayer networks. Science 247 (4945), 978–982, 247 (4945).

Prasad, V., Schley, M., Russo, L. P., Wayne Bequette, B., 2002. Product property and production rate control of styrene polymerization. Journal of Process Control

12 (3), 353–372, 12 (3).

Principe, J. C., Euliano, N. R., Lefebvre, W. C., 2000. Neural and adaptive systems. Wiley New York, Wiley New York.

Qin, S. J., 1997. Neural networks for intelligent sensors and control - practical issues and some solutions. Neural Systems for Control, 213234Neural Systems for Control.

Qin, S. J., 1998. Recursive pls algorithms for adaptive data modeling. Computers and Chemical Engineering 22 (4-5), 503–514, 22 (4-5).

Qin, S. J., McAvoy, T. J., 1992. Nonlinear pls modeling using neural networks. Computers & Chemical Engineering 16 (4), 379–391, 16 (4).

Qin, S. J., Yue, H., Dunia, R., 1997. Self-validating inferential sensors with application to air emission monitoring. Ind. Eng. Chem. Res 36, 1675–1685, 36.

Radhakrishnan, V. R., Mohamed, A. R., 2000. Neural networks for the identification and control of blast furnace hot metal quality. Journal of Process Control 10 (6), 509–524, 10 (6).

Rao, M., Corbin, J., Wang, Q., 1993. Soft sensors for quality prediction in batch chemical pulping processes. In: Intelligent Control, 1993., Proceedings of the 1993 IEEE International Symposium on. pp. 150–155, pp. 150–155.

Rnnar, S., MacGregor, J. F., Wold, S., 1998. Adaptive batch monitoring using hierarchical pca. Chemometrics and Intelligent Laboratory Systems 41 (1), 73–81, 41 (1).

Rong, H. J., Sundararajan, N., Huang, G. B., Saratchandran, P., 2006. Sequential adaptive fuzzy inference system (safis) for nonlinear system identification and prediction. Fuzzy Sets and Systems 157 (9), 1260–1275, 157 (9).

Rotem, Y., Wachs, A., Lewin, D. R., 2000. Ethylene compressor monitoring using model-based pca. AIChE Journal 46 (9), 1825–1836, 46 (9).

Ruta, D., Gabrys, B., 2000. An overview of classifier fusion methods. Computing and Information Systems 7 (1), 1–10, 7 (1).

Ruta, D., Gabrys, B., 2005. Classifier selection for majority voting. Information Fusion 6 (1), 63–81, 6 (1).

Schafer, J. L., Graham, J. W., 2002. Missing data: Our view of the state of the art. Psychological Methods 7 (2), 147–177, 7 (2).

Scheffer, J., 2002. Dealing with missing data. Research Letters in the Information and Mathematical Sciences 3 (1), 153–160, 3 (1).

Serneels, S., Verdonck, T., 2008. Principal component analysis for data containing outliers and missing elements. Computational Statistics and Data Analysis 52 (3), 1712–1727, 52 (3).

Smith, H., Fingar, P., 2003. Business process management: The third waveBusiness process management: The third wave.

Stanimirova, I., Daszykowski, M., Walczak, B., 2007. Dealing with missing values and outliers in principal component analysis. Talanta 72 (1), 172–178, 72 (1).

Su, H. B., Fan, L. T., Schlup, J. R., 1998. Monitoring the process of curing of epoxy/graphite fiber composites with a recurrent neural network as a soft sensor. Engineering Applications of Artificial Intelligence 11 (2), 293–306, 11 (2).

Tzanakou, E., Michalak, R., Harth, E., 1979. The alopex process: Visual receptive fields by response feedback. Biological Cybernetics 35 (3), 161–174, 35 (3).

Valentini, G., Masulli, F., 2002. Ensembles of learning machines. In: 13th Italian Workshop on Neural Nets. Vol. 2486 of Series Lecture Notes in Computer Sciences. Springer-Verlag, p. 322, p. 322.

van Sprang, E. N. M., Ramaker, H. J., Westerhuis, J. A., Gurden, S. P., Smilde, A. K., 2002. Critical evaluation of approaches for on-line batch process monitoring. Chemical Engineering Science 57 (18), 3979–3991, 57 (18).

Van sprang, E. N. M., Ramaker, H. J., Westerhuis, J. A., Smilde, A. K., Wienke, D., 2005. Statistical batch process monitoring using gray models. AIChE Journal 51 (3), 931–945, 51 (3).

Vapnik, V. N., 1998. Statistical learning theory. Wiley New York, Wiley New York.

Venkatasubramanian, V., Rengaswamy, R., Kavuri, S. N., 2003a. A review of process fault detection and diagnosis part ii: Qualitative models and search strategies. Computers and Chemical Engineering 27 (3), 313–326, 27 (3).

Venkatasubramanian, V., Rengaswamy, R., Kavuri, S. N., Yin, K., 2003b. A review of process fault detection and diagnosis part iii: Process history based methods. Computers and Chemical Engineering 27 (3), 327–346, 27 (3).

Venkatasubramanian, V., Rengaswamy, R., Yin, K., Kavuri, S. N., 2003c. A review of process fault detection and diagnosis part i: Quantitative model-based methods. Computers and Chemical Engineering 27 (3), 293–311, 27 (3).

Vilalta, R., Drissi, Y., 2002. A perspective view and survey of meta-learning. Artificial Intelligence Review 18 (2), 77–95, 18 (2).

Walczak, B., Massart, D. L., 1995. Robust principal components regression as a detection tool for outliers. Chemometrics and Intelligent Laboratory Systems 27 (1), 41–54, 27 (1).

Walczak, B., Massart, D. L., 2001a. Dealing with missing data part i. Chemometrics and Intelligent Laboratory Systems 58 (1), 15–27, 58 (1).

Walczak, B., Massart, D. L., 2001b. Dealing with missing data: Part ii. Chemometrics and Intelligent Laboratory Systems 58 (1), 29–42, 58 (1).

Wang, L., Shao, C., Wang, H., Wu, H., 2006. Radial basis function neural networks-based modeling of the membrane separation process: Hydrogen recovery from refinery gases. Journal of Natural Gas Chemistry 15 (3), 230–234, 15 (3).

Wang, S., Cui, J., 2005. Sensor-fault detection, diagnosis and estimation for centrifu-

gal chiller systems using principal-component analysis method. Applied Energy 82 (3), 197–213, 82 (3).

Wang, S., Xiao, F., 2004. Ahu sensor fault diagnosis using principal component analysis method. Energy & Buildings 36 (2), 147–160, 36 (2).

Wang, X., Kruger, U., Irwin, G. W., 2005. Process monitoring approach using fast moving window pca. Industrial & Engineering Chemistry Research 44 (15), 5691–5702, 44 (15).

Wang, Y., Rong, G., 1997. A self-organizing neural-network-based fuzzy system. Artificial Neural Networks, Fifth International Conference on (Conf. Publ. No. 440), 106–110Artificial Neural Networks, Fifth International Conference on (Conf. Publ. No. 440).

Warne, K., Prasad, G., Rezvani, S., Maguire, L., 2004a. Statistical and computational intelligence techniques for inferential model development: a comparative evaluation and a novel proposition for fusion. Engineering Applications of Artificial Intelligence 17 (8), 871–885, 17 (8).

Warne, K., Prasad, G., Siddique, N. H., Maguire, L. P., 2004b. Development of a hybrid pca-anfis measurement system for monitoring product quality in the coating industry. In: Systems, Man and Cybernetics, 2004 IEEE International Conference on. Vol. 4. Vol. 4.

Weiss, S., Kulikowski, C., 1991. Computer systems that learn: classification and prediction methods from statistics, neural nets, machine learning, and expert systems. Morgan Kaufmann Publishers Inc. San Francisco, CA, USA, Morgan Kaufmann Publishers Inc. San Francisco, CA, USA.

Welch, G., Bishop, G., 2001. An introduction to the kalman filter. An introduction to the kalman filter.

Widmer, G., Kubat, M., 1996. Learning in the presence of concept drift and hidden

contexts. Machine Learning 23 (1), 69–101, 23 (1).

Wold, S., Geladi, P., Esbensen, K., Ohman, J., 1987. Multi-way principal components and pls analysis. Journal of Chemometrics 1 (1), 41–56, 1 (1).

Wold, S., Kettaneh, N., Fridn, H., Holmberg, A., 1998. Modelling and diagnostics of batch processes and analogous kinetic experiments. Chemometrics and Intelligent Laboratory Systems 44 (1-2), 331–340, 44 (1-2).

Wold, S., Sjstrm, M., Eriksson, L., 2001. Pls-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems 58 (2), 109–130, 58 (2).

Wolpert, D. H., 1992. Stacked generalization. Neural Networks 5 (2), 241–259, 5 (2).

Yan, W., Shao, H., Wang, X., 2004. Soft sensing modeling based on support vector machine and bayesian model selection. Computers and Chemical Engineering 28 (8), 1489–1498, 28 (8).

Yang, S. H., Chen, B. H., Wang, X. Z., 2000. Neural network based fault diagnosis using unmeasurable inputs. Engineering Applications of Artificial Intelligence 13 (3), 345–356, 13 (3).

Yang, Y., Chai, T., 1997. Soft sensing based on artificial neural network. In: American Control Conference, 1997. Proceedings of the 1997. Vol. 1. Vol. 1.

Zadeh, L. A., 1996. Fuzzy sets. World Scientific Series In Advances In Fuzzy Systems, 19–34World Scientific Series In Advances In Fuzzy Systems.

Zamprogna, E., Barolo, M., Seborg, D. E., 2004a. Development of a soft sensor for a batch distillation column using linear and nonlinear pls regression techniques. Control Engineering Practice 12, 917–929, 12.

Zamprogna, E., Barolo, M., Seborg, D. E., 2004b. Estimating product composition profiles in batch distillation via partial least squares regression. Control Engineering Practice 12 (7), 917–929, 12 (7).

Zhang, H., Lennox, B., 2004. Integrated condition monitoring and control of fed-

batch fermentation processes. Journal of Process Control 14 (1), 41–50, 14 (1).

Zhao, L., Chai, T., 2004. Adaptive moving window mpca for online batch monitoring. In: Control Conference, 2004. 5th Asian. Vol. 2. Vol. 2.

Publication	Applied method(s)	Applic. type	Process description	Process type
Casali et al. (1998)	SRM (ARMAX)	OP	particle size estimation in a grinding plant	Cont.
Park and Han (2000)	PCA/PLS+LWR	OP	toluene composition in a splitter column, diesel temperature in crude oil column	Cont.
Kadlec and Gabrys (2008a)	MLR ensemble	OP	industrial drier	Cont.
Devogelaere et al. (2002)	MLP	OP	sugar quality estimation	Cont.
Jos de Assis and Maciel Filho (2000)	MLP, FPM, eKF	OP	biomass estimation in a fermentation process	Batch
Qin (1997)	MLP, NNPLS	OP	refinery	Batch
Meleiro and Finho (2000)	MLP	OP	control loop support of ethanol production support	Batch
Park and Han (2000)	MLP+Expert System	OP	silica content control in the steel production	Cont.
Fortuna et al. (2005)	MLP	OP	C4 and C5 concentration prediction in a debutanizer refinery process	Cont.
Desai et al. (2006)	MLP, RBFN, SVR	OP	two simulated biochemical processes	Batch
Wang et al. (2006)	RBFN	OP	membrane separation process modelling	Cont.
Kadlec and Gabrys (2008b)	MLPs ensemble	OP	industrial drier	Cont.
James et al. (2002)	MLP, RBFN, Hybrid (MLP/RBFN+FPM)	OP	biomass concentration prediction	Batch
Su et al. (1998)	RNN+FPM	OP	degree-of-cure prediction in epoxy/graphite fiber composites process	Cont.
Chen et al. (2004a)	RNN	OP	biomass concentration prediction	Batch
Chen et al. (2004b)	RNN	OP	melt-flow-length prediction in injection molding process	Cont.
Yang and Chai (1997)	RNN	OP	three simple simulated processes	Cont.
Fellner et al. (2003)	generalised ANN	OP	diacetyl concentration prediction	Batch
Lin et al. (2007)	PCA	OP	product estimation in cement kiln, NO_X monitoring	Cont.
Qin et al. (1997)	PCA	OP, SFD	air emission monitoring	Cont.
Zamprogna et al. (2004b)	PLS, PCA	OP	simulated distillation column	Batch
Dayal and MacGregor (1997)	EWPLS	OP	stirred reactor, flotation circuit	Cont.
Qin (1998)	RPLS	OP	research octane number prediction in a refinery process	Cont.

Feng et al. (2003)	LS-SVM	OP	gasoline absorbing rate in FCC	Cont.
Yan et al. (2004)	LS-SVM	OP	light diesel freezing point detection in FCC	Cont.
Merikoski et al. (2001)	ANFIS	OP	rubber viscosity estimation	Cont.
Warne et al. (2004a)	PCA+ANFIS	OP	polymeric-coated substrate anchorage	Cont.
Luo and Shao (2006)	NFS + GA	OP	light diesel freezing point detection in FCC	Cont.
Arazo-Bravo et al. (2004)	NFS	OP	penicillin production bioprocess	Batch
Wang and Rong (1997)	NFS	OP	propylene purity prediction in a distillation column	Cont.
Macias and Zhou (2006)	evolving NFS	OP	crude oil distillation in refinery process	Cont.
Dong et al. (1995)	NLPCA+NNPLS	OP	NO_x prediction in exhaust gas	Cont.
Li et al. (2005)	PSO+MLP	OP	ethylene distillation column	Cont.
Kalos et al. (2003)	analytic NN+SVM+GP	OP	interface level estimation in a neutralization unit	Cont.
Chen et al. (2000)	FPM+RBFN	OP	microbial population in a bioreactor	Batch
Rao et al. (1993)	Intelligent Soft Sensor	OP	sulphite pulping system	Batch
Gonzalez et al. (2003)	SRM, TS, FC, PLS, WBM, ANN	OP	copper concentrate grade in a rougher flotation bank process	Cont.
Li et al. (2000)	RPCA	PM	rapid thermal annealing process	Batch
Nomikos and MacGregor (1995b)	PCA	PM	polymerisation process	Batch
Rotem et al. (2000)	MBPCA	PFD	ethylene compressor	Batch
Wang et al. (2005)	FMWPCA	PM	simulated FCC unit process	Cont.
Amazouz and Pantea (2006)	PCA+PLS	PM	Lumber drying	Batch
Zhang and Lennox (2004)	PLS	OP, PM	simulated penicillin production process	Batch
He et al. (2005)	FDA	PM	quadruple tank process; polyester film manufacturing process	Cont.
Alhoniemi (1999)	SOM	PM, OP	cont. pulp digester; steel production process; pulp and paper industry	Cont.
Yang et al. (2000)	FPM+ANN	PFD	FCC reactor	Cont.
Kampjarvi et al. (2008)	PCA, SOM, RBFN	PM, PFD	ethylene cracking process	Cont.

Marjanovic et al. (2006)	MPLS	PM	process end point detection	Batch
Dunia and Qin (1998a)	PCA	PFD, SFD	boiler process	Cont.
Lee et al. (2004)	TLPCA	PFD, SFD	polymerisation process	Batch
Wang and Cui (2005)	PCA	SFD	centrifugal chiller process	Cont.
Wang and Xiao (2004)	PCA	SFD+PM	air handling unit	Cont.

Table 1

List of the presented Soft Sensor publications (for the list of abbreviations see Table 2)

Abbreviation	Explanation
Methods:	
AnaNN	Analytical Neural Network
ANN	Artificial Neural Networks
eKF	enhanced Kalman Filter
EWPLS	Exponentially Weighted Partial Least Squares
FC	Fuzzy Combinational
FMWPCA	Fast Moving Window Principle Component Analysis
FPM	First Principle Model
GP	Genetic Programming
LWR	Locally Weighted Regression
MBPCA	Model-Based Principle Component Analysis
MLP	Multi-Layer Perceptron
MPLS	Multi-way Partial Least squares
MLR	Multiple linear regression
NFS	Neuro-Fuzzy System
NLPCA	Non-Linear Principle Component Analysis
NNPLS	Neural Network Partial Least Squares
PCA	Principle Component Analysis
PCR	Principle Component Regression
PLS	Partial Least Squares
PLSR	Partial Least Squares Regression
PSO	Particle Swarm Optimization
RBFN	Radial Basis Function Network
RNN	Recurrent Neural Network
RPCA	Recursive PCA
SOM	Self-Organizing Network
SRM	Stepwise Regression Method
SVM	Support Vector Machines
SVR	Support Vector Regression
TLPCA	time lagged Principle Component Analysis
TS	Takagi and Sugeno model
WBM	Wavelet-Based Model
Application Types:	
OP	On-line Prediction
PFD	Process Fault Detection
PM	Process Monitoring
SFD	Sensor Fault Detection

Table 2

List of abbreviations