Adaptive PLS inferential soft sensor for continuous online estimation of NOx emission in industrial water-tube boiler

S H Hasnen1*, H Zabiri1,2, K K Prakash2 and T T Mat2

1 Chemical Engineering Department, Universiti Teknologi PETRONAS, Seri Iskandar, Perak, Malaysia.
2 Group Technical Solutions (GTS), PETRONAS, Kuala Lumpur, Malaysia.

*E-mail: haslindazabiri@utp.edu.my

Abstract. In common industrial application, the use of a linear and static PLS soft sensor for online prediction and monitoring of industrial boiler is often preferred due to its simple and intuitive framework. However, process dynamics and time-variant factors can negatively affect the accuracy and reliability of PLS soft sensor over its long-term application in process industries. In this paper, development of adaptive soft sensor based on dynamic PLS method has been applied to an industrial water-tube boiler for continuous online prediction of Nitric Oxides emission. In the case study, it is found that the adaptive PLS soft sensor which includes lagged measurements of NOx emission in the model input can significantly improve the prediction accuracy and reliability by 72.7% relative to the performance of linear and static PLS soft sensor when tested on online dataset containing gradual and abrupt changes in the process operating conditions.

1. Introduction

Chemical process industries are faced with pressing responsibilities of devising their production policies to adhere accordingly with safety rules and environmental regulations. For process involving combustion, it has been made compulsory for all parties involved to monitor flue gases in the exhaust streams due to stringent air regulations [1,2]. In general practice, the industries rely mostly on hardware analysers, like Continuous Emission Monitoring System (CEMS), to be installed at every air emission release point to give measurement of the exhaust gas composition. Despite so, the hardware systems are complex and costly to install and maintain [3,4]. On the other hand, Predictive Emission Monitoring (PEMS) or soft sensors can offer a few benefits in place of hardware sensors in the industrial applications [5].

Soft sensors are inferential model that use information from easily measured online variables, such as temperature, pressure, flowrate and level, to estimate a target variable(s) that is significantly harder to make continuous measurement in online environment, such as feed or product’s composition or concentration. The target variable is often informative of the state of production quality and is therefore used as a reference for estimating and monitoring of quality of an operational production and environmental regulation. These soft sensors application field is further on referred to as on-line prediction [4,6]. The data-driven algorithm applied for soft sensor development are either based on statistical methods or machine learning approaches for linear and non-linear processes. The linear modelling methods like Multivariate Linear Regression (MLR), Principal Component Analysis (PCA) and Partial Least Squares (PLS) are mostly applied in the industries, i.e.- they account for more than 90% of the soft sensors used in chemical plants [5].
Despite the simplicity to develop a linear and static soft sensor model based on the PLS method, its accuracy and validity usually suffer declination when changes happen in the process behaviour. This is so because process dynamics and other time-variant factors have caused the process operating conditions to deviate from the ranges that the trained soft sensor have been initially modelled to [6]. To update a linear and static PLS model with the new process behaviour, it is important to appropriately adapt the model with updating methods that are suitable with the PLS model structure. There are a significant number of adaptive soft sensors that are based on PLS modelling or its extension to dynamic and non-linear form, such as dynamic PLS, kernel PLS, neural network PLS, moving-window PLS, recursive PLS, and etc [7].

Furthermore, the ensemble-model based adaptive methods such as local learning, state-partitioning and Just-in-Time(JIT) have paved ways for PLS to do model regression for each local region defined in the process data, which is found to be more efficient in predicting abrupt changes in the process data, compared to adaptation based on global model structure [8-12]. For instance, a recently published work on highly-overlapped, recursive PLS soft sensor with state partitioning via local variable selection adaptation methods have been applied to industrial dataset of a thermal oxidizer unit for online prediction of NOx emission [2].

In this paper, an industrial case study on the development of adaptive PLS soft sensor based on dynamic PLS method for continuous online prediction of NOx emission in the industrial water-tube boiler derived from actual industrial data of the system is briefly discussed. There are three key areas that have been highlighted in the development of the adaptive PLS soft sensor: - suitable selection of model inputs from available industrial process variables based on process knowledge related to the combustion process and NOx formation mechanisms in the industrial water-tube boiler, PLS model training based on normal operating region of the industrial water-tube boiler system, and consequent adaptation of PLS model with online industrial dataset based on dynamic PLS updating approaches in response to gradual and abrupt changes in the operating regions of the industrial water-tube boiler system.

2. Background

2.1 Partial Least Squares (PLS)

The inferential model is based on PLS model structure, which is essentially a mathematical equation based on the relationships between independent key process variables (X) in the boiler system that are related to the NOx emission (Y) in the boiler and it is derived from real measurements made in the boiler system. The detailed theory and derivation of PLS outer and inner regression models can be found in several influential works by Wold and Geladi [13].

\[ Y = Y(\text{mean}) + X.W.C + F \]  

(1)

The PLS basic regression model represented in equation (1) above is the inferential model for prediction of NOx (Y) in which Y(mean) is the average value of NOx emission, W*C are the loading weights or coefficients for respective key process parameters (X), and F is the residual between predicted and actual NOx value from the regression. PLS is a supervised regression method that uses Y variable to look for correlations between systematic variations in X and variations in Y in the process data.

Among the linear regression methods, PLS has rapidly increased in popularity due to several practical reasons. First, the PLS method is better at handling the commonly existed data co-linearity and measurement noise in industrial dataset than the MLR method [7]. Secondly, while the PCA method can only transform the dataset into several latent variables based on only correlations among the process variables, the PLS method can derived latent projections from the dataset transformation as such that the covariation between the process variables to the target variable is maximized [10-11]. This can make the prediction of target variable from the process variables become more accurate. Finally, PLS model has a
transparent model structure that can be easily understood by engineers and process operators when performing data analysis and soft sensor development [3,11].

2.2 NOx Emission Formation
The boiler produced harmful emissions from the combustion process, especially NOx and CO gases, of which formations depend mainly on reaction rate in the boiler’s combustion chamber. In the case of gas-fired combustion application, the dominant mechanism is mostly attributed to the thermal NOx route, as the range can account from 80 to 95% of the NOx emission [1]. NOX emission is correlated to the actual flame temperature in the combustion zone and several process parameters such as fuel flow, air flow, combustion air temperature, pressure, humidity and more. Hence a data-driven soft sensor can be developed based on these process parameters for the online prediction of NOX emission.

2.3 Case Study : Industrial Water-tube Boiler
The studied industrial boiler is a water-tube type boiler, fired by natural gas. The objective of the boiler is to generate high pressure (HP) steam for plant-wide utilization. The boiler by design can generate maximum 83 tons per hour of steam at 107.9 barg and 515 °C. Fuel gas feed input to the boiler mainly consists of methane. The main parts of the boiler consist of the upper and lower drums, combustion chamber and superheaters. Figure 1 illustrates the schematic diagram of the industrial boiler used in the case study.

Figure 1. Schematic Diagram of Industrial Process Boiler.

3. Methodology
The procedure for soft sensor development based on PLS for prediction of NOx concentration in the boiler stack is illustrated in Figure 2. In summary, the development of the inferential soft sensor model begins with offline data pre-processing and variable selection. Then it is continued with offline PLS model training and validation analysis development before it is ready for online prediction and online
model adaptation studies. All the research works were conducted in the SIMCA-p data analysis tools and the MATLAB R2013a commercial software.

![Figure 2. Soft Sensor methodology flowchart.](image)

The model development begins with data collection of samples of available industrial data of the water-tube boiler, which is a set of data consisted of 27 process variables (26 X and 1 Y). Process measurements in the data set were recorded within a duration of 2 weeks and taken at every one-minute sampling threshold. Next, the collected industrial data need to be pre-processed through a series of statistical analysis to filter out missing values and critical outliers. The total numbers of samples after data pre-processing are around 13000, and they needed to be normalized and scaled to zero mean and one-unit variance, before these data were ready to be used in PLS modelling. For data selection, the offline dataset used for model training formed 70% of the whole dataset, while the other 30% of the pre-processed dataset is dedicated for online prediction testing and model adaptation.

Normally the PLS regression performance increase with the amount of input variables used in model training as PLS is relatively capable of handling collinearity and noises such as missing values and incomplete variables. However, incorporating high number of model inputs can implicate the process of model maintenance. Therefore, the theoretical causes for emergence of NOx in the boiler system, as discussed in Section 2.2, are used as the main criteria for the variable selections. Consequently, the key process variables (X) were finalized and then used as the model inputs to predict Y in PLS model training.

The execution of PLS outer and inner regression calculation on a pair of data matrices \{(X, Y)\} which has \(m\) input variables, \(p\) output variables and \(n\) samples produce results of the corresponding scores, loadings and regression coefficient matrices which can be represented based on equation (2) as follow:

\[
(X, Y)_{\text{OFFLINE}} \xrightarrow{\text{PLS}} (T, W, P, B, Q)_{\text{OFFLINE}}
\]  

The static and linear PLS model’s accuracy is prone to deterioration after online implementation due to changes in the process operating conditions. In addition, development of inferential model with static data does not give good representation of the studied chemical process due to dynamic characteristics generally observed in chemical processes. For instance, the future state of Y can be due to cumulating effects of the current and previous operating conditions in the process. To update the offline PLS model with new process behaviours, it is beneficial to use adaptive methods that are suitable with the PLS model structure.

When considering both process dynamics that are represented in the online dataset and a suitable adaptive method for PLS-based model, the selected input variables for model training as well as the previous measurements of NOx emission can be used as new model inputs (Z) for PLS transformation to its dynamic version. The PLS model is converted into a DPLS model by adding lag of X and Y
variables together with X variables as the new model inputs (Z), notated in equation (3) to train the DPLS model with offline dataset as in equation (4), and consequently validate its adaptability on online dataset as follows:

\[ Z = [X_1, X_1(n-1), X_2, X_2(n-1), \ldots, Y(n-1)] \]  

(3)

\[ (Z,Y)_{\text{OFFLINE}} \xrightarrow{DPLS} (T, W, P, B, Q)_{\text{OFFLINE}} \]  

(4)

In actual practice, data-driven inferential model can only be updated when a validated or actual measurement of the boiler emission by hardware analyser such as gas chromatography become available. If the actual measurement of NOx emission is available after a sampling threshold of 10 minutes, the PLS model can be extended to dynamic PLS model by including the dataset of X and Y with the lag period of 10 minutes, for example X(n-1) until X(n-10) and Y(n-10), to the initial PLS model inputs accordingly.

4. Results and Discussion

4.1 Variable selection based on process knowledge

From all of 26 process variables that are measured around the industrial water-tube boiler system in Figure 1, there are only 5 of them that are directly associated with the combustion reaction, particularly the NOx formation in the boiler through the thermal NOx route. The list of the key predictors is given in table 1. Data set of the finalized key variables are then gathered together to form a set of model inputs (X) for the development of PLS model. The variable selection analysis helped in making the optimal selection of the model inputs for future works in PLS model development.

**Table 1.** List of key variables based on combustion theory

| Important Variables                          |
|---------------------------------------------|
| x1- Combustion air pressure                 |
| x2- Flue gas Economizer Inlet temperature   |
| x3- Combustion air flowrate source 1        |
| x4- Fuel gas flowrate                       |
| x5- Fuel Latent Heat Value                  |

4.2 Comparison between PLS model and DPLS model on their adaptive capability for online prediction.

Following the variable selection, the soft sensor model development continues with offline model training of the PLS and DPLS model from offline data, which is consisted of process states and corresponding NOx emission measurements that are within the normal range of the boiler operating conditions. Next, to be able to see capability of the offline PLS and DPLS models to adapt to the changes in the process, the online testing data were chosen such that it contained several process states that were operated outside of the normal range.

The prediction performance of the developed model is calculated based on the Root Mean Square Error (RMSE) statistic parameter to find error between predicted Y and actual Y values. Furthermore, visual interpretation method is also used to help with analysis of the online prediction of NOx emission in the water-tube boiler.
Table 2. RMSE for PLS and DPLS models on online dataset

| Model | RMSE   |
|-------|--------|
| PLS   | 0.8882 |
| DPLS  | 0.2480 |

Two sets of inferential soft sensor models based on PLS and DPLS methods have been developed from the offline dataset and tested on the online dataset to compare both of their accuracy and adaptive capability after significant changes have been made to the industrial water-tube boiler operating conditions. In Table 2, the differences between RMSE values calculated from the prediction error made by PLS model and DPLS model relative to the actual NOx emission measurements in the online dataset indicate that the adaptive DPLS model has significantly better prediction accuracy compared to the non-adaptive PLS model by 72.7%.

Furthermore, the visual interpretation of prediction performance of PLS and DPLS models on the online test data as shown in Figure 3 and Figure 4 respectively also reflects that the adaptive DPLS model has very good tracking ability in response to gradual process change (from time 1:2500) and abrupt process change (from time 2501:5000) that can be observed in the actual NOx emission levels in the online dataset. Based on these two criteria to evaluate adaptive capability of the inferential models, the results have proved that DPLS model is significantly more accurate in predicting NOx emission and better at tracking changes in the boiler operating changes for long-term implementation in process industry.

![Figure 3. Comparisons between actual NOx trend (blue line) and prediction of NOx emissions (green line) made by the PLS model on the online test data.](image-url)
Figure 4. Comparisons between actual NOx trend (blue line) and prediction of NOx emissions (purple line) made by the adaptive Dynamic PLS model on the online test data.

The significant improvement in the prediction and adaptive performance of the DPLS model can be associated to several reasons. First, the information of the previous Y variables helps DPLS to also capture serial correlations between lagged Y(n-10) and the current Y(n) observed in the training dataset, besides the spatial correlations between X(n) and Y(n) for prediction of future Y(n+1). In addition, statistical analysis such as corresponding regression coefficient and variable importance projection (VIP) of each variable in the DPLS model inputs can be used to determine the most important process variable that can explain majority of the variation observed in NOx emission level.

Upon further inspection of the model regression coefficients and the correlation value, it could be found that lagged of NOx emission measurements significantly dominates other process variables in the DPLS model inputs. Hence, it could also be suggested here that the lagged of NOx emission measurement is responsible for the superior prediction and highly adaptive capability of DPLS model for prediction of NOx emission in the industrial water-tube case study.

5. Conclusion
Inferential model which is the main body of a soft sensor for continuous online prediction of NOx emission has been developed based on actual industrial data of a gas-fired, water-tube boiler. The model inputs for the inferential model development have been selected based on theoretical knowledge related to combustion reaction and thermal NOx formation. Based on the results from the online testing made on the developed inferential models, the PLS model seems not able to properly adapt to the changes in the process operating conditions. On the other hand, the DPLS model is found to be significantly more superior than the PLS model in terms of its prediction performance and adaptive capability for continuous online estimation of NOx emission in the industrial boiler by 72.7%. It has also been suggested that the lagged of NOx emission measurement Y(n-10) could be responsible for the significant advantage in prediction accuracy and highly adaptive capability of DPLS model for continuous prediction of NOx emission in the industrial water-tube case study.

Acknowledgements
The authors would like to thank Group Technical Solutions (GTS) PETRONAS and PETRONAS University of Technology for the industrial data and facilities provided in carrying out this research.
References

[1] Nielsen JS and Weel M 2015 First Principle Models for Emission and Performance Monitoring. Abu Dhabi International Petroleum Exhibition and Conference: Society of Petroleum Engineers.

[2] Poerio DV and Brown SD 2018 Highly-overlapped, recursive partial least squares soft sensor with state partitioning via local variable selection. Chemometrics and Intelligent Laboratory Systems 175 104-15.

[3] Kano M and Fujiwara K 2012 Virtual sensing technology in process industries: trends and challenges revealed by recent industrial applications. Journal of chemical engineering of Japan. 12 167

[4] Kadlec P, Gabrys B and Strandt S 2009 Data-driven soft sensors in the process industry Computers & Chemical Engineering 33(4) 795-814

[5] Souza FA, Araujo R and Mendes J 2016 Review of soft sensor methods for regression applications Chemometrics and Intelligent Laboratory Systems 152 69-79

[6] Kadlec P, Grbić R and Gabrys B 2011 Review of adaptation mechanisms for data-driven soft sensors Computers & Chemical Engineering 35(1) 1-24

[7] Qin SJ 2012 Survey on data-driven industrial process monitoring and diagnosis Annual Reviews in Control 36(2) 220-34

[8] Shao W and Tian X 2015 Adaptive soft sensor for quality prediction of chemical processes based on selective ensemble of local partial least squares models Chemical Engineering Research and Design 95 113-32

[9] Shao W and Tian X 2017 Semi-supervised selective ensemble learning based on distance to model for nonlinear soft sensor development Neurocomputing 222 91-104

[10] Ge Z and Song Z 2010 A comparative study of just-in-time-learning based methods for online soft sensor modeling Chemometrics and Intelligent Laboratory Systems 104(2) 306-17

[11] Kim S, Okajima R, Kano M and Hasebe S 2013 Development of soft-sensor using locally weighted PLS with adaptive similarity measure Chemometrics and Intelligent Laboratory Systems 124 43-9

[12] Ni W, Brown SD and Man R 2014 A localized adaptive soft sensor for dynamic system modeling Chemical Engineering Science 111 350-63

[13] Wold S, Sjöström M and Eriksson L 2011 PLS-regression: A basic tool of chemometrics Chemometrics and Intelligent Laboratory Systems 58(2) 109-30