Accepted manuscript

Korpela, T., Kumpulainen, P., Majanne, Y., Häyrinen, A., Lautala, P. 2017. Indirect NO_x emission monitoring in natural gas fired boilers. *Control Engineering Practice*. 65, 11–25. DOI: 10.1016/j.conengprac.2017.04.013.

http://dx.doi.org/10.1016/j.conengprac.2017.04.013

Indirect NO_x Emission Monitoring in Natural Gas Fired Boilers

Timo Korpela[1*], Pekka Kumpulainen[1], Yrjö Majanne[1], Anna Häyrinen[2], Pentti Lautala[1]

[1]Tampere University of Technology, Automation and Hydraulic Engineering
P.O. BOX 692, 33101 Tampere, Finland. *Corresponding author, +358 40 849 0063.
timo.korpela@tut.fi, pekka.kumpulainen@tut.fi, yrjo.majanne@tut.fi, pentti.lautala@tut.fi

[2]Helen Ltd, 00090 Helen, anna.hayrinen@helen.fi

Abstract

New emission regulations will increase the need for inexpensive NO_x emission monitoring solutions also in smaller power plants. The objective in this study is to find easily maintainable and transparent but still valid models to predict NO_x emissions in natural gas fired hot water boilers utilizing existing process instrumentation. With a focus on long-term applicability in practical installations, the performance of linear regression is compared in two municipal 43 MW boilers with three widely used nonlinear methods: multilayer perceptron, support vector regression, and fuzzy inference system. The linear models were the most applicable providing the best estimation results (relative error of <3 % in all cases), generalizability and simplicity. Therefore, the approach fulfils the requirements of the Industrial Emission Directive and is valid to be applied as a soft sensor in PEMS[1] applications in practise. However, each boiler model should be identified individually.

Keywords: Monitoring, estimation, modeling, soft sensor, NO_x, natural gas, combustion

Highlights

- Trial runs with two fairly similar natural gas fired municipal boilers were conducted
- Linear multivariate regression models outperform the applied nonlinear models
- Linear models provide NO_x estimates with relative error of <3 % in all cases
- The models have long term applicability in PEMS applications
- Every boiler model must be identified and maintained individually

1. Introduction

There is an increasing demand to protect the environment from harmful emissions. One of the main sources of air pollution are combustion processes, which emit sulphur dioxide (SO_2), particulate matter, carbon monoxide (CO) and dioxide (CO_2), unburned hydrocarbons (C_XH_Y), and nitrogen oxides (NO_x). NO_x emissions are considered as primary pollutants, since they can cause health issues in addition to environmental problems, e.g. photochemical smog, acid rain, tropospheric ozone and ozone layer depletion. (Skalska *et al.*, 2010)

As energy production is a major source of emissions, authorities have set further tightening emission limits to power plants. In the European Union (EU), flue gas emissions, i.e. SO_2, NO_x and dust, are restricted by the Industrial Emission Directive (IED), which came into effect in 2016. It is applicable in all existing power plants exceeding 50 MW. Additionally, the combination of plants where flue gases from multiple individual plants with rated thermal input over 15 MW could be discharged through a common stack are considered as a single combustion plant and their capacities are combined when calculating the total thermal input. This aggregation rule extended the scope of the IED to combustion plants with rated thermal power between 15–50 MW. Especially in Nordic countries, these <50 MW boilers typically have low operation hours. They are typically remotely operated and built for peak load and reserve capacity generation in district heating networks. Their role, however, might change in the future, as district heating networks and their components become more intensively interlinked with renewable energy generation systems.

According to the IED, the concentrations of SO_2, NO_x and dust in flue gases must be measured continuously in all combustion plants exceeding total thermal capacity of 100 MW. Otherwise these emissions (and, CO for gas fired plants) must be measured periodically at least once every 6 months. However, the IED provides an alternative to discontinuous SO_2 and NO_x measurements through other procedures if they are verified and approved by a competent authority. Such procedures must rely on CEN or other international standards, which ensure the provision of data of scientific quality.

[1] Predictive Emission Monitoring System

There are three options to monitor flue gas NO_x emissions from a combustion unit, i.e. by periodic measurements, CEMS (Continuous Emission Monitoring System) or PEMS (Predictive Emission Monitoring System). Periodic measurements are typically performed with calibrated equipment and conducted by an emission-testing laboratory with moderate costs. However, it is not guaranteed that the emission levels are valid in actual operation between the campaigns. As illustrated e.g. by (Korpela *et al.*, 2016b), heat only boilers may increasingly be operated with fast power transients to stabilize the district heating networks when nearby combined heat and power (CHP) plants contribute to balancing of the electric power system. In these cases, the periodic measurements are not able to determine the actual NO_x emissions (although this is not yet required in IED). In CEMS, the emission monitoring equipment is installed on-site and is active during plant operation. CEMS provides online and actual information on emissions, when maintenance and calibrations are conducted according to standard procedures. However, purchase and maintenance costs of CEMS are relatively high, especially compared to their benefits in boilers that have low power outputs and few operation hours. PEMS combines the positive properties of the former methods, as PEMS is a software based monitoring system, which utilizes existing process measurements and calculation models to estimate power plant emissions. There is no accepted CEN standard for PEMS at present but preparation is under way by CEN/TC 264/WG 37.

Indirect measurements, such as those calculated in PEMS, require a model to be driven with the online measurements. The model can be based on first principles and the physical properties of the boiler, measurement data correlations, or the combination of them. Derivation of an accurate first principle model for NO_x emissions is a complex task due to several factors, for example, complex nitrogen reactions. Hence, data based models are dominating in NO_x emission estimation. The purpose of NO_x emission estimation is primarily to study and predict the emissions in different operating conditions from process perspective, to derive indirect NO_x emission monitoring solutions to provide analytical redundancy to online measurements, or to estimate the emissions when no NO_x measurement is available. Numerous data based NO_x emission models have been presented in literature for various combustion applications (e.g. pulverized coal boilers, fluidised bed boilers, grate boilers, gas boilers, gas turbines, engines) in various power ranges and fuels (e.g. coal, biomass, waste, oil, and gas). These aspects significantly affect NO_x emissions and therefore also NO_x emission estimation. These NO_x models typically utilize existing process measurements. The amount, quality and type, and the maintenance procedures of the measurements that can be utilized in NO_x estimation applications are dependent directly or via environmental permits on the power rating of the plant, its location and the operation purpose (e.g. hot water or steam only, power generation). Therefore, the variety of applied modelling methods in data based NO_x estimation is vast, covering linear multivariate regression and nonlinear multivariate methods, such as radial basis function (RBF), multilayer perceptron (MLP), partial least squares (PLS), least squares support vector machine (LSSVM), Fuzzy Inference Systems (FIS), Kohonen's self-organising maps (SOM), and so on. The NO_x models are mainly static, but dynamic models also exist when the dynamics have a significant role in NO_x formation.

Data based models are, in general, only valid in the operating conditions which exist in the identification data used in their training. Therefore, the collection of the training data should be carefully considered. The data for model derivation can be generated by measuring normal operation of the combustion plant, by conducting separate trial runs to stimulate the processes, or by simulating the process with some other models. The first approach typically requires input data collection over a long period; however, the data may then not cover all potential operating conditions. Separate trial runs may improve this situation significantly, but the amount of data collected with this approach is usually much smaller than in the case of online process data collection. In the third approach, the input data to the models may be derived from complex combustion models, e.g. Computational Fluid Dynamic (CFD) models, but this approach is hardly a general solution in PEMS applications. In this study, separate and comprehensive trial runs were conducted to enable a reliable identification of the models. The models were validated by data collected from normal process operation but with increased excitement of the processes by frequent set point changes. This allows the performance of the models to be evaluated in realistic conditions, which is a prerequisite for the presented approach.

In summary, there are numerous approaches and NO_x models for different kinds of combustor applications. Selection of suitable NO_x modelling methods is case specific taking into account all relevant aspects determining the appropriate modelling approach for the task. This paper focuses on indirect NO_x emission estimation in natural gas (NG) fired boilers. The applicable literature on the topic is reviewed next. In (Iliyas *et al.*, 2013), a three-dimensional (3D) CFD model for a 160 MW gas boiler was developed to produce data for computational NO_x and O_2 sensors. The system utilized 6–8 input variables and RBF and MLP neural networks, of which the RBF model with six input variables performed best. (Ferretti & Piroddi, 2001) utilized a 3D CFD model and developed a neural network-based strategy utilizing two different learning strategies to NO_x emission estimation for an oil and methane fired 320 MW thermal power plant. Eng-genes and MLP neural networks were applied to the same power plant by (Li *et al.*, 2005), utilizing Arrhenius type equations in a semi-empirical model. The last two papers utilized cell temperatures derived with the CFD model as model inputs, which are usually not available in real systems (Li *et al.*, 2005). Another semi-empirical model was presented in (Bebar *et al.*, 2002), where simplified kinetic equations utilizing six input variables describing the formation of nitrogen oxides were developed for a testing facility. There, the average estimation error was 9.1%, with the maximum of 25%. In another study, a NO_x soft sensor for NG fired water tube boiler was developed by (Shakil *et al.*, 2009). This model utilized static and dynamic neural networks. A principal component analysis (PCA) was used to reduce model inputs from 9 to 6, and genetic algorithms were used to identify the system's time delays. The estimation accuracies of 83% and 99 % were obtained with static and dynamic models, respectively. In another case, application of linear models was studied with full or limited operation regions (Korpela *et al.*, 2015) and linear and non-linear models were compared (Kumpulainen *et al.*, 2015) in natural gas fired 43 MW hot water boilers.

Though linear models were considered in the last two articles, the literature on prediction of NO_x emission in NG combustion is focused on nonlinear and multiple input variable models. These models can provide quite accurate NO_x predictions but are arguably not generally applicable nor inexpensive solutions to PEMS applications. As this paper is a significant extension to the last two articles, there are no other published NO_x models meeting the objectives of this paper. The objective of this study is to evaluate the accuracy, generalizability and long-term applicability of indirect NO_x estimates in existing NG fired boilers, to apply the methods in practise in a cost effective way. For the study, trial runs were conducted with two similar 43 MW NG fired hot water boilers. The studied boilers are structurally relatively simple with a single burner with fixed air distribution simplifying the modelling task. The goal is to find maintainable and transparent but still valid models for NO_x estimation and to evaluate the performance in varying operating conditions. Requiring easy maintainability and transparency promotes the application of simple models, e.g. linear regression models, which require tuning of only a few parameters. As the performance drop cannot be too significant, the simple models have to be compared with models that are more complex, to estimate the differences in performance. Hence, the linear multivariate regression models are compared with commonly used nonlinear models, which are identified with automatic procedure without any manual fine-tuning. Additionally, the number of input variables and selection of data sets are examined together with model sensitivity analysis.

2. NO_x formation, control and identification

The abbreviation NO_x refers to nitrogen monoxide (NO) and nitrogen dioxide (NO_2), which are present in the flue gas because of chemical reactions of nitrogen and oxygen. Approximatively 95% of the NO_x emitted from combustion processes consists of NO and c. 5% of NO_2. As the majority of NO reacts to NO_2 in the atmosphere, the environmental effects are practically the same for both of the NO_x components (Van Loo & Koppejan, 2008), and hence the emission limits for NO_x are typically set for NO_2, e.g. in IED.

This section provides some background information about NO_x formation. Primary reduction methods are also briefly discussed, as these methods are actively in use and they provide information on which parameters need to be considered while predicting the NO_x emissions. Additionally, identification of NO_x emissions is discussed. In-depth presentations of NO_x formation and control can be found from literature, e.g. from (Hill & Smoot, 2000) and (Skalska *et al.*, 2010), respectively.

2.1 NO_x formation mechanisms and primary reduction methods

Element nitrogen contributing to NO_x emissions originates from two sources, i.e. bound organic nitrogen from fuel and elemental nitrogen (N_2) from air. Based on these nitrogen sources, nitrogen oxides are formed in combustion by four different mechanisms: 1) fuel NO_x mechanism 2) thermal NO_x mechanism, 3) prompt NO_x mechanism, and 4) via nitrous oxide (N_2O). Fuel NO_x is formed in reactions after the release of fuel bound nitrogen when the fuel is heated during volatilization, and is therefore to major extent dependent on the content of chemically bound nitrogen in the fuel. Fuel NO_x is mainly reduced by providing a fuel rich combustion environment by air or fuel staging in the primary combustion zone. Thermal NO_x formation requires sufficiently high temperature (>1300 °C) and residence time for dissociation of the atmospheric diatomic molecules N_2 and O_2. These can, for example, be reduced by lowering reaction times at high temperatures and especially by reducing the peak temperatures below the threshold, e.g. by favouring long and radiative flames, recirculating flue gas, reducing air preheating, injecting water or steam, and by reducing excess air (Kilpinen, 2002). Prompt NO_x formation takes place in reactions of atmospheric nitrogen and hydrocarbon radicals in fuel-rich regions. The mechanism is less dependent on temperature and significantly faster than thermal NO_x (Turns, 2000). NO_x formation via N_2O produces NO_x from atmospheric nitrogen. The mechanism is significant with high excess air and pressure, and low temperature (Kilpinen, 2002); (Turns, 2000).

2.2 NO_x emissions in natural gas fired burner

Combustion system design, fuel type, process operation and prospective NO_x abatement techniques significantly influence NO_x emissions (Skalska *et al.*, 2010); (Bělohradský & Kermes, 2012); (Habib *et al.*, 2008); (Ilbas *et al.*, 2016); (Skryja *et al.*, 2015). When analysing one specific boiler, the interest focus on the dominating NO_x formation mechanism. In NG fired boilers, the contributions of prompt and N_2O based NO_x mechanisms are typically less than 5 % each

(Kilpinen, 2002). Nitrogen included in the NG is in elementary form (N_2), which behaves similarly to atmospheric nitrogen (Kilpinen, 2002), and therefore fuel NO_x formation is insignificant in NG combustion. However, the presence of chemically bound nitrogen in the gaseous fuel influences the formation of fuel based NO_x (Skryja *et al.*, 2014). This must also be considered if other gases than pure NG are combusted and especially if the fuel quality changes. The dominant NO_x formation mechanism in NG combustion is the thermal one (Turns, 2000). However, the dominant mechanism may change depending on the operating point, so the trial runs for NO_x identification must be designed with care.

In thermal NO_x formation, the primary variables on NO_x yields are time, temperature, and oxygen availability (Turns, 2000). Therefore, any actions affecting flows and especially flame temperatures (e.g. via fuel power, air distribution, active flame cooling, air preheating, and boiler water temperature) or flue gas oxygen contents must be considered. Potential primary NO_x control methods aim at preventing the formation of NO_x emissions through these means. The potential methods include low-NO_x burner structures, reduction of excess air, air or fuel staging, water or steam injection, and flue gas recirculation. Therefore, the NO_x control actions are only relevant to the indirect NO_x emission estimation if they affect the combustion conditions actively. In these cases, the effect of these control variables must be included in NO_x models. Secondary NO_x reduction methods, such as selective catalytic and non-catalytic reduction systems (SCR, SNCR), that aim at reducing already produced NO_x emissions have not been applied, at least yet, to boilers of the considered category.

2.3 NO_x emission identification

When identifying a NO_x emission model based on measurement data, it is vital to consider all the affecting variables. Ferretti & Piroddi (2001) concluded that NO_x correlations in literature can be expressed in general form:

$$[NO_x] = f(p, \tau, T, \lambda), \tag{1}$$

where NO_x emissions are expressed as a function of pressure, residence time, temperature and excess air ratio, of which the first three are related to fuel flow rate and therefore to fuel power. Depending on the structure of the plant however, the number and quality of measurements and controllable parameters may vary significantly. Fig. 1 presents a general framework for NO_x estimation with various affecting controllable and uncontrollable variables, though some other variables may exist especially in large power plants. In practise, some of the measurements may be missing or otherwise unavailable. In that case, the control signals of actuators can be utilized instead, but then the potential nonlinearity of the characteristics of the actuator is not included in the NO_x model. Additionally, it should be noted that not all the variables are free variables, in other words, they might be controlled or linked to other variables e.g. by a control chart linkage. It is beneficial to conduct initial tests systematically and then to analyse the effect of predetermined variables on NO_x emissions, which is illustrated e.g. in (Bělohradský & Kermes, 2012); (Skryja *et al.*, 2015). After the main affecting input variables are known, the identification and validation tests can be conducted. However, if the boiler or burner are modified or the operation strategy of the boiler changes significantly, the process should be thoroughly identified again.

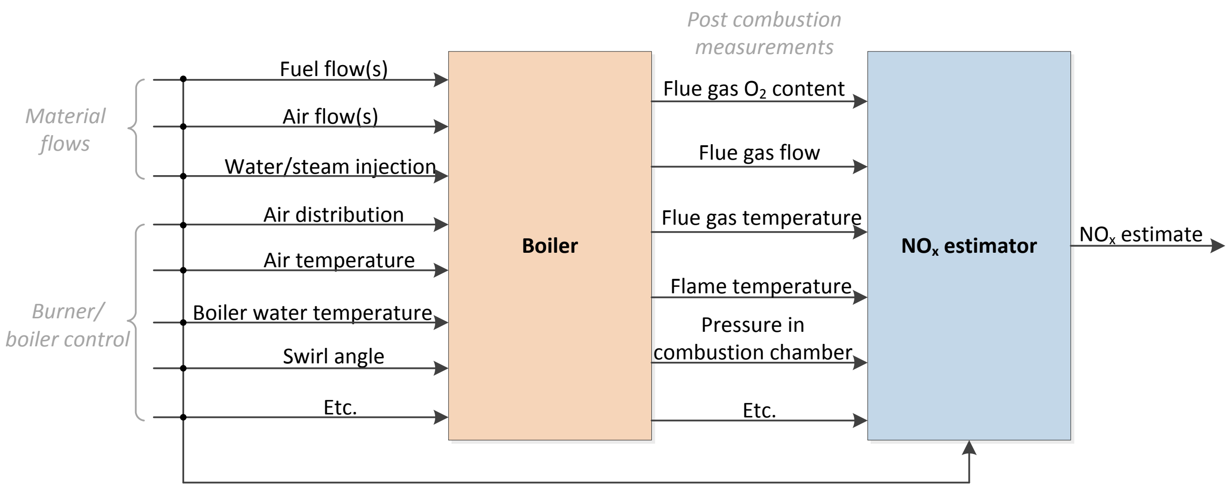

Fig 1. A generic framework for NO_x estimation.

3. Test setup and trial runs

The test plant was a municipal 258 MW_{fuel} back-up heating plant in Helsinki, Finland. The plant consists of six 43 MW_{fuel} hot water boilers connected to a common stack. The plant is a remotely operated back up and peak load plant. Typical annual operation hours of the boilers are relatively low, but still the plant is a solid part of the district heating network of Helsinki. Three of the boilers can be equipped with either NG or heavy fuel oil (HFO) burners and three with only HFO burners. The trial runs presented here were conducted with two NG fired boilers (i.e. Boiler A and B). All the boilers are identical, and the burners (one at a time in a boiler) are located at the ceiling of the vertical combustion chamber with downward flames. The gas burners were originally similar, but due to ongoing burner modifications, they were slightly dissimilar with respect to air feed structure. Therefore, the NO_x emissions of the burners behave to some extent differently and represent two similar but not identical cases. Both burners had fixed air distribution structure that can be changed only mechanically. NG came from Siberia, Russia, via gas pipe. The nitrogen content of the NG is in elementary form (N_2) and is less than 1 % of the gas volume.

The trial runs were conducted during four separate measurement campaigns with similar outdoor conditions. Measurements were provided by the normal measurement setup present in each boiler, including fuel flow, flue gas O_2 content, temperatures of flue gas, feed water, return water, and boiler room, and pressures of combustion chamber and combustion air. Additionally, a portable flue gas analyser was used that consists of a) Horiba PG-350 SRM portable flue gas analyser with NO_x, SO_2, CO, CO_2 and O_2 measurements, b) gas conditioning, drying and sampling system PSS-5, and c) portable gas sample probe PSP4000-H. The data from separate sources, i.e. the automation system and external NO_x analyser, were synchronized and combined to a data set with 15-second sampling interval. Due to the standard procedure of drying of sample gas that is fed to the analyser, the analyser measured the gas properties (i.e. O_2, CO & NO_x) from dry flue gas (marked as dry basis, d.b). In comparison, the fixed O_2 sensors of the boilers measured oxygen content directly from moist flue gas (marked as wet basis, w.b.). The difference between dry and moist flue gas contents depends on the fuel hydrogen content and is $X_{d.b.} = 1.233\ X_{w.b}$ for NG in stoichiometric conditions, i.e. no excess air, but the factor 1.233 is lowered when the oxygen excess is increased. This fraction is included in the parameters of the data based models, but it should be kept in mind when interpreting the O_2 levels from different measurements. Additionally, the gas properties are converted by standard procedure to represent the emission levels in reference O_2 level, which is calculated by

$$f_{O_2} = \left(20.95 - X_{O_2,measured,dry}\right) / \left(20.95 - X_{O_{2,reference}}\right). \qquad (2)$$

$X_{O_{2,reference}}$ refers to O_2 reference, a fuel dependent parameter which is 3 % for gaseous fuels. In summary, the calculated NO_x estimates are presented in the form that is applied in the IED, i.e. [mg NO_2/Nm^3, O_2=3 %, d.b.], where Nm^3 stands for normalized volume in standard temperature and pressure, i.e. T = 0°C & p = 101 325 Pa. The emission values required by the IED are hourly averages, but here the applied sampling and estimation interval was 15 seconds and no averages were calculated. This approach was used to study the momentary emission conditions in the process and to evaluate how correctly the method is able to estimate the emissions in all the relevant process conditions.

The manipulated variables of the boilers are fuel flow (controlled primarily based on power requirements), airflow (controlled in relation to fuel flow and is fine-tuned based on flue gas O_2 content), and boiler-water recirculation flow (controlled by flue gas temperature and boiler return water temperature that is 60 °C). Combustion air is taken from the boiler room without air preheating, except that the incoming air to the boiler room is heated to be clearly above 0 °C. Hence, the combustion air temperatures are typically in the range 5–15 °C. The temperature of feed water to the district heating network varies between 110–120 °C, but is typically controlled to 110 °C. The pressures of combustion air and combustion chamber are measured but their values are fixed in relation to the NG flow, so they are not independent variables in practise. To conclude, as the boiler water and combustion air temperatures have very limited operating ranges, air distribution is mechanically fixed and there is no water injection to boilers, the only controllable parameters of the boiler are air and fuel flows. The gas flow directly determines the fuel power in NG combustion. Within a fuel power level, alterations to the airflow are followed by changes to flue-gas oxygen contents. Therefore, the fuel flow and flue-gas oxygen content are of special interest when conducting the tests. This is also supported by the NO_x theory presented in Section 2 and the results from preliminary tests, which have been excluded here for convenience. However, other variables, for example, flue gas temperature, pressure in combustion chamber, or return water temperature, might fluctuate and have an effect on NO_x emissions. However, as these variables cannot directly be affected, the identification of their contribution is more difficult to determine. Still, their contributions are studied in Section 5.4.

For NO_x model identification, two separate trial runs were conducted with both boilers: *identification* and *extra validation* runs. The identification runs used constant power levels within the normal operating ranges. They are referred to as *train* sections as they relate to model training. At every power level, air feed steps to vary oxygen levels were performed exceeding the typical operating regions. The sections of constant power levels were used for identification of the NO_x models. The power transition sections, called *test* sections, were run in normal automation mode with power level set point changes and with constant O_2 set points. The test sections were used for validating the model performance within the identification day. The ultimate target is that the trial runs in practical cases could be conducted in a day for each boiler and contain both identification (i.e. *train*) and validation (i.e. *test*) sections. In this work, however, extra validation runs with both boilers were conducted in order to estimate repeatability, the effect of varying process conditions, and the contribution of normal automatic operation in contrast to the manual runs. These extra validation runs were executed in normal automation mode, i.e. feedback control loops were active, with set point changes to flue-gas oxygen content and to fuel power in Boiler A and heating power in Boiler B. The trial runs of Boiler A took place on adjacent days with similar winter conditions. The trial runs of Boiler B, however, were conducted in winter (identification run) and in autumn (extra validation run) with a 7-month time difference. The processes were unchanged between the runs. Only the boiler under study was in operation during each run, all the others were shut down, except in the beginning of the identification run of Boiler B.

In the identification runs, step changes to air feed were conducted at different power levels in manual mode. The responses to fuel flow and flue gas oxygen content, which are the process measurements used for model inputs, are presented for Boilers A & B in left of Fig. 2 and 3, respectively. As the lower heat value of the NG is 36 MJ/Nm^3, the constant fuel powers represent 25.0, 35.5, and 18.5 MW for boiler A and 37.4, 30.4, 24.6, and 17.2 MW for boiler B. Oxygen level 2.2 % O_2 (w.b) is a typical set point used in remote operation.

The extra validation runs of Boilers A and B are presented in Fig. 2 & 3 (right). The runs were conducted in normal automation mode, and the changes to process states were conducted by fuel power and O_2 controller set point changes. With Boiler B, the values of the controlled variables fluctuated significantly during the trial run, due to exceptionally high heat power and temperature fluctuations in the district heating network and limited control performance of the boiler. Therefore, the run is operated in close to extreme operating conditions, which would indicate that if the NO_x estimation succeeds in the run, it is expected to perform at least as well in normal operation conditions.

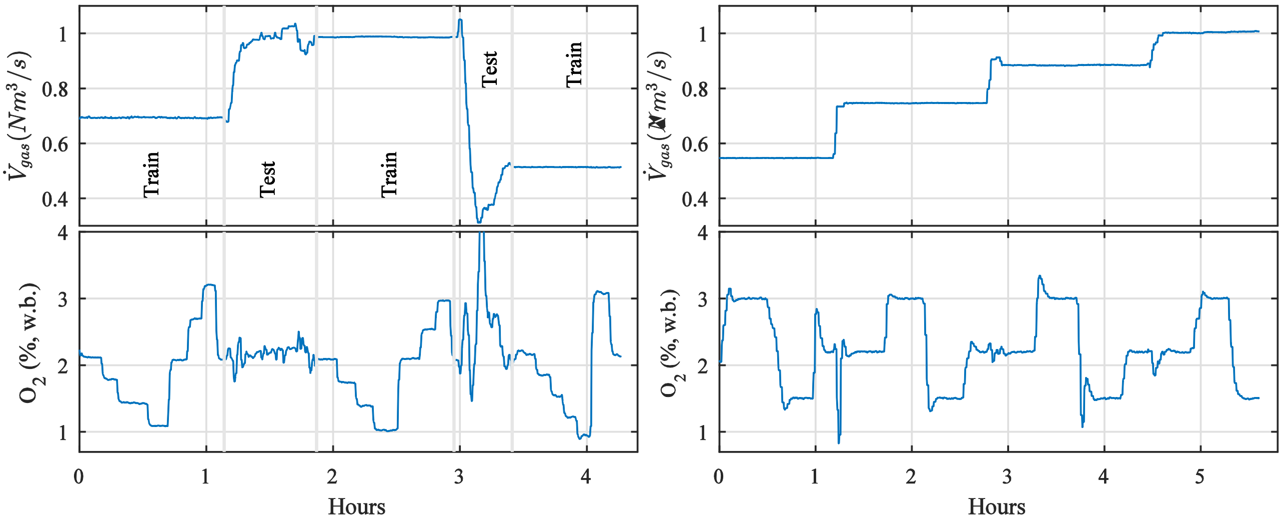

Fig. 2. Measured model inputs of Boiler A in the identification run (left) and extra validation run (right), i.e. gas flow (top) and moist flue-gas oxygen content (bottom).

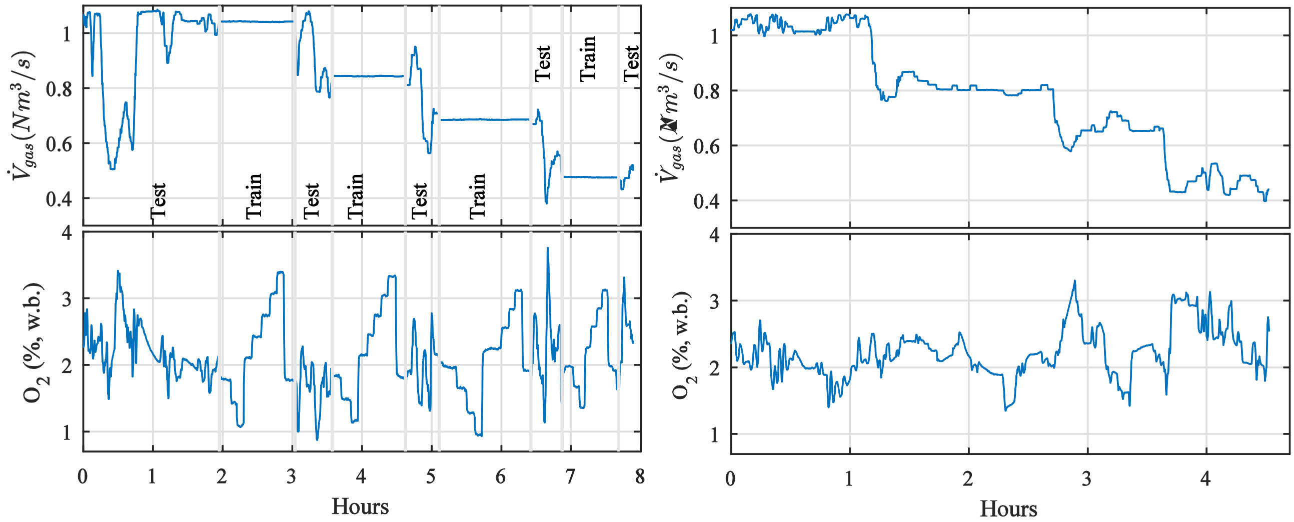

Fig. 3. Similar to Fig. 2 in case of Boiler B.

Fig. 4 and Fig. 5 present scatter plots of NO_x and CO emissions as a function of flue-gas oxygen contents for Boilers A and B, respectively. Fig. 4 illustrates that the NO_x data points at small and medium power levels are linear and almost linear at high power level. However, the respective plot of Boiler B in Fig. 5 indicates that the NO_x data points are almost linear at two of the lowest power levels, but significant nonlinearity and some variance can be seen at two of the highest power levels and especially at oxygen levels higher than 3.0 % (d.b.). Therefore, it is expected that linear models are not fully applicable in the whole oxygen range in the case of Boiler B.

The typical set point for O_2 used in normal remote operation is 2.2 % O_2 (w.b) which corresponds to 2.7 % (d.b.). Therefore, CO scatter plots presented in Fig. 4 & 5 indicate that there is potential to reduce the oxygen set point without the threat of increasing CO, which would be beneficial from the perspective of efficiency, NO_x emission level, and NO_x response linearity.

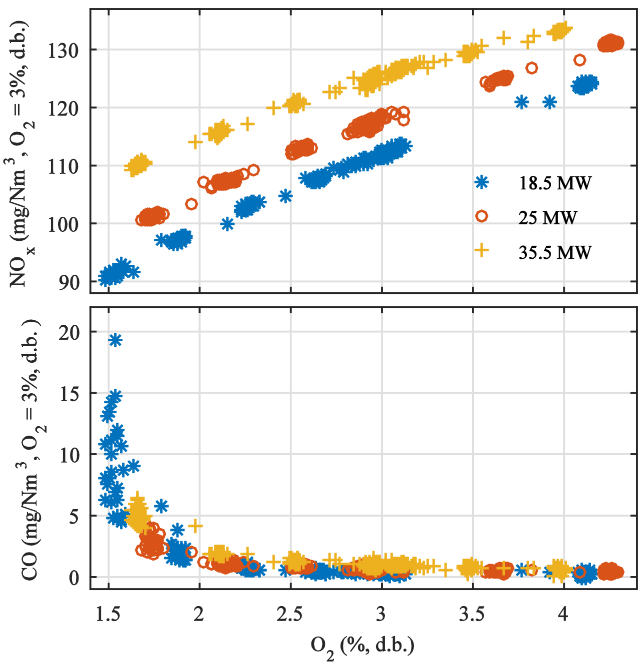

Fig. 4. Measured NO_x (top) and CO (bottom) emission of Boiler A as functions of O_2 level at constant power levels. The data is collected from the train sections of the identification run.

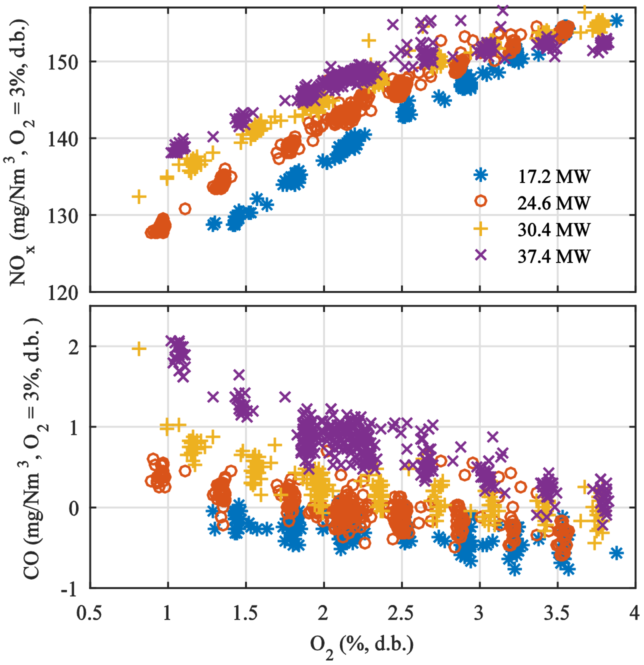

Fig. 5. Similar to Fig. 4 in case of Boiler B. The CO curve is slightly biased due to a calibration error.

4. Methods

As was pointed out in the Introduction, quite a few data based modelling methods can be applied in NO_x emission estimation. The objective of this study was to find practical solutions for NO_x emission monitoring in the existing case processes in the long-term, which introduces additional requirements to the applied methods. In addition to prediction accuracy, the model must be reasonably straightforward to identify and maintain in varying process conditions, and the selected modelling methods should preferably be widely in use and easily available. Based on these criteria, four modelling methods were selected for the study, of which one is linear (affine) and the others nonlinear. Some other relevant methods may exist, but the selected methods should reveal most of the modelling potential. All the chosen methods are static since the dynamics of NG combustion is very fast.

The four applied methods are presented next. Linear multivariate regression is the most basic data based modelling method (Draper & Smith, 1998). It is straightforward to identify without iterative optimization of parameters and is easy to update and maintain. Multilayer perceptron (MLP) neural networks (Haykin, 1994) with one hidden layer containing a sufficient number of neurons provides a universal approximation. Hence, they are very commonly applied nonlinear regression models. In this work, the MLP networks were trained using the Levenberg-Marquardt method presented in (Nørgaard, 2003). Fuzzy inference system (FIS) of Sugeno type is commonly applied for nonlinear regression (Sugeno, 1985). The final output is a weighted average of the constant or linear output functions, weighted by the membership functions of the input variables. The membership functions can be identified from the data by ANFIS (adaptive-network-based fuzzy inference system), a hybrid-learning algorithm, which combines the least-squares method, and the backpropagation gradient descent method (Jang, 1993). In this study, an ANFIS model with linear output functions and bell-shaped membership functions were used for input variables. Support Vector Regression (SVR), such as ν-SVR (Schölkopf *et al.*, 2000), can be used for nonlinear modelling of continuous data. The parameter ν controls the number of support vectors in the model. In this study, radial Basis Function (RBF) kernels $K(\mathbf{x}_i,\mathbf{x}_j) = \exp(-\gamma\|\mathbf{x}_i - \mathbf{x}_j\|^2)$, $\gamma > 0$ were used, which is a versatile method and works well in most applications (Hsu *et al.*, 2016). For that, a software package LIBSVM (Chang & Lin, 2011) was used in this study.

It should be noted that the estimation results of MLP and FIS are to some extent random due to randomly selected initial values. Therefore, the modelling results of these methods might be slightly dissimilar when the identification is repeated. The models presented in this article were identified several times and the models converged well to similar results. In comparison, linear regression and ν-SVR provide deterministic estimates.

The performance of the NO_x emission estimates were analysed by comparing absolute and relative RMSE (Root Mean Squared Error) values, that are based on absolute (ϵ_i) and relative (δ_i) errors, expressed as

$$\epsilon_i[mg/Nm^3] = NO_{x,est,i} - NO_{x,meas,i} \quad (3)$$
$$\delta_i[\%] = 100 \cdot \epsilon_i / NO_{x,meas,i} \quad (4)$$
$$RMSE[mg/Nm^3] = \sqrt{\frac{1}{N}\sum_{i=1}^{N} {\epsilon_i}^2} \quad (5)$$
$$Relative\ RMSE[\%] = \sqrt{\frac{1}{N}\sum_{i=1}^{N} {\delta_i}^2}, \quad (6)$$

where N denotes the number of samples.

5. Results

In this section, the NO_x models and their estimation results are presented. However first, the model derivation procedures, including model structure selections, are presented. The idea is that the NO_x emission models should be identified with standard procedures without any manual fine-tuning in order to enable straightforward implementation of PEMS applications by operators that might be unfamiliar with the modelling methods. In Section 5.1, the data is first scaled and then the structures of the nonlinear models with two input variables are detected by cross validations. After fixing the model structures, the NO_x models are identified and presented for Boilers A and B in Sections 5.2 and 5.3, respectively. Section 5.4 examines the effect of additional input variables with selected model structures. Section 5.5 discusses the selection of model identification data, and finally, Section 5.6 discusses the model sensitivity to errors in the measurements used as model inputs.

5.1 Data pre-processing and cross validation

The input variables were scaled before applying the nonlinear modelling methods. The normal operating range of each input variable was scaled to [-1 1]. The first part of the section concentrates on utilizing two model input variables that are fuel flow and flue gas O_2 content (w.b.). The selected operational ranges were [0 1.2] Nm^3/s (corresponding to [0 43.2] MW) for fuel flow and [0 6] % for O_2. The target variable, NO_x was scaled similarly from operational range [70 180] mg NO_2/Nm^3. In comparison, the linear multivariate regression models were identified with original non-scaled data.

In order to be applied generally and to avoid overfitting of the models, the optimal parameters for all nonlinear models were selected by cross validation with the training data sets. The optimal number of neurons in the hidden layer of the MLP network was determined by 10-fold cross validation, including ten Monte Carlo repetitions for each partitioning. Fig. 6 indicates that the minimum cross validation RMSE errors were achieved with two neurons for Boiler A and four neurons for Boiler B. The cases with 6 and 7 hidden layer neurons are examples where the model has been overfitted producing very high validation errors.

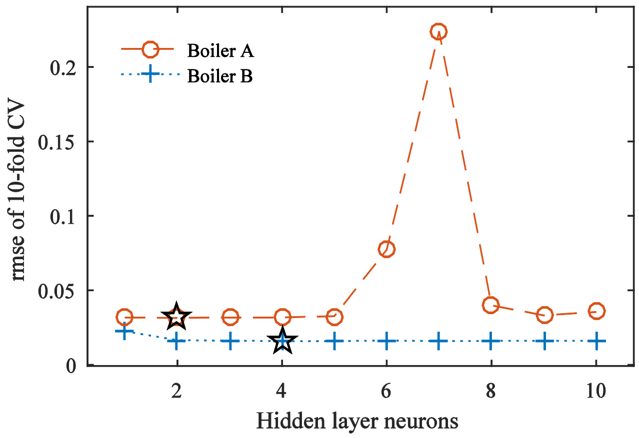

Fig. 6. Ten-fold cross validation errors of MLP.

The numbers of bell shaped membership functions for inputs in the FIS were selected by ten-fold cross validation, including ten Monte Carlo repetitions for each partitioning. The minimum cross validation errors we achieved with two functions for Boiler A and three functions for Boiler B, which can be seen in Fig. 7.

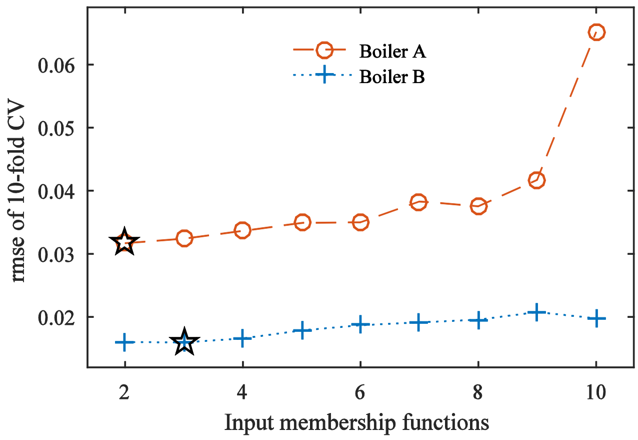

Fig. 7. Ten-fold cross validation errors of FIS.

The optimal values for the parameters ν and γ for ν-SVR were also selected by ten-fold cross validation. The cross validation errors are presented in Fig. 8, where the darker colours indicate higher error. The minimum error occurs for Boiler A with parameters $\nu = 0.7$ and $\gamma = 2$ (i.e. $\log_2(\gamma) = 1$), and respectively for Boiler B with parameters $\nu = 0.75$ and $\gamma = 4$ ($\log_2(\gamma) = 2$).

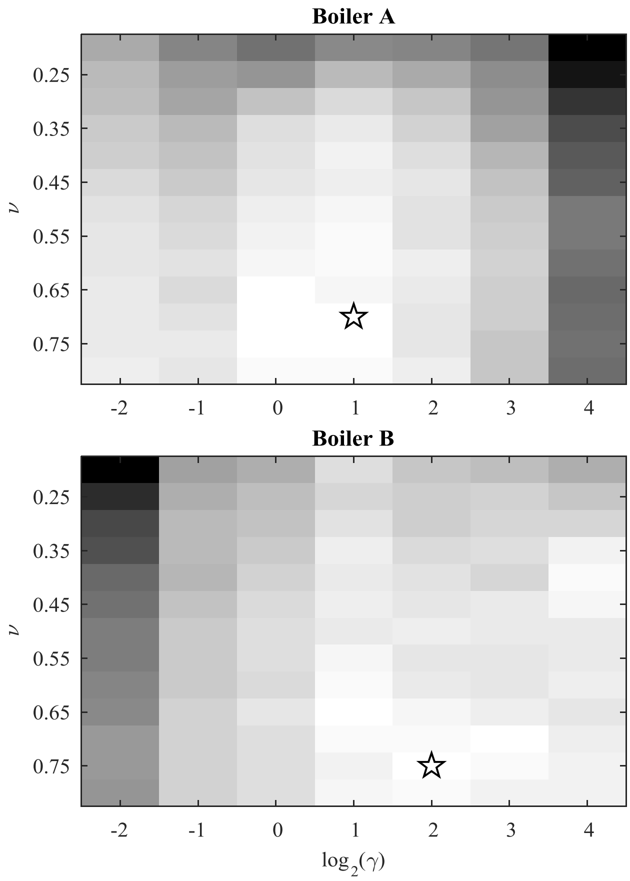

Fig. 8. Ten-fold cross validation errors of ν-SVR regression.

5.2 NO_x estimates with two input variables – Boiler A

Based on measured data, the following linear NO_x emission model [mg NO_2/Nm^3, O_2=3 %, d.b.] for Boiler A was identified:

$$NO_{x,est} = 13.8 \cdot O_2 + 29.6 \cdot \dot{V}_{gas} + 66.6, \tag{7}$$

where O_2 denotes measured moist oxygen content and $\dot{V}_{gas}$ gas flow to the burner as presented previously in Fig. 2. The respective measured and estimated NO_x emissions are presented in Fig. 9 along with MLP, SRV and FIS models identified in the previously described manner. In the figure, the training and testing sections of the identification run are separated by vertical lines. The linear model performs well in general but has some difficulties in the extreme O_2 levels, which is expected based on the scatterplot presented in Fig. 4. Other than that, all models have essentially equal performance. Notable differences are a slightly inferior performance of FIS in the first test section, and the peak in the second test section, which SVR fails to estimate. Still, the performance during the test sections is very similar to those of MLP and SVR. Therefore, the models are able to predict the NO_x emission response fairly well in all sections.

Figure 10 presents NO_x estimates of Boiler A in the extra validation run. As can be seen, the linear model performs very well, except for minor biases. Both MLP and SVR perform generally even better, but they underestimate the NO_x level at the end of the run at the highest power level. The performance of FIS is inadequate, as there is a significant bias at all power and O_2 levels.

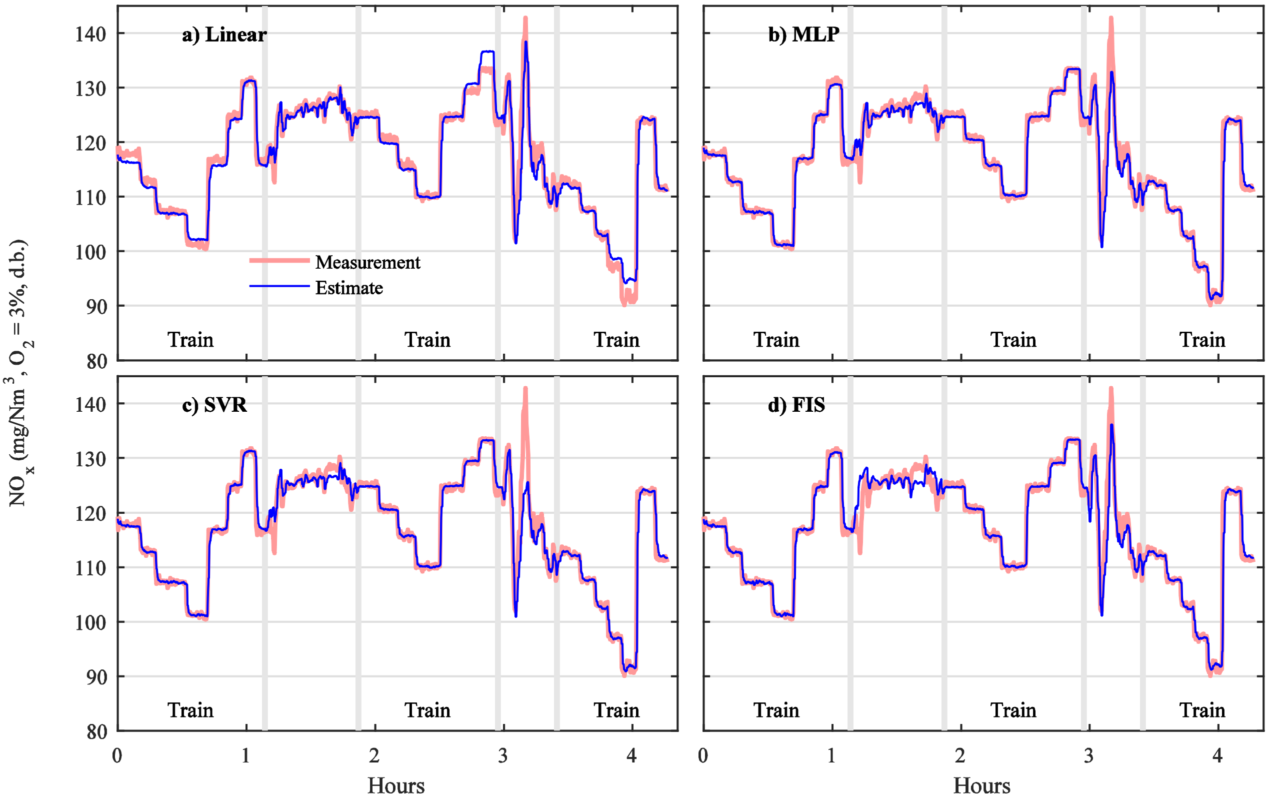

Fig. 9. NO_x estimates of Boiler A in the identification run. Training sections are marked with 'Train', and the remaining sections are Test sections.

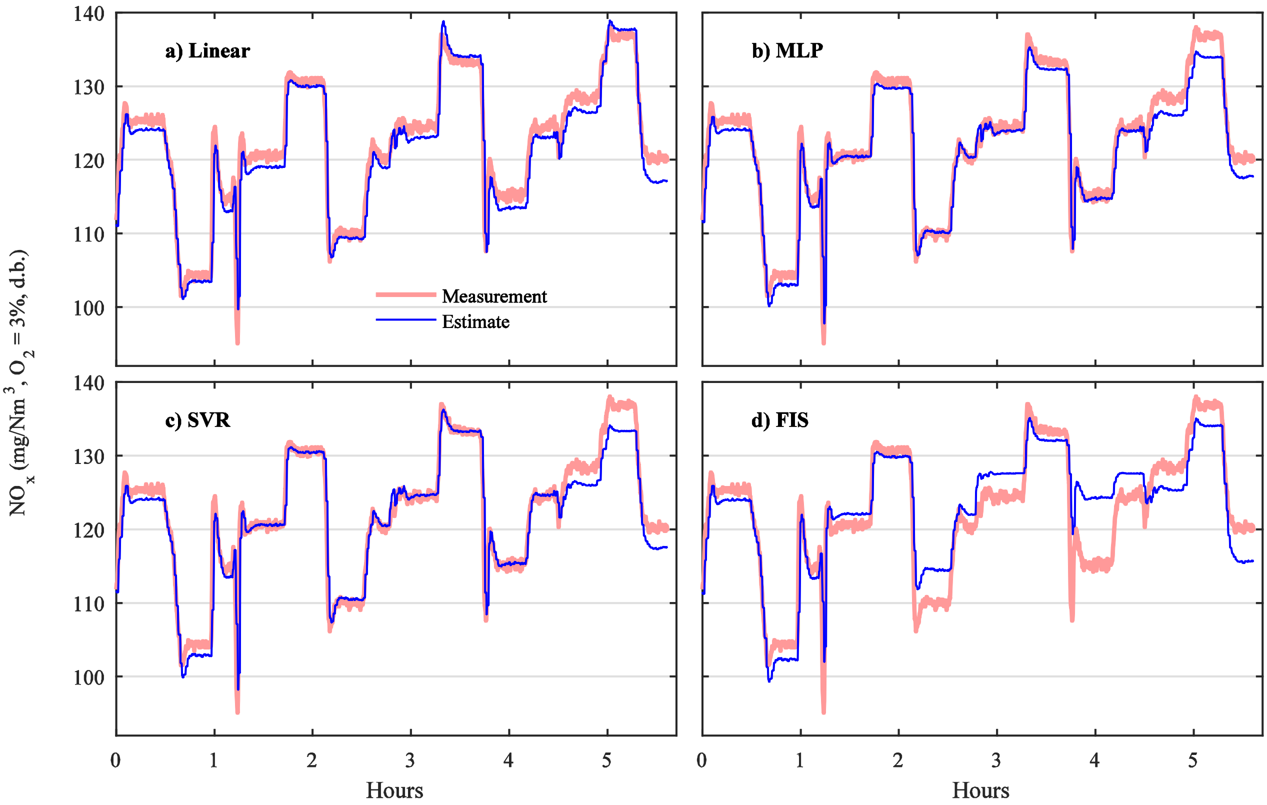

Fig. 10. NO_x estimates of Boiler A in the extra validation run.

Figure 11 sums up the estimation errors of NO_x models of Boiler A in the identification run (left) and extra validation run (right). The figures verify that the estimation errors are typically modest and significant biases exist only with FIS and at high oxygen contents with the linear model. However, there are some more or less significant error peaks in both runs, which occur during fast transients after the changes in control signals. Such errors, typically active for a few samples and usually followed by opposite error samples, are primarily due to a minor time shift between the portable gas analyser and process measurements. Thus, the two measurements represent slightly different process conditions. However, some faulty and especially opposite samples have a minor contribution to an hourly average that is applied in the IED. Typically, boilers of this type are operated with slow changes in set points even in the new energy systems, indicating that the influence of such errors in normal operation should be insignificant.

Fig. 11. Estimation errors of NO_x models of Boiler A in the identification run (left) and extra validation run (right).

Table 1 presents the absolute and relative RMSE values of Boiler A in train and test sections of the identification run and in the extra validation run. Based on the results, the linear model is the best in test sections and performs equally well in comparison to MLP and SVR in the extra validation run. The FIS fails in all test and validation sets. As a result, the linear model is the best two input variable model for Boiler A, predicting the NO_x emissions with a relative RMSE of 2.2 % in the test section of the identification run and 1.9 % in the validation run. The minor differences of the RMSE values in test and validation runs can be explained by different operating points and different operating conditions. Thus, the effect of additional input variables is presented in Section 5.4. Still, the linear model with two input variables performs very well in Boiler A, and can hence be utilized in PEMS applications.

Table 1. The absolute and relative RMSE values of Boiler A.

	Absolute RMSE (mg/Nm^3)			Relative RMSE (%)		
Model	Train	Test	Validation	Train	Test	Validation
Linear	2.03	2.65	2.27	1.78	2.16	1.91
MLP	1.71	2.90	2.21	1.46	2.33	1.86
SVR	1.71	3.57	2.25	1.46	2.76	1.88
FIS	1.70	3.42	3.89	1.45	2.80	3.35

5.3 NO_x estimates with two input variables – Boiler B

Based on measured data, the following linear NO_x emission model [mg $\mathrm{NO_2/Nm^3}$, $\mathrm{O_2}$=3 %, d.b.] for Boiler B was identified:

$$NO_{x,est} = 9.1 \cdot O_2 + 10.4 \cdot \dot{V}_{gas} + 117.4. \tag{8}$$

where O_2 and $\dot{V}_{gas}$ are presented in Fig. 3. The respective measured and estimated NO_x emissions are presented in Fig. 12 along with MLP, SRV and FIS models. As previously, the training and testing sections of the identification run are separated by vertical lines. All nonlinear models perform well in training sections. The linear model has difficulties in the first training section when both, the power and O_2, have the highest or the lowest levels as presented in Fig. 3. However, the performance during test sections is very similar to those of MLP and SVR. The first test section, when other boilers were also in operation, is best predicted by the MLP model. The performance of FIS is excellent in training sections, but the predictions fail totally at some points in all three test sections.

Fig. 13 presents the NO_x estimates of Boiler B in the extra validation run. The shapes of the estimates mostly equals the measurement apart from the FIS. However, all the estimates include noticeable and similar biases highlighted in Fig. 14, which presents the NO_x estimation errors of Boiler B. This indicates that the outdoor, process, or sensor conditions have slightly changed between the two trial runs conducted with a 7-month time difference. The cause for the change is examined in the next section by studying the effect of additional variables. However, such biases are easily removed by calibration of sensors if the process is not changed. This is especially the case with linear models, which require only few measurement points to find the new parameter values for the equations.

Table 2 collects the absolute and relative RMSE values of all distinct sections. The MLP and SVR provide the best estimates in test sections of identification run, but the linear model performs best in the extra validation run. Still, the modest bias and the relative RMSE of 2.7 % after 7 months from identification in different operating conditions indicate that the modelling principle is robust.

Table 2. The absolute and relative RMSE values of Boiler B.

	Absolute RMSE (mg/Nm^3)			Relative RMSE (%)		
Model	Train	Test	Validation	Train	Test	Validation
Linear	2.26	2.70	3.80	1.58	1.84	2.70
MLP	0.86	2.35	5.08	0.60	1.58	3.58
SVR	0.86	2.22	5.10	0.60	1.49	3.60
FIS	0.85	7.72	9.30	0.59	5.28	6.48

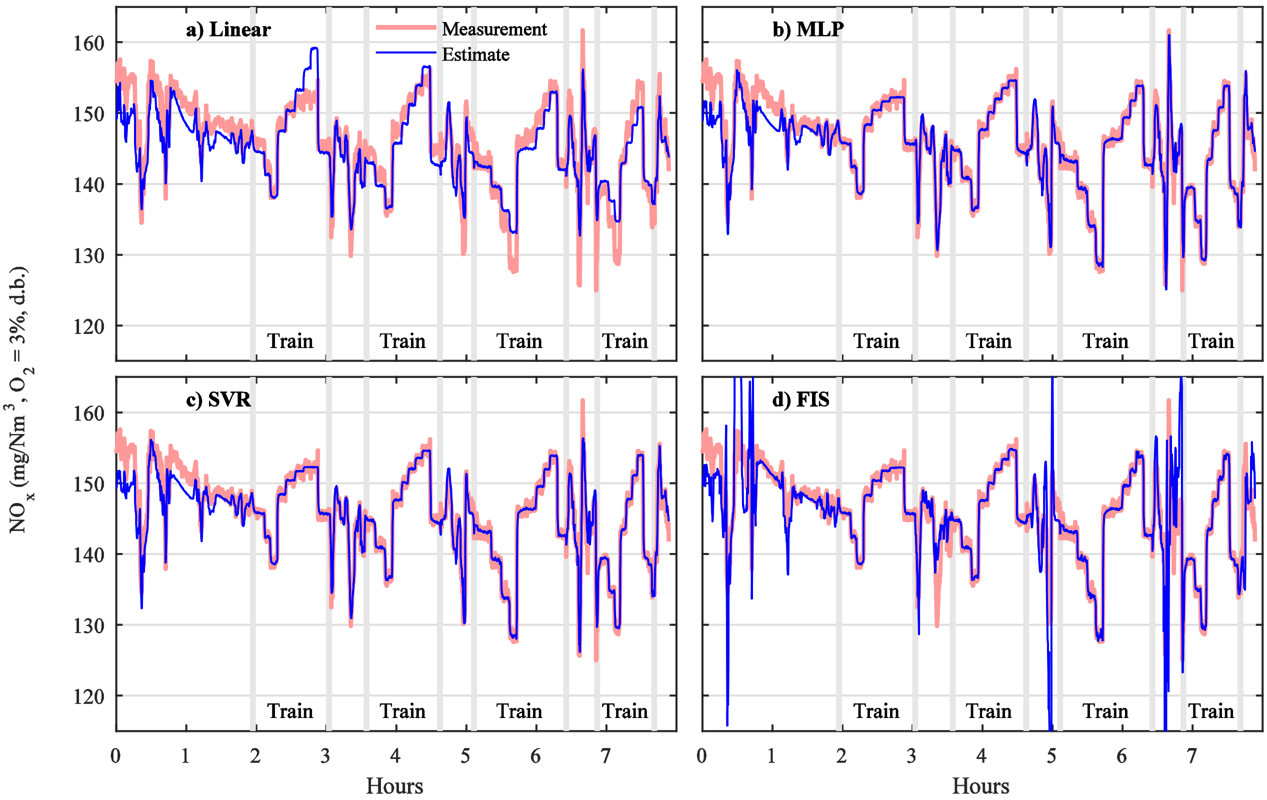

Fig. 12. NO_x emission estimates for Boiler B in the identification run. Training sections are marked with 'Train', and the remaining sections are Test sections.

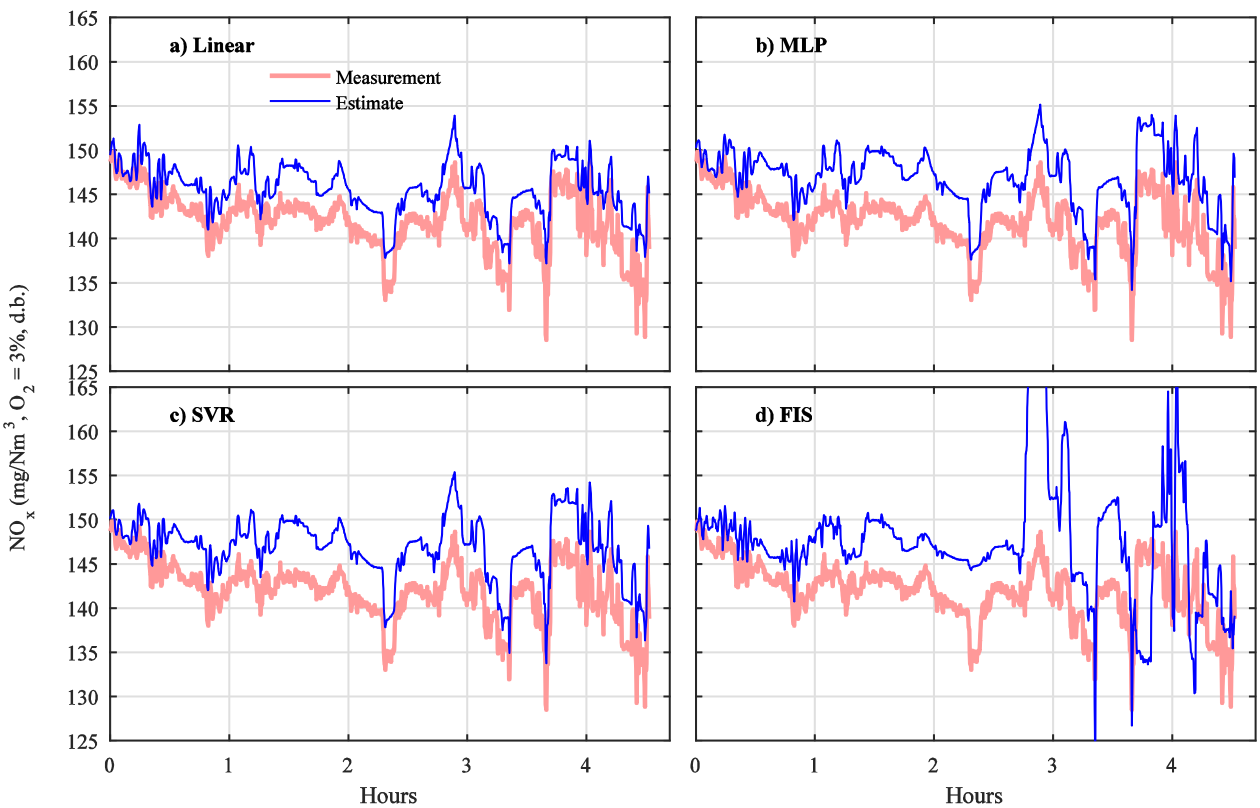

Fig. 13. NO_x emission estimates for the validation run of Boiler B.

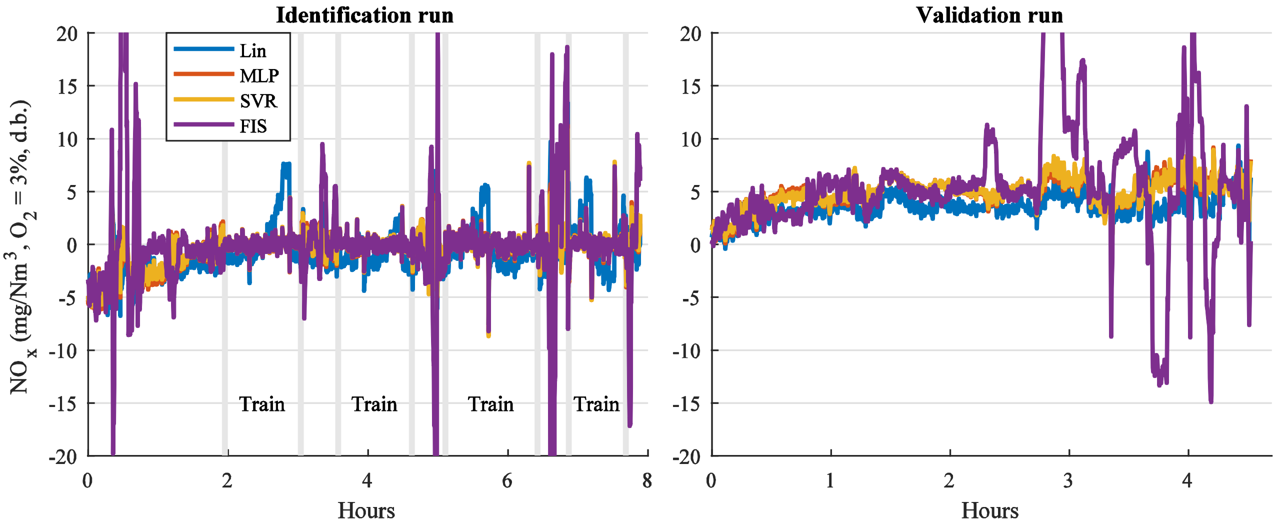

Fig. 14. Estimation errors of NO_x models of Boiler B in the identification run (left) and extra validation run (right).

As shown in Table 2, the linear model has the best generalization ability with respect to the extra validation run. However, the performance was lower in the test section of the identification run in comparison to MLP and SRV. As was already pointed out, and illustrated in Fig 5, the largest estimation errors in test sections with linear model occurred with the largest oxygen contents at the highest power levels. In practise, however, these regions are not typically used due to emission and especially efficiency related aspects. Thus, the performance of the linear model could possibly be improved by excluding these sections from input data without significant impairing model utilization. A new linear model, excluding the time instants in training sections where the oxygen level exceeds 3.0 % (d.b.) corresponding to 2.5 % (w.b.), was calculated:

$$NO_{x,est} = 10.9 \cdot O_2 + 13.4 \cdot \dot{V}_{gas} + 112.0. \tag{9}$$

Fig 15 presents the scatter plots of linear regression models with full (Eq. 8) and limited (Eq. 9) input data at four constant power levels existing in training sections of the identification run. The figure illustrates that the model with limited range has a smaller constant parameter and steeper gradient, with the result that the measured samples fit better with the model than the one that also covers the highest oxygen levels. Therefore, it is expected that the model with limited oxygen range has a better performance at lower oxygen levels, although the difference between the model outputs is rather small. However, this benefit is not fully exploited in the extra validation run, as the run was conducted in a normal operating region avoiding the lowest oxygen levels.

Table 3 presents the RMSE values for the linear and MLP models both for the original (Data included: All) and reduced (Limited) training data. The results verify that reduced training data lowers the RMSE values especially with the linear model in the test section, but performance is lowered in the extra validation run partly due to different operating regions. The RMSE of the MLP is lower with the identification data as expected, but the performance improvement within the test section is negligible. Though not the case here, this indicates that limiting the operating region could improve the estimation performance with linear regression models in some cases.

In conclusion, the linear multivariate regression model performed the most consistently when identifying the NO_x emissions of Boiler B. The nonlinearity of NO_x response seen in Figure 5 suggests the utilization of models with limited oxygen range, but this did not improve the estimation performance consistently. Therefore, simple linear model is the best. However, the constant bias of ca. 3 mg/Nm^3 suggests that the two input variables do not fully explain the long-term NO_x emission behaviour. Hence, the next section studies the effect of additional input variables to the estimates.

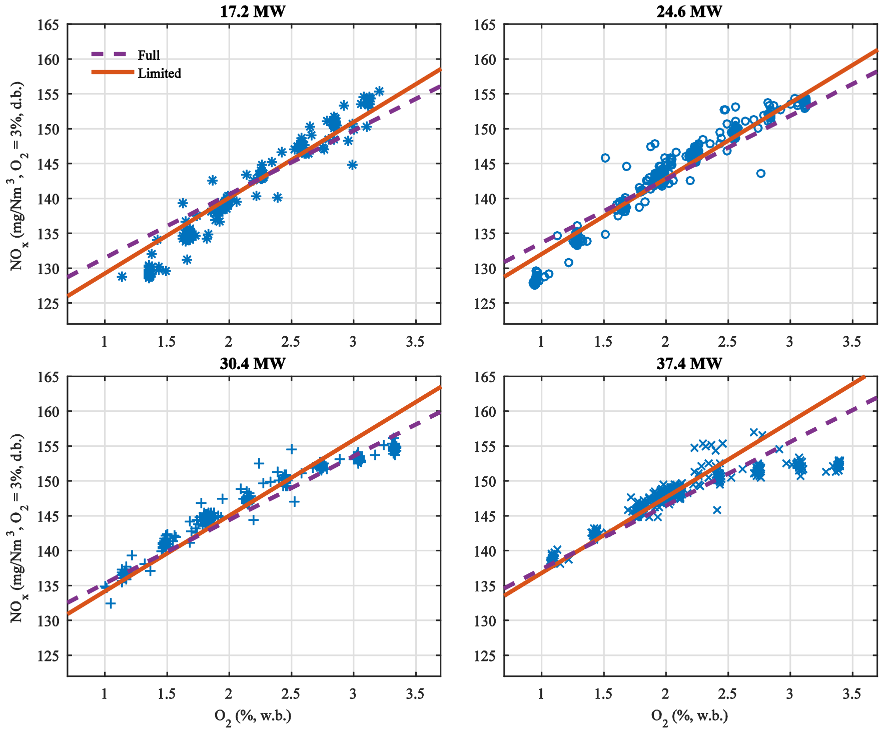

Fig. 15. Scatter plots of linear regression models with full (Eq. 8) and limited (Eq. 9) input data at four constant power levels existing in training sections of the identification run.

Table 3. The absolute and relative RMSE values of models of Boiler B that have limited oxygen range.

	Absolute RMSE (mg/Nm³)			Relative RMSE (%)		
Model	Train	Test	Validation	Train	Test	Validation
Linear, all	2.26	2.70	3.80	1.58	1.84	2.70
Linear, lim.	1.64	2.05	4.53	1.15	1.42	3.08
MLP, all	0.86	2.35	5.08	0.60	1.58	3.58
MLP, lim.	1.17	2.00	4.98	0.81	1.38	3.39

5.4 NO_x estimates with additional input variables

In the previous sections, the performance of NO_x emission estimates was analysed with models with two controllable input variables. Next, the contribution of three additional and uncontrollable input variables is tested with linear multivariate regression and MLP using the same data as previously. The selected input variables are flue gas temperature, boiler feed water temperature and pressure in the combustion chamber. The input variables were scaled in the same way as the two initial input variables, thus the normal operating range of each input variable was scaled to [-1 1]. The operating ranges of these additional input variables were [70 200] for flue gas temperature (°C), [50 150] for boiler feed water temperature (°C), and [-10 20] for pressure (Pa) in the combustion chamber.

Linear stepwise regression was used to select input variables to the linear multivariate model. The O_2, fuel flow, and pressure of the combustion chamber were the first variables to be included in both boilers, respectively. For Boiler A, the remaining two input variables were not considered significant, resulting in a three input variable model. For Boiler B, the stepwise procedure included all five input variables. The identification procedure of MLP models was equal to the one with the two input variables. Ten-fold cross validation with ten Monte Carlo repetitions was performed to select the number of hidden layer neurons. For Boiler A, the best number of neurons was two for all the variable combinations. The minimum errors for Boiler B were achieved with four, three and three neurons for two, three- and five- input variable models, respectively.

Next the results for three combinations of input variables for both linear and MLP models are presented. The two input variable models with all data are equal to the ones presented in the previous sections, containing the fuel flow and O_2.

The model with three input variables contains the pressure of the combustion chamber, and the model with five input variables includes the flue gas temperature and feed water temperature. The RMSE values of the results are presented in Table 4. With Boiler A, the additional input variables reduce the model performance of the linear model in test sections but slightly improve the performance in validation runs when three inputs are used. In comparison, the MLP results are poor, in particular with the test data with three input variables. This is due to two brief sections where the pressure of the combustion chamber is slightly higher than elsewhere but still well within the normal operating range. As a result, the NO_x estimates of these sections are out of the bounds, yielding very high RMSE values. The estimation performance is not significantly improved and the addition of extra input variables is not justified with Boiler A. With Boiler B, the additional input variables impair the estimation results of the linear model significantly in test sections but improve the performance noticeably in the extra validation run. On the other hand, the results of the MLP models of the test data outperform the linear model and even improve slightly with additional variables. However, the MLP presents poor results in the extra validation run and the performance is even worsened with the additional model inputs. Therefore, the additional input variables do not significantly improve the NO_x estimates, behave consistently nor clearly explain the bias in the validation run, so their addition to the models is not reasonable. Moreover, the addition of input variables may improve the estimation performance slightly or weaken it significantly depending on the case. In conclusion, the addition of variables should be considered only in special cases with special care.

Table 4. RMSE (mg/Nm^3) values of models with 2, 3, and 5 input variables for linear and MLP models.

	Number	**Boiler A**			**Boiler B**		
Model	of variables	Train	Test	Validation	Train	Test	Validation
Linear	2	2.03	**2.65**	**2.27**	2.26	**2.70**	**3.80**
Linear	3	1.99	3.01	**2.19**	1.89	4.40	**3.51**
Linear	5	1.92	3.51	**2.40**	1.79	5.01	**3.35**
MLP	2	1.71	2.90	**2.21**	0.86	**2.28**	5.05
MLP	3	1.59	38.0	5.98	0.84	**2.16**	5.53
MLP	5	1.56	3.56	3.19	0.81	**2.40**	5.80

5.5 NO_x estimates with extended input data

All the models presented so far have been identified based on data from training sections of the identification runs, as the idea was to cost effectively identify the models with training data and validate the model with test data from the same trial run. The results indicate that the nonlinear models do not generally improve the estimation results, but rather the opposite. However, due to the nonlinear behaviour of Boiler B, an additional calculation was performed. Here, the whole identification run, including the transients was used as identification data while the extra validation run data was used for validation. Table 5 presents the results, which indicate that the estimation results are worse when compared to the original procedure. Therefore, increasing the input data did not improve the estimation.

Table 5. The absolute and relative RMSE values of models of Boiler B identified with train and additional identification data (+).

	Absolute RMSE (mg/Nm^3)		Relative RMSE (%)	
	Identification	Validation	Identification	Validation
Linear		3.80		2.70
Linear+	2.30	4.49	1.59	3.18
MLP		5.08		3.58
MLP+	1.31	5.68	0.90	4.01

5.6 Model sensitivity analysis and sensor quality control

One important aspect when utilizing data based NO_x emission models in PEMS applications is the model sensitivity to measurement errors. If the uncertainty of the NO_x analyser is excluded, the model uncertainty will originate from the error in the measurements used as model inputs. The uncertainty can originate from unfavourable sensor locations and from sensor fouling or wearing, or from random noise. The effect of location is primarily compensated for in model derivation if the sensor locations are fixed, however, the influence of sensor fouling or wearing on measurement error can change over time. In the presented cases, the measurements used as model inputs are fuel flow and flue gas oxygen content. When considering fuel flow measurements, the error of ±10 % is not uncommon. The nominal power of both boilers is 43 MW, which corresponds to NG flow of 1.2 Nm^3/s. Considering the linear regression models presented in Eq. 7 & 8, such an error would contribute to the maximum error of $\epsilon_{\dot{V}_{gas},\text{ Boiler A}} = 0.1 \cdot 1.2 \cdot 29.6 \approx 3.6\ (\text{mg NO}_2/\text{Nm}^3)$ and $\epsilon_{\dot{V}_{gas},\text{ Boiler B}} = 0.1 \cdot 1.2 \cdot 10.4 \approx 1.3\ (\text{mg NO}_2/\text{Nm}^3)$ for Boiler A and B, respectively. Similarly, an error of 0.3 percentage points in oxygen measurement is possible between calibrations, which could contribute the errors by $\epsilon_{O_2,\text{ Boiler A}} = 0.3 \cdot 13.8 \approx 4.1\ (\text{mg NO}_2/\text{Nm}^3)$ and $\epsilon_{O_2,\text{ Boiler B}} = 0.3 \cdot 9.1 \approx 2.8\ (\text{mg NO}_2/\text{Nm}^3)$. The contribution of additional variables can be similarly examined. Thus, the influence of measurement errors on the linear models is relatively modest. However, even small errors may have a large influence when nonlinear models are used. This was illustrated in Section 5.4 where a small deviation in the pressure of the combustion chamber caused a significant error in the NO_x estimate. The total error can be calculated by summing the root mean squares of the error components, including the model errors.

The quality of measurements used in PEMS applications must be controlled regularly. Advanced sensor quality control procedures, e.g. as presented in (Korpela *et al.*, 2016a), cannot be applied considering the cost effective requirements. Instead, the effect of measurement uncertainty can be handled by regular sensor calibrations and hardware or software redundancy, and monitoring of the process measurements and NO_x estimates compared to history data, e.g. Nikula *et al.* (2016). These actions are especially important in PEMS applications, where the NO_x models utilize process measurements that have not been previously actively maintained.

6. Discussions

The literature review proposes to use nonlinear models for NO_x monitoring that provide accurate NO_x predictions, which is supported by the complexity of NO_x formation. Sensors, including the ones used in model inputs, may drift in the long term, so operation personnel of the plants should be able to tune the PEMS software on demand. Nonlinear models are difficult to update, as the influence of parameter adjustments cannot be easily deduced. Therefore, automatic procedures should be favoured. However, because the behaviour of the processes changed slightly between the trial runs and process operating principle to another, so too good a model does not necessarily contribute to the best long-term generalization ability. Therefore, the models should be as simple as possible with good maintainability and transparency, especially when considering that the models should be identified and validated cost effectively.

In case there are excess of suitable and independent measurements in the process, the application of robust estimation methods is enabled. However, this condition was not met in the tested boilers. There is, though, a flue-gas temperature measurement available, whose signal typically correlates with O_2 measurement signal. However, the correlation is not consistent in this case, because the flue gas temperature is controlled by boiler-water recirculation flow. The situation with measurement redundancy could be improved if there were an additional O_2 measurement or airflow or flue-gas flow measurements present that could be used to improve the quality of measurements used as model inputs. However, the application of such methods do not remove the uncertainty of NO_x model in changing process conditions, and therefore they are not enough to guarantee the validity of the estimate. On the other hand, the adaptation of the model is hard to be convinced to the authorities, because the input-output relation is not predictable anymore.

The results presented in the previous sections indicate that linear regression models perform typically at least as well as the nonlinear models in both boilers. The linearity of the tested boilers can partly be explained by homogenous fuel (NG), a simple burner structure, and especially the fixed air distribution structures. Still, some effort is required to apply linear models to almost linear processes. The nonlinearity must be carefully considered in the modelling stage by conducting trial runs that cover the full operating region for all the controlled input variables. Scatter plots (e.g. Fig 3&4) can reveal nonlinearities easily and linear models can be tuned to consider limited operating regions, as was tested with Boiler B. If errors are small with relevant data, the model is able to predict NO_x behaviour and the model can be utilized in normal operation. If no process modifications are made, the linearity of the process should remain. Therefore, retuning linear models (e.g. after recalibration of sensors) is straightforward and can be conducted with a simple trial run with a few power and oxygen levels. Notably, constant biases in the whole operating region, which were noticed in the case of Boiler B, can easily be removed by correcting the constant parameter of the linear model by a measurement with two measurement points. This kind of model validation can be conducted after the introduction of the PEMS regularly with an interval of a few months with modest costs. Therefore, PEMS with the NO_x estimation model could be a cost effective solution. However, the results indicate that each boiler should be identified and modelled separately. This is understandable due to the sensitiveness of NO_x emission formation to burner and furnace structures, which is also in line with the results presented in (Pulles & Heslinga, 2004). Moreover, the combustion and hence NO_x emission formation is sensitive to changes from NG to biogas (Ilbas *et al.*, 2016) and to potential enrichment of oxygen content in air (Riahi *et al.*, 2016). Hence, if these kinds of changes are applied, the NO_x full modelling procedure should be redone.

If the performance of a linear model is substantially lower than that of some nonlinear one, there are still some options to avoid the usage of nonlinear models. This can be achieved e.g. by applying the model in a limited operating region, analysed in this paper in the case of Boiler B. Alternatively, piecewise linear functions could be applied, or alternatively a nonlinear polynomial could be used to model the curvature as seen in the scatter diagrams. On the other hand, the challenge is further increased when the number of controllable model inputs increases, as the length of trial runs is substantially lengthened, especially if the effect of the variables is nonlinear. Additionally, some excluded variables might also affect NO_x emission formation, even if the results in this case indicated that their contribution did not improve the estimation results consistently. Still, their role should be carefully considered if the long-term generalization ability of the model is not good enough. However, the cost reduction potential of PEMS is easily lost if the effort required to identify and maintain the models is substantially increased. Therefore, applying PEMS to relatively simple processes should be most beneficial.

7. Conclusions

This paper compared NO_x emission estimation models in two similar 43 MW natural gas fired hot water boilers. The models utilize online process measurements that are always available even in relatively low-instrumented set-ups. The objective of the work was to find maintainable and transparent but still valid models for NO_x estimation, in order to evaluate the potential of PEMS application for NO_x emission monitoring in long-term operation in practical installations. The performance of linear regression models was compared with widely used nonlinear models, i.e. MLP, SVR and FIS, which were identified with automatic procedures without any fine-tuning. Most of the analysis concentrated on the models with two input variables, i.e. flue gas oxygen content and fuel flow. They are in practise the only controllable variables affecting NO_x emissions in the tested and many similar boilers that now fall under the jurisdiction of the Industrial Emission Directive. Additionally, the influence of additional, uncontrollable input variables and selection of data sets was considered in order to determine the best available model. In the study, two separate trial runs with two different operating modes (manual and automatic) were conducted in order to evaluate the long-term operation of the boilers and hence the performance of the modelling approach.

The RMSE values indicate that the linear multivariate regression models have the most consistent long-term performance. The performance of MLP and SVR varied significantly between the boilers and the test sections. Despite the cross validation, the FIS was overfitted, performing very well with training sets but failing in the validation set. Its performance could probably have been improved by fine-tuning; however, this would violate the requirement for easy maintainability. The extra measurements increased the accuracy of the linear models to some extent in some cases but not consistently. In conclusion, the results indicate that the linear model performs nearly as well as the best nonlinear models with both boilers. If the performance of the model is good enough, the simpler model structure should be favoured, which also promotes the easy model maintainability. However, the results indicate that the identification and maintenance of the models must be conducted separately for each boiler due to the differences in burner structures and instrumentations. Moreover, when the process itself or its parameters are changed, the validity of the models should be carefully reconsidered.

In conclusion, linear models with two input variables presented the most feasible approach in PEMS applications to estimate NO_x emissions in the types of boilers considered here. As presented, these models were insensitive to simultaneous operation of multiple boilers and to significant drift in measurement error in long-term operation. The linear regression models provided accurate estimates with relative RMSE values of less than 3 % in all the analysed cases, which is a good result also in comparison to previously published results. Online NO_x estimates provide reliable estimates of total NO_x emissions released in various operating conditions. This is a major benefit of PEMS applications and cannot be achieved with intermittent measurements carried out with portable NO_x analysers. Additionally, the estimates provide prospects to NO_x emission optimization in boiler operation. The proposed approach is best feasible for boilers in range 15–100 MW, as in that range continuous emission monitoring solutions are not required, the instrumentation level is generally low and the process designs are typically relatively simple. Naturally, with more available instrumentation and more complex processes and hence more degrees of freedom, the feasibility of the solution should be reconsidered, especially if the model accuracies are not at an adequate level.

Additionally, the NO_x estimates derived with the method can be used to improve the control of ammonia/urea dosing in SCR and SNCR (Selective (Non-) Catalytic Reduction) NO_x reduction method, if further tightening emission limits require such actions to be applied also in such boilers in the future.

Despite the four separate trial runs, future work should include long-term verification of the models in various weather conditions. Additionally, the effect of fuel quality changes due to changed gas type e.g. via LNG or upgraded biogas in NG network, should be researched, especially if PEMS is to be widely applied. Moreover, the approach could be tested in other types of boiler setups.

Acknowledgements

This work was conducted in Measurement, Monitoring and Environmental Assessment (MMEA) program managed by CLEEN Ltd and primarily financed by Tekes – the Finnish Funding Agency for Innovation (Grant no. 3155/31/2009). The paper was finished in Academy of Finland project EL-TRAN (Grant no. 293437). Financers and project partners, especially Helen, are gratefully acknowledged.

References

Bebar, L., Kermes, V., Stehlik, P., Canek, J., Oral, J. (2002). Low NO_X Burners - Prediction of Emissions Concentration Based on Design, Measurements and Modelling. *Waste Manage., 22*(4), 443-451. DOI: 10.1016/S0956-053X(02)00028-4.

Bělohradský, P., Kermes, V. (2012). Experimental Study on NOx Formation in Gas-Staged Burner Based on the Design of Experiments. *Chem. Eng. Trans., 29*, 79-84. DOI: 10.3303/CET1229014.

Chang, C., Lin, C. (2011). LIBSVM: A Library for Support Vector Machines. *ACM Trans.Intell.Syst.Technolog., 2*(3)DOI: 10.1145/1961189.1961199.

Draper, N. R. and Smith, H. (1998). *Applied regression analysis* (3. ed. ed.). New York: Wiley.

IED: Industrial Emissions Directive (2010/75/EU) by European Union. 2010.

Ferretti, G., Piroddi, L. (2001). Estimation of NOx Emissions in Thermal Power Plants using Neural Networks. *J.Eng.Gas Turbines Power, 123*(2), 465-471. DOI: 10.1115/1.1367339.

Habib, M.A., Elshafei, M., Dajani, M. (2008). Influence of Combustion Parameters on NOx Production in an Industrial Boiler. *Comput. Fluids, 37*(1), 12-23. DOI: 10.1016/j.compfluid.2007.04.006.

Haykin, S. (1994). *Neural networks: A comprehensive foundation*. New York: Macmillan College Publishing Company.

Hill, S.C., Smoot, L.D. (2000). Modeling of Nitrogen Oxides Formation and Destruction in Combustion Systems. *Prog.Energy Combust.Sci., 26*(4), 417-458. DOI: 10.1016/S0360-1285(00)00011-3.

Hsu,C., Chang, C., Lin, C. A Practical Guide to Support Vector Classification. http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf

Ilbas, M., Sahin, M., Karyeyen, S. (2016). Combustion Behaviours of Different Biogases in an Existing Conventional Natural Gas Burner: An Experimental Study. *Int.J.Renew.Energy Res., 6*(3), 1178-1188.

Iliyas, S.A., Elshafei, M., Habib, M.A., Adeniran, A.A. (2013). RBF Neural Network Inferential Sensor for Process Emission Monitoring. *Control Eng.Pract., 21*(7), 962-970. DOI: 10.1016/j.conengprac.2013.01.007.

Jang, J.-.R. (1993). ANFIS: Adaptive-Network-Based Fuzzy Inference System. *IEEE Trans.Syst.Man Cybern., 23*(3), 665-685. DOI: 10.1109/21.256541.

Kilpinen, P. (2002). Typen oksidien muodostaminen ja hajoaminen. In R. Raiko, J. Saastamoinen, M. Hupa & I. Kurki-Suonio (Eds.), *Poltto ja palaminen* 2nd ed., pp. 300-342. Helsinki: International Flame Research Foundation - Suomen kansallinen osasto.

Korpela, T., Suominen, O., Majanne, Y., Laukkanen, V., Lautala, P. (2016a). Robust Data Reconciliation of Combustion Variables in Multi-Fuel Fired Industrial Boilers. *Control Eng.Pract., 55*, 101-115. DOI: 10.1016/j.conengprac.2016.07.002.

Korpela, T., Kaivosoja, J., Majanne, Y., Laakkonen, L., Nurmoranta, M., Vilkko, M. (2016b). Utilization of district heating networks to provide flexibility in CHP production. *The 15th International Symposium on District Heating and Cooling,* September 4-7, 2016. Seoul, Republic of Korea. pp. 1-8. DOI: 10.1016/j.egypro.2017.05.077.

Korpela, T., Kumpulainen, P., Majanne, Y., Häyrinen, A. (2015). Model Based NO_X Emission Monitoring in Natural Gas Fired Hot Water Boilers. *IFAC-PapersOnLine, 48*(30), 385-390. DOI: 10.1016/j.ifacol.2015.12.409.

Kumpulainen, P., Korpela, T., Majanne, Y., Häyrinen, A. (2015). Modelling of NO_X Emissions in Natural Gas Fired Hot Water Boilers. *Commun. Comput. Info. Sci., 517*, 100-108. DOI: 10.1007/978-3-319-23983-5_10.

Li, K., Peng, J., Irwin, G.W., Piroddi, L., Spinelli, W. (2005). Estimation of NOX emissions in thermal power plants using eng-genes neural networks. *Ifac Wc 2005,* 3 - 8 July, 2005. Prague. pp. 115-120.

Nikula, R., Ruusunen, M., Leiviskä, K. (2016). Data-Driven Framework for Boiler Performance Monitoring. *Appl.Energy*, 183, 1374-1388. DOI: 10.1016/j.apenergy.2016.09.072.

Nørgaard,M. NNSYSID Toolbox - for use with MATLAB. Version 2. http://www.iau.dtu.dk/research/control/nnsysid.html

Pulles, T., Heslinga, D. (2004). On the Variability of Air Pollutant Emissions from Gas-Fired Industrial Combustion Plants. *Atmos.Environ., 38*(23), 3827-3838. DOI: 10.1016/j.atmosenv.2004.03.060.

Riahi, Z., Mergheni, M.A., Sautet, J.-., Ben Nasrallah, S. (2016). Experimental Study of Natural Gas Flame Enriched by Hydrogen and Oxygen in a Coaxial Burner. *Appl.Therm.Eng., 108*, 287-295. DOI: 10.1016/j.applthermaleng.2016.07.079.

Schölkopf, B., Smola, A.J., Williamson, R.C., Bartlett, P.L. (2000). New Support Vector Algorithms. *Neural Comp., 12*(5), 1207-1245.

Shakil, M., Elshafei, M., Habib, M.A., Maleki, F.A. (2009). Soft Sensor for NOx and O2 using Dynamic Neural Networks. *Comput.Electr.Eng., 35*(4), 578-586. DOI: 10.1016/j.compeleceng.2008.08.007.

Skalska, K., Miller, J.S., Ledakowicz, S. (2010). Trends in NOx Abatement: A Review. *Sci.Total Environ., 408*(19), 3976-3989. DOI: 10.1016/j.scitotenv.2010.06.001.

Skryja, P., Bělohradský, P., Hudák, I. (2014). The Influence of Bound Nitrogen Content in the Gaseous Fuel on the Formation of NOx. *Chem. Eng. Trans., 39*(Special Issue), 1321-1326. DOI: 10.3303/CET1439221.

Skryja, P., Bělohradský, P., Hudák, I., Jurena, T. (2015). Experimental Study on NOx Formation in Gas-Staged Burner. *Chem. Eng. Trans., 45*, 997-1002. DOI: 10.3303/CET1545167.

Sugeno, M. (1985). *Industrial applications of fuzzy control*. New York: Elsevier Science Inc.

Turns, S. R. (2000). *An introduction to combustion: Concepts and applications* (2nd ed.). Boston: McGraw-Hill.

Van Loo, S. and Koppejan, J. (Eds.). (2008). *The handbook of biomass combustion and co-firing.* Earthscan. London. 442 p.