Development and Assessment of Predictive Emission Monitoring Systems (PEMS) Models in the Steel Industry

Conference Paper · May 2006

CITATION
1

READS
535

7 authors, including:

Norbert Link
VDEh-Betriebsforschungsinstitut
20 PUBLICATIONS **34** CITATIONS
SEE PROFILE

John Niska
Swerea - Swedish Research
23 PUBLICATIONS **147** CITATIONS
SEE PROFILE

Inigo Unamuno
Sidenor R&D
30 PUBLICATIONS **295** CITATIONS
SEE PROFILE

Joao F Gomes
Instituto Politécnico de Lisboa
236 PUBLICATIONS **1,399** CITATIONS
SEE PROFILE

Development and Assessment of Predictive Emission Monitoring Systems (PEMS) Models in the Steel Industry

Mansour Saiepour and Neil Schofield
Swinden Technology Centre, Corus UK Ltd, Rotherham, United Kingdom
Tel: +441709 825262
Fax: +441709 825400
E-mail: mansour.saiepour@corusgroup.com

Bo Leden and John Niska
Metallurgical Research Institute AB, MEFOS, Lulea, Sweden
Tel: +46-920-201977
Fax: +46–920-255059
E-mail: bo.leden@mefos.se

Norbert Link
BFI Betriebforschungsinstitut GmbH, Dusseldorf, Germany
Tel: +49 211 67 07- 245
Fax: +49 211 67 07-310
E-mail: Norbert.link@bfi.de

Iñigo Unamuno
Sidenor I & D, Basauri, Spain
Tel: +34 94 487 1878
Fax: +34 94 487 1844
E-mail: inigo_unamuno@sidenor.com

João Gomes
Instituto de Soldadura e Qualidade (ISQ), Lisbon, Portugal
Tel: +351 21 3850991
Fax: +351 21 3850991
E-mail: ip241783@ip.pt

Keywords: PEMS, Emissions Monitoring, Soft Sensing, Predictive Monitoring, Model Development, Environmental Emission

INTRODUCTION

The conventional method for continuously monitoring emissions to air is by the use of hardware-based analysers commonly known as continuous emission monitoring systems (CEMS). It is likely that in Europe, as is already the case in the USA, environment regulatory authorities will require major process plants to be equipped with CEMS. For instance, the UK 2004 IPPC guidance notes indicate that continuous monitoring would be expected where the releases are significant and where it is needed to maintain good control (1, 2).

Although CEMS technologies have improved significantly over the past 15 years, they can still be relatively expensive to purchase, install and maintain. A complete system can cost more than €50,000 to purchase, and frequent costly calibrations are also required. The total cost of ownership of CEMS has been a major driver in the search for a more cost-effective way to meet regulations for emissions monitoring. To this end, software-based predictive emissions monitoring systems (PEMS) were proposed as an alternative to hardware-based CEMS.

PEMS indirectly infer emission values by using process parameters as model inputs. PEMS use software-based sensor (soft sensor) approaches, which have been used for quality control, condition monitoring and process control in many different industries (3, 4, 5). Other than lower costs, PEMS have two distinct advantages over hardware-based CEMS. Firstly, they have the ability to provide useful information on source emissions under different operating conditions and identify contributing factors. Secondly, the prediction can be used for inferential control and process optimisation.

Since the early 1990's, the US Environmental Protection Agency (EPA) has allowed PEMS as an alternative to CEMS for use on for example some steam generating units to predict NO_X emission rates (6). In Europe, environment regulatory authorities have started to consider the acceptance of PEMS as an alternative to CEMS. For example, although they differ by country, many European national regulations already call for software-based surrogate emission monitoring systems where the CEMS are taken off-line for maintenance or calibration.

The technologies, that PEMS models may be based on, are:

- Analytical physical equations modelling that describe the process dynamics
- Empirical (statistical) modelling based on linear or non-linear methods and using historical process data (7).

As part of a multi-partner research project, supported by the European Commission through the Research Fund for Coal and Steel (RFCS), aimed at reducing emissions from iron and steel making processes, PEMS models have been developed to provide a cost-effective alternative to hardware-based CEMS. The modelling work has been carried out by the project partners at the electric steelmaking Sidenor plant (Spain) using the BFI DataTools software, SSAB Tunnplåt AB reheat furnace (Sweden) using the Perceptive Engineering MonitorMV software and a coke oven plant operated by Corus in the UK also using the MonitorMV software (8).

This paper describes a procedure followed for developing the PEMS models, at the Corus coke oven plant in the UK, and highlights certain factors that may improve the performance of these models. The performance of a developed model has then been assessed against the US EPA performance specifications for PEMS.

MULTIVARIATE STATISTICAL TECHNIQUES

The methods for monitoring plant data have traditionally been limited to statistical process control (SPC) or univariate statistical analysis, which have been used to identify key variables that cause certain process occurrences. An important limitation of the classical univariate statistical methods is that the process variables are considered individually and their interactions with each other are ignored.

In contrast, multivariate statistical techniques rely on the concept of cross-correlation in order to capture the underlying relationships between the different process variables that exist during normal process operation (9). Therefore, when used to monitor a process, multivariate statistical modelling can more correctly detect the statistically significant events or abnormalities.

Multivariate statistical modelling forms new variables, called principal components or latent variables, which are linear combinations of the existing plant variables, but are selected to provide independent unique features or variability in the data. A very large portion of the variability in plant variables can be summarised and compressed into a relatively small number of latent variables.

Principal Component Analysis (PCA) is a method of extracting the majority of information from a set of measured process variables, and expressing it using a reduced number of variables, known as principal components or scores. Recursive Least Squares (RLS) is a proven regression technique for identifying dynamic cause-effect predictive models. The major advantage of RLS over Ordinary Least Squares (OLS) is that it will deal with highly correlated data because it avoids matrix inversion. Partial Least Squares or Projection to Latent Structures (PLS) is a method of identification that is similar to both PCA and RLS techniques. However, unlike PCA, PLS divides the variables into cause (input) and effect (output). It then identifies the primary features in the cause variables that describe the variation in the effect variables. Hence, PLS may be regarded as equivalent to both PCA and RLS combined and can be used for both condition monitoring and prediction. As with the RLS approach, PLS avoids high correlation problems.

LINEAR DYNAMIC SYSTEM MODELS

For a PEMS model, in which the output variable needs to be predicted in the absence of the measured output values, the commonly used dynamic system equation (8) known as Auto-Regressive with eXogeneous variable (ARX) will be reduced to the Finite Impulse Response (FIR) model shown below:

$$\hat{y}(t) = b_1 u(t-d) + b_2 u(t-d-1) + \ldots + b_m u(t-d-m+1) + e(t) \quad (1)$$

where,

$u(t)$ is the input (cause) variable measurement at time t

$\hat{y}(t)$ is the value of the effect variable that is predicted by the model at time t

$e(t)$ is the prediction error or difference between predicted and measured output variables

m is the delay-spread or number of predecessor time-points from the measured input variable to be considered

d is the time delay of the system (in samples) and determines the time taken before a change in the inputs will affect the output. In the MonitorMV software this is referred to as the prediction interval.

$b_1, b_2, \ldots, b_m$ are the coefficients that are related to the dynamics of the particular system and the aim is to determine these coefficients to achieve an accurate model.

In a multivariable system the above equation will take a multi-dimensional matrix form. A multivariate statistical technique such as RLS or PLS, which considers the correlations between all the variables simultaneously, is then used to identify the most accurate representation for the coefficients of the associated matrices.

PROCESS DESCRIPTION AND PLAN OF WORK

The research was conducted using the process variables measured at a coke oven plant operated by Corus in the UK. The plant consists of three batteries, each containing 25 coke ovens to convert coal to coke (carbonisation or coking process) for use in the production of steel. The coking process produces raw coke oven gas, which is cleaned in the by-products plant. Some of the cleaned coke oven gas is then used as fuel in the coke oven heating chambers. The waste gas resulting from burning of the coke oven fuel gas is directed through ducts leading to the main stack. This waste gas contains oxides of nitrogen (NO_X), sulphur dioxide (SO_2) and carbon monoxide (CO) amongst other species. The direction of flow of the fuel gas around each oven is reversed every 20 minutes to preheat the combustion air in the regenerative process. Reversals also allow uniform heat transfer for the carbonisation process. The duration of the reversal time between each heating cycle is approximately 12 seconds. The total carbonisation time for converting coal to coke is about 18 hours.

This plant was used for the research because it was likely that the environment agency would require the plant to continuously monitor NO_x and SO_2 emissions, since they were measured quarterly over only 1 or 2 hour durations. However, CO gas emitted from the plant main stack was continuously measured with CEMS and the values were being logged along with other process variables. Hence, the large quantity of continuous data, required for the model development, could be provided. Therefore, it was decided to develop the models based on the CO emissions and build knowledge and experience on all the factors in this process that may affect the PEMS model development. The concept will then be applied, to a few days' NO_x and SO_2 data, in order to develop these models.

In order to carry out the research three months of process and CO emissions data, during which the plant was known to be operating normally, were used. From this data 15 days were used for model training, 3 days for intermediate tests and 3 days for final evaluation. The data were pre-processed to remove outliers such as random spikes. Good training data were then selected, from the 15 days' data, following the procedure described in the next section. The most relevant variables were then selected for the model to predict CO emissions. The US EPA performance specifications were then used to assess the performance of the off-line models. Finally, the developed models were tested on-line using real-time data.

COLLECTION OF REPRESENTATIVE DATA

The integrity of the developed model depends on the quality of the data used for the analysis and model development. Special care was therefore taken in the collection, preparation and selection of the data, and the following aspects were considered:

a) Variability– the process variables must not stay constant (zero variance) and should move around throughout the data range,
b) Richness – the data should include all known normal operating regions of the process. If all the regions are not included, the model will provide a wrong prediction in an unrecognised operating region,
c) Consistency – the data must be consistent throughout and care must be taken to avoid data in which the process is operating abnormally, as this can affect the quality of the model produced.

A large amount of continuous data is necessary in order to select the required quality data for training the models over all the different normal operating regions. Also, the defined ranges of process variables that would be established during the model development (parameter operating envelope) must be maximised, so that the parameter inputs during evaluation would not fall outside the training data operating envelope, which would lead to unacceptable emission predictions.

Hardware and software connections have been completed between the process data computer and a separate on-site computer, and the process and emission data were logged at a sampling rate of one per minute and saved as MS Excel files.

The data used for training the models were over 15 days in September 2004 and for evaluating the models were over 3 consecutive days in October 2004. Additional 3 days data from August 2004 were also used for the intermediate tests during model development.

Data Pre-processing

Visual examination of the data showed that random spikes were present in the signals. In addition, inspection of the Fast Fourier Transform Power Spectrum showed the presence of high frequency components. It was, therefore, necessary to remove these spikes without adversely affecting the characteristics of the data. For this purpose, a Median Filter (10 steps to resolve) was applied to all the data that were selected for the training and evaluation. This pre-processing procedure had to be carried out for any new data that were collected for use in the model development.

Selection of Training Data

It is important to remove any data that is non-representative of normal process operation, before carrying out any model development. Such data are usually outliers or spikes, missing values, constant values (zero variance) and abnormal operating modes such as start-ups and shut-downs or the presence of faults.

Also, the data that is representative of normal operation must not be deleted; otherwise it would be more likely for the model to see more unrecognisable operating regions during evaluation and provide an invalid prediction. To avoid the deletion of representative data, because they may be shown as outliers by the statistical distributions, process engineers were consulted to check if the data related to any normal process conditions.

In order to select the representative data regions for training, a PCA model with fifty nine process variables was computed on all the data and the following procedure was then followed:

a) Inspection of Score plot contour charts

A plot of any two scores against each other, for each sample of the analysed data, produces a cluster chart. These charts can provide important information relating to several process conditions shown as different clusters.

Contour charts are based on the cluster charts and are concerned with the probability details associated with probability density function (PDF) sets. They include boundary limits that indicate the regions within which 95% or 99% of the points should be under normal operating conditions. Any excursions from these limits would then indicate a plant condition that should be analysed further. In this study, these limits have been calculated using Gaussian distribution assumptions, in which case ellipses are generated that identify the boundaries within which the data should be under normal conditions.

Figure 1 shows a contour chart for the plot of score 1 versus score 3. The red crosses outside the 99% boundary limit show that there may be either a different normal condition or an abnormal condition. Since this chart only shows the single plane of score 1 versus score 3, a complete picture of the multivariate situation is not presented. For example, the introduction of a fault condition may result in an excursion from normal operating conditions, but it may not be revealed in one contour chart.

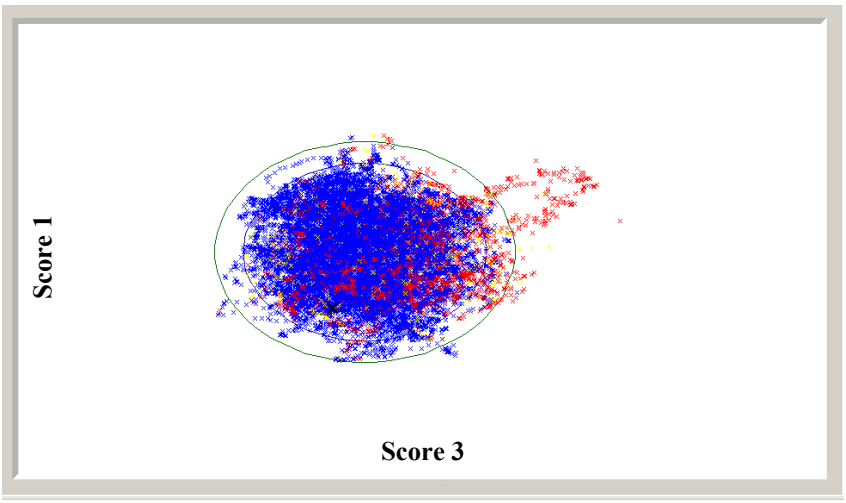

Figure 1 Contour chart for a plot of Score 1 versus Score 3

b) Inspection of Hotelling's T^2 chart

Hotelling's T^2 (extension of the univariate t-test) is a statistical measure of the multivariate distance of each observation from the centre of the data set. The T^2 chart provides an overall measure of how close the model scores are to their mean or typical value under normal operating conditions. Therefore, any significant deviations from the representative training region should be detected by analysing this single chart.

Figure 2 shows the absolute T^2 statistic trend at the top and a normalised score probability metric trend with upper 95% and 99% boundaries at the bottom. The colouring of the T^2 trend corresponds to the colouring of the score contour charts discussed above. In this application, only one class has been assigned to the training data.

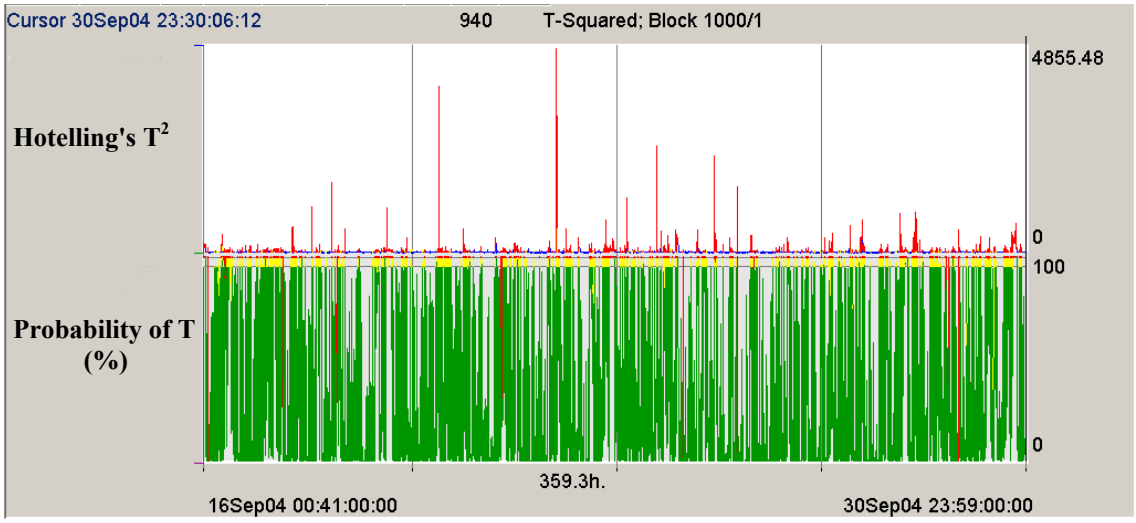

Figure 2 T^2 chart for all the data before selection of good training data

Examination of the T^2 chart indicated that some data were outside the 99% boundary and were statistically not representative. It was necessary to investigate these data further, since all the normal conditions must be included and all the unacceptable data must be removed. For example the large peak (duration of 12 minutes) at the centre of the T^2 trend was examined. A close-up view of this peak is shown in Figure 3. This large excursion was also indicated as significant in the squared prediction error (SPE) chart discussed below.

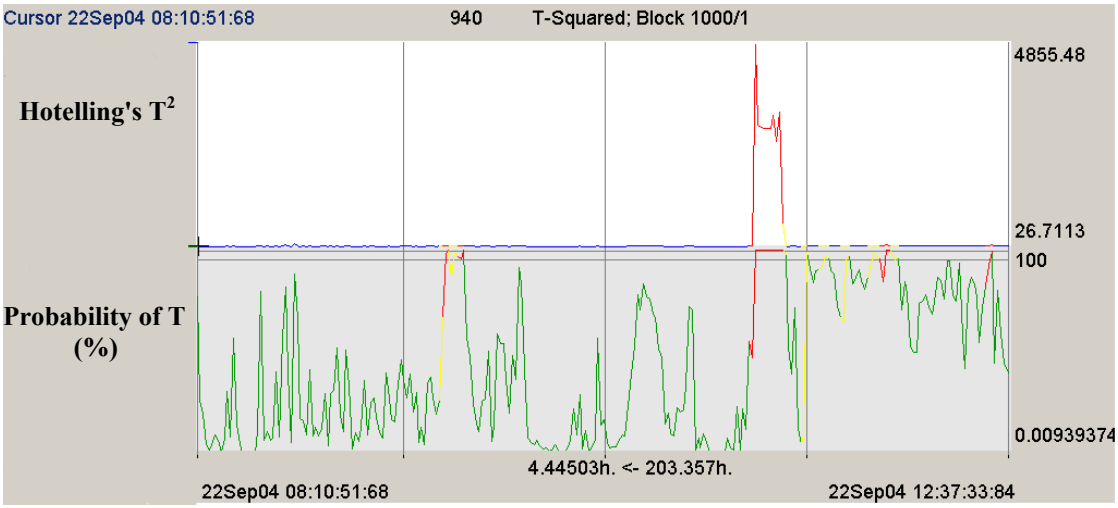

Figure 3 A close-up view of the large peak in the T^2 chart of Figure 2

c) Squared Prediction Error (SPE) chart

The SPE chart provides a measure of the error between the measured and predicted values of all the variables (as opposed to the T^2 chart which relates to scores). Therefore, any significant prediction errors should be detected by analysing this single chart.

Figure 4 shows the SPE trend (top) and a normalised probability error metric trend with upper 95% and 99% boundaries (bottom).

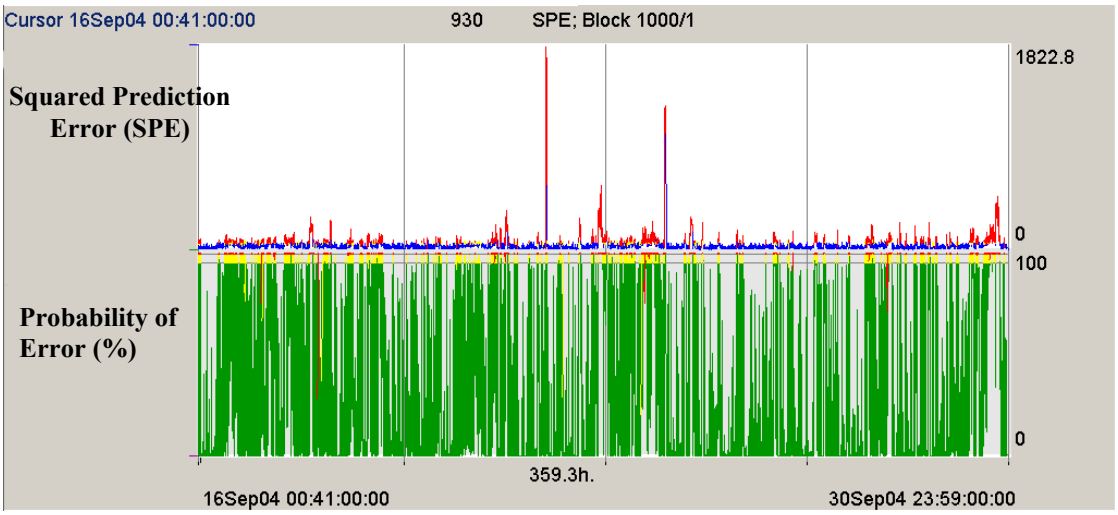

Figure 4 SPE chart for all the data before selection of good training data

Examination of the SPE chart showed that this short duration excursion was significantly different to what the model considered as normal and could not be predicted by the PCA model. Inspection of the individual prediction error trends showed that all the variables could not be accurately predicted during this time.

d) Score Contribution chart

The score prediction contribution chart shows for the selected sample the extent to which each plant variable contributes to the estimated scores or T^2 statistic at that sample (and two samples before and after). The variables that have contributed the most to the T^2 chart at a given time instant will have relatively large contributions (large bars) whilst those variables that have made little contribution to the T^2 chart, will have smaller bars. This chart provides diagnostic information relating to the particular variable(s) that may be responsible for the statistics exceeding their limits.

Figure 5 shows that four variables have a much greater contribution to the large excursion observed in the T^2 chart than all the other variables. These are related to Battery 2, indicating that there may have been an unusual activity undertaken on this Battery.

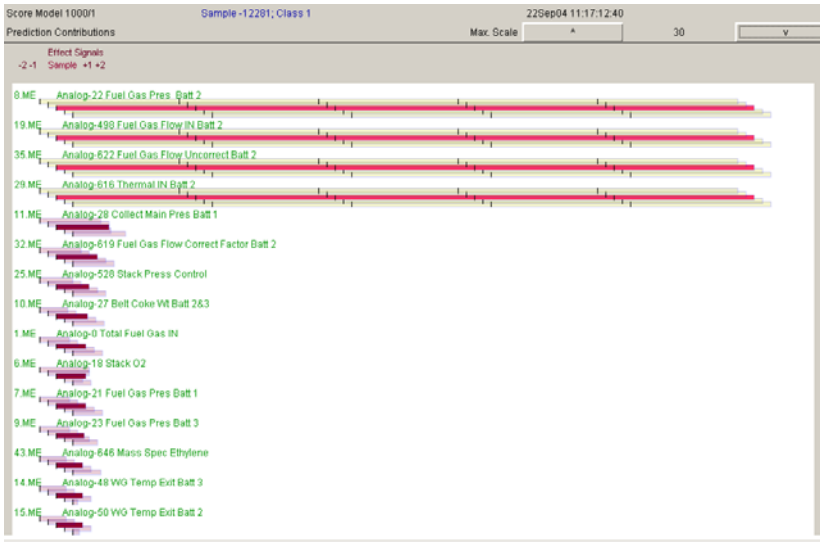

Figure 5 Score Prediction Contribution chart showing the only four variables responsible for the large excursion in the T^2 chart

e) Inspection of Variables

The 4 variables identified in the score contribution chart were inspected for that duration. Figure 6 shows the trends, in a univariate SPC chart, for the above variables with their sudden and large drop in value.

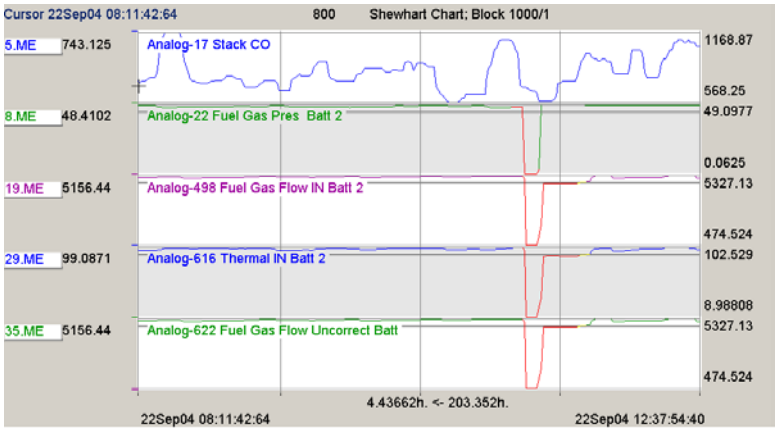

Figure 6 A univariate SPC chart showing trends of the only 4 variables responsible for the large excursion

The cause of this behaviour was then investigated and confirmed with the process engineers to be related to the instrumentation. The fuel gas flow into Battery 2 (19.ME) is controlled with a control valve. There are two different pressure transducers at either side of the valve. One of them measures the fuel gas pressure at the inlet to the valve (8.ME) and the other measures the pressure at the outlet of the valve, which is at the inlet to Battery 2. These two signals are fed to a flow meter that measures the pressure drop across the valve by subtracting the two pressure values from each other and calculating the flow rate. In this situation the valve inlet sensor (8.ME) was removed for maintenance, so it gave an approximately zero reading. The other three signals 19.ME, 29.ME and 35.ME, which are directly calculated from 8.ME, would also disappear. Hence, it was confirmed that the data in this region was abnormal and must be removed.

When a fault or abnormal process condition was confirmed, such as the one discussed above, it was masked out from the training data using the data selector tool provided within the MonitorMV software.

Input Variable Selection

The input variable reduction was necessary in order to reduce model complexity, which affects the robustness (consistent performance on new and unseen data) of the model (10). The disadvantage of having a large number of variables is that it:

- would be more likely to cause overfitting of the model, resulting in a poor generalisation on new data
- increases the risk of more sensors malfunctioning and misleading the model predictions
- increases the likelihood of a parameter falling outside the operating envelope established during training.

The presence of strong correlations in the data is needed to identify an accurate model of the process. Before carrying out any model development the correlations that exist between the process variables needed to be examined. After selecting the representative training data, the PCA model discussed above was retrained and a correlation analysis for the fifty nine variables was computed.

In conjunction with process engineers the coke making process was analysed and relationships between process variables and emissions were investigated. This knowledge of the process was needed to complement the correlation analysis.

The process variables with high cross-correlations (greater than 0.9) were grouped and only the most highly correlated variable with the output (Stack CO) was selected. This procedure was carried out so that the selected process variables will be more widely differing, hence providing information on the influence of more of the variables on Stack CO. After removing the highly cross-correlated variables, 20 of the remaining input variables that had the highest correlation coefficients with Stack CO were selected. Table I shows the 20 selected input (cause) variables in the order of their correlation coefficients with the output (effect) variable – Stack CO (5.ME). It also shows the signal IDs used in the modelling software, variable descriptions and their units. The mass spectrometer variables are the measurements made on the fuel coke oven gas compositions.

Table I Input variables and emission (Stack CO) variable used in the models

Signal ID	Variable Description	Units	Correlation Coeff. with Stack CO
6.ME	Stack O2	%	-0.645
20.ME	Fuel Gas Pressure IN Battery 2	MBAR	0.358
52.ME	Mass Spec Air Required	%	0.338
53.ME	Mass Spec CV	CV	0.336
41.ME	Mass Spec Argon	%	-0.302
47.ME	Mass Spec Benzene	%	0.298
2.ME	Stack Temperature	DEG C	0.265
18.ME	Fuel Gas Pressure IN Battery 1	MBAR	0.239
39.ME	Mass Spec H2	%	-0.237
22.ME	Fuel Gas Pressure IN Battery 3	MBAR	-0.233
15.ME	WG Temperature Exit Battery 2	DEG C	0.224
16.ME	WG Draught Exit Battery 2	MBAR	0.22
37.ME	Mass Spec CO	%	0.21
51.ME	Mass Spec Propane	%	-0.188
43.ME	Mass Spec Ethylene	%	0.187
45.ME	Mass Spec O2	%	-0.178
7.ME	Fuel Gas Pressure Battery 1	MBAR	0.175
49.ME	Mass Spec Propylene	%	-0.158
31.ME	Fuel Gas Flow Correction Factor Batt 1	NONE	-0.152
29.ME	Thermal Input Battery 2	GJ/HR	0.142

It was important that the computed correlation coefficient was not used in isolation as an indicator of high correlation with Stack CO and that the scatter plot was examined and process knowledge applied.

MODEL DEVELOPMENT

The aim of model development is to identify the coefficients in the equation that describes the process dynamic system (described earlier) to achieve an accurate predictive model.

To develop a reliable model, it needs to be structured correctly. For the PEMS model a linear absolute data format was necessary in order to base the predictions on absolute values rather than incremental values. As mentioned above, to have a soft sensor model that predicts the output variable from the input variables without any terms that include the measured output values, an FIR model for the process needs to be used. In order to configure the PEMS model to use this equation, the order of dynamics was set to zero. Under these conditions, a prediction interval of 1 sample (1 minute) was found to be the most appropriate. This value determines the time taken before a change in the inputs will affect the output.

To allow the model to deal with the dynamic nature of the plant data, a suitable duration (delay-spread) of predecessor samples needed to be considered. In order to find the optimum delay-spread, RLS modelling was used because it enables quicker testing of prediction performance without being concerned about the optimum number of scores, which would be the case with PLS models. RLS models were trained on the selected data using delay-spread values from 0 to 20 and for each value the models were tested on 3 days' data (August 2004), and the root mean squared error (RMSE) and correlation coefficient between measured and predicted stack CO values were recorded. Results of this investigation showed that although for training, the RMSE and correlation coefficient values improved as the delay-spread was increased, there was an optimum value of 5 for testing, which meant that 5 samples (5 minutes) in the past were considered at any one time.

PLS, which provides information on the process condition as well as predictions, was used to train the final model on the selected training data with the delay-spread of 5 determined above. The number of selected scores was reduced till the RMSE and correlation coefficient values on both training and test data were just below those obtained from the RLS model above.

In order to maximise model robustness a low number of scores that provide a cumulative contribution to describing the process of higher than 90% is recommended (10). The number of scores could not be reduced below 20 (equal to the original variables) because the total contribution was only 61%. Fewer variables with stronger correlations with stack CO and a high total contribution of scores (above 90%) would have been needed to further reduce the number of scores. This would have further improved the robustness of the model.

In some applications, it may be sensible to define several models each of which would be developed for a particular process operating condition (cluster). These models would be switched between as the plant moves from one operating regime to the other.

Finally, the developed model was evaluated on the 3 days' evaluation data (October 2004).

MODEL PERFORMANCE ASSESSMENT

Figure 7 shows the measured (blue) and predicted (brown) Stack CO values along with seven cause variables and their predictions for evaluation of the final PLS model over 24 hours data. To the right of each trace are shown the maximum and minimum values of the measured variable and Figure 8 shows the scatter plot, indicating the amount of correlation, of the measured and predicted values.

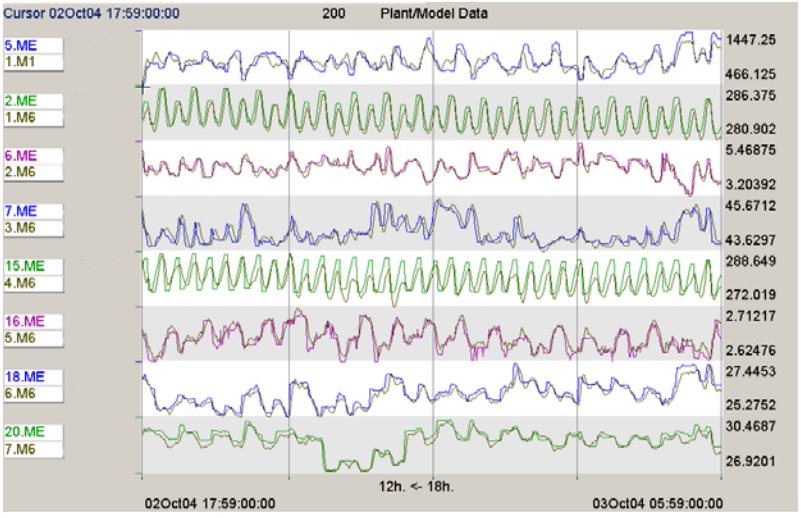

Figure 7 Measured and predicted Stack CO (5.ME and 1.M1) and other variables (see Table I) with their predictions (1.M6 – 7.M6) on twelve hours of the evaluations data

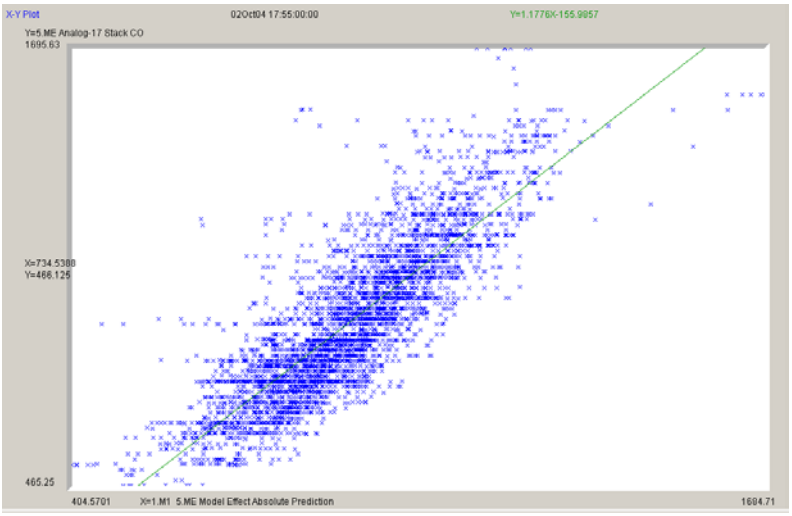

Figure 8 Scatter Plot of measured and predicted Stack CO for the complete three days of evaluation data

These Figures show that the performance of the developed models on evaluation unseen data, in predicting the Stack CO emissions, was good with correlation coefficients of higher than 0.8. Also, close examination of the results showed that the models had good robustness over the whole evaluation period.

PEMS Performance Specification Tests

Detailed performance specifications (PS-16) for PEMS are set out by the US EPA (11), which states that for a continual compliance PEMS the following factors must be considered:

a) Relative Accuracy (RA), which quantifies the deviation of the PEMS from the gas analyser measurement or Reference Method (RM) relative to the emission levels occurring at the time of the tests. For PEMS installed to meet New Source Performance Standards, the RA of the PEMS must be no greater than 10 percent when based upon the average RM data (which must be measured in the units of the emission standard). For emissions below 25 percent of the emission standard, 20 percent RA based upon the emission standard may be used. For emissions below 10 percent of the emission standard, average PEMS measurements within 2 ppm of the RM mean value constitutes an acceptable RA test.

b) Bias Test to determine if a bias exists between the RM and PEMS.

c) F-Test to determine if the RM and PEMS variances differ by more than might be expected from chance.

d) Correlation Analysis to determine how well the RM and PEMS correlate. A correlation coefficient of higher than 0.8 is required.

The US EPA performance specifications provide guidance for calculating the above factors.

In order to calculate the above factors, it is a regulatory requirement that a minimum of 27 sets of PEMS and RM data with the same duration are recorded for each test period (9 at low, 9 at normal and 9 at high emission levels). In this work, 10 representative samples of data sets, one hour duration each, were considered from each day for low, normal and high CO emission levels. The RM tests were carried out using a calibrated on-line CO gas analyser.

Table II shows the means and standard deviations of prediction error, measured and predicted values, and the confidence coefficient.

Table II Means and Standard deviations

Emission Level (ppm)	Mean of Error Means (ppm) $\overline{d}$	Standard Deviation of Error Means (ppm) S_d	Confidence Coefficient CC	Mean of Measured values (ppm) $\overline{RM}$	Mean of Predicted values (ppm)	Standard Deviation of Mean Measured (ppm)	Standard Deviation of Mean Predicted (ppm)
Low	-25.50	29.93	21.41	813.96	839.46	54.05	66.20
Normal	7.99	61	43.92	927.30	919.31	111.35	68.62
High	49.44	34	24.41	1055.08	1005.64	78.85	63.90

Table III shows the results of the calculations of RA and the required statistical tests for the PEMS model. The results show that, the evaluation RA values are less than 10%, correlation coefficients greater than 0.8 and F-values well below the critical F-value of 3.197, which are well within the required US EPA limits. A bias correction adaptor would be required if the emission levels deviate from the normal level values.

Table III PEMS Performance Specification Test Results

Emission Level (ppm)	Relative Accuracy RA (%)	Correlation Coefficient	Bias Test	F-value (critical value = 3.197)
Low	5.76	0.89	Needs bias correction	1.50
Normal	5.60	0.91	OK	0.38
High	7.00	0.88	Needs bias correction	0.66

ON-LINE APPLICATION

On-line tests have shown that the models can be applied for real-time prediction of CO emissions from the Corus coke oven plant. Alarms can be set to indicate when data is detected to be outside the training region or when there are large prediction errors using the T^2 and SPE charts.

Another model, based on the multivariate techniques discussed above, has been applied to on-line NO_x emission monitoring in a 300 t/hr steel slab reheating furnace at SSAB Tunnplåt AB, Borlänge. Both the dynamic variations and the mean level of the NO_x concentration are predicted using a linear FIR model. Good correlation between predicted and measured NO_X levels has been obtained from the modelling of the furnace parameters. Bias correction has been applied to the non-zero mean of the prediction error by using a bias adaptor. The application is also being further developed to control the NO_x emissions. Details of this furnace and the on-line system are given in a previous publication (8)

CONCLUSIONS

As part of a multi-partner European research project (RFCS), aimed at reducing emissions from iron and steel making processes, PEMS models have been developed to provide a cost-effective alternative to CEMS. A procedure has been presented for the development of PEMS models to predict CO emissions from a Corus coke oven plant in the UK.

PCA was applied to select the best training data. The suspect data were investigated by inspecting the score contour charts, T^2 chart, SPE chart and the score contribution chart. The variables were then inspected in conjunction with the process engineers. In this way, all the unacceptable data were removed and care was taken not to remove any useful data.

PCA was then used on the selected good training data and a correlation analysis was carried out. The relevant variables for predicting the stack CO were selected by examining the correlation coefficients, scatter plots and applying process knowledge.

Linear FIR RLS models were then trained and tested to determine the optimum delay-spread in order to deal with the dynamic nature of the plant data. This enabled quicker testing without being concerned with the optimum number of scores. Finally, PLS models were trained, based on the RLS model results, and the optimum number of scores were determined.

Performance specification test results showed RA values lower than 10%, F-values lower than the critical F-value and correlation coefficients between RM and PEMS values of greater than 0.8. The bias test showed that a bias correction adaptor may be required to correct for the non-zero mean of the prediction error. Although these results are well within the US EPA requirements, the model robustness could be further improved by a more complete coverage of the different process variations. A lower number of scores in the PLS model with a higher cumulative contribution to process description would also improve the model robustness.

On-line tests have shown that these models can be applied for real-time prediction of CO emissions from the Corus coke oven plant. In addition, the application in the SSAB Tunnplåt AB reheat furnace (Sweden) is being further developed to control the NO_X emissions.

ACKNOWLEDGEMENTS

The authors would like to thank the European Commission for their support to enable the work carried out with a financial grant from the European Research Fund for Coal and Steel as part of a project, *Emissions reduction through analysis, modelling and control*, with contract number *RFS-CR-03001*.

REFERENCES

1. UK Environment Agency, "Coke, Iron and Steel Sector Guidance Note IPPC S2.01," Issue 1, June 2004.

2. UK Environment Agency, "Hot Rolling Sector Guidance Note IPPC S2.04," Issue 1, February 2004.

3. A. Salgado, R. Folly, B. Valdman and F. Valero, "Model Based Soft-Sensor for On-Line Determination of Substrate," *Applied Biochemistry and Biotechnology*, vol. 113, no. 1-3, pp. 137-144, April 2004.

4. M. Brdys and Z. Lisiak, "Robust On-Line Monitoring of Mixing Quality in Water Supply Networks," *Water Resources Planning and Management Conference 99 - Preparing for the 21st Century*, Proceedings of the 26th Annual Water Resources Planning and Management Conference held in Tempe, Arizona, June 6-9, 1999.

5. N. Wickström, M. Taveniku, A. Linde, M. Larsson and B. Svensson, "Estimating pressure peak position and Air-Fuel ratio using the ionization current and artificial neural networks," *Proceedings of IEEE Conference on Intelligent Transportation Systems (ITSC'97)*, Boston, November 9-12, 1997.

6. J. J. Macak, III, "The Pros and Cons of Predictive, Parametric, and Alternative Emissions Monitoring Systems for Regulatory Compliance," *AWMA paper 96-WP92.02, presented at the AWMA 89th Annual Meeting and Exhibition*, Nashville, TN, June 1996.

7. H. H. Yue, S. Valle, S.J. Qin and J. Ghosh, "Comparison of Several Methods of Multivariate Soft Sensor for Emission Monitoring," *The University of Texas, Austin*, USA, May 1998

8. M. Saiepour, J. Niska, B. Leden, N. Link, J. Gomes and I. Unamuno, "Soft Sensing Techniques for Emissions Monitoring and Control in the Steel Industry," *CLEAN AIR 2005 - Eighth International Conference on Energy for a Clean Environment*, Lisbon, Portugal, 23 - 27 June 2005, Paper No. 25.24

9. J. MacGregor and T. Kourti, "Statistical process control of multivariate processes," *Control Engineering Practice*, 3(3), pp. 403-414, 1995.

10. P. Bonissone and K. Goebel, "When will it break? A Hybrid Soft Computing Model to Predict Time-to-break Margins in Paper Machines," *Proc. SPIE 2002*, pp. 53-64, Aug. 2002, Seattle, WA.

11. Performance Specifications for Predictive Emission Monitoring Systems (PS-16), *Federal Register*, Vol. 70, No. 151, pp. 45608-45625, August 8, 2005, Proposed Rules, expected to be finalised after Summer 2006.