Pollutant air emissions prediction by process modelling in the iron and steel industry

Deepak Bhuyan*, Suchismita Swain, Narayan Parida

Department of Mechanical Engineering
Gandhi Institute for Education and Technology, Baniatangi, Bhubaneswar

Mr. Rakesh kumar sahu, Department Of Mechanical Engineering,
Capital Engineering College, Bhubaneswar

Abstract

Monitoring air pollutant emissions of large industrial installations is necessary to ensure compliance with environmental legislation. Most of the available measurement techniques are expensive, and measurement conditions such as high-temperature emissions, difficulty of access, are often difficult. That is why legislation can not impose a permanent emission monitoring in many countries. The possibility to replace it with predictive models based on the routine measurements of the main control parameters of the installation is analysed in this paper. In order to identify these models, a special measurement campaign of emissions must be performed or, alternatively, a deterministic modelling of the process can be developed. This study was carried out in the case of a real installation in the steel industry i.e. a billet re-heating furnace. Physicalphenomena involved in combustion within the furnace were complex enough to prefer an empirical *black-box* modelling of the furnace overa deterministic approach. A 3-week monitoring campaign of fume emissions at the stack was performed; furnace process parameters duringthe same period were available. The relationship between CO_2 emissions and furnace process parameters could successfully be expressed linearly, while NO_2 emission modelling required a non-linear model. Artificial neural networks modelling revealed a good ability to predictNO_2 and CO_2 emissions.

Keywords: Fume emissions; NO_2; CO_2; Steelworks process modelling; Artificial neural networks; Multiple linear regression; Correlation method

1. Introduction

Environmental legislation is not uniform across the world, or even within the EU Community. Uniform emission standards are not generally adopted. Some countries sustain the air quality standards, e.g. in the United States air quality standards are based on the belief that the purpose of pollution control is to prevent targets from being put at risk. Viewed in this way, emission standards will therefore vary from place toplace. Uniform emission standards follow the precautionary principle that we should emit the least possible quantity of pol-lutant that best available technology (BAT) permits, according to the Integrated Pollution Prevention and Control (IPPC) Directive in the EU.

The United States use a combination of air pollution source regulations (e.g. power plants) and general ambient air quality and they require permanent continuous emission monitoring devices.

At present, in France, the most important installations are held to declare their releases; according to this declaration and to the current legislation, the installation's owner has to pay a tax on air pollution. The French environmental legisla- tion offers four main possibilities to evaluate the releases to be declared. The first method is the permanent monitoring of emissions; although the most precise, this is in general not possible because measurement techniques are expensive and measurement conditions are often difficult, due to the high-temperature of emissions and to the difficulty of access.

In the absence of permanent monitoring, three other alternatives are accepted. The most commonly-used are the mass balance or the estimation based on an emission factor, characteristic of each type of installation and each pollutant. According to their complexity, the mass balance can not be used for all industrial processes, but only for combustion (BOMET, 1991); this approach is based on the chemical reactions involved in combustion. The last possibility is the so-called *correlation method*, based on a supplementary campaign of measurements permitting to establish relevant correlations between emissions and process parameters. These correlations are then used for emission evaluation from real-time monitor- ing of the control installation parameters.

This paper focuses on the development of a *correlation method* in the case of a real installation of the French iron and steel industry i.e. a billet re-heating furnace. This study was performed within the frame of the AI/EX[1] European Pro- ject (see Schofield et al., 2002).

Section 2 presents a general overview of the re-heating furnace and the database built with the process parameters which are continuously monitored and stored for the furnace optimisation and control, and with the results of a monitoring campaign specially designed to assess measurements of several fume parameters at the stack. An intermediate stage, data pre-processing, is described in Section 3, while emission modelling is presented in Sections 4 and 5: first, a linear, simpler modelling in Section 4, giving satisfactory results for CO_2, and then, in Section 5, a more complex non-linear modelling of CO_2 and NO_2, using artificial neural networks. Finally, the possibility to use the methodology presented in Section 5 as a *correlation method* for emission estimation is discussed.

2. Data collection

2.1. Re-heating furnace description

The billet re-heating furnace of a merchant and strip bar mill studied here is a high-capacity one (150 billets hour^{-1} of 600 12 12 cm^3), with 3 heating zones (3 16 burners), resulting in a 30 MW total power. Globally, the volume inside the furnace is about 24 m long, 7 m large and 1.5 m high, while the stack is 35 m high. The heating efficiency (60%) is assumed to be a good value. Operation is discontinuous, depending on the billet production. The fuel used in combustion is Groningen (the Netherlands) natural gas. A furnace layout is presented in Fig. 1 and the list of the monitored parameters, in Table 1.

There is a gas circuit (G) providing the combustion gas for zones 1, 2 and 3. Combustion air e circuit (A) e consists of fresh air re-heated in the heat exchanger HE (for a better combustion efficiency) and then distributed to zones 3, 2 and 1. Fume e circuit (F) e is collected from all three zones and passes through the heat exchanger HE, releasing heat to the fresh air, further used as combustion air; the fume is finally

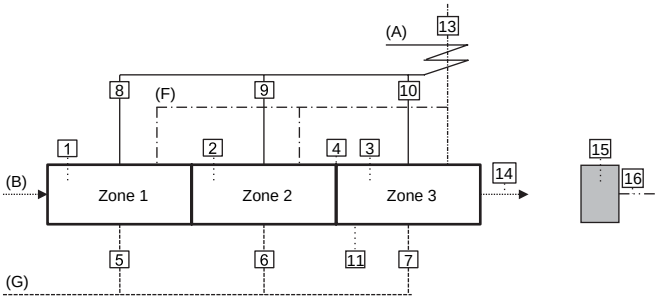

Fig. 1. Billet re-heating furnace layout.

released into the atmosphere. The billets, about 1 ton each, are initially at ambient temperature. They are introduced in the furnace e circuit (B) e where they move at a 12 m s^{-1} speed, passing through zones 1, 2 and 3; when leaving the furnace, the billets have a temperature of 800e1000 °C. The first zone, is a pre-heating one; here, temperature varies between 870 and 1200 °C. The second zone plays the role of a temperature equaliser (at about 1200 °C) and the third one, is the ''high fire'' zone; here, temperature is quasi-stable (about 1200 °C with variation of maximum 40 °C). The com- bustion is controlled for stoechiometric proportions; there is an oxygen analyser between zones 2 and 3 in order to obtain an optimal ratio air/gas flows. Oxygen injection (at the furnace temperature) was designed for this purpose. Combustion is performed with air in excess.

Emissions in fume are traditionally not monitored, while the furnace is not equipped with a gas analysis system at the stack. A gas analysis system coupled to a data sampler was then implemented to monitor gas pollutants from air emissions (point 13) during the measurement campaign specially designed for this study's purposes and realised by the LECES[2], the laboratory that is in charge of the data collection from the society operating the furnace, and which performed the additional measurements. Thus, an emission database was built, containing the concentrations of O_2, SO_2, NO_2, CO and CO_2 in fume. The process parameters, continuously monitored for optimisation and control, are available from the plant process database.

2.2. Furnace database

Process parameters are 1-min sampled and the same sampling period was adopted for the emission monitoring, which lasted for three weeks. Process parameters (21 variables) and gas emissions were gathered, leading to 25,540 recordings of a 25-variable database. Unfortunately, the 26th variable, CO was not included in the database, because it was truncated by the analyser's range.

3. Data pre-processing

Data pre-processing can improve the quality of the data, having a significant effect on model performance or on data mining results. There are a number of pre-processing

[1] Artificial Intelligence/EXpert systems for steelworks pollution controls, ESCS-STEEL C, 7210-PR/076.

[2] LECES Environnement, Voie Romaine, Maizières-les-Metz, France.

Table 1
Re-heating furnace monitoring parameters and their basic statistics: mean, standard deviation, minimum and maximum

Location	Parameter	Symbol	Statistics			
			Mean	Standard deviation	Minimum	Maximum
1/2/3	Temperature inside the furnace (°C) e zone 1/2/3	T_1	1110	42.3	870	1197
		T_2	1199	30.3	1017	1278
		T_3	1195	14.7	1089	1237
1/2	Pressure inside the furnace (mbar) e zone 1/2	P_1	9.64	1.13	1.30	14.32
		P_2	10.62	1.42	0.87	16.55
4	Oxygen concentration (%) e between zone 2e3	$O2_{23}$	2.73	2.18	0.05	14.24
	Oxygen temperature (°C) e between zone 2e3	TO_2	1209	28.3	1070	1276
5/6/7	Gas flow burners ($Nm^3\ h^{-1}$) zone 1/2/3	Qg_1	916	292.6	246	1190
		Qg_2	961	435.8	304	1785
		Qg_3	194	103.4	63	4217
8/9/10	Air flow burners ($Nm^3\ h^{-1}$) zone 1/2/3	Qa_1	8118	2564	1080	10,920
		Qa_2	8541	3841	2580	16,880
		Qa_3	1719	922	232	4217
11	Temperature under the furnace (°C) n°1/n°2	Tsf_1	87	12.5	63	128
		Tsf_2	112	19.3	82	191
12	Temperature of the combustion air after the heat exchanger (°C)	Tac	294	48	153	384
	Fume temperature after the heat exchanger (°C)	Tse	391	35	159	554
13	Fume temperature at surface outlet (°C)	Tf_{sf}	689	61	499	857
	Oxygen in fume e concentration (%)	O_2	7.38	1.87	3.7	15.7
	SO_2 in fume e concentration (mg Nm^{-3})	SO_2	13.19	15.05	1	122
	NO_2 in fume e concentration (mg Nm^{-3})	NO_2	71.70	30.13	4.6	195.5
	CO_2 in fume e concentration (%)	CO_2	7.97	1.09	3.3	10.1
14	Billet production time between two consecutive billets (s)	CAD	49.5	11.98	29	99
15	Motor intensity	e				
16	Billet surface temperature (°C)	TB	860	103	800	1084

Parameters at points 1e12 and 14e16 concern the process and they are measured continuously for the furnace control (process parameters). At point 13, measurements of emissions in fume are not traditionally realised, they were carried out during the monitoring campaign only.

techniques, such as data cleaning, data transformation (normalisation), data reduction etc.

3.1. Data cleaning

Determination of outliers is one of the first steps in cleaning the database. In the present study, some outliers were identified to instrument calibration periods (characterised by abnormal peaks) or to instrument drift during the stand-by regime of the furnace (furnace is not working during the week-end).Data was checked also for other inconsistencies (discrep- ancies) which were also removed, resulting in more missing data in the initial database, which was already incomplete.

3.2. Missing data handling

Almost all large empirical data sets suffer from periods of missing data caused by instrument failure or human error. A lot of attention has been given to this problem by researchers (Kolehmainen et al., 2001; Schlink et al., 2003; Elkamel et al., 2001; Abdul-Wahab and Al-Alawi, 2002; Andretta et al., 2000) and various treatments have been proposed to recon- struct missing data. These treatments always involve some modelling of the data.

Thus, in order to minimise the effect of *a priori* assumptions on the results, in this work, a discrimination between short and long gaps was done according to the dynamics ofthe different variables. Short and long gaps were then treated differently. For short gaps (4 consecutive missing values at most), missing data were replaced by linear interpolation. For longer gaps, data was not reconstructed and the corre- sponding intervals were ignored.

The working database, after cleaning outliers and ignoring the recordings containing at least one missing value, was reduced from 25,540 to 13,947 recordings, corresponding to a loss of 45.4%; there is one exception however, for the

SO_2 variable, where only 6763 values remained after data cleaning.

3.3. Descriptive statistics

In any model development process, familiarity with the data is of the utmost importance. A summary of all param- eters monitored during the campaign is given in Table 1.Their variability is further discussed according to the ratio (expressed in percentage) of the standard deviation to themean.

One can notice that the standard deviation is very low with respect to the mean (std/mean less than 5%) for some parameters such as: T_1, T_2, T_3, TO_2 which are fixed to set-points, the most severely constrained being temperature in zone 3. Pressures (P_1 and P_2) are controlled too, but variations are more important than those for the temperatures (10%e15%). Gas and air flows are highly correlated; there are weaker variationsin zone 1 (about 30%), but they increase progressively to about 55% in zone 3, where the temperature is allowed very little variation. High variability can be noticed for the combustion oxygen ($O2_{23}$) (80%) and above all, for SO_2 concentration in fume (114%).

SO_2 is not correlated to the furnace variables (correlation coefficients range between 0.11 and 0.08). The high numberof values not validated and this lack of correlation can be due to a measurement problem; arguably, SO_2 levels are at or below the detectability of the analyser since natural gas does not contain appreciable sulphur (about 10 mg Nm^{-3}). In this context, there is no sense in trying to model SO_2 emissions.

4. Linear modelling of emissions

Before using any non-linear model, it is judicious to test firstly the performance of a simple, linear one. If the model output can be satisfactorily explained by a linear combination of the inputs, there is no sense in using a more complicated non-linear model.

4.1. Multiple Linear Regression

Multiple linear regression MLR (see Saporta, 1990; Agirre-Basurko et al., 2006) was carried out to find the percentage of variance of the parameter to be estimated, explained by the best linear combination of the measured variables. A good indicator of the modelling quality is the root mean squared error which can be calculated as

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_p - x_m\right)^2} \qquad (1)$$

where x_p represents the value predicted by the model, x_m the corresponding measurement and N the number of measures.

It appeared that CO_2 can be satisfactorily estimated by a linear regression (Fig. 2a); 83.4% of its variance can be explained linearly and the RMSE corresponding to this linear modelling is 0.45%; compared to the mean value of CO_2, RMSE represents a relative error of 5.58%.

Meanwhile, for NO_2 (Fig. 2b) only 51.6% of the variance can be explained linearly and the RMSE corresponding to this linear modelling is 20.55 mg Nm^{-3}; compared to the mean value of NO_2, RMSE represents a relative error of 28.66%.

Thus, for CO_2, the error estimation is below the instrument accuracy, but this is not the case for NO_2. For the latter, a non-linear modelling is necessary for a better estimation.

The previous results were obtained using all 21 furnace process variables.

When expressing CO_2 by a linear combination of the process parameters, the number of explicative variables was varied from 1 to 21 and each time the best linear combination was selected. All the parameters are not (linearly) significant for the CO_2 variability; starting from 10 variables, the R^2coefficient is rather constant. In addition, with only 5 parameters (oxygen concentration $O2_{23}$, air flow in the first zone Qa_1, gas flow in the third zone Qg_3, temperature in the first zone T_1, andfume temperature Tf_{sf}) one can explain the 80.6% obtained forthe CO_2 variance and 96% of the maximum linear variance, using all 21 parameters. The fume temperature at the furnace outlet (Tf_{sf}) can explain by itself 66.9% of the CO_2 variance. This means that the fume temperature best characterises the combustion process from which the CO_2 concentration is issued; however, the fume temperature is not a control parameter, but only the final result of the combustion process; nevertheless, it can be considered as a raw indicator of CO_2. Among the process parameters, those which are the most related to CO_2 emissions are $O2_{23}$, Qa_1, Qg_3 and T_1.

5. Neural networks modelling of emissions

5.1. Neural network modelling

Neural networks (NN) are capable of modelling highlynon-linear relationships (Gardner and Dorling, 1999, 2000; Abdul-Wahab and Al-Alawi, 2002). The greatest advantageof a neural network is its ability to model a complex non-linear relationship without *a priori* assumptions on its nature (Bu-Hamra et al., 2003).

In a comprehensive review of applications of artificial intelligence in combustion systems, Kalogirou (2003) presents a summary of 22 applications of NNs in combustion, published between 1995 and 2002. In two of them, the authors used this technique for emission monitoring; the application developed by Tronci et al. (2002) concerned the combustion chambers, while Ferretti and Piroddi (2001) estimated NO_x emissions in power plants. In (Zhou et al., 2004), the authors evaluated NO_x emissions of a coal-fired boiler by CFD and by NNs; according to the authors, the second technique was much easier than the first one. As a general conclusion of all the cited papers, NN are a good candidate for modelling complex industrial installations.

For function approximation, the most suitable architectures are considered to be the multilayer perceptron (MLP) (see

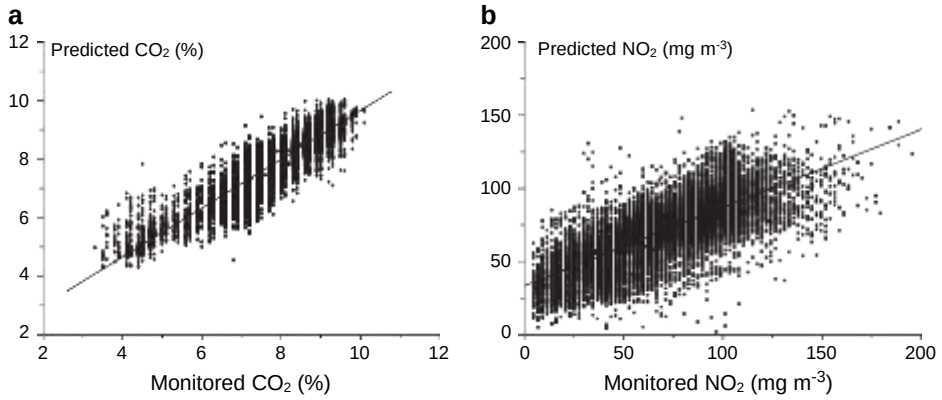

Fig. 2. (a) Predicted *versus* monitored CO_2. MLR result: predicted CO_2 is calculated *via* a linear combination of all the process parameters. (b) Predicted *versus* monitored NO_2. MLR result: predicted NO_2 is calculated *via* a linear combination of all the process parameters.

Abdi, 1994; Fausett, 1994; Bishop, 1995; Ripley, 1996), which is also the most popular architecture for NN, and the Radial Basis Neural Networks.

MLP neural networks are universal approximators (Hornik et al., 1989) i.e. they can approximate any smooth, measurable function. They also possess the remarkable property of parsimony, i.e. for a similar accuracy, NN require less fitting parameters than the universal approximators commonly used; more precisely, their number varies linearly with the number of degrees of freedom, while it varies exponentially for most other approximators (Hornik et al., 1994).

MLP non-linearity is achieved by using a non-linear activation function and by including at least one hidden layer in the network's architecture. A good description of the MLP (schematic architecture and equations) is given by Agirre-Basurko et al. (2006). The nature of the functional relationship between inputs and outputs is learnt during a supervised training process directly from the data.

5.2. Training algorithms

During the supervised training procedure, series of input and associated output data are repeatedly presented to thenetwork in order to learn to model the relationship and to ac- curately generalise when presenting new, unseen data. This learning phase corresponds mathematically to the optimisationof a cost function in the weights space; the weights character- ise the importance of each connection between neurons from different layers, and they represent the fitting parameters ofthe NN model.

The cost function is chosen according to the performance criteria. If performance is estimated in terms of prediction accuracy, this corresponds to the minimisation of a function of the error signal, defined as the difference between the desired and the actual output of the network (Schlink et al., 2003), e.g. the sum of the squared errors.

Prediction accuracy is affected by the optimisation algorithm. Unfortunately, the error surface is often complex and contains many local minima (Comrie, 1997); if the optimisation algorithm is trapped in a local minimum, the final MLP model may be sub-optimal. Global methods of optimisation are attractive, but their convergence is slow (Maier and Dandy, 1998) and their implementation in complex cases is rather difficult (Lu et al., 2003).

Typically, local optimisation methods are preferred, although the global minimum is not reached; a good local one is usually treated as an acceptable solution (Gardner and Dorling, 2000). In this study, several local optimisation algorithms were tested.

5.2.1. Generalisation ability

Generalisation ability is defined as the model's potential to perform well on data that were not used to calibrate it (Cheng and Titterington, 1994). For the purpose of forecasting, the most important property of an algorithm is its ability to generalise and filter out the noise. Overtraining occurs when the model learns the noisy details in the training data, which results in the model having poor generalisation capabilities when presented with new data.

In this study, the purpose is not to get a model reproducing as well as possible the measurements during the monitoring campaign, but a model able to predict plausible results when applied to other measurements of the furnace, thus becoming a reliable tool for emission estimation.

Generalisation ability is a function of the ratio of the number of training samples to the number of connection weights (Maier and Dandy, 2000). Different ratios are proposed empirically by (Masters, 1993; Weigend et al., 1990; Amari et al., 1997). If this ratio is too small, continued training can result in overfitting of the training data. Traditionally, optimal geometries have been found by trial and error, but a number of systematic approaches for determining optimal network geometry were also proposed, including pruning or constructive algorithms (Bebis and Georgiopoulos, 1994).

In order to avoid overtraining, and to restrict model complexity, a regularisation technique, known as early stopping can be used. Early stopping (ES) is based on dividingthe data into 3 subsets (Gardner and Dorling, 1999). The first subset is the training set used for computing the gradient and updating the network parameters; the second subset is the validation set. The error on the validation subset is monitored during the training process. The validation error will normally

decrease during the initial phase of training, as does the training set error; when the network begins to overfit the data, the error on the validation set typically begins to rise, then training is stopped and the network parameters at the minimum of the validation error are returned. Finally, the performance of the network is tested on the third subset, the test set.

Bayesian regularisation (BR) (Buntine and Weigend, 1991) is another theoretical robust mechanism which can be used to avoid network overfitting, whilst allowing the available data to be split into only two sets, the training set and the test set. A regularisation term is incorporated into the cost function in order to penalise overly complex models (Dorling et al., 2003).

It is possible to improve generalisation modifying the performance function by adding a term that consists of the mean of the sum of squares of the network weights and biases. Using this performance function weights and biases are smaller, forcing the network response to be smoother and less likely to overfit. The problem with regularisation is that it is difficult to determine the optimum value for the performance ratio parameter i.e. too large, leads to overfitting, whereas if it is too small, the network doesn't adequately fit the training data.

In this study, several network geometries were tested varying the number of hidden layers (1 or 2) and the number of neurons in each hidden layer. For each simulation, either ES or combined BR/ES were implemented in order to avoid overtraining.

5.2.2. Data pre-processing

Generally, different variables span different ranges. In order to ensure that all variables receive equal attention during the training process, they should be standardised (Maier and Dandy, 2000). In addition, the variables have to be scaled in such a way to commensurate with the output range of the activation function (Fausett, 1994). It should be noted that if the transfer function in the output layer is unbounded (e.g. lin-ear), scaling is not strictly required, but it is still recommended (Maier and Dandy, 2000).

The most common activation functions are the sigmoidal-type ones, such as the logistic and the hyperbolic tangent, ranging between 0 and þ and 1 and 1, respectively.They are both monotonically increasing, and possess simple derivatives. Very often the identity function is used for the output layer, presenting a special interest when it is necessary to extrapolate beyond the range of the training data (Maier and Dandy, 2000).

When using the logistic sigmoid, scaling between 0 and þis recommended. Data can be transformed by determining the minimum and the maximum values of each variable over the whole data period and calculate normalised variables using the formula (Elkamel et al., 2001):

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (2)$$

Dandy, 1998). When using tanh, Gardner and Dorling (1999) propose calculating normalised variables using the formula:

$$x_{norm} = 2\frac{x - x_{min}}{x_{max} - x_{min}} - 1 \quad (3)$$

5.3. NO_2 and CO_2 modelling by NN

In this study, input and output data were normalised (Maier and Dandy, 1998, 2000):

$$x_{norm} = \frac{x - \bar{x}}{s} \quad (4)$$

The maximum absolute value of x_{norm} was 8 for NO_2. Each normalised variable was then scaled, in order to range in an interval included in ½1 þ ; scaling consisted in the division of each variable by an *S*-value determined as:

$$S = a^{-1}\max(|x_{norm}|) \quad (5)$$

where a was taken 0.8, less than 1, in order to avoid quasi-null values of the activation function's derivative.

It is important to notice that only linear transformations were applied to data *prior* to the model identification and these transformations do not influence the results of a non-linear regression. Data were later returned to original units using the corresponding inverse formulae.

Models selected for NO_2 and CO_2 prediction consisted of 3- or 4-layer feed-forward neural networks (MLP with one or two hidden layers). The first layer (the input layer) was composed of one neuron for each input (21 neurons), thelast one (the output layer), consisted of a single neuron, corresponding to the output (NO_2 or CO_2 values), while the hidden layers were composed of a variable number of neu- rons. The following activation functions have been selected: tanh for the hidden layers, and the unbounded identity func- tion (linear) for the output layer. Several training algorithms were tested for mainly two network architectures and the results are given in Table 2. To improve generalisation, the initial data set, after putting samples in a random order, was split in 3 data subsets: training (60%), validation (20%) and test (20%). Validation set could be used for early stopping. Bayesian regularisation (BR) combined with Levenberg-Marquardt (LM) algorithm was tested too for im- proving generalisation.

The best results for NO_2 correspond to the LM4 simula- tion (see Table 2) and they are presented in Fig. 3a,c,e. The network architecture consists of a 21-neuron input layer, 2 hidden layers with 40 and 20 neurons respectively and an output layer consisting of a single neuron. The training algorithm was based on the Levenberg-Marquardt optimisation method. Early stopping (ES) was implemented in order to

Hyperbolic tangent (tanh) ranges between between –1 and þ1. The gain of the hyperbolic tangent is greater than the sigmoidal one. As a result, one would expect training to be slower when the sigmoidal transfer function is used (Maier and avoid overtraining. The RMSE obtained in this case was 7.48 mg Nm^{-3} for the training set, 10.56 mg Nm^{-3} for the validation set and 10.39 mg Nm^{-3} for the test set. In terms of relative error compared to the NO_2 mean, the error

Table 2
NO_2 modelling by different NN architectures and using different training algorithms

Simulation	Architecture	Training Algorithm	Regularisation	Epochs	RMSE		
					Training	Validation	Test
GD1	21-40-1	GD	e	10,000*	21.28	21.08	21.24
GD2	21-40-20-1	GD	e	10,000*	23.86	24.04	24.32
GD3	21-40-1	GDM (momentum ¼ 0.6)	e	10,000*	21.19	21.67	21.16
GD4	21-40-20-1	GDM (momentum ¼ 0.5)	e	25,000*	18.08	17.98	17.96
GD5	21-40-1	GDA	ES	557	26.3	28.52	26.93
GD6	21-40-20-1	GDA	ES	155	22.12	22.07	21.98
GD7	21-40-1	GDX	ES	3401	17.69	17.99	17.65
GD8	21-40-20-1	GDX	ES	1588	18.58	18.72	18.42
GD9	21-40-1	RBP	ES	978	14.32	14.91	14.60
GD10	21-40-20-1	RBP	ES	716	12.55	13.60	13.10
CG1	21-40-1	CGB	ES	277	14.69	15.03	14.75
CG2	21-40-20-1	CGB	ES	242	14.23	14.67	14.48
CG3	21-40-1	CGF	ES	246	15.84	15.94	15.59
CG4	21-40-20-1	CGF	ES	276	14.46	14.97	14.60
CG5	21-40-1	CGP	ES	146	17.64	17.54	17.46
CG6	21-40-20-1	CGP	ES	288	14.45	14.96	14.73
CG7	21-40-1	SCG	ES	100	18.94	18.82	18.65
CG8	21-40-20-1	SCG	ES	74	18.72	18.90	18.32
CG9	21-60-30-1	SCG	ES	171	15.93	16.08	15.94
QN1	21-40-1	BFGS	ES	625	11.33	12.55	12.09
QN2	21-40-20-1	BFGS	ES	212	12.21	13.17	12.75
QN3	21-40-1	OSS	ES	592	15.86	16.31	15.89
QN4	21-40-20-1	OSS	ES	334	15.73	16.03	15.92
LM1	21-40-1	LM	ES	64	8.77	10.96	11.15
LM2	21-60-1	LM	ES	53	9.12	11.91	12.30
LM3	21-40-20-1	LM	ES	31	8.00	11.20	11.10
LM4	21-40-20-1	LM	ES	30	7.48	10.56	10.39
LM5	21-60-30-1	LM	ES	28	7.77	11.69	11.62
LM6	21-40-1	LM	BR/ES	48	10.96	11.94	11.79
LM7	21-40-20-1	LM	BR/ES	94	8.94	11.05	11.49

Activation functions were: tanh for the hidden layers and unbounded identity (linear) for the output layer. *Maximum number of epochs chosen for training;ES: Early Stopping; BR: Bayesian Regularisation; GD: Gradient Descent; GDM: Gradient Descent with Momentum; GDA: Gradient Descent with Variable Learning Rate; GDX: Gradient Descent with Momentum and Variable Learning Rate; RBP: Resilient Backpropagation; CGB: Conjugate Gradient withPowell-Beale Restarts; CGF: Conjugate Gradient with Fletcher-Reeves Update; CGP: Conjugate Gradient with Polak-Ribiere Update; SCG: Scaled Conjugate Gradient; BFGS: Quasi-Newton (Broyden, Fletcher, Goldfarb and Shanno algorithm); OSS: One Step Secant (Quasi-Newton secant algorithm); LM: Levenberg- Marquardt.

estimation is 10.43% for the training set, 14.72% for the validation set and 14.49% for the test set. This error is comparable to the measurement error (10e12%).

For CO_2, the results obtained by MLR were already comparable to the measurement error (5.6%). In order to improve CO_2 estimation, a NN model was tested; the NN architecture that gave the best results for NO_2 was em- ployed. The results are presented in Fig. 3b,d,f; in terms of RMSE they showed: for the training set 0.24% (relative error compared to the CO_2 mean 3%), for the validation set 0.31% (relative error 3.9%) and for the test set 0.29% (rel- ative error 3.6%). It results that the difference from 3.6% to 5.6% corresponds to a part of non-linearity in the CO_2 variation.

A joint model for NO_2 and CO_2 simultaneous simulation using a 21-40-2 architecture and the Levenberg-Marquardt algorithm with Early Stopping led to very comparable results with the previous individual models; the RMSE for NO_2 was 9.04 mg Nm^{-3} for training, 11.25 mg Nm^{-3} for validation and 10.47 mg Nm^{-3} for the test, while for CO_2, the results presented in the same order were 0.26%, 0.28% and 0.29%.

6. Discussion and conclusions

The most important point to discuss is the possibility and the advantages in using the methodology developed in this study, the so-called *correlation method*.

It was assessed by neural networks modelling that: (i) the RMSE for NO_2 modelling is about 14.5% of the NO_2 mean (71 mg Nm^{-3}), which is slightly higher than the measure-ment error 10e12% and (ii) the RMSE for CO_2 modelling is about 3.6% of the CO_2 mean (3%), sensibly lower than the measurement error. Unfortunately, the generalisation ability of the models could not be tested on other databases. One can suppose that if the working regime of the installa- tion does not change significantly, the model should give similar performances. It is important that the measuring campaign used for the model identification to be designed such as to include all the different working regimes. In these

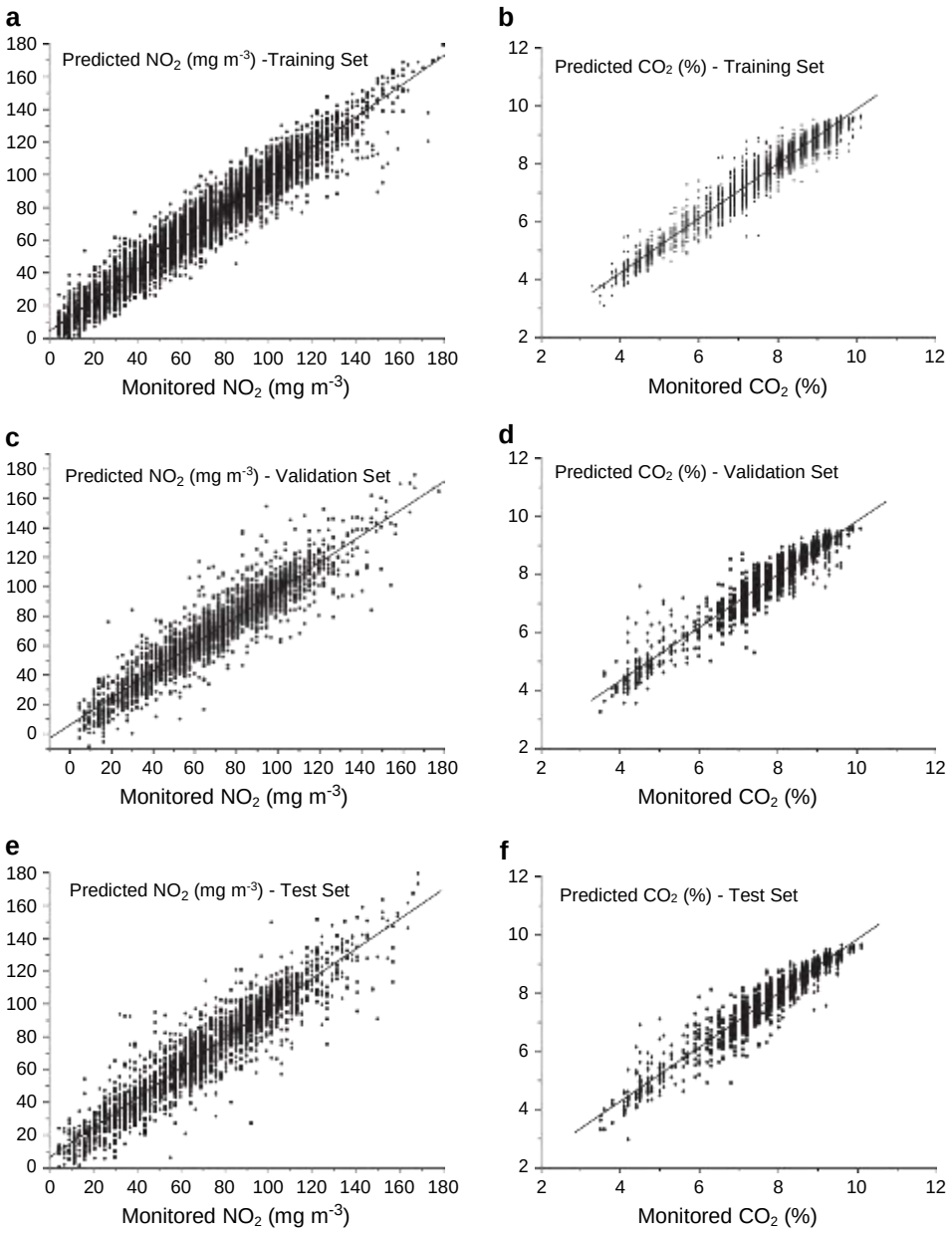

Fig. 3. Monitored and predicted NO_2 for: (a) the training (RMSE ¼ 7.48, R^2 ¼ 0.94), (c) the validation (RMSE ¼ 10.56, R^2 ¼ 0.88) and (e) the test set (RMSE ¼ 10.39, R^2 ¼ 0.88) e see Table 2, LM4. Monitored and predicted CO_2 for: (b) the training (RMSE ¼ 0.24, R^2 ¼ 0.95), (d) the validation (RMSE ¼ 0.31, R^2 ¼ 0.92) and (f) the test set (RMSE ¼ 0.29, R^2 ¼ 0.93).

conditions, the neural networks modelling can be considered a reliable *correlation method* as well for NO_2 as for CO_2. Moreover, in the case of CO_2, a simple linear model gives less efficient results than the neural networks (5.6%), but is still comparable to the measurement error.

The other two methods accepted by the French environ- mental legislation based on the global balance of the whole installation or on the emission factor are either too difficultto perform (the first case) or provide a very global estima- tion (the second one). An example of comparison is pro- posed here between NO_2 measured emissions and those calculated using the emission factor. The emission factor proposed by the present legislation for NO_2 in the case of a re-heating furnace in the steel industry is 170 g equivalent NO_2 per ton of steel (BOMET, 1991). Using the gas flow values and the combustion coefficients characteristic to the gas burnt here, one can calculate the fume flow and then,the total quantity of NO_2 from their measured concentra- tions; it results in about 350 kg of NO_2 emitted during 233 h when the furnace was working. From the billet pro- duction, the total mass of steel can be estimated, and consid- ering 170 g NO_2/ton of steel, it results in about 1800 kg of NO_2, which represents about 5 times more than the measure- ments. By comparison, over the same period, emissions cal- culated by the best available NN amounted to a result closer than 1% to measurements, because the algebraic mean error monitoring-prediction is very low: 0.12 mg Nm^{-3} for thetraining set, 0.37 mg Nm^{-3} for the validation set, and 0.09 mg Nm^{-3} for the test set. The quasi-null value of the mean error reveals that this method is very successful for calculating global emitted amounts, i.e. emissions over a lon-ger period.

One can notice that the modelling developed in this study was not parsimonious. Indeed, all the variables were used as inputs, for the following reasons:

- all of them are permanently available (for the processcontrol);
- the main purpose was to achieve an estimation of the emis-sions as good as possible;
- input selection was not crucial, as neural networks belong to the class of data-driven approaches (Maier and Dandy, 2000).

However, this strategy also presents some drawbacks. The network's size increases artificially; consequently, processing speed decreases and the amount of data required to estimate the connection weights efficiently becomes larger. Input selec-tion is important also for finding the most influential variableson the emissions and defining thus a more efficient control of the process. For this purpose, a post-modelling analysis can also be done. Indeed, neural network modelling permits to assess the importance of each of the input variables by using the network weights (Garson, 1991; Goh, 1995; Abdul-Wahaband Al-Alawi, 2002). Parameter selection before and aftermodelling in order to achieve a more efficient control on the installation and to reduce emissions is the subject for some future work.

In conclusion, in the present work, a *correlation method* based on neural networks is proposed for CO_2 and NO_2 emis- sion estimation, with an error comparable to the measurement one. The method is simpler than the global balance of the whole installation, more precise than the emission factor method and requires a short, but well-designed monitoring campaign of emissions.

Acknowledgements

The authors are grateful to all the participants to the AI/EX project and in particular to Mr. Ph. Le Louër, from LECES Environnement, for putting at our disposal all themeasurements, for the fruitful collaboration between the LECES and the CERTES, and for the financial support.

References

Abdi, H., 1994. Les réseaux de neurones. Presse Universitaire de Grenoble.

Abdul-Wahab, S.A., Al-Alawi, S.M., 2002. Assessment and prediction of tropospheric ozone concentration levels using artificial neural networks. Environmental Modelling & Software 17, 219e228.

Agirre-Basurko, E., Ibarra-Berastegi, G., Madariaga, I., 2006. Regression and multilayer perceptron-based models to forecast hourly O_3 and NO_2 levels in the Bilbao area. Environmental Modelling and Software 21,430e446. Amari, S.I., Murata, N., Müller, K.-R., Finke, M., Yang, H.H., 1997. Asymp- totic statistical theory of overtraining and cross-validation. IEEE Transac- tions on Neural Networks 8 (5), 985e996.

Andretta, M., Eleuteri, A., Fortezza, F., Manco, D., Mingozzi, L., Serra, R., Tagliaferri, R., 2000. Neural networks for sulphur dioxide ground level concentrations forecasting. Neural Computing & Applica- tions 9, 93e100.

Bebis, G., Georgiopoulos, M., 1994. Feed-forward neural networks: Why network size is so important. IEEE Potentials, 27e31. October/ November.

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

BOMET, 1991. Circulaire du 24 décembre 1990 relative aux installations classées pour la protection de l'environnement. Taxe parafiscale sur la pollution atmosphérique (BOMET n° 588-91/14 du 20 mai 1991). avail- able at <http://aida.ineris.fr/cadre_rech.htm> (Page consulted on July 19, 2006).

BuHamra, S., Smaoui, N., Gabr, M., 2003. The Box-Jenkins analysis and neu-ral networks: prediction and time series modelling. Applied Mathematical Modelling

27 (10), 805e815.
Buntine, W.L., Weigend, A.S., 1991. Bayesian back-propagation. Complex Systems 5, 603e643.
Cheng, B., Titterington, D.M., 1994. Neural networks: A review from a statis-tical perspective. Statistical Science 9 (1), 2e54.
Comrie, A.C., 1997. Comparing neural networks and regression models for ozone forecasting. Journal of the Air and Waste Management Association 47, 653e663.
Dorling, S.R., Foxall, R.J., Mandic, D.P., Cawley, G.C., 2003. Maximum likelihood cost functions for neural networks models of air quality data. Atmospheric Environment 37, 3435e3443.
Elkamel, A., Abdul-Wahab, S., Bouhamra, W., Alper, E., 2001. Measure- ment and prediction of ozone levels around a heavily industrialized area: a neural network approach. Advances in Environmental Research5, 47e59.
Fausett, L., 1994. Fundamentals of Neural Networks. In: Architectures, Algo- rithms and Applications. Prentice Hall, Englewood Cliffs, NJ 07632.
Ferretti, G., Piroddi, L., 2001. Estimation of NO_x emissions in thermal power plants using artificial neural networks. Journal of Engineering Gas Turbines Power: Trans ASME 123 (2), 465e471.
Gardner, M.W., Dorling, S.R., 1999. Neural network modelling and predictionof hourly NO_x and NO_2 concentrations in urban air in London. Atmo- spheric Environment 33, 709e719.
Gardner, M.W., Dorling, S.R., 2000. Statistical surface ozone models: an improved methodology to account for non-linear behaviour. Atmospheric Environment 34, 21e34.
Garson, G.D., 1991. Interpreting neural-network connection weights. AI Expert 6 (7), 47e51.
Goh, A.T.C., 1995. Back-propagation neural networks for modelling complexsystems. Artificial Intelligence in Engineering 9, 143e151.
Hornik, K., Stinchombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Networks 2, 359e366.
Hornik, K., Stinchombe, M., White, H., Auer, P., 1994. Degree of approxima-tion results for feedforward networks approximating unknown mappings and their derivatives. Neural Computation 6, 1262e1275.
Kalogirou, S.A., 2003. Artificial intelligence for the modeling and control of combustion process: a review. Progress in Energy and Combustion Science29, 515e566.
Kolehmainen, M., Martikainen, H., Ruuskanen, J., 2001. Neural networks andperiodic components used in air quality forecasting. Atmospheric Environ-ment 35, 815e825.
Lu, W.Z., Fan, H.Y., Lo, S.M., 2003. Application of evolutionary neural network method in predicting pollutant levels in downtown area of Hong Kong. Neurocomputing 51, 387e400.
Maier, H.R., Dandy, G.C., 1998. The effect of internal parameters and geom- etry on the performance of back-propagation neural networks: an empiricalstudy. Environmental Modelling & Software 13, 193e209.
Maier, H.R., Dandy, G.C., 2000. Neural networks for the predictionand forecasting of water resources variables: a review of modelling issues and applications. Environmental Modelling & Software 15, 101e124.
Masters, T., 1993. Practical Neural Network Recipes in Cþþ. Academic Press, San Diego, CA.
Ripley, B.D., 1996. Pattern Recognition and Neural Networks. Cambridge University Press.
Saporta, G., 1990. Probabilités, analyse des données et statistique. Editions Technip.
Schlink, U., Dorling, S., Pelikan, E., Nunnari, G., Cawley, G., Junninen, H., Greig, A., Foxall, R., Eben, K., Chatterton, T., Vondracek, J., Richter, M., Dostal, M., Bertucco, L., Kolehmainen, M., Doyle, M., 2003. A rigorous inter-comparison of ground-level ozone predictions.Atmospheric Environnement 37, 3237e3253.
Schofield, N., Le Louër, P., Mirabile, D., Hubner, R., 2002. Primary Steelmak-ing. In: European Commission EUR 20, 501. Artificial Intelligence/Expert (AI/EX) Systems for Steelworks Pollution Control, Technical Steel Research, 92-894-4387-1, 153 p.
Tronci, S., Baratti, R., Servida, A., 2002. Monitoring pollutant emissions in a 4.8 MW power plant through neural network. Neurocomputing 43, 3e15.
Weigend, A.S., Rumelhart, D.E., Huberman, B.A., 1990. Predicting the future: a connectionist approach. International Journal of Neural Systems 1 (3), 193e209.
Zhou, H., Cen, K., Fan, J., 2004. Modelling and optimization of the No_x emi-sion characteristrics of a tangentially fired boiler with artificial neural net-works. Energy 29 (1), 167e183.