Turk J Elec Eng & Comp Sci
(2019) : –

doi:10.3906/elk-1807-87

Predicting CO and NO_x emissions from gas turbines: novel data and a benchmark PEMS

Heysem KAYA[1]*, Pınar TÜFEKCİ[1], Erdinç UZUN[1]
[1]Department of Computer Engineering, Çorlu Faculty of Engineering
Namık Kemal University, 59860, Çorlu, Tekirdağ, TURKEY

ORCID iD (HK): https://orcid.org/0000-0001-7947-5508
ORCID iD (PT): https://orcid.org/0000-0003-4842-2635
ORCID iD (EU): https://orcid.org/0000-0003-4351-2244

Received: .201 • **Accepted/Published Online:** .201 • **Final Version:** ..201

Abstract: Predictive Emission Monitoring Systems (PEMS) are important tools for validation and backing up of costly continuous emission monitoring systems used in gas turbine based power plants. Their implementation relies on the availability of appropriate and ecologically valid data. In this paper, we introduce a novel PEMS dataset collected over five years from a gas turbine for the predictive modeling of the CO and NO_x emissions. We analyze the data using a recent machine learning paradigm, and present useful insights about emission predictions. Furthermore, we present a benchmark experimental procedure for comparability of future works on the data.

Key words: Predictive Emission Monitoring Systems, CO, NO_x, exhaust emission prediction, gas turbines, extreme learning machine, database

1. Introduction

The increasing demand for energy had a doubly negative impact on the environment both through increasing deforestation and increasing carbon and flue gas emissions. Thanks to the increased environmental awareness in both public opinion and political circles, the Paris Convention on Climate Change[1] was adopted by 196 participating nations. The convention aimed to reduce the global greenhouse emissions and already the signing parties passed rigorous environmental laws, including tax implementation for carbon emissions[2].

Being the core issue of the convention, air pollution poses a vital threat. The term "air pollutant" covers all substances which may harm living beings. The combustion processes of fossil fuels used in power plants and vehicles comprise the major portion of air pollution. NO_x ($NO_x = NO_2 + NO$) are considered the primary pollutants of the atmosphere, since they are responsible for environmental problems such as photochemical smog, acid rain, tropospheric ozone, ozone layer depletion, and eventually global warming [1]. In addition to these environmental catastrophes, they cause various health problems in humans exposed to high concentrations of these gases [1].

An important source of harmful pollutants (NO_x and CO) released in the atmosphere is the combustion

*Correspondence: kaya.heysem@gmail.com
[1]Framework Convention on Climate Change (2015). Adoption of the Paris Agreement [online]. Website http://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf [accessed 26 July 2019].
[2]Carbon Tax Center [online]. Website http://www.carbontax.org/ [accessed 26 July 2019].

process in the power industry [1, 2]. Therefore, there is a special concern on reducing the emissions from power plants. These emissions are restricted within certain limits by rigid environmental rules in different parts of the world [3]. European Union (EU) restricts the flue gas emissions (i.e. NO_x, CO and dust) by Large Combustion Plant (LCP) directive and its replacement Industrial Emissions Directive (IED) that took effect in 2016 for power plants having higher than 50 MegaWatt (MW) power. According to IED, the flue gas concentrations of NO_x, and CO must be measured continuously from each combustion plant exceeding a total capacity of 100 MW [2]. Moreover, NO_x and CO emissions are limited to 25 ppmdv (parts per million by dry volume) by the EU when natural gas is used as fuel [4].

It is important to monitor the NO_x and CO pollutants emitted during combustion operations in a power plant. To this end, three solutions are developed to monitor flue gas emissions from a combustion unit: i) periodic measurements, ii) CEMS (Continuous Emission Monitoring System) or iii) PEMS (Predictive Emission Monitoring System), respectively. Periodic measurements are typically performed with calibrated equipments by emission testing laboratories, with moderate costs. In CEMS, the emission monitoring equipment (e.g. sensor set) is installed on-site. CEMS provides real-time information on emissions gathered directly from sensors. The validity of these measurements depend on proper maintenance and calibrations that should be conducted according to the standard procedures. PEMS, on the other hand, is an expert system that takes as input some process variables (e.g. ambient temperature, turbine inlet temperature, etc.) and uses predictive models trained on past data to estimate emission components. Despite increasing research, there is no accepted CEN[3] standard for PEMS yet, but preparations are known to be underway by CEN/TC 264/WG 37 [2].

In this paper, we present the following contributions on PEMS. We first introduce a novel flue gas emission dataset collected from five years (from 01/01/2011 to 31/12/2015), in order to boost research on PEMS. In terms of time span, this is the largest dataset collected for this task. Using this dataset, we develop a benchmark PEMS using an extreme learning machine (ELM) regressor. This machine learning paradigm is applied to the PEMS problem for the first time in the literature. We further provide a clear experimental procedure to ensure repeatability of our work, and to allow fair comparisons with alternative approaches. The predictions of the PEMS are generally used to validate the outputs of CEMS during regular operation. Thus PEMS can be configured to raise maintenance triggers when the absolute difference between the predictions and the sensor measurements are higher than a predefined threshold. In addition to this, PEMS can be used to estimate the periods of CEMS maintenance and can serve as backup of CEMS during failure/maintenance.

The paper is organized as follows. In the next section, we provide a literature review and some background on the learning method we have used. The collected dataset is introduced in Section 3 with a brief statistical analysis of its features. Experimental results are presented in Section 4. Section 5 discusses our findings, and concludes with future directions.

2. Background and Related Work

In this section, we provide background on flue gas emission monitoring systems and on the machine learning paradigm we have used for prediction.

2.1. Related Work

PEMS are used for continuous monitoring of emissions at stationary sources as an alternative and backup for CEMS. Both in practice and in the literature, they are mainly used to predict NO_x emissions from combustion

[3]European Committee for Standardization [online]. Website https://www.cen.eu [accessed 26 July 2019].

processes [1]. PEMS define a mapping between a set of characteristic process parameters of an emission source (e.g. air pressure, turbine after temperature) and the corresponding flue gas emission. If the variables of the combustion process are continuously monitored and recorded, it is possible to predict the emission concentration after combustion. PEMS are plant-specific emission monitoring systems and vary in terms of methodology and design, ranging from relational models to neural network models.

Computational models based on process history (i.e. measurement data) are typically used as the central component of PEMS. For example, methods such as neural networks (NN) are widely used as the basis of predictive emission monitoring systems [2, 3, 5–10]. Recently, Korpela et al. present models based on linear regression and nonlinear neural networks, to estimate NO_x emission in two similar natural gas-fired municipal hot water boilers [2]. Dynamic neural networks are used in [3] to develop soft sensors for the NO_x and O_2 emission due to combustion operation in industrial boilers. The early use of NN to predict real-time engine emissions (HC, NO_x, CO, and CO_2) from process variables dates back to 1999 [5]. This work is followed by an NN based NO_x reduction system [6]. In [7], a neural network structure is proposed for predicting NO_x and O_2 of an industrial boiler. Lv et al. propose a least squares support vector machine-based ensemble learning paradigm to predict NO_x emission of a coal-fired boiler using real operation data [8]. For the NO_x prediction of a coal-based boiler, Smrekar et al. [9] compare linear (AutoRegressive/Moving Average - ARX/ARMAX) and nonlinear (e. g. Support Vector Regressor - SVR) modeling approaches. However, they observe that nonlinear models do not improve over linear ARX models [9]. Ciric et al. [10], employ three NN alternatives (feed forward NN, recurrent NN, and a hybrid neuro-fuzzy estimator, respectively) for the estimation of CO_2 emissions in Serbia. Lazaretto and Toffolo [11] fit a model to experimental data for predicting performance and emissions of a two-shaft gas turbine, based on emission formulations proposed by Rizk and Mongia [12]. Dragomir and Oprea introduce a multi-agent system for monitoring urban air pollution due to power plant operation [13]. Idzwan et al. used Support vector machines for predicting NO_x emission of a power generation plant in Malaysia [14]. Liukkonen and Hiltunen presented a novel emission monitoring platform based on the self-organizing maps [15].

While numerous NN approaches were used for prediction of exhaust gas emissions, this work is the first attempt to use extreme learning machine classifiers (ELM) for this problem. We now give a brief background in ELMs.

2.2. Extreme Learning Machines

ELMs [16–18] are recently used in many machine learning applications for classification and regression, including wind-speed forecasting [19] and audio-visual emotion recognition [20, 21].

Initially, ELM is proposed as a fast learning method for Single Hidden Layer Feedforward Networks (SLFN), as an alternative to back-propagation [22]. ELM proposes the random generation of the hidden node output matrix $\mathbf{H} \in \mathbb{R}^{N \times K}$, where N and K denote the number of instances and the hidden neurons, respectively. The hidden node activation via randomly generated mapping matrix $\mathbf{W}$ and bias vector $\mathbf{b}$ is defined as in regular SLFN:

$$\mathbf{H}_{n,k} = h_k(\mathbf{x^n}) = g(\mathbf{x^n}, \mathbf{w_k}, b_k), n = 1, ..., N, k = 1, ..., K, \tag{1}$$

where n and k index the feature vectors and the hidden nodes, respectively; and nonlinear activation function $g()$ can be any infinitely differentiable bounded function [16]. A common choice for $g()$ is the sigmoid function:

$$g(\mathbf{x}, \mathbf{a}, b) = \frac{1}{1 + exp\left(-(\mathbf{w} \cdot \mathbf{x} + b)\right)}, \tag{2}$$

where $\mathbf{w}$ and b correspond to weight vector and bias value, respectively. The actual learning takes place in the second layer between $\mathbf{H}$ and the label matrix $\mathbf{T} \in \mathbb{R}^{N \times L}$, where L is the number of outputs. $\mathbf{T}$ is composed of continuous annotations in the case of regression, and therefore, is an $N \times 1$ dimensional vector of outputs. In the case of L-class classification, $\mathbf{T}$ is represented with one vs. all coding:

$$\mathbf{T}_{n,l} = \begin{cases} +1 & \text{if } y^n = l, \\ -1 & \text{if } y^n \neq l, \end{cases} \tag{3}$$

where n and l index the rows (instances) and columns of the target matrix $\mathbf{T}$, and y^n denotes the categorical class label for n^{th} instance. The second level weights $\beta \in \mathbb{R}^{K \times L}$ are learned by least squares solution to a set of linear equations $\mathbf{H}\beta = \mathbf{T}$. The output weights can be learned via:

$$\beta = \mathbf{H}^{\dagger}\mathbf{T}, \tag{4}$$

where $\mathbf{H}^{\dagger}$ is the Moore-Penrose generalized inverse [23] that gives the minimum L_2 norm solution to $||\mathbf{H}\beta - \mathbf{T}||$, simultaneously minimizing the norm of $||\beta||$. To improve robustness and generalization, a regularization term C is introduced to the optimization problem, yielding the regularized ELM learning rule [18]:

$$\beta = \mathbf{H}^T\left(\frac{\mathbf{I}}{C} + \mathbf{H}\mathbf{H}^T\right)^{-1}\mathbf{T}, \tag{5}$$

where $\mathbf{I}$ is the $N \times N$ identity matrix. In our experiments, we use regularized ELM learning rule given in eq.(5). To improve stability of ELM, the input weight matrix $\mathbf{W}$ is orthonormalized using the orthogonal projection method, such that $\mathbf{W}^T\mathbf{W} = \mathbf{I}$ [18]. Similarly, L_2 normalization is applied to the bias vector $\mathbf{b}$. Finally, for a test set instance $\mathbf{x}^t$, the prediction is given as $\hat{y}^t = g(\mathbf{x}^t, \mathbf{W}, \mathbf{b})\beta$.

3. Exhaust Emission Dataset

The dataset[4] is composed of hourly average sensor measurements of eleven variables (nine input and two target variables). There are a total of 36 733 instances collected over five years. The nine input measurements (independent variables) can be grouped into two as ambient variables (e.g. temperature, humidity, pressure) and process parameters (e.g. turbine energy yield, air filter difference pressure). The names, abbreviations and basic statistics of the variables used in the study are summarized in Table 1. The data are collected in an operating range between partial load (75%) and full load (100%). Histograms for Carbon Monoxide (CO) and Nitrogen Oxides (NO_x) are given in Figure 1. In Figure 2, the locations of sensors and sources of turbine parameters are shown on the illustration of the gas turbine.

3.1. Analysis of Features

In our former work, we have carried out predictive modeling for net energy yield using Ambient Temperature, Ambient Humidity, Ambient Pressure and Exhaust Vacuum, using data collected from the same power plant [24,

[4]Exhaust Emission Dataset [online]. Website http://www.e-adys.com/datasets/pp_gas_emission.zip [accessed 26 July 2019].

Table 1. Basic statistical information of data used in the study.

Variable	Abbr.	Unit	Min	Max	Mean
Ambient Temperature	AT	°C	-6.23	37.10	17.71
Ambient Pressure	AP	mbar	985.85	1036.56	1013.07
Ambient Humidity	AH	(%)	24.08	100.20	77.87
Air Filter Difference Pressure	AFDP	mbar	2.09	7.61	3.93
Gas Turbine Exhaust Pressure	GTEP	mbar	17.70	40.72	25.56
Turbine Inlet Temperature	TIT	°C	1000.85	1100.89	1081.43
Turbine After Temperature	TAT	°C	511.04	550.61	546.16
Compressor Discharge Pressure	CDP	mbar	9.85	15.16	12.06
Turbine Energy Yield	TEY	MWH	100.02	179.50	133.51
Carbon Monoxide	CO	mg/m3	0.00	44.10	2.37
Nitrogen Oxides	NO_x	mg/m3	25.90	119.91	65.29

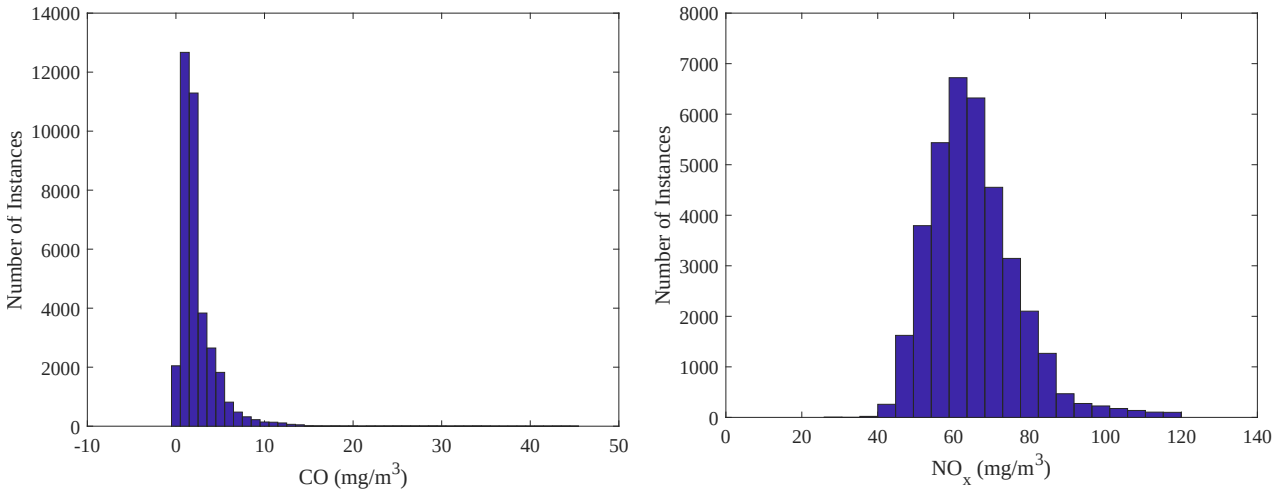

Figure 1. Histograms for CO and NO_x.

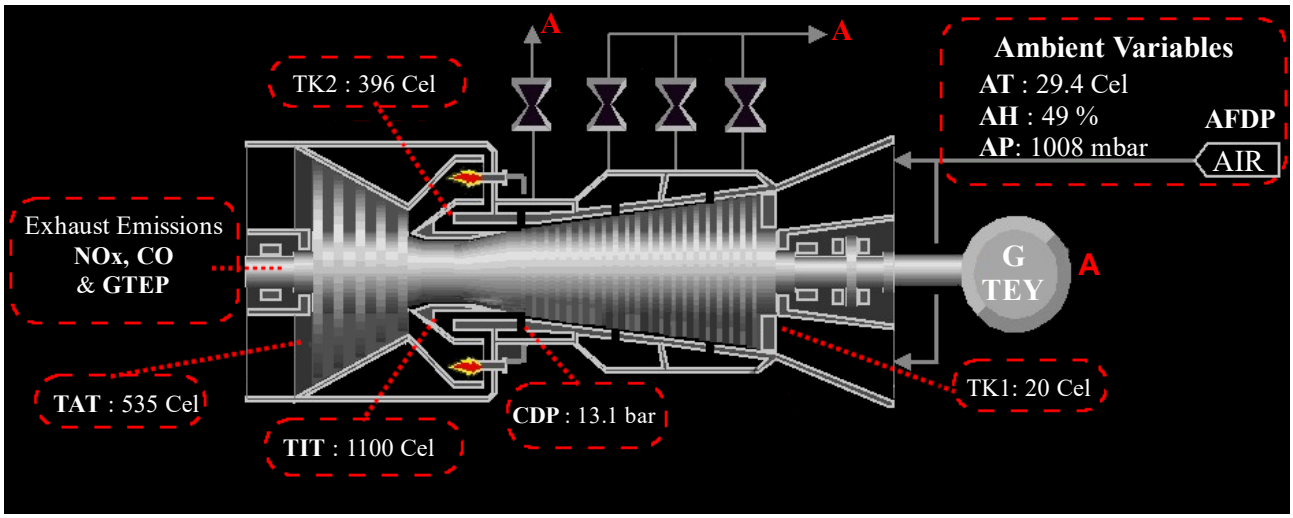

Figure 2. The sensor locations/parameter sources for all input and output variables used in the study. The parameters used are shown in dashed red rectangles, whereas the dashed white arrows show the sensor locations.

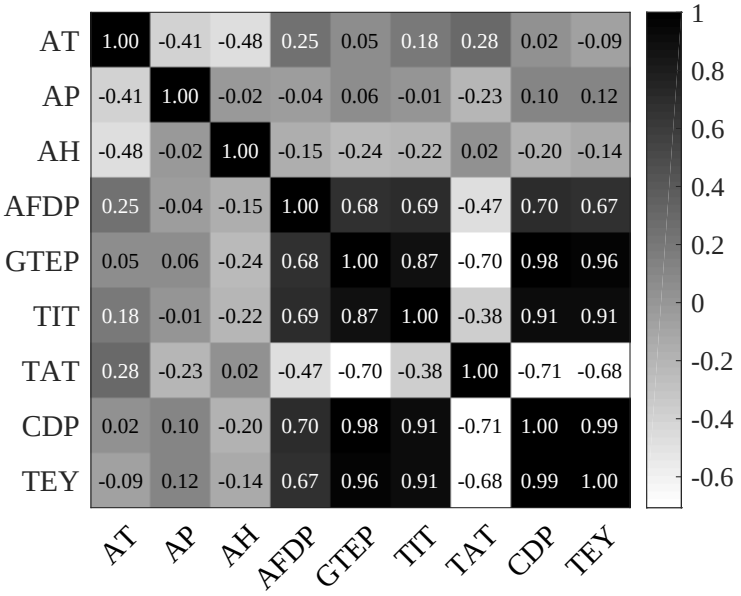

Figure 3. Correlation matrix of input variables

25]. In the following, we provide second order statistical analysis of the features and target variables, in order to have insight about the data. In Figure 3 and Table 2, the linear dependency between two random variables x and y is measured via Pearson's Correlation (PC):

$$corr(x,y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - m_x)(y_i - m_y)/(\sigma_x \sigma_y), \tag{6}$$

where N denotes the number of observations, m_x and σ_x are mean and std. statistics for x, respectively.

Figure 3 reveals the existence of a very strong linear dependency among the input variables, particularly between compressor discharge pressure (CDP) and turbine energy yield (TEY) (0.99), similarly CDP and gas turbine exhaust pressure (GTEP) (0.98). GTEP has also very strong correlation with TEY (0.96). This shows that some of the features may contain *redundant* information, and thus can be eliminated during model learning. Moreover, we see that the five turbine parameters (namely GTEP, CDP, AFDP, TIT and TAT) have stronger correlations with TEY, compared to the three ambient variables (AT, AP and AH) used as features in [24, 25]. Although the current findings show benefits in predictive modeling, this would not be of practical use for predicting the next day's hourly energy output, as process variables are obtained from real-time operation, and thus can not be estimated earlier. They can be used, however, in a machine learning model to cross-validate the sensor measurements.

Table 2. Pairwise correlation between features and the two target variables.

Feature	**CO**	**NO_x**
AT	-0.174	-0.558
AP	0.067	0.192
AH	0.107	0.165
AFDP	-0.448	-0.188
GTEP	-0.519	-0.202
TIT	-0.706	-0.214
TAT	0.058	-0.093
CDP	-0.551	-0.171
TEY	-0.570	-0.116

When we analyze Table 2, we see stronger correlations between CO and the turbine parameters that characterize the process, compared to NO_x. Therefore, in the regression experiments, we expect higher predictive accuracy with CO. Note that all correlations are statistically significantly different than zero ($p < 10^{-6}$).

When correlations are analyzed between individual features and target variables, we clearly see the effect of turbine inlet temperature (TIT) on CO (Pearson corr. -0.706). It is a known fact that CO is produced more when incomplete combustion with lower inlet temperature occurs. Regarding NO_x, we see the strongest correlation with the ambient temperature (-0.558), which suggests working in higher temperatures is more appropriate to reduce this exhaust emission.

4. Experiments

4.1. The Experimental Procedure

To ensure a clear separation of the training/test data and to allow comparability in future studies, we split the dataset into *train* (data from 2011-2012), *validation* (year 2013 data) and *test* sets (data from 2014 and 2015).

Moreover, to avoid over-fitting, we sequestered the test set and the hyper-parameters of the models trained on the train set are optimized on the validation set. To overcome the variance of the models due to the random first layer projection and also to benefit from diverse decisions, we train an ensemble of Regularized ELM models and fuse them. We use three fusion strategies: simple (unweighted) averaging and stacking to a Random Forest (RF) and basic ELM. For the latter two, the models are learned on the validation set predictions. The features are z-normalized using mean and std. statistics estimated from the corresponding training set. Mean Absolute Error (MAE) and R^2, which can be calculated as square of PC in eq. 6, are used to report performance.

4.2. Experimental Results

The overall pipeline of the PEMS modeling using the data and explained above is given in Figure 4. In the pipeline, predictions on the test set provide an unbiased estimate of the real-life performance.

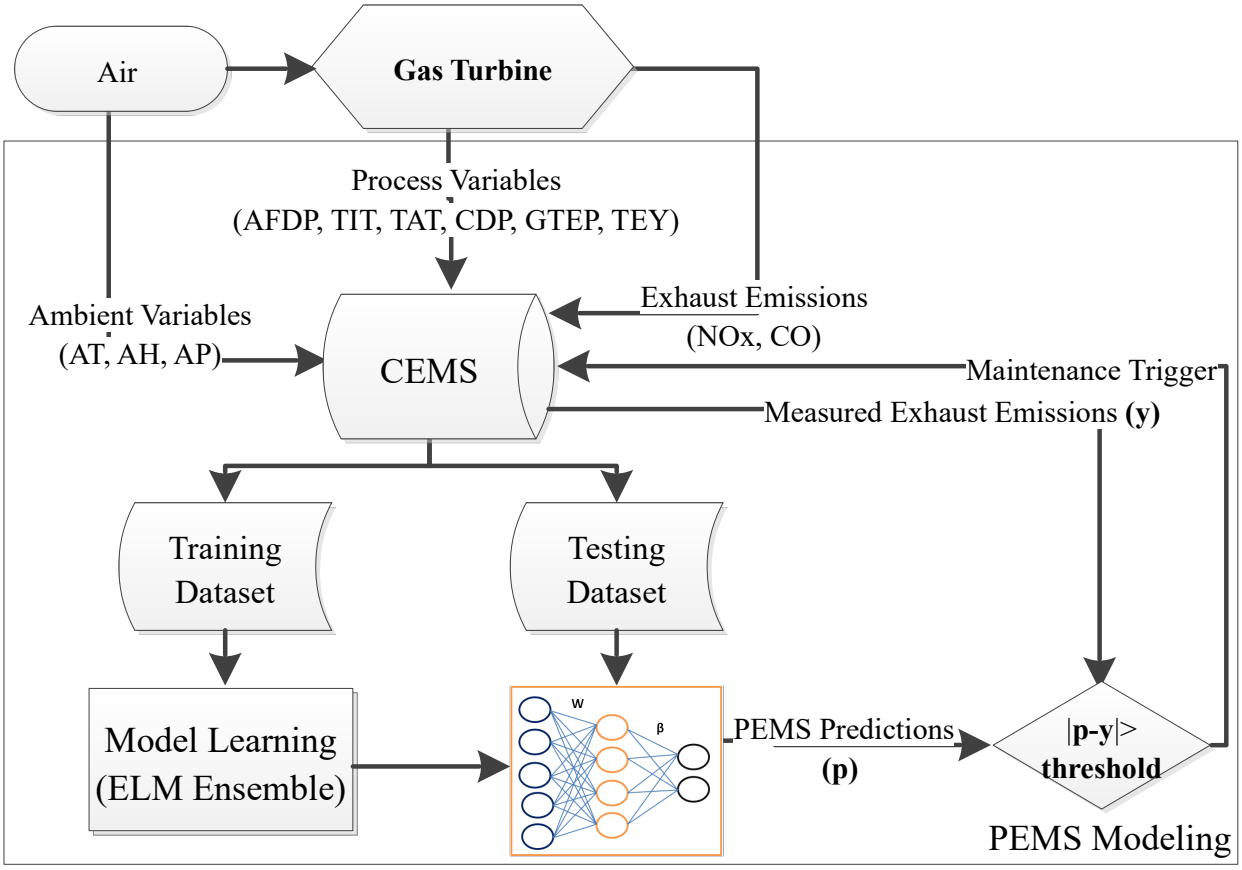

Figure 4. Pipeline of the PEMS proposed in the study.

The error in regression analysis can be decomposed into bias and variance terms, where the latter term can be minimized via combining multiple learners [26]. In order to reduce the variance of the ELM models stemming from random generation of the first layer weight matrix, we employ an ensemble learning approach. Given the hyper-parameters (i.e. the number of hidden nodes K and the regularization parameter C), we train 50 models and subsequently apply three fusion strategies: a) we take the average of predictions as the ensemble

output b) we stack the predictions to a basic ELM with 50 hidden neurons and also to c) a Random Forest (RF) regressor with 50 trees. RF is also an ensemble learner, where each tree is grown with a random subset of features and a random subset (with replacement) of instances [27].

To optimize the two hyper-parameters, we apply a grid search with exponential steps for both. In the preliminary experiments, we use $K = 2^{\{5,6,\ldots,15\}}$ and $C = 10^{\{-6,-5,\ldots,3\}}$. Analyzing the results of these preliminary experiments, we narrow the search sets into $K = 2^{\{9,10,\ldots,13\}}$ and $C = 10^{\{-3,-2,-1,0\}}$ for the remaining experiments. Prior to our experiments with stacking, we carry out a preliminary experiment using a 3-fold Leave-One-Year-Out (LOYO) cross validation with combination of training and validation sets.

4.2.1. Preliminary Experiments using LOYO Cross-Validation and Feature Selection

Table 3. LOYO cross-validation R^2/MAE performances for flue gas emission concentration prediction with varying number of hidden neurons (K) and the complexity parameter (C). Best results are shown in **bold**.

	CO				NO_x			
K/C	0.001	0.01	0.1	1	0.001	0.01	0.1	1
512	0.42/1.33	0.44/1.19	0.48/1.15	0.47/1.23	0.54/6.11	0.64/4.90	**0.67/4.57**	0.66/4.59
1024	0.44/1.23	0.50/1.05	0.52/1.07	0.48/1.18	0.57/5.65	0.65/4.76	0.66/4.64	0.65/4.71
2048	0.45/1.15	0.52/1.02	0.53/1.02	0.50/1.09	0.61/5.21	0.66/4.66	0.66/4.64	0.64/4.81
4096	0.48/1.07	**0.55/0.96**	0.54/1.00	0.48/1.08	0.63/4.93	0.66/4.62	0.65/4.70	0.61/4.97
8192	0.52/1.02	**0.56/0.97**	0.54/1.01	0.46/1.11	0.64/4.80	0.65/4.64	0.63/4.81	0.56/5.27

The overall LOYO results of CO and NO_x emission concentration predictions are shown in Table 3. We see that CO has a lower MAE, since CO emissions are concentrated close to zero. The best results for CO (MAE 0.96) are obtained with $C = 0.01$ and $K = 4096$. On the other hand, the results and optimal parameter pairs differ greatly for NO_x. The best MAE performance on the 3-fold LOYO cross-validation is 4.57, using 512 hidden nodes and $C = 0.1$. Considering the two tables, we observe that while the MAE performance on CO prediction is better than NO_x, the R^2 measure between the ground truth and the predictions is higher in NO_x, although the MAE is also high. This is particularly due to higher mean and range statistics of the NO_x emissions (see Table 1). After optimizing the hyper-parameters, we apply the corresponding models to the test set, in order to estimate the real-life performance. This rendered test set R^2/MAE performances of 0.494/1.32 and 0.608/10.37 for CO and NO_x, respectively.

To obtain a ranking of features and to improve the regression performance after feature selection, we employed a fast and effective method based on Canonical Correlation Analysis (CCA) [28]. CCA is a statistical method that aims to find linear projection bases that maximize mutual correlation among two sets of variables [29, 30]. It can be seen as a matrix extension of univariate correlation analysis. CCA is a powerful, fast and stable method commonly used for feature extraction and selection [28, 30]. In short, the idea used in [28] is to apply CCA between features and the target (response) variables, and subsequently rank the features with respect to the absolute value of the linear projection vector.

We apply the CCA based ranking proposed in [28] separately for CO and NO_x. We then take the top ranking p^* features, which cumulatively account for 95% of the sum of weights. This approach selects 5 and 7 features for CO and NO_x, respectively. The normalized weights and ranking of features are shown in Figure 5. Analyzing the table, we observe that CDP has a dramatically high normalized weight (0.753) for prediction of CO. This makes the number of features needed to account for 95% of the total weight less in CO compared to

NO_x, where the highest weight is 0.202. Note that, although the feature rankings differ for the two tasks, TEY and CDP are both found to be highly important (ranking in top three) for predicting CO and NO_x.

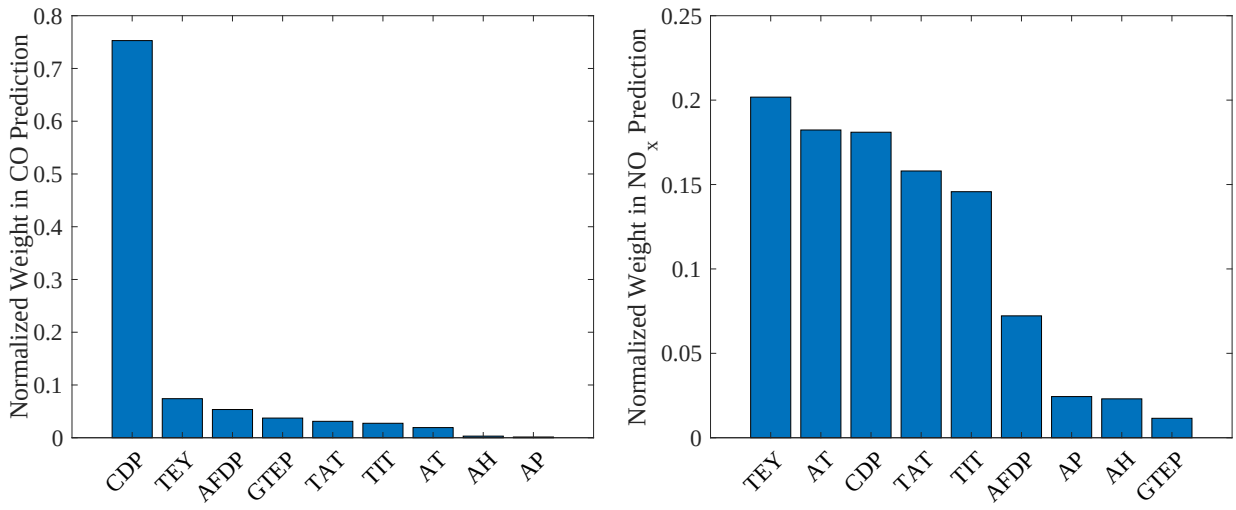

Figure 5. Feature importance weights and corresponding ranking for CO and NO_x.

Using the selected features and the hyper-parameters optimized on the whole set of features, we re-train the ELM models and cast predictions on the test set and finally average the predictions. The comparative test set results are summarized in Table 4, where we can see that feature selection clearly improves the prediction performance.

Table 4. Test set MAE performance of best models optimized on the training set with and without feature selection.

Target	K	C	Features	MAE
CO	4096	0.01	All	1.32
			CDP, TEY, AFDP, GTEP, TAT	**0.89**
NO_x	512	0.1	All	10.37
			All except AH and GTEP	**7.27**

4.2.2. Comparative Experiments with Stacking

To compare the performance of simple averaging and further improve our benchmark PEMS using a stacking framework, we switch to the train, validation and test procedure with features selected in the preliminary experiments. In Meta ELM, we stack predictions of the 50 Regularized ELM models to a single basic ELM with 50 hidden neurons. In the RF based stacking framework, we use 50 trees to combine predictions from the ensemble Regularized ELM models.

Comparative validation set results for CO and NO_x are presented in Table 5. From the table, we observe that compared to simple averaging of the predictions Meta ELM performs similar for CO prediction, however way poorer for NO_x prediction. On the other hand, RF based fusion outperforms the other two fusion schemes on the validation set for both correlation and MAE performance measures. For CO emission prediction, both

fusion schemes yield the best results with 512 hidden neurons with the base ELM models, while they slightly differ with the complexity hyper-parameter. For NO_x emission prediction however, we see that optimal hyper-parameter pairs are different for averaging and RF based fusion schemes. Note that, in both tables there may be ties in terms of correlation or MAE performance. In such cases, we favor simpler models (smaller K and C) due to parsimony principle (Ockham's Razor), and MAE has priority over R^2.

Table 5. Comparative validation set R^2/MAE performances for flue gas emission prediction with varying number of hidden neurons (K) and the complexity parameter (C). Best results of each method/target are shown in **bold**.

	CO				NO_x			
	Fusion Scheme: Averaging				Fusion Scheme: Averaging			
K/C	0.001	0.01	0.1	1	0.001	0.01	0.1	1
512	0.50/1.18	0.65/1.02	0.66/0.96	**0.64/0.99**	0.61/5.75	0.63/5.52	0.63/5.48	0.64/5.43
1024	0.55/1.14	0.67/1.01	0.66/0.97	0.62/1.02	0.62/5.64	0.63/5.50	0.63/5.46	0.64/5.42
2048	0.60/1.10	0.67/1.00	0.64/1.00	0.60/1.03	0.62/5.57	0.63/5.49	0.64/5.44	**0.64/5.41**
4096	0.65/1.06	0.67/0.98	0.64/1.01	0.59/1.03	0.63/5.53	0.63/5.48	0.64/5.43	0.64/5.41
8192	0.67/1.04	0.66/0.99	0.62/1.03	0.56/1.03	0.63/5.51	0.63/5.47	0.64/5.42	0.64/5.42
	Fusion Scheme: Meta ELM				Fusion Scheme: Meta ELM			
K/C	0.001	0.01	0.1	1	0.001	0.01	0.1	1
512	0.63/0.95	0.65/0.92	0.64/0.93	0.63/0.96	0.01/8.96	0.03/8.86	0.02/8.87	0.01/8.95
1024	0.51/1.05	0.61/0.98	0.61/0.99	0.57/1.07	0.08/8.61	0.09/8.59	0.08/8.65	0.08/8.68
2048	0.60/0.95	0.63/0.94	0.60/1.03	0.58/1.05	0.18/8.21	0.14/8.38	0.21/7.84	0.10/8.45
4096	0.61/0.97	0.57/1.03	0.57/0.99	0.58/1.00	0.26/7.90	0.25/7.86	0.21/7.93	0.19/8.01
8192	**0.69/0.81**	0.67/0.85	0.64/0.92	0.64/0.95	**0.40/6.86**	0.36/7.11	0.36/7.10	0.29/7.61
	Fusion Scheme: Random Forest				Fusion Scheme: Random Forest			
K/C	0.001	0.01	0.1	1	0.001	0.01	0.1	1
512	0.78/0.64	0.79/0.62	0.79/0.61	**0.80/0.61**	0.70/4.79	**0.71/4.74**	0.70/4.79	0.70/4.81
1024	0.76/0.65	0.78/0.63	0.79/0.62	0.80/0.62	0.70/4.87	0.70/4.82	0.69/4.86	0.69/4.90
2048	0.76/0.65	0.78/0.63	0.79/0.62	0.79/0.64	0.69/4.93	0.68/4.96	0.68/4.98	0.68/5.00
4096	0.77/0.65	0.77/0.64	0.79/0.64	0.78/0.66	0.68/5.00	0.68/5.00	0.67/5.08	0.67/5.08
8192	0.77/0.66	0.77/0.66	0.78/0.67	0.77/0.67	0.67/5.11	0.67/5.13	0.66/5.16	0.66/5.16

With the hyper-parameters optimized on the validation set (corresponding to **bold** items in Table 5), we apply the base models trained on the train set and the fusion scheme (for RF) trained on the validation set predictions on the held-out test set. The final results are summarized in Table 6. As expected, the performance on the test set is lower compared to the validation set. On the other hand, a use of 512 neurons for the base Regularized ELM models seems optimal for both tasks, while a use of 2048 neurons results in over-fitting, considering the degraded correlation performance. In CO emission prediction task, the RF based fusion gives the best test set results (MAE 0.927, R^2 0.584). On the other hand, in NO_x emission prediction task, simple averaging of 512-neuron ELM models' predictions outperforms RF based system (MAE 7.91, R^2 0.638). Hence, we conclude that the optimal fusion scheme depends on the target task. On the other hand, Meta ELM (stacking regularized ELM predictions to a second level ELM) yields very poor test set performance.

In order to have an insight about the test set performances of the fusion systems provided in Table 6, the predictions of each system is plotted against the ground truth sensor measurements (see Figure 6). If the regressors could be perfect, the points of a scatter plot would lie on y=x line. For CO prediction, we observe that regressors' suffer from noisy examples. For NO_x, the outputs except Meta ELM are more centered around y=x line, thus the correlation is higher, however the larger range of this target variable leads to a higher MAE.

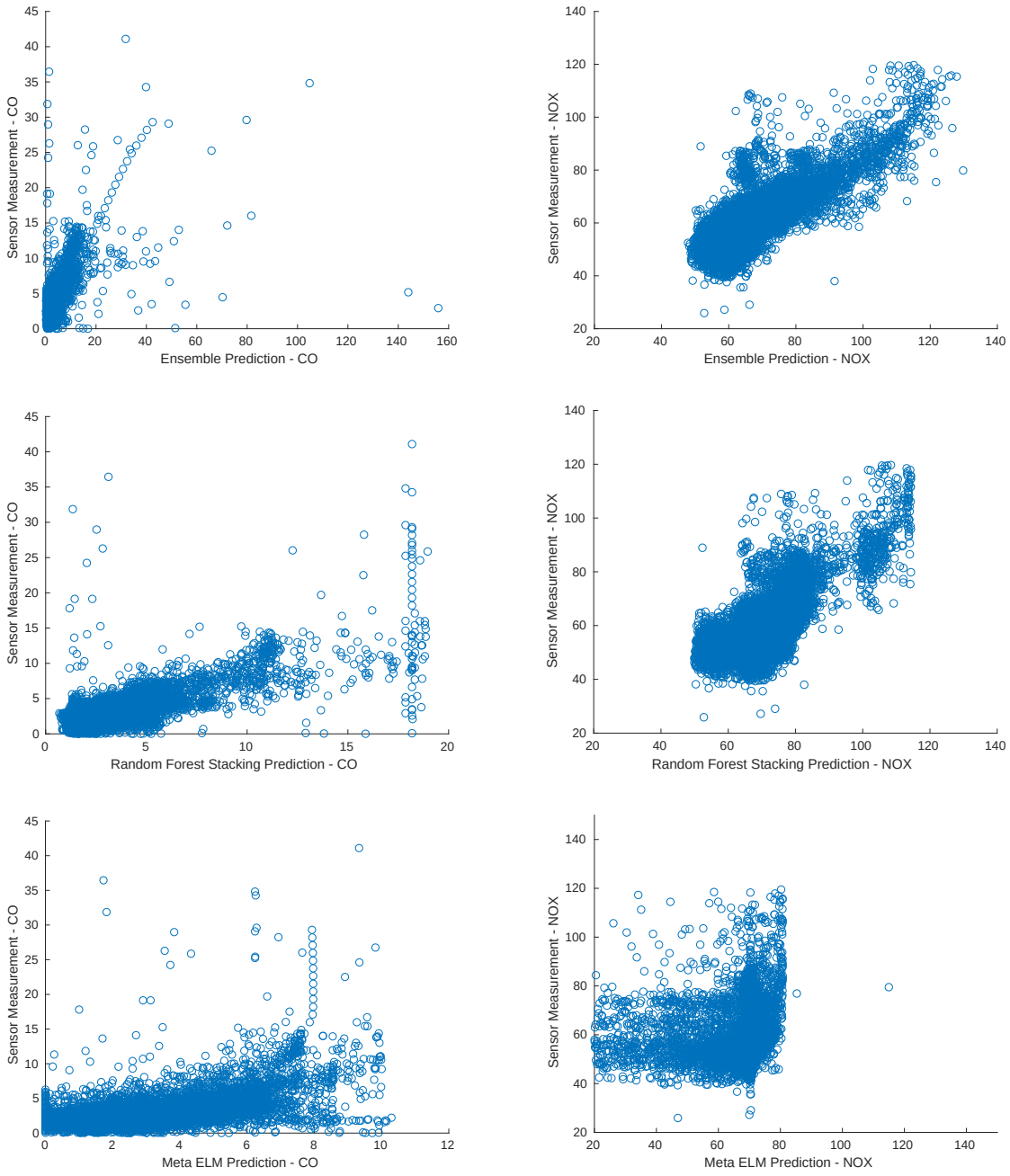

Figure 6. Scatter plots of test set predictions for three fusion schemes and two target variables. All base regularized ELM models are trained with K=512 hidden nodes. Complexity hyper-parameter C is set to 1 and 0.01 for CO and NO_x, respectively.

Table 6. Test set performances of three fusion schemes of ELM predictions. Best results are shown in **bold**.

			Averaging		Random Forest		Meta ELM	
Task	K	C	MAE	R^2	MAE	R^2	MAE	R^2
CO	512	0.1	1.05	0.43	1.05	0.55	1.34	0.31
	512	1	1.14	0.37	**0.93**	**0.58**	1.26	0.32
NO_x	512	0.01	**7.91**	**0.64**	11.29	0.53	24.05	0.00
	2048	1	10.77	0.16	11.91	0.12	6.64×10^4	0.00

5. Conclusions and Outlook

In this paper, we presented a novel, publicly available exhaust gas emission dataset for future use and comparative analyses. We first gave a statistical summary and correlation analysis of the features, and then defined a procedure to avoid over-fitting and to allow comparability of future works on the data. In addition to that, we applied ELMs to provide benchmark regression performance under the defined procedure, for the first time on these prediction tasks.

Our results show that feature selection based on linear projection weights markedly improves prediction performance on the test set. An analysis of feature importance weights, which was obtained from the ranking, reveals that TEY and CDP are highly important for predicting both CO and NO_x.

The experimental results corroborate what the linear correlation analyses suggests, namely, that the best predictive performance obtained with CO is higher compared to NO_x. This is partly because of the data distribution (CO emission concentration is accumulated near zero, as can be seen in Figure 1), and partly because the features used to characterize the process have stronger correlations with CO.

We applied and compared three decision fusion schemes, namely averaging, Meta ELM and stacking to Random Forests (RF). The experimental results have shown that while RF yields good results on both prediction tasks with the validation set, the best test set results are obtained with different fusion schemes, while Meta ELM exhibits a clear case of over-fitting. The over-fitting of Meta ELM, while observing relatively good performance of stacking to Random Forest indicates that when the base and second layer learners are of the same type, the bias may increase while variance does not decrease. This goes in line with the machine learning practices summarized as "one size does not fit all".

This work took a machine learning (regression) approach to the problem, providing a data-driven model for predicting exhaust CO and NO_x emission concentrations. It is also possible to look at the issue from a thermo-chemical reactions perspective, which was beyond the scope of this paper.

References

[1] Skalska K, Miller JS, Ledakowicz S. Trends in NO_x abatement: a review. Science of the Total Environment 2010; 408 (19): 3976–3989. doi: 10.1016/j.scitotenv.2010.06.001

[2] Korpela T, Kumpulainen P, Majanne Y, Häyrinen A. Model based NO_x emission monitoring in natural gas fired hot water boilers. IFAC-PapersOnLine 2015; 48 (30): 385–390. doi: 10.1016/j.ifacol.2015.12.409

[3] Shakil M, Elshafei M, Habib MA, Maleki F, Soft sensor for NO_x and O_2 using dynamic neural networks. Computers & Electrical Engineering 2009; 35 (4): 578–586. doi: 10.1016/j.compeleceng.2008.08.007

[4] Fichet V, Kanniche M, Plion P, Gicquel O. A reactor network model for predicting NO_x emissions in gas turbines. Fuel 2010; 89 (9): 2202–2210. doi: 10.1016/j.fuel.2010.02.010

[5] Traver ML, Atkinson RJ, Atkinson CM. Neural network-based diesel engine emissions prediction using in-cylinder combustion pressure. SAE Transactions 1999; 108: 1166–1180. http://www.jstor.org/stable/44716748

[6] Radl BJ. Neural networks prove effective at NO_x reduction. Modern Power Systems 2000; 20 (5): 59–62.

[7] Iliyas SA, Elshafei M, Habib MA, Adeniran AA. RBF neural network inferential sensor for process emission monitoring. Control Engineering Practice 2013; 21 (7): 962–970. doi: 10.1016/j.conengprac.2013.01.007

[8] Lv Y, Liu J, Yang T, Zeng D. A novel least squares support vector machine ensemble model for NO_x emission prediction of a coal-fired boiler. Energy 2013; 55: 319–329. doi: 10.1016/j.energy.2013.02.062

[9] Smrekar J, Potočnik P, Senegačnik A. Multi-step-ahead prediction of NO_x emissions for a coal-based boiler. Applied Energy 2013; 106: 89–99. doi: 10.1016/j.apenergy.2012.10.056

[10] Ćirić I, Ćojbašić Z, Nikolić V, Živković P, Tomić M. Air quality estimation by computational intelligence methodologies. Thermal Science 2012; 16: S493–S504. doi: 10.2298/TSCI120503186C

[11] Lazzaretto A, Toffolo A. Prediction of performance and emissions of a two-shaft gas turbine from experimental data. Applied Thermal Engineering 2008; 28 (17-18): 2405–2415. doi: 10.1016/j.applthermaleng.2008.01.021

[12] Rizk N, Mongia H. Semianalytical correlations for NO_x, CO, and UHC emissions. Journal of Engineering for Gas Turbines and Power 1993; 115 (3): 612–619. doi: 10.1115/1.2906750

[13] Dragomir EG, Oprea M. A multi-agent system for power plants air pollution monitoring. IFAC Proceedings Volumes 2013; 46 (6): 89–94. doi: 10.3182/20130522-3-RO-4035.00017

[14] Saiful Idzwan B, Phing CC, Kiong TS. Prediction of NO_x using support vector machine for gas turbine emission at Putrajaya power station. Journal of Advanced Science and Engineering Research 2014; 4 (1): 37-46.

[15] Liukkonen M, Hiltunen T. Monitoring and analysis of air emissions based on condition models derived from process history. Cogent Engineering 2016; 3 (1): 1174182. doi: 10.1080/23311916.2016.1174182

[16] Huang GB, Zhu QY, Siew CK. Extreme learning machine: theory and applications. Neurocomputing 2006; 70 (1-3): 489–501. doi: 10.1016/j.neucom.2005.12.126

[17] Huang GB, Wang DH, Lan Y. Extreme learning machines: a survey. International Journal of Machine Learning and Cybernetics 2011; 2 (2): 107–122. doi: 10.1007/s13042-011-0019-y

[18] Huang GB, Zhou H, Ding X, Zhang R. Extreme learning machine for regression and multiclass classification. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 2012; 42 (2): 513–529. doi: 10.1109/TSMCB.2011.2168604

[19] Liu H, Tian H, Li Y. Four wind speed multi-step forecasting models using extreme learning machines and signal decomposing algorithms. Energy Conversion and Management 2015; 100: 16–22. doi: 10.1016/j.enconman.2015.04.057

[20] Kaya H, Karpov AA, Salah AA. Robust acoustic emotion recognition based on cascaded normalization and extreme learning machines. In: International Symposium on Neural Networks; St. Petersburg, Russia; 2016. pp. 115–123. doi: 10.1007/978-3-319-40663-3_14

[21] Kaya H, Gürpınar F, Salah AA. Video-based emotion recognition in the wild using deep transfer learning and score fusion. Image and Vision Computing 2017; 65: 66–75. doi: 10.1016/j.imavis.2017.01.012

[22] Huang GB, Zhu QY, Siew C. Extreme learning machine: a new learning scheme of feedforward neural networks. In: IEEE 2004 International Joint Conference on Neural Networks; Budapest, Hungary; 2004. pp. 985–990.

[23] Rao CR, Mitra SK. Generalized Inverse of Matrices and Its Applications. Vol. 7, New York, NY, USA: Wiley, 1971.

[24] Kaya H, Tüfekci P, Gürgen FS. Local and global learning methods for predicting power of a combined gas & steam turbine. In: International Conference on Emerging Trends in Computer and Electronics Engineering; Dubai, UAE, 2012. pp. 13–18.

[25] Tüfekci P. Prediction of full load electrical power output of a base load operated combined cycle power plant using machine learning methods. International Journal of Electrical Power & Energy Systems 2014; 60: 126–140. doi: 10.1016/j.ijepes.2014.02.027

[26] Alpaydin E. Introduction to Machine Learning, 2nd ed. Massachusetts, MA, USA: The MIT Press, 2010.

[27] Breiman L. Random forests, Machine Learning 2001; 45 (1): 5–32. doi: 10.1023/A:1010933404324

[28] Kaya H, Eyben F, Salah AA, Schuller BW. CCA based feature selection with application to continuous depression recognition from acoustic speech features. In: IEEE 2014 International Conference on Acoustics, Speech, and Signal Processing; Florence, Italy; 2014. pp. 3757–3761. doi:10.1109/ICASSP.2014.6854298

[29] Hotelling H. Relations between two sets of variates, Biometrika 1936; 28: 321–377.

[30] Hardoon DR, Szedmak S, Shawe-Taylor J. Canonical correlation analysis: an overview with application to learning methods, Neural Computation 2004; 16 (12): 2639–2664. doi: 10.1162/0899766042321814