Development of NO$_x$ Emission Model for a Tangentially Fired Acid Incinerator

Elangeshwaran Pathmanathan, Rosdiazli Ibrahim, Vijanth Sagayan Asirvadam

Abstract—This paper aims to develop a NO$_x$ emission model of an acid gas incinerator using Nelder-Mead least squares support vector regression (LS-SVR). Malaysia DOE is actively imposing the Clean Air Regulation to mandate the installation of analytical instrumentation known as Continuous Emission Monitoring System (CEMS) to report emission level online to DOE. As a hardware based analyzer, CEMS is expensive, maintenance intensive and often unreliable. Therefore, software predictive technique is often preferred and considered as a feasible alternative to replace the CEMS for regulatory compliance. The LS-SVR model is built based on the emissions from an acid gas incinerator that operates in a LNG Complex. Simulated Annealing (SA) is first used to determine the initial hyperparameters which are then further optimized based on the performance of the model using Nelder-Mead simplex algorithm. The LS-SVR model is shown to outperform a benchmark model based on backpropagation neural networks (BPNN) in both training and testing data.

Keywords—artificial neural networks, industrial pollution, predictive algorithms, support vector machines

I. INTRODUCTION

POWER plants, chemical plants, government utilities and petroleum refineries typically produce more than the plant’s capacity in order to meet demands. Hence these industries must continuously monitor their exhaust stacks for primary pollutants such as nitrogen oxides, carbon monoxide, sulfur dioxide, and carbon dioxide under regulations promulgated under the New Source Performance Standards (U.S. Environmental Protection Agency (EPA) PA 40 CFR Part 60) or the Clean Air Act Amendments Title IV (40 CFR Part 75)[1] and Malaysian Environmental Quality Act, 1978[2]. CEMS utilize a sample probe, umbilical and sample conditioning system to extract a representative sample of the stack gas exhaust stream to provide a continuous flow for direct measurement of the pollutant concentration on individual analyzers[3]. A data acquisition system (DAS) is typically used to collect, calculate emission rates, alarm, and store historical data from these continuous emissions monitoring systems (CEMS)[4]. The promulgated regulations in U.S. EPA 40 CFR covering the monitoring of combustion units for primary and other pollutants allow for the use of predictive emissions monitoring systems (PEMS) in lieu of CEMS. PEMS is sophisticated software based on a sensor/prediction system that is directly interfaced with the process control system and inputs from the combustion or pollution control process. The U.S. EPA have revealed the devastating effects of the continuous release of carbon dioxide, methane, nitrous dioxide and other greenhouse (heat-trapping) gasses to the atmosphere[5]. Climate change induced by these gases can cause damage to human health, agriculture, natural ecosystems, coastal areas and other climate sensitive systems. Currently, emission monitoring is done via analytical instruments which are very expensive to install and maintain; the cost for an online NOx analytical and monitoring system is around $100,000 to $200,000 and the cost for maintenance is approximately $15,000 per year[6].

Several works have been done to develop predictive systems for industrial emissions. One of the earlier ideas was presented by G. Baines[7]; a consultant from Fisher-Rosemount Solutions. The author proposed the idea of predicting the stack contaminants in real time by correlating emissions with key unit parameters like fuel type, air and fuel flow and combustion temperature. S.S.S Chakrarthy, A.K Vohra and B.S Gill[8] have developed a PEMS for industrial process heaters. The authors have used heuristic optimizer genetic algorithm (GA) to tune the NOx kinetic parameters. J. Deng and R. Stobart[9] incepted the idea of using a hybrid clustering technique that involves the modifications of Fuzzy C-means method and using neural networks to achieve the best performance of the fuzzy system; for a diesel engine emission model. L.Zheng, S. Yu, M. Yu[10] used generalized regression neural network (GRNN) to establish a non-linear model between the parameters of the boiler (of 300MW steam capacity) and the NOx emissions. Later, the same authors proposed the idea of replacing the existing generalized regression neural network PEMS with least square support vector regression (LS-SVR) model. The authors claim that this new algorithm yields better accuracy of a coal combustion unit[11]. The main objective of this paper is to develop NOx emission model of an acid gas incinerator based on least squares-support vector regression algorithm with Nelder-Mead optimization of its hyperparameters. This paper is organized as follows: Section II presents a brief literature on artificial neural network and least-squares support vector regression. Section III describes the modeling background in terms of the parameters used, case study of acid gas incinerator and the dataset used. Section IV presents the results and analysis and Section V gives a brief conclusion.

R Rosdiazli Ibrahim is with the Universiti Teknologi PETRONAS Bandar Seri Iskandar, 31720 Tronoh, Perak, Malaysia (phone: 605-368-7821; fax: 605-365-7443; e-mail: rosdiazli@petronas.com.my).

Vijanth Sagayan Asirvadam is with the Universiti Teknologi PETRONAS Bandar Seri Iskandar, 31720 Tronoh, Perak, Malaysia (phone: 605-368-7781; fax: 605-365-7443; e-mail: vijanth_sagayan@petronas.com.my).

Elangeshwaran Pathmanathan with the Universiti Teknologi PETRONAS Bandar Seri Iskandar, 31720 Tronoh, Perak, Malaysia (e-mail: elanz117@gmail.com).
II. LITERATURE REVIEW

A. Artificial Neural Network (ANN)

An ANN is represented as a non-linear interconnected layers of processing nodes which are normally referred to as neurons, a term borrowed from neurobiology [12]. The most common class of ANN is the multilayered feedforward network which primarily consist of input layer, one or more hidden layer and an output layer. The input layer holds the data and distributes them into the network via interconnections to neurons in the hidden layer(s) where they are processed by the activation function to obtain the output signal. This type of network is also known as multilayer perceptron (MLP)[13]; a typical MLP network is shown in

\[
y = A \\
y = \frac{1}{1 + e^{-A}} \\
y = \frac{2}{1 + e^{-2A}} - 1
\]

Where \( A \), is the element activation function and it is represented by (4):

\[ A = \sum_{i=1}^{n} w_i x_i \]  

(4)

The variable \( w_i \) is the weight term and \( x_i \) is the network input. The network uses these weights to identify the strength of the interconnections between neurons. These weights are adjusted throughout the learning process. For MLP network, learning algorithms based on the gradient or Jacobian of the network error with respect to the weights is preferred because of their superior performance. A common Jacobian-based algorithm is the backpropagation algorithm. The backpropagation neural network however has been widely used to develop soft sensors for prediction of NOx[14]. However, BPNN has some weaknesses, including the need for numerous controlling parameters, difficulty in obtaining a stable solution and the danger of overfitting. The solution shown by Zheng et al [10] points to the fact that BPNN is unreliable even if all of the network objects are pre-determined.

B. Least Squares Support Vector Regression (LS-SVR)

Standard support vector machine formulation leads to a quadratic programming (QP) problem with linear constraints. The size of the matrix involved in the QP problem is directly proportional to the size of the training data[11]. Therefore to reduce the complexity of the optimization problem, Suykens et al[15] introduced a modified version of SVM called least squares – SVM. LS-SVM formulation results in a set of linear equations instead of a quadratic programming problem. LS-SVM is used for both classification and regression problems. The formulation for LS-SVR starts by taking a training set and estimate using a non-linear function:

\[ f(x) = w^T \phi(x) + b \]  

where \( \phi(x) : R^n \rightarrow R^k \) is a mapping function to a high dimensional and potentially infinite dimensional feature space; in this paper, there are 22 dimensions in the low dimensional space and 484 dimensions in high dimension space. Next, the optimization is formulated in primal weight space;

\[ \min \Phi(w,e) = \frac{1}{2} w^T w + \gamma \sum_{i=1}^{l} e_i^2 \]

Subject to;

\[ y_i = w^T \phi(x_i) + b + e_i \quad i = 1, \ldots, l \]

The optimization formulation in (5) is ridge regression cost function formulated in feature space. Constructing the Lagrangian of the problem, the dual space problem is derived;

\[ L(w, b, e; \alpha) = \Phi(w,e) - \sum_{i=1}^{l} \alpha_i [w^T \phi(x_i) + b + e_i - y_i] \]

The conditions for optimality are given by[15];

\[
\begin{align*}
\frac{\partial L}{\partial w} &= 0 & \Rightarrow & w = \sum_{i=1}^{l} \alpha_i \phi(x_i) \\
\frac{\partial L}{\partial b} &= 0 & \Rightarrow & \sum_{i=1}^{l} \alpha_i = 0 \\
\frac{\partial L}{\partial \alpha_i} &= 0 & \Rightarrow & \alpha_i = y_i & i = 1, \ldots, l \\
\frac{\partial L}{\partial e_i} &= 0 & \Rightarrow & w^T \phi(x_i) + b + e_i - y_i & i = 1, \ldots, l
\end{align*}
\]

Upon elimination of the variables \( w \) and \( e \) and solving in \( \alpha \) and \( b \) the following solution in dual space is obtained

\[
\begin{bmatrix}
0 \\
1_y \\
\Omega + 1/\gamma
\end{bmatrix}
\begin{bmatrix}
\alpha \\
1_y
\end{bmatrix}
= \begin{bmatrix}
0 \\
y
\end{bmatrix}
\]

(8)

where \( y = [y_1; \ldots; y_l] \), \( 1_y = [1, \ldots, 1] \) and \( \alpha = [\alpha_1; \ldots; \alpha_l] \). The “kernel trick”[16] is applied here as shown;

\[
\Omega_{ij} = \phi(x_i)^T \phi(x_j) = K(x_i, x_j) & i, j = 1, \ldots, l
\]

(9)

Hence the resulting LS-SVR model becomes;
from metallurgy where it is a global optimizer for a large search space. SA is designed to find an optimized solution in a given time, rather than finding the best possible solution[19], this saves computation time and it is suitable for finding initial values for further optimizations. The performances of the models are gauged using standard performance functions in the form of correlation factor (R) and root means square error (RMSE);

\[ \text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{X}_i - X_i)^2} \]  

\[ R = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(\hat{X}_i - \bar{\hat{X}})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(\hat{X}_i - \bar{\hat{X}})^2}} \]  

where \( X_i \) is the predicted value, \( \hat{X}_i \) is the true value and \( n \) is the number of testing samples.

C. Case Study

In this paper, the NO\textsubscript{x} emission from a tangentially fired acid gas incinerator is modeled. The incinerator train is part of an LNG complex. The plant is divided into four sections consisting of upstream facilities, gas treating, liquefaction and storage/terminal. The first part of the gas treating section deals with the acid gas in the feed. CO\textsubscript{2} and H\textsubscript{2}S are removed to meet product specification. Solvent with absorbed acid gases are regenerated in regenerator column at high temperature and low pressure. The acid gases are then incinerated before being discharged safely to the environment. In this LNG complex, acid gases are incinerated using two methods; one is using steam boilers and the other is using acid gas thermal incinerators [20]. The process flow scheme of the incinerator train is shown in Fig.2. The main constituents to the incinerator are acid gas, flash gas, fuel gas and combustion air. The flue gas that are released to the stack consist of NO\textsubscript{x}, SO\textsubscript{2}, SO\textsubscript{3}, CO, CO\textsubscript{2} and H\textsubscript{2}S.

In this paper only NO\textsubscript{x} emission will be modeled based on plant input and output process variable data. The data is obtained from the LNG complex incinerator train DCS’ historical dataset taken from April 2010 to November 2010 and contains the data from 29 input process variables and 7 output process variables. The most relevant process variables are selected by reviewing the design and operational specifications in the incinerator material balance sheet. Initial correlation analysis was also performed to the dataset to determine the variables that have the strongest variance with each other. As a result, 22 input variables and one output variable (i.e. flue gas NO\textsubscript{x} content) are chosen with 1068 number of samples, taken at 30 minutes apart for the whole month of October, 2010. Tables 1 and 2 outline the descriptions of the chosen input process variables the seven output variables respectively.
Before data are used for modeling, it is first pre-processed to remove bad inputs, non-values and outliers.

For both models, 70% of the data is reserved for training, 10% for validation (i.e. crossvalidation for LS-SVR) and 20% for testing.

IV. RESULTS AND ANALYSIS

The individual model performance of BPNN and LS-SVR will be presented in this section. Tables 3 and 4 presents a comparison of the performance of the models in terms of correlation, root means squared error and accuracy between real and predicted values. The performance of both models in terms of computational time is also shown. Fig. 3, 4, 5 and 6 show how each model is able to track the real value in both training and testing data. This will be followed by a brief analysis of the results obtained.
The performance of the LS-SVR model depends greatly on the values of the hyperparameters selected. The initial values of the hyperparameters are: $\gamma = 34.84$ and $\sigma = 11.89$, and was computed using SA. These values were then used for further optimization using Nelder-Mead simplex algorithm.

The objective function is to minimize model error (MSE) based on several constraints such as maximum number of optimization steps, maximum number of function evaluations, stopping criterion based on the relative change in value of the function in each step and stopping criterion based on the change in the minimizer in each step. Table 5 shows the simplex computation for this model.

The optimized values for the hyperparameters are:

- $\gamma = 32.96$
- $\sigma = 11.43$

### Table V

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Function Evaluation</th>
<th>min(f(x))</th>
<th>logf(x)</th>
<th>logλ(x)</th>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>3.14159265-001</td>
<td>3.5508</td>
<td>2.4753</td>
<td>initial</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>3.14159265-001</td>
<td>3.5508</td>
<td>2.4753</td>
<td>contract inside</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>3.14159265-001</td>
<td>3.5508</td>
<td>2.4753</td>
<td>contract outside</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>3.14159265-001</td>
<td>3.5508</td>
<td>2.4753</td>
<td>contract inside</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>3.14159265-001</td>
<td>3.5508</td>
<td>2.4753</td>
<td>reflect</td>
</tr>
<tr>
<td>6</td>
<td>12</td>
<td>3.14159265-001</td>
<td>3.5508</td>
<td>2.4753</td>
<td>contract inside</td>
</tr>
<tr>
<td>7</td>
<td>14</td>
<td>3.14159265-001</td>
<td>3.5508</td>
<td>2.4753</td>
<td>contract outside</td>
</tr>
<tr>
<td>8</td>
<td>16</td>
<td>3.14159265-001</td>
<td>3.5508</td>
<td>2.4753</td>
<td>contract inside</td>
</tr>
<tr>
<td>9</td>
<td>18</td>
<td>3.14159265-001</td>
<td>3.5508</td>
<td>2.4753</td>
<td>reflect</td>
</tr>
<tr>
<td>10</td>
<td>20</td>
<td>3.14159265-001</td>
<td>3.5508</td>
<td>2.4753</td>
<td>contract inside</td>
</tr>
<tr>
<td>11</td>
<td>22</td>
<td>3.14159265-001</td>
<td>3.5508</td>
<td>2.4753</td>
<td>contract outside</td>
</tr>
<tr>
<td>12</td>
<td>24</td>
<td>3.14159265-001</td>
<td>3.5508</td>
<td>2.4753</td>
<td>contract inside</td>
</tr>
<tr>
<td>13</td>
<td>26</td>
<td>3.14159265-001</td>
<td>3.5508</td>
<td>2.4753</td>
<td>reflect</td>
</tr>
<tr>
<td>14</td>
<td>28</td>
<td>3.14159265-001</td>
<td>3.5508</td>
<td>2.4753</td>
<td>contract inside</td>
</tr>
<tr>
<td>15</td>
<td>30</td>
<td>3.14159265-001</td>
<td>3.5508</td>
<td>2.4753</td>
<td>contract outside</td>
</tr>
</tbody>
</table>

The NOx emission as shown in Fig. 3, 4, 5 and 6 presents a highly non-linear distribution; hence it takes a model with the capabilities to map these nonlinearities precision and accuracy.

As shown in Tables III and IV, the performance of the LS-SVR model is better in both training and testing data compared to the BPNN model. The graphs also show that the LS-SVR model is able to track and predict NOx emission with much better accuracy than BPNN model. This could be because LS-SVR is a much more ‘transparent’ algorithm where the resultant model has weights and biases that were optimized to suit the particular dataset. Another plausible explanation is that, with BPNN, the initial weights are chosen randomly and hence the performance varies significantly with the number of simulations as shown by Zheng et al [10]. In contrast, LS-SVR initial parameters are optimized via SA. The solutions provided by support vector machines are unique and robust in terms of the class of the problem that it can handle, be it classification or complex regression with multiple dimensions. However, one of the drawbacks of LS-SVR includes heavy computational dependencies when it comes to large datasets. Nevertheless this could be improved using the latest techniques available such as quadratic Renyi entropy and sequential learning LS-SVM. Overall, the primary objective of the paper has been achieved, and in the future a hybrid model that involves both neural network and support vector machines will be developed in order to further improve the predictive accuracy and computational efficiencies.

### References


Sequential Learning with LS-SVM for World Largest Tangentially Fired Heaters.


