IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 2, 2019, accepted July 15, 2019, date of publication July 24, 2019, date of current version August 28, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2930555

Development of Predictive Emissions Monitoring System Using Open Source Machine Learning Library – Keras: A Case Study on a Cogeneration Unit

MINXING SI[1,2], TYLER J. TARNOCZI[3], BRETT M. WIENS[3], AND KE DU[2]

[1]Tetra Tech Canada Inc., Calgary, AB T2C 2X5, Canada
[2]Department of Mechanical and Manufacturing Engineering, University of Calgary, Calgary, AB T2N 1N4, Canada
[3]Cenovus Energy Inc., Calgary, AB T2P 0M5, Canada

Corresponding author: Ke Du (kddu@ucalgary.ca)

This work was supported in part by the Tetra Tech Canada Inc., and Cenovus Energy Inc.'s Internal Research and Development Program, and in part by the Natural Sciences and Engineering Research Council (NSERC) under Grant CRDPJ535813-18.

ABSTRACT The study provides an overview of Predictive Emissions Monitoring System's (PEMS) research, application, installation, and regulatory framework as well as develops predictive models for NO_x emissions from a natural gas fired cogeneration unit using an open source machine learning library, Keras, and open source programming languages, Python and R. Nine neural network based predictive models were trained with 12 086 examples and tested with 3020 examples. The neural network-based models use eight process parameters as inputs to predict NO_x emissions. All models meet the regulatory requirements for precision. The best model (32-64-64-64) has four hidden layers and uses the Nadam method for optimization. The best model has a mean absolute error of 0.5982, r-value of 0.9451, and a difference of 0.14% between the measured and predicted emission values using the test dataset. The study demonstrated the feasibility of using open source machine learning library in PEMS development. It also provides guidance to facility operators to develop their own PEMS models for monitoring emissions.

INDEX TERMS Air emissions monitoring, environmental monitoring, Keras, machine learning, NO_x, PEMS, predictive emissions monitoring system, tensorflow.

I. INTRODUCTION

The Predictive Emissions Monitoring System (PEMS) was developed as an alternative to overcome the drawbacks of the Continuous Emissions Monitoring System (CEMS), such as high initial capital cost, high operating cost, maintenance and operator training [1].

A PEMS relies on using the operating parameters of combustion facilities through first principle, statistical or Artificial Intelligence (AI) methods to build a model that can predict emissions. Historical paired emissions and selected process data (e.g., load, fuel composition, flow, pressure and temperature data, environmental conditions, turbine and boiler settings) are used to generate a model to determine the plant emissions.

The associate editor coordinating the review of this manuscript and approving it for publication was Canbing Li.

The capital costs for PEMS are estimated to be 50% less than for CEMS, and the operations and maintenance costs are approximately 10–20% of the CEMS cost [2], [3]. The application of the PEMS includes a) compliance reporting, b) offline what-if analysis, c) analyzer availability enhancement and d) continuous estimating when CEMS is offline [4].

There are a few commercial PEMS providers, such as Rockwell Automation, ABB and CMC Solutions. However, on-site training is required, and ongoing support from the software providers is often needed to adjust PEMS. The software licensing can be complicated or cost prohibitive for wide installations of PEMS.

This paper compiles PEMS research and application and summarizes the regulatory framework for PEMS in the United States and Europe. This paper also presents a case study of model development for monitoring oxides of nitrogen (NO_x) using an open source machine learning library,

Keras. Keras is a high-level neural network Application Programming Interface (API). The objective of this case study is to provide enough knowledge to facility operators to develop their own PEMS. The study is the first to test the predictive models with Rectified Linear Units (ReLU) as the activation function in accordance with the regulatory requirements for precision. Most existing neural network based PEMS models use Sigmoid as the activation function, which results in slow computation and vanishing gradient problems [5].

Section II of this paper provides an overview of PEMS research, application and installations to date based on publicly available information. Section III presents a case study of PEMS development using Keras and its Python and R interfaces. Section IV presents the results from the case study. Section V gives conclusions on the performance of the predicted models and discusses the challenges of developing a PEMS.

II. PEMS OVERVIEW

A. METHODS FOR PEMS DEVELOPMENT

To date, the methods used to develop PEMS can be grouped into three broad categories:

1. First Principle: This method uses analytical physical equations from thermodynamics, kinetics, mass and energy balance.
2. Statistical: This method explores the statistical relationships between operating parameters and corresponding pollutants emissions. Process dynamics (first principle) is often considered in the statistical model. It is sometimes referred to as the Statistical Hybrid Model.
3. Machine Learning: This method uses machine learning techniques to build predictive models using operating parameters. Neural network is the most widely used machine learning algorithm for PEMS development.

The statistical method focuses on confirmatory inference, which achieves emissions prediction by creating and fitting an emissions probability model [6]. The model finds relationships between variables and predicted emissions, as well as computes a quantitative emission rate of confidence. In statistical modeling, the operating and emissions data guide the selection of a stochastic model that serves as the abstraction for making emissions predictions.

Machine learning is a subset of AI; it gives computers ''the ability to learn without being explicitly programmed,'' defined by Arthur Samuel in 1959 [7]. In contrast to the statistical method, the machine learning method uses learning algorithms to find regularities in the historical paired operating and emissions data to predict emissions. PEMS is one of the first AI applications in environmental monitoring.

Both statistical and machine learning methods use historical data to develop prediction models. In some research papers, the two methods are combined into one category, referred to as the data-driven method [8], [9].

B. REGULATORY FRAMEWORK

Although PEMS is used for environmental monitoring worldwide, the major regulatory frameworks for PEMS were developed by the United States Environmental Protection Agency (US EPA) and the European Union. Many countries outside of the US and Europe have adopted these standards, especially the US EPA's PEMS certification standards. The main US federal regulation that accepts PEMS as an emission monitoring and reporting tool is Title 40 of the Code of Federal Regulations (CFR 40):

- Part 60 - Standards of Performance for New Stationary Sources
- Part 61 - National Emission Standards for Hazardous Air Pollutants
- Part 63 - National Emission Standards for Hazardous Air Pollutants for Source Categories

In March 2009, the Performance Specification (PS16) for PEMS was published by the US EPA [10]. The PEMS performance must meet the requirement set out in the PS16. In addition, an electrical generating facility can apply PEMS as an alternative monitoring system under CFR 40, Part 75, Subpart E, §75.40–75.48. The requirements in PS16 and Part 75 are consistent.

PEMS predicted values (e.g., NO_x emission rates) need to be compared with the data obtained from the reference method (RM) to demonstrate the model accuracy. The RM includes CEMS or field measurement. The US EPA's requirements for PEMS, including precision, reliability, accessibility and quality assurance and quality control (QA/QC), are summarized in Table 1.

If the emission rates change significantly from the previous PEMS training conditions (e.g., operating conditions of the equipment) or if a PEMS fails a relative accuracy test audit (RATA), the US EPA requires the PEMS to be re-trained and re-certified [11], [12].

In Europe, The Netherlands, Denmark and the United Kingdom accept using PEMS for emissions monitoring and compliance reporting [9]. A technical specification (TS) for PEMS applicability, execution and quality assurance (CEN/TS 17198:2018) was published by European Committee Standardization (CEN) in August 2018 [13]. The functionality requirements for PEMS operation and quality assurance are summarized in Table 2.

The TS outlines the performance requirements for the emission model's applicability, completeness, uncertainty, sensitivity, validation, availability and reproducibility. The uncertainty requirements for PEMS are not included in the TS, and it refers to legislative requirements for a specific plant. The only specified PEMS performance requirements in the TS are greater than or equal to 95% of availability and 8% of emission limit value for reproducibility. The requirements for applicability, completeness and sensitivity are based on the PEMS manufacturer's specifications.

TABLE 1. Requirements for PEMS by EPA under 40 CFR Part 75.

Requirement	Criteria
Precision	t-test : $d_{avg} \leq \|cc\|$ Mean difference between RM and PEMS (d_{avg}) is less than or equal to the absolute value of confidence coefficient (CC) at 97.5% one-sided confident interval. Otherwise, a bias factor needs to be applied to the PEMS predicted values.
	Coefficient correlation $r \geq 0.8$ Correlation between RM and PEMS must be 0.8 or greater.
	F-test, F- value $\leq F_{critical}$ The variance ratio of PEMS and RM must be less or equal to $F_{critical}$ value.
Reliability	Availability $\geq 95\%$
Accessibility and Timeliness	Record keeping according to 40 CFR Part 75 Subpart F Reporting according to 40 CFR Part 75 Subpart G
QA	PEMS input parameters must be operating within permitted operating ranges.
	Daily QA/QC check
	Instrumentations of the input parameters must be maintained in accordance with the manufacturer's recommendation.
	Alarm system for PEMS out of control Relative Accuracy (RA), less than 10% of average RM values.

TABLE 2. Functionality requirements for PEMS by CEN.

Requirement	Criteria
Functionality	Reading and validating sensors; performing emission model integrity tests; predicting emissions; displaying PEMS operating status, sensor readings and sensor validation; equipping with an alarm system
Capabilities	Hardware failure, power outrage, and communication line failures
Protection	Unauthorized access and log for modification

C. PEMS RESEARCH

The first PEMS was developed in 1973 and published in 1975 [14]. It used operating parameters of combustion facilities, through thermodynamics, to construct a polynomial equation. The model was used to monitor NO_x emissions from three gas turbines. In early 1980s, the model was coded into a computer program to simulate a NO_x control system for three 7 MW gas turbines and was run in parallel with a certified CEMS for over 10 years [15]. Since 1990, the PEMS studies and commercial applications for various sources and industries have been widely researched.

Kamas and Keeler [16] tested two PEMS on two cement kilns in 1995 by comparing PEMS with CEMS. One PEMS was used to monitor NO_x and SO_2 at kiln No. 1, which used coal fuel. The other PEMS monitored NO_x at kiln No. 2, which used waste-derived fuel. The R^2 values between predicted and CEMS measured NO_x data were 0.91 and 0.87 for the two kilns. The R^2 value for SO_2 prediction was 0.8.

Harnevie *et al.* [17] studied the first principle based PEMS on a biomass-fired power plant in Sweden in 1996, and concluded that a PEMS is suitable for emission monitoring for biomass-fired plants, even if the biomass fuel has varying water content.

Cooper and Andreasson [18] monitored NO_x for one year in 1997–1998 using a PEMS in parallel with a CEMS for a passage ferry in Sweden. The PEMS was developed using the first principle method, and the relative accuracy was 6.7%, compared to the CEMS monitoring data.

Faravelli *et al.* [19] used a kinetic model to predict NO_x from industrial boilers in 2000. The predicted NO_x value has a 2% difference from the measured value.

Chien *et al.* [20] analyzed NO_x and CO using a two-layer neural network PEMS for a simulated coal-fired boiler. The PEMS was used to explore the process optimization for NO_x and CO reduction.

Lee *et al.* [21] used integrated multiblock partial least-squares model for NO_x prediction to identify the root causes of excessive NO_x emissions and heater malfunction in large-scale multi-heater systems in 2005. The model used 187 process parameters and was developed based on 9,182 gas samples at sampling intervals of 5 minutes. The root-mean-squared error of predictions by the model was 5.640 ppm, or 3.7% of the average value.

Reifman *et al.* [22] investigated the application of PEMS for controlling the spatial distribution and total rate of injection of natural gas of the fuel-lean gas reburn system for NO_x control in coal-fired boilers in 2000. The neural network based PEMS was used to understand the relationships between the distribution of the injected natural gas in the upper region of the furnace and the average NO_x exiting the furnace. The PEMS model was developed based on 20 tests and could predict NO_x within the measurement uncertainties.

Chien *et al.* [20] developed a PEMS using the first principle method and tested it on a cogeneration unit at Hsinta power station in Taiwan for NO_x and O_2 monitoring in 2003. The cogeneration unit contained three gas turbines (90 MW) equipped with one heat recovery steam generator (HRSG) (170 MW). The model did not meet the EPA's draft PS16 requirement for PEMS. In 2005, Chien *et al.* updated the PEMS using the statistical method and tested it on the same cogeneration units. The relative accuracies (RA) are 3.38–14.77% and the correlation coefficients are 0.955–0.989 between the best model and the CEMS [23]. This statistical model met the draft PS16 requirements. In 2010, Chien *et al.* applied the same model to the same size and type cogeneration unit but at a different power station, Nanpo station, and concluded that the statistical model could not be transferred

from one piece of equipment to another due to differences in equipment condition, operating model, maintenance and other parameters. [24].

Saiepour *et al.* [8] developed PEMS models using multivariate statistical techniques for CO monitoring from a coke oven plant in the UK in 2004. The model was based on over 15 days of operating data and met the US EPA requirements. Bias adjustments were needed for low emission level and high emission level. No bias adjustment was needed for normal emission level.

Ross *et al.* [25] tested PEMS for NO_x monitoring at a glass manufacturing furnace based on periodic monitoring with a portable chemical cell gas analyzer in 2008. The predictive calculations were accepted by air emission control agencies for demonstrating regulation compliance.

Botros *et al.* [26]–[28] developed neural network based NO_x predictive models for three gas turbine engines in natural gas compressor stations in Alberta, Canada, from 2008 to 2010. The models include four or seven process parameters as inputs, one single hidden layer with two units, and one output for NO_x emissions. The Sigmoid activation function was used for the models. The purpose of the NO_x predictive models was to compare the predictive NO_x values with the US EPA's NO_x emission factors.

Tan *et al.* [29] developed a NO_x prediction model using an extreme learning machine (ELM) algorithm for coal-fired boilers at a power station in China in 2016. The mode was then used to optimize operation for NO_x reduction. The ELM algorithm is a single hidden layer feedforward neural network. The model was developed based on 10 days of operating data. The NO_x emissions predicted by the model had a mean absolute error (MAE) of 1.4%, a mean square error (MSE) of 62.1 and an r-value of 98% compared to the measured NO_x values.

Vanderhaegen *et al.* [30] compared a neural network based PEMS with a CEMS for NO_x and CO monitoring from a gas turbine for over one year to test whether PEMS can be a backup system when a CEMS malfunctions. Adaptive modeling techniques were used on the neural network PEMS. The model was re-trained and re-calibrated at certain intervals. The optimal interval was reported to be 10 days. The adaptive model was proven to predict NO_x and CO accurately in different operating regimes with large variations of NO_x.

Cuccu *et al.* [31] tested multiple methods for NO_x prediction from ten large-scale gas turbines. The methods tested include linear regression, kernel ridge regression, support vector regression and neural network with different optimization methods. The models were trained using 1,000 examples and tested by 1,000 examples. Neural network and support vector regression produced the highest precision for the NO_x predictions.

In summary, PEMS models have evolved from first principle based models to AI based models in the past five decades. Due to rapid development in computing technology, AI based models can process more data than ever. This makes them more flexible in selecting and processing process data than the first principle or statistical models because the latter must use process data related to NO_x formation. Nevertheless, developing PEMS or predictive models still has challenges:

1) A PEMS is often an equipment-specific system. Transferring a model from one piece of equipment to another requires re-training the model.
2) Models need to be adjusted when equipment operational conditions change.
3) Models do not perform well under abnormal operational conditions, e.g., equipment startup and shutdown period.

D. PEMS INSTALLATION

In the US, the first PEMS was installed in 1990 in a 14,100 HP gas turbine at a gas compressor station in Washington State [15]. This PEMS was equipped with an operator interface and could display the runtime of emission monitoring, print daily reporting and store historical data. The installed PEMS was integrated into the facility's control system.

The first regulatory approved PEMS was installed on a 233 GJ/h gas fired boiler at Arkansas Eastman, near Batesville, Arkansas in 1993 using 22 process parameters for NO_x emissions monitoring. The PEMS was approved by the US EPA after passing a RATA in June 1993 [16].

A PEMS was installed on eight ethylene furnaces in Texas in 1995 to monitor NO_x and O_2 for compliance reporting. The PEMS was certified by passing RATAs and operated for compliance reporting [32].

ABB installed a PEMS in a turbo-compressor station for gas injection in the United Arab Emirates in 2008. The PEMS monitored emission from four gas turbines and was certified in accordance with the US EPA requirements [33].

A major European oil refinery installed a PEMS as a backup to the existing CEMS. The main purpose of the installation was to increase the emission monitoring system's online availability to over 97.5%. The PEMS provides redundant values of SO_2, CO, NO, O_2, flue gas flow and particulate emissions from Fluid Catalytic Cracking and the Sulfur Recovery Units. The PEMS provided over 99% continuous monitoring with ±5% accuracy, compared to physical instrumentation measurements [2].

Between 2001 and 2018, 26 electrical generation units were approved to use a PEMS to fulfill emission monitoring requirements pursuant to CFR 40, Part 75 [34]; 22 of the electrical generation units were gas turbine units, and 17 of the gas turbines were equipped with HRSGs.

The PEMS installations pursuant to CFR 40, Parts 60, 61 and 63 are not publicly available. Based on the discussion with CMC Solution, a major commercial PEMS provider, more than 300 PEMS in the US have been installed for compliance reporting or as backup systems under CFR 40, Part 60 [35].

III. PEMS DEVELOPMENT FOR A COGENERATION UNIT USING KERAS

In Canada, PEMS is an accepted alternative monitoring method to CEMS at federal and provincial levels [36]–[38].

However, PEMS implementation and use in Canada, particularly in Alberta, remains low. The major reasons and challenges include the following:

1) Industry awareness. Few companies are aware of this alternative monitoring method. Even though PEMS was developed in the late 1970s, is widely used in various sectors and countries and is accepted by provincial and federal regulations, the benefits of PEMS have not been realized by the industry in Alberta and Canada.
2) Software licensing. On-site training is required, and ongoing support from software companies is often needed to adjust PEMS. The closed software system does not offer the flexibility desired by reporting personnel.
3) Lack of clarity in the regulatory approval process for PEMS installation and application. Although PEMS is accepted by federal and provincial regulations, the regulatory framework and approval processes are not clear in the federal and provincial regulations. Requests for PEMS application are reviewed case by case.

This case study was conducted in collaboration with Alberta industry and researchers, and aims to

1) Increase industry awareness through collaboration;
2) Develop predictive models using open source libraries and programming language to avoid licensing costs; and
3) Provide a showcase to federal and provincial regulatory agencies in Canada and demonstrate the performance of AI based predictive models for emissions monitoring.

A. EQUIPMENT DESCRIPTION AND DATA PREPARATION

A cogeneration unit at one of the Cenovus oil facilities was selected for the study. The cogeneration unit contains one 45.9 MW gas turbine and one 117 MW HRSG. The cogeneration unit is equipped with one CEMS on the HRSG stack. The CEMS continuously monitors stack gas flow rate, temperature, and NO_x flow rate. The HRSG in the cogeneration unit does not have additional air intake. All combustion air in the HRSG comes from the gas turbine exhaust gas (TEG). The gas turbine in the cogeneration unit uses pipeline natural gas. The HRSG has duct burners to provide supplemental heat for steam production. The duct burners use a mixture of pipeline natural gas and produced gas from oil production processes. As outlined in the Alberta CEMS monitoring requirements and the facility's operational permit, hourly exhaust temperature (°C), exhaust gas flow (m^3/s) and NO_x flow rate (kg/h) are submitted to the regulatory agency for compliance reporting. Note that Cenovus is not currently using PEMS to meet any of its regulatory requirements.

B. DATA PREPARATION

Turbine-related process data were collected at 5 second intervals from 149 process tags from 07:00 on September 1, 2013, to 06:00 on September 1, 2015. The hourly NO_x emission data from 2013 to 2015 were retrieved from the facility's emissions compliance reports.

The following steps were taken to select process parameters.

1. Duplicate tags for the same process parameters were removed. For example, the parameter ''turbine exhaust gas temperature'' is recorded in 14 process tags. Data from only one process tag for the TEG temperature was used for the model development.
2. Process parameters that are not related to NO_x formation were removed based on thermodynamics.

After the above steps, 20 process tags or process parameters were selected. The 5 second interval process data were then averaged into hourly data to pair with the hourly NO_x data. The following steps were taken to further select the process parameters.

3. Process parameters with too many missing values were removed because missing values may be caused by instrumentation failure and removal or installation of instruments. The predictive model should be developed based on reliable and continuously monitored process parameters.
4. Process parameters with near-zero variance were removed because constant values add little value to prediction.
5. The dependency of targets on predictors and independency of the predictors on each other were analyzed. Process variables with the least redundancy were selected.

Eight process parameters were selected after the above steps for the model development: generator power output, fuel gas temperature, turbine exhaust temperature, O_2 concentration in the turbine exhaust, fuel gas flow, heat balance in the exhaust, turbine speed and compressor air temperature.

The process data and NO_x data were then tabulated into a matrix of 17,520 by 9. The process data from 07:00 on September 1, 2013, to 06:00 on September 1, 2015, were paired with the emission data in the corresponding period. Each row in the matrix represents one example. The nine columns contain eight process parameters and one CEMS measured NO_x values. After the data were tabulated, the following steps were taken:

6. Equipment downtime was removed. The cogeneration unit was down for maintenance for 1,101 hours in the specified period.
7. The rows with substituted NO_x data were removed. The substituted data were not measured by the CEMS; they were calculated based on the regulatory agency's prescribed methods to backfill the missing data when the CEMS was offline or out of control.
8. Rows with missing values were removed because the remaining complete data could fully train the models.
9. NO_x values for partial hours or outliers were removed because the study focused on normal operation.
10. Values corresponding to possible instrumentation noise (e.g., negative readings) were discarded.

As a result of steps 6 to 10, 7.6% of NO_x values were removed—4.3% due to data substitution (Step 7), 0.9% due to missing values in the process variables (Step 8), 0.6% due

to outliers of NO_x values (Step 9) and 1.8% due to negative values in the process data (Step 10).

The dataset was then standardized using Z-score before training for the machine learning process as expressed in (1).

$$z = \frac{x - u}{\sigma} \tag{1}$$

where
z is the standardized value of process parameter x
x is the original data of process parameter x
u is the mean value of process parameter x
σ is the standard deviation of process parameter x

C. METHOD

NO_x predictive models were developed using open source machine learning library, Keras, and its R and Python interfaces. Keras is a neural network library. It was the second most referenced deep learning framework in scientific research, next to Google's TensorFlow, in 2017 [39]. Keras is now the official frontend of TensorFlow. Python and R are open source programming languages.

Feedforward network architecture with a backpropagation algorithm was used to construct the NO_x predictive models. The network used is fully connected so that every single unit (node) in a layer is connected to each unit (node) in the following layer through weights and biases. Data from the input layer are passed through the network until units in the output layer are reached. An example of feedforward network structure is presented in Fig. 1.

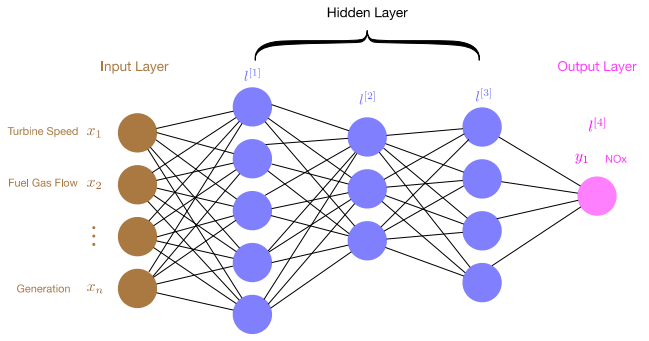


FIGURE 1. Structure of feedforward neural network.

A backpropagation algorithm optimizes weights and biases to minimize the difference between actual measured values and predicted values [40].

The dataset was randomly shuffled and then divided into a training set (80%), which was the data used build models and optimize weights and biases for certain network structures, and a test set (20%), which was data that had not been run through the model before testing. A holdout validation strategy was used to assess the models and select the best one. The training set was further divided by 90/10 for training and holdout validation. The validation dataset was used multiple times to assess different network structures. The test dataset was only used once and gave an unbiased evaluation of the final model's performance. The evaluation was to test the machine learning model and its ability to provide sensible answers to new inputs [41].

The final examples used for training, validation and evaluation are presented in Table 3.

TABLE 3. Training, validation and test sets.

Dataset	Number of Examples (rows)	Use
Training	10,877	Model Development
Validation	1,209	Assess performance among different model structures to select the best model
Total - Training	**12,086**	
Test	3,020	Evaluate the performance of the best model, such as generalization and predictive power
Total - Data	**15,106**	

The following hyperparameters were tuned to construct different network structures to select the best model:

- Number of units
- Number of layers
- Optimization methods

The optimization methods tested included Nesterov-accelerated Adaptive Moment Estimation (Nadam) [42], Adaptive Moment Estimation (Adam) [43], and RMSProp [44].

The following hyperparameters were set to be constant:

- Loss function: MSE. Loss function is used to compute the difference between the actual output (CEMS measured NO_x value) and predicted output (model predicted NO_x value).
- Learning rate: 0.001. The learning rate scales the magnitude of weight updates to minimize the network's loss function. The initial learning rate is set to be 0.001. If the models do not meet the regulation requirements for precision, smaller learning rates, such as 0.0005 and 0.0001, will be applied [45].
- Activation function: ReLU. Activation function is a nonlinear transformation that is applied to an input signal of a unit in a network to convert the input to an output.
- Metrics to validate and evaluate the accuracy of models: MAE.
- Batch size: 32. Batch size is the number of examples that an algorithm works through before updating the model parameters.

Five randomly selected network structures and three optimization methods for 15 models were tested. The five network structures are

1) 32-64;
2) 16-32-32;
3) 16-64-32;
4) 16-64-64; and
5) 32-64-64-64.

The structure refers to the number of units and layers for the hidden layers. For example, 32-64 structure represents two hidden layers with 32 units in the first hidden layer and 64 units in the second hidden layer. The input layer has 8 units, which represent 8 selected process parameters. The output layer has one unit to be NO_x emission in kg/h.

Neural networks use randomness to effectively learn the function being approximated for the problem, for example, random initialization of the network weights. The five selected model structures with three optimization methods (15 models total) were run 30 times (30 repeats). The average MAE and MSE by the validation dataset in the 30 repeats were used to select models for further assessment.

The selected models were further assessed using the K-fold cross-validation method, then made predictions on the entire training dataset of 12,086 training examples, including 10,877 examples in the training and 1,209 examples in the validation set. In this study, 6-fold was used to validate the models. Statistical tests were then conducted in accordance with the US EPA's precision requirements as outlined in Table 1.

The models that passed the statistical tests were evaluated using the test dataset. The statistical tests were conducted on the test dataset.

To evaluate the models, the CEMS measured NO_x, and predicted values in the test dataset were normalized using Min-Max scaling method as expressed in (2).

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{2}$$

where

x_{norm} is the normalized value

x is the original predicted or CEMS measured values

x_{min} is the minimum value of the predicted or CEMS measured values in the dataset

x_{max} is the maximum value of the predicted or CEMS measured values in the dataset

Standard residual is also used to evaluate model performance; it is defined as the residual divided by the standard deviation of the residuals.

IV. RESULTS

A. MODEL SELECTION

The average MAE and MSE using the validation dataset by five network structures and three different optimization methods are presented in Fig. 2. The networks with four hidden layers have the smallest errors. The network of 32-64-64-64 with Adam optimizer has the smallest MAE and MSE. For all 15 models, the average MAE in the 30 repeats ranges from 0.5877 to 0.7852, and the average MSE in the 30 repeats ranges from 0.8084 to 1.4689.

Nine models that have the minimum average MSE and MAE in the 30 repeats were chosen for statistical tests using the entire training dataset of 12,086 training examples (rows). All models pass the requirements for r-value and F-value set out by the US EPA in Table 1. The network 32-64-64-64 with

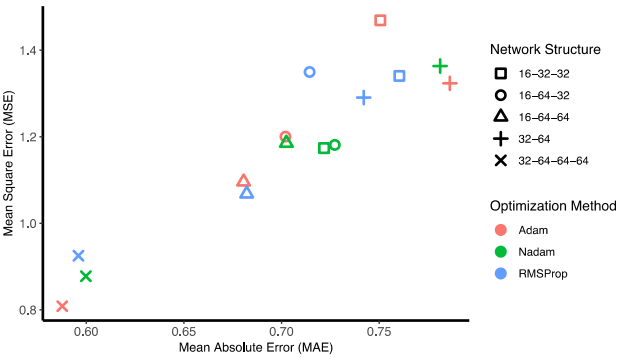


FIGURE 2. Average MAE and MSE in the repeats using the validation dataset.

Adam optimizer has the highest r-value of 0.9744. The r-values between the predicted values and CEMS values range from 0.9279 to 0.9744. F-values of test statistics range from 0.8331 and 0.9520 (Fig. 3). The critical F-value is 1.0304 for the 12,086 training examples.

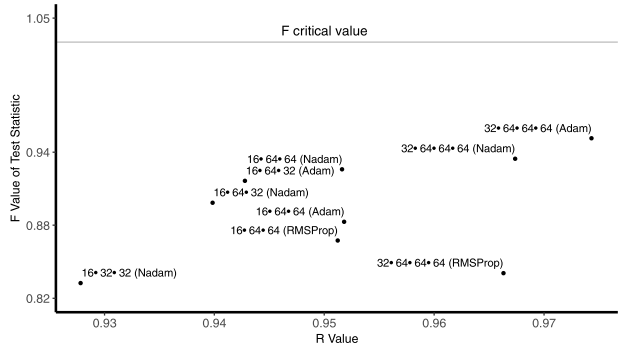


FIGURE 3. F and R values using the training dataset. based on 12,086 training examples.

The mean difference between CEMS and predicted values in six models is greater than the absolute value of confidence coefficient. The percentage difference between the CEMS measured and predicted values ranges from −0.4% to 1.9% in the 12,086 training examples.

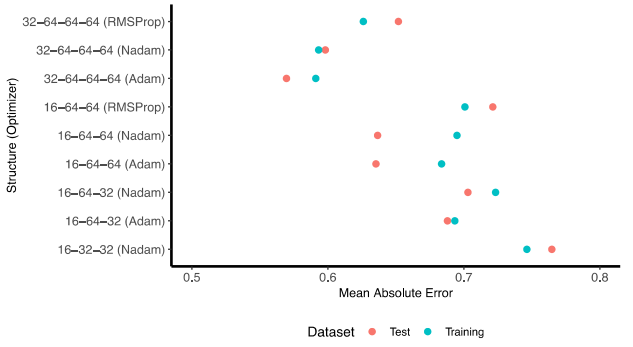


FIGURE 4. Mean absolute error in the training and test dataset. based on 12,086 training examples and 3,020 test examples.

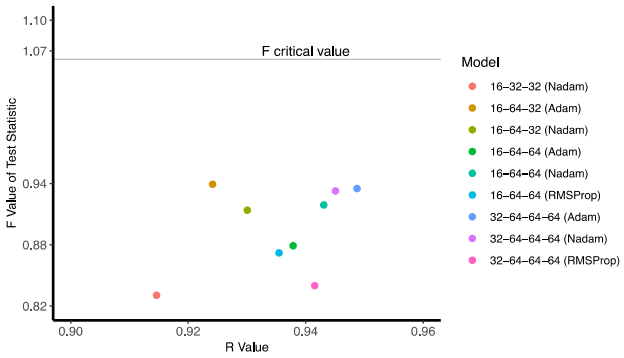


FIGURE 5. R and F values using the test dataset.

TABLE 4. Statistical tests using the training dataset.

Model[1]	d_{avg}	|cc|	Difference
32-64-64-64 (Nadam)[2]	0.0170	0.0122	0.1%
16-32-32 (Nadam)	-0.0189	0.0180	-0.1%
16-64-32 (Nadam)[2]	0.1172	0.0165	0.7%
16-64-64 (Nadam)	-0.2030	0.0148	-1.2%
32-64-64-64 (Adam)[2]	0.0518	0.0108	0.3%
32-64-64-64 (RMSProp)[2]	0.1610	0.0126	0.91%
16-64-32 (Adam)[2]	0.1458	0.0161	0.8%
16-64-64 (Adam)	-0.0758	0.0148	-0.4%
16-64-64 (RMSProp)[2]	0.3284	0.0149	1.9%

1: Model name is defined as Structure (Optimization Method).
2: Bias adjustment is needed for these models.
3: The training dataset has 12,086 examples.

All nine models pass the statistical tests using the training dataset, and three out of nine do not need bias adjustment based on the precision requirement listed in Table 1, i.e., $d_{avg} < |cc|$ (Table 4).

B. MODEL EVALUATION

The nine selected models were evaluated using the test dataset. The MAEs between the training and test datasets are presented in Fig. 4. The MAE for the training dataset is the average from the 6-fold cross-validation. The MAE in the test set ranges from 0.5696 to 0.7647, compared to 0.5491 to 0.7464 in the training set.

The r-values and F-values between predicted values and CEMS measured values in the training and test datasets are presented in Fig. 4. All r-values using the test dataset are less than the values using the training dataset. The r-values ranged from 0.9146 to 0.9487 in the test, compared to 0.9279 to 0.9744 in the training. The F-values of the test statistic for all nine models are less than the F-critical value, which is 1.0617 for the 3,020 test examples (Fig. 5).

The mean difference between CEMS and predicted values and the confidence coefficient using the test dataset is presented in Table 5. Five of the nine models needed bias adjustment.

The adjusted predicted value is calculated by the original predicted values multiplied by the bias factor. The absolute percentage of the difference between the CEMS and predicted values ranges from 0.06% to 2.01% in the 3,020 test examples.

Comparing predictions to observations., the predicted values by the 32-64-64-64 (Nadam) and 32-64-64-64 (Adam)

TABLE 5. Statistical tests using the test dataset.

Model	d_{avg}	|cc|	Bias[1]	Difference
32-64-64-64 (Nadam)	0.0239	0.0316	N/A	0.14%
16-32-32 (Nadam)	0.0103	0.0390	N/A	0.06%
16-64-32 (Nadam)	0.1496	0.0355	1.0086	0.85%
16-64-64 (Nadam)	-0.1933	0.0321	N/A	-1.10%
32-64-64-64 (Adam)	0.0623	0.0305	1.0035	0.35%
32-64-64-64 (RMSProp)	0.2099	0.0326	1.0120	1.19%
16-64-32 (Adam)	0.1633	0.0371	1.0093	0.93%
16-64-64 (Adam)	-0.0797	0.0334	N/A	-0.45%
16-64-64 (RMSProp)	0.3541	0.0341	1.0205	2.01%

1: Bias adjustment factor, calculated based on US EPA PS16.
2: The test dataset has 3,020 examples.

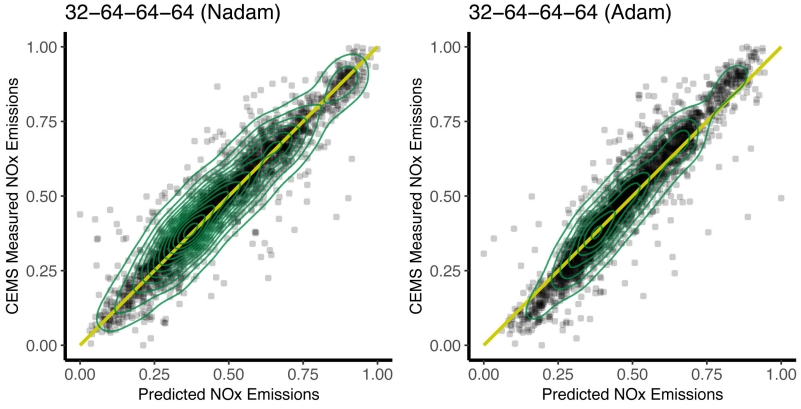


FIGURE 6. CEMS Measured NO_X versus Predicted Emissions. Based on 3,020 test examples. Values on both axes are normalized between 0 and 1 using Min-Max scaling method. The yellow line represents the CEMS values that are equal to the predicted values. The green circles represent the density of the data.

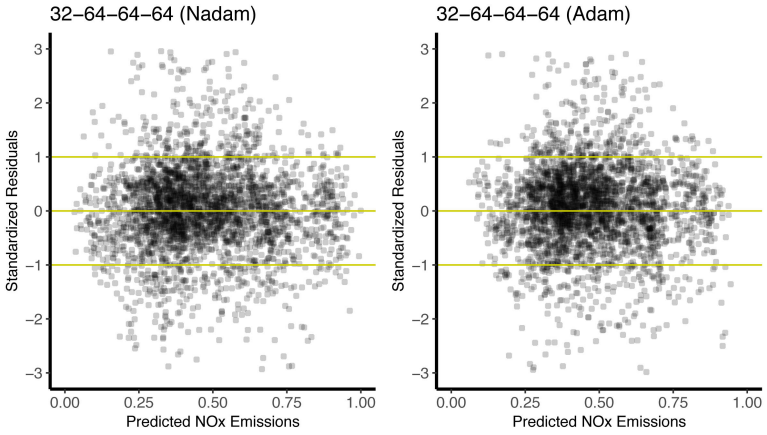


FIGURE 7. Standardized Residuals versus Predicted NO_x emissions. Based on 3,020 test examples. Values of predicted NO_x emissions are normalized between 0 and 1 using Min-Max scaling method.

models have better alignment with the CEMS measured values than the other models (Fig. 6). Moreover, 53.6% of the standardized residuals from the 32-64-64-64 (Nadam) model and 53.9% of the standardized residuals from the 32-64-64-64 (Adam) model fall in the [−1.1] range (Fig. 7). The percentages of standardized residuals in the [−1.1] range from the other seven models are below 53%. In addition, these two models have the minimum MAEs in the test dataset; the MAEs are 0.5982 for 32-64-64-64 (Nadam) and 0.5696 for 32-64-64-64 (Adam).

The comparison of the performance for all nine models is provided in Appendix.

V. CONCLUSION AND DISCUSSION

Nine neural networks based on models with different structures and optimization methods were tested to predict NO_x emissions using eight process parameters. The models were constructed using Keras and its Python and R interfaces.

The models were trained with 12,086 training examples and tested with 3,020 test examples. All models passed the requirements for r- and F-values set out by the US EPA. Bias adjustment was needed for five models. The test shows that the maximum difference between the sum of CEMS values and predicted values is 2.01%, produced by the 16-64-64 (RMSProp) model. The 16-32-32 (Nadam) model produced the minimum difference of 0.06% between the sum of CEMS values and predicted values.

According to the assessment of MAE and standardized residuals, the 32-64-64-64 (Nadam) and 32-64-64-64 (Adam) models have the best performance of the nine tested models. In addition, the 32-64-64-64 (Nadam) model does not need bias adjustment for the predicted values, and the 32-64-64-64 (Nadam) and 16-64-32 (Adam) models produce similar MAE using the training dataset and test dataset (Fig. 1), which indicates that the models are not overfitting or underfitting. Therefore, the 32-64-64-64 network structure using Nadam for optimization is considered to be the best model.

Although learning rate and learning schedule are considered the most important hyperparameters, the initial learning rate of 0.001 was proven to be sufficient, because all the models in this study met the regulatory requirements for precision with that learning rate. The study also showed that models with more complex structures had better performance than the models with simpler structures. The more complex models exceeded the precision requirements set out by the US EPA without overfitting. It is possible that even more complex models or smaller learning rates would achieve even better results.

The purpose of this study was to demonstrate whether an open source library can be used to develop a PEMS model that still meets the standards set out by the US EPA. Exhaustively testing many possible architectures or hyperparameters for the absolute best predictive value was beyond the scope of this study but would merit further investigation.

With Keras and R and/or Python, a facility operator can develop their own PEMS for emission monitoring and control analysis with minimal or no cost. The predictive emissions model can help facility operators meet emissions compliance and support equipment adjustment decisions, process optimization for emissions control and maintenance scheduling.

For compliance monitoring, additional functionality needs to be included to meet the requirements of US EPA PS16 or CEN/TS 17198:2018, such as daily checks and a PEMS system alarm.

Data preparation is critical for model development. Close attention should be paid to the following items when developing predictive models:

1) Instrumentation noise: Instrumentation noise needs to be closely examined and carefully selected. In this study, negative values were treated to be instrumentation noise, and invalid data were removed for modeling.
2) Missing values: Methods to deal with missing values in the process parameters need to be established

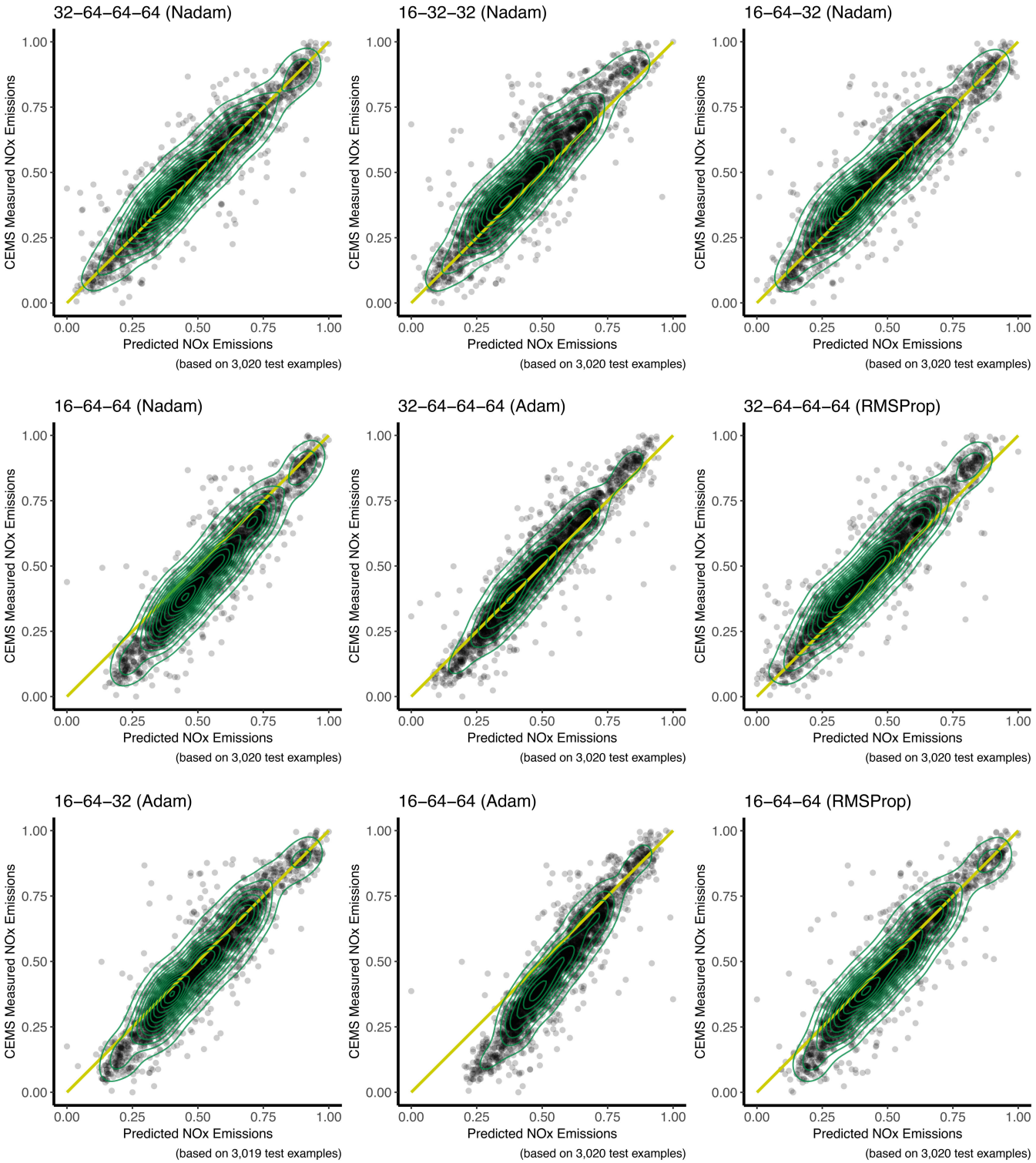


FIGURE 8. CEMS Measured versus Predicted NO_x Emissions Using the Test Dataset. Values on both axes are normalized between 0 and 1. One negative predicted value was removed in the Model 16-64-32 (Adam).

and examined. In this study, some process parameters were monitored by multiple instruments. For example, TEG temperature is monitored by 13 temperature transmitters and recorded in 13 different process tags. If TEG temperature is missing from one process tag, it can be supplemented with the data from other process tags. In this study, missing values were simply removed.

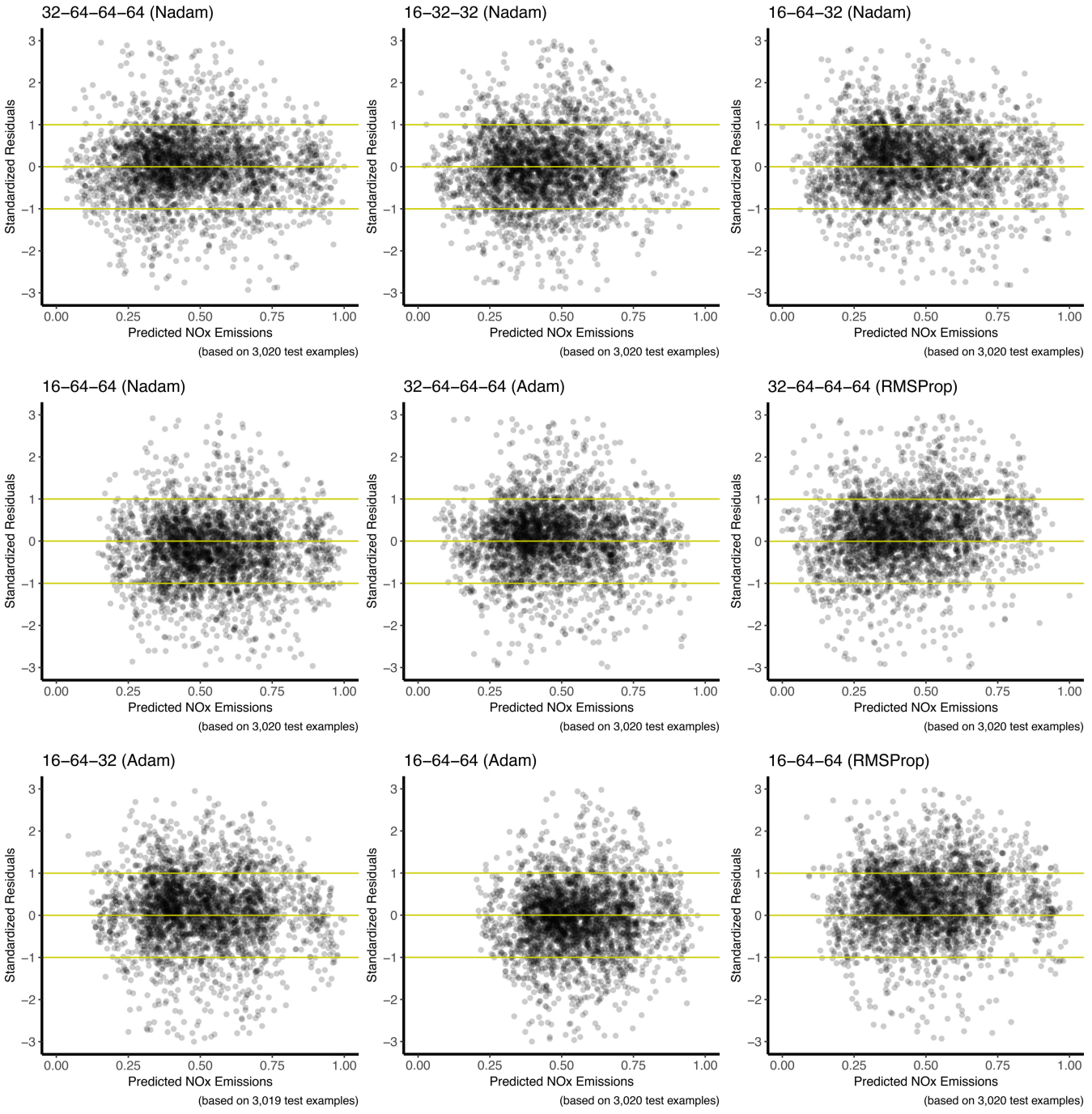


FIGURE 9. Standardized Residuals versus Predicted NO_x emissions. Values of predicted NO_x emissions are normalized between 0 and 1. One negative predicted value was removed in the Model 16-64-32 (Adam).

3) Non-normal operation: Partial hours and periods of equipment startup and shutdown need to be examined. Predicted models do not perform well during abnormal operations. In this study, if partial hours were included for NO_x prediction, all nine models generated negative values. Further studies are needed for non-normal operation periods.

Compared to the first principle method and the statistical method, the neural network based approach relies less on the number and type of specific process parameters used for model development. In other words, certain process parameters, such as combustion temperature, are essential to develop predictive models for NO_x emissions using the first principle and statistical methods. However, with a neural

network based approach, facility operators have the flexibility to choose process parameters based on data availability at their own facility.

This study is the first to assess the predictive models with ReLU as the activation function in accordance with the regulatory requirements for precision outlined by US EPA's CFR 40, Part 75 and PS16. Most existing neural network based PEMS models use Sigmoid as the activation function. When plotted, the Sigmoid function forms the shape of an S. Because of the geometrical shape of the Sigmoid activation function, it has major drawbacks, such as slow computation and vanishing gradient problems [5]. The vanishing gradient causes the algorithm to stop the neural network from further training.

In addition, this study is the first to use an open source library to develop a predictive model, which could be more widely used by industry than proprietary models. Users can easily repeat the simulation and validate the model by themselves with the open source library.

Moreover, this study presents the most thorough comparison to date of different network structures regarding performance and optimization methods in accordance with the US EPA's regulatory requirements.

Finally, the study contributes to wider use of predictive models and to cost reductions for industry since anyone can use the open source library to repeat this study's tests and build their own models by following the procedure described in the paper.

APPENDIX

- Appendix: Fig.8: CEMS Measured NO_x versus Predicted Emissions
- Appendix: Fig.9: Standardized Residuals versus Predicted NO_x emissions.

ACKNOWLEDGMENT

The authors thank Environment and Water Group and Jessica Coles in Tetra Tech Canada Inc. for editing this paper.

REFERENCES

[1] M. Roth and P. Lawrence, "A cost-effective alternative to continuous emission monitoring systems," *Environ. Sci. Eng. Mag.*, pp. 56–57, May 2010.

[2] G. Ciarlo and F. Callero, "Predictive emission monitoring systems (PEMS) Deploying software-based emission monitoring systems for refining processes," 2015. Accessed: May 30, 2018. [Online]. Available: https://library.e.abb.com/public/310e4b7c45a742578722e61894999b4e/AT_ANALYTICAL_029-EN_A.pdf

[3] *Software CEM Predictive Emissions Monitoring System*, Rockwell Automat., Milwaukee, WI, USA, 2009.

[4] T. Eisenmann, D. R. Bianchin, and D. Triebel, "Predictive Emission Monitoring (PEM): Suitability and application in view of U.S. EPA and European regulatory frameworks," presented at the 11th Int. Conf. Exhib. Emissions Monit. (CEM), Istanbul, Turkey, 2014, p. 15.

[5] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, "Activation functions: Comparison of trends in practice and research for deep learning," Nov. 2018, *arXiv:1811.03378*. [Online]. Available: https://arxiv.org/abs/1811.03378

[6] D. Bzdok, N. Altman, and M. Krzywinski, "Points of significance: Statistics versus machine learning," *Nature Methods*, vol. 15, no. 4, pp. 233–234, Apr. 2018.

[7] Z. Ge, Z. Song, S. X. Ding, and B. Huang, "Data mining and analytics in the process industry: The role of machine learning," *IEEE Access* , vol. 5, pp. 20590–20616, 2017.

[8] M. Saiepour *et al.*, "Development and assessment of predictive emission monitoring systems (PEMS) models in the steel industry," in *Proc. Iron Steel Technol. Conf.*, Cleveland, OH, USA, vol. 2, 2006, pp. 1121–1132.

[9] O. Antson and T. Pellikka, "Predictive emission monitoring systems, PEMS and their acceptance and use in Europe," VTT Tech. Res. Centre Finland, Espoo, Finland, Res. Rep. VTT-R-04313-14, Oct. 2014.

[10] U.S. Environmental Protection Agency, "Performance Specification 16 for predictive emissions monitoring systems and amendments to testing and monitoring provisions," *Federal Register*, vol. 74, no. 56, pp. 12575–12591, Mar. 2009.

[11] U.S. Environmental Protection Agency. (Mar. 16, 2015). *Request for Changes to the PEMS Approval for Units G1 through G7 at the Arkansas Electric Cooperative Corporation's Harry L. Oswald Generating Station (Facility ID (ORISPL) 55221)*. Accessed: Nov. 27, 2018. [Online]. Available: https://www.epa.gov/sites/production/files/2015-05/documents/harry_l._oswald_generating_station.pdf

[12] U.S. Environmental Protection Agency. (Mar. 31, 2006). *Approval of the Predictive Emission Monitoring System Installed on Unit 3 at Braintree Electric Light Department's Norton P. Potter II Station (ORISPL 01660)*. [Online]. Available: https://www.epa.gov/sites/production/files/2015-05/documents/potter.pdf

[13] *Stationary Source Emissions—Predictive Emission Monitoring Systems (PEMS)—Applicability, Execution and Quality Assurance*, Eur. Committee Standardization, Brussels, Belgium, Aug. 2018.

[14] W. S. Y. Hung, "An experimentally verified NO_x emission model for gas turbine combustors," in *Proc. Int. Gas Turbine Conf. Products Show (ASME)*, Mar. 1975, Art. no. V01BT02A009.

[15] W. S. Y. Hung and F. Langenbacher, "PEMS: Monitoring NO_x emissions from gas turbines," in *Manufacturing Materials and Metallurgy; Ceramics; Structures and Dynamics; Controls, Diagnostics and Instrumentation*; vol. 5. Houston, TX, USA: IGTI Scholar Award, 1995, Art. no. V005T15A016.

[16] J. Kamas and J. Keeler, "Predictive emissions monitoring systems: A low-cost alternative for emissions monitoring [in cement industry]," in *Proc. IEEE Cement Ind. Tech. Conf. 37th Conf. Rec.*, San Juan, PR, USA, Jun. 1995, pp. 497–509.

[17] H. Harnevie, L. Sarkoezi, and S. Trenkle, "Predictive emission monitoring system (PEMS) for emission control in biomass fired plants," Stiftelsen foer Vaermeteknisk Forskning, Stockholm, Sweden, Tech. Rep. SVF- -575, 1996.

[18] D. A. Cooper and K. Andreasson, "Predictive NO_x emission monitoring on board a passenger ferry," *Atmos. Environ.*, vol. 33, no. 28, pp. 4637–4650, Dec. 1999.

[19] T. Faravelli, L. Bua, A. Frassoldati, A. Antifora, L. Tognotti, and E. Ranzi, "A new procedure for predicting NO_x emissions from furnaces," *Comput. Aided Chem. Eng.*, vol. 8, pp. 859–864, Jan. 2000.

[20] T. W. Chien, H. Chu, W. C. Hsu, T. K. Tseng, C. H. Hsu, and K. Y. Chen, "A feasibility study on the predictive emission monitoring system applied to the Hsinta power plant of taiwan power company," *J. Air Waste Manage. Assoc.*, vol. 53, no. 8, pp. 1022–1028, Aug. 2003.

[21] Y.-H. Lee, M. Kim, and C. Han, "Application of multivariate statistical models to prediction of NO_x emissions from complex industrial heater systems," *J. Environ. Eng.*, vol. 131, no. 6, pp. 961–970, Jun. 2005.

[22] J. Reifman, E. E. Feldman, T. Y. C. Wei, and R. W. Glickert, "An intelligent emissions controller for fuel lean gas reburn in coal-fired power plants," *J. Air Waste Manage. Assoc.*, vol. 50, no. 2, pp. 240–251, Feb. 2000.

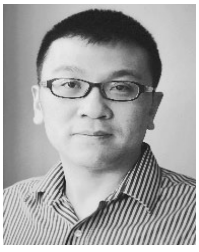
[23] T. W. Chien, H. Chu, W. C. Hsu, Y. Y. Tu, H. S. Tsai, and K. Y. Chen, "A performance study of PEMS applied to the Hsinta power station of Taipower," *Atmos. Environ.*, vol. 39, no. 2, pp. 223–230, Jan. 2005.

[24] T.-W. Chien, H.-T. Hsueh, H. Chu, W.-C. Hsu, Y.-Y. Tu, H.-S. Tsai, and K.-Y. Chen, "A feasibility study of a predictive emissions monitoring system applied to Taipower's Nanpu and Hsinta power plants," *J. Air Waste Manage. Assoc.*, vol. 60, no. 8, pp. 907–913, Aug. 2010.

[25] C. P. Ross, D. Russell, and J. Mino, *An Effective PEMS Alternative to CEMS for Quantifying Glass Furnace NO_x Emissions* (Ceramic Engineering and Science Proceedings), J. Kieffer, Ed. Hoboken, NJ, USA: Wiley, 2008, pp. 153–167.

[26] K. K. Botros, C. Selinger, and L. Siarkowski, "Verification of a neural network based predictive emission monitoring module for an RB211-24C gas turbine," in *Proc. Microturbines Small Turbomachinery; Oil Gas Appl.*, Orlando, FL, USA, vol. 5, 2009, pp. 431–441.

[27] K. K. Botros and M. Cheung, "Neural network based predictive emission monitoring module for a GE LM2500 gas turbine," in *Proc. 8th Int. Pipeline Conf.*, Calgary, AB, Canada, vol. 3, 2010, pp. 77–87.
[28] K. K. Botros, C. Williams-Gossen, S. Makwana, and L. Siarkowski, "Predictive emission monitoring (PEM) systems development and implementation," presented at the 19th Symp. Ind. Appl. Gas Turbines Committee, Banff, AB, Canada, Oct. 2011.
[29] P. Tan, J. Xia, C. Zhang, Q. Fang, and G. Chen, "Modeling and reduction of NO_x emissions for a 700 MW coal-fired boiler with the advanced machine learning method," *Energy*, vol. 94, pp. 672–679, Jan. 2016.
[30] E. Vanderhaegen, M. Deneve, H. Laget, N. Faniel, and J. Mertens, "Predictive emissions monitoring using a continuously updating neural network," in *Proc. ASME*, vol. 2, 2010, pp. 769–775.
[31] G. Cuccu, S. Danafar, P. Cudré-Mauroux, M. Gassner, S. Bernero, and K. Kryszczuk, "A data-driven approach to predict NO_x-emissions of gas turbines," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Boston, MA, USA, Dec. 2017, pp. 1283–1288.
[32] R. L. Croy, C. W. Simpson, D. B. Tipton, P. Sinha, and M. E. Lashier, "NO_x PEMS experience on operating ethylene furnaces," *Environ. Prog.*, vol. 16, no. 4, pp. 297–300, Dec. 1997.
[33] ABB S.p.A. (2014). *Predictive Emission Monitoring Systems the New Approach for Monitoring Emissions From Industry*. Accessed: Aug. 22, 2018. [Online]. Available: https://library.e.abb.com/public/4128e88396a14d83b8116a2a71b5b88f/PB_PEMS-EN.pdf
[34] *Performance Specification 16-Specifications and Test Procedures for Predictive Emission Monitoring Systems in Stationary Sources*, U.S. Environ. Protection Agency, Washington, DC, USA, Aug. 2018.
[35] B. Swanson, private communication, Nov. 2018.
[36] *Continuous Emissions Monitoring System (CEMS) Code*, Alberta Environ. Protection, Red Deer, AB, Canada, May 1998.
[37] *Multi-Sector Air Pollutants Regulations*, Environ. Climate Change Canada, Govern. Canada, Ottawa, ON, Canada, Jul. 2018.
[38] *Protocols and Performance Specifications for Continuous Monitoring of Gaseous Emissions*, Environment Canada, Ottawa, ON, Canada, Jun. 2012.
[39] *Why Use Keras?*, RStudio, San Francisco, CA, USA, 2018.
[40] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, pp. 533–536, Oct. 1986.
[41] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.
[42] T. Dozat, "Incorporating Nesterov Momentum into Adam," in *Proc. ICLR Workshop*, 2016, pp. 1–6.
[43] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," presented at the 3rd Int. Conf. Learn. Represent., San Diego, CA, USA, Dec. 2014.
[44] S. Ruder, "An overview of gradient descent optimization algorithms," Sep. 2016, *arXiv:1609.04747*, [Online]. Available: https://arxiv.org/abs/1609.04747
[45] A. Klein, K. Murphy, E. Christiansen, and F. Hutter, "Towards reproducible neural architecture and hyperparameter search," presented at the ICML-RML Workshop, Stockholmsmässan, Sweden, 2018, p. 6.



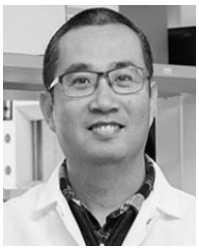
MINXING SI received the bachelor's degree in agriculture from Shenyang Agricultural University, Shenyang, China, in 2004, the master's degree in agriculture from Shanghai Jiaotong University, Shanghai, China, in 2007, and the master's degree in natural resources management from the University of Manitoba, Winnipeg, Canada, in 2011. He is currently pursuing the Ph.D. degree with the University of Calgary, Calgary, Alberta, Canada.

He is a Team Lead of the Climate Change and Sustainability Services with Tetra Tech Canada Inc. He works with a wide range of industries for their air emissions and greenhouse gas monitoring, reporting, and verification. He is a Professional Engineer (P.Eng.) with the Association of Professional Engineers and Geoscientists of Alberta, Canada.

TYLER J. TARNOCZI received the B.Sc. degree in environmental science, with a concentration in chemistry, from the University of Calgary, in 2007, the master's degree in natural resources management from the University of Manitoba, in 2009, and the MBA degree, with a specialization in finance, from the Haskayne School of Business, University of Calgary, in 2015.

He is currently a Senior Advisor at Cenovus Energy Inc., Calgary, Alberta, Canada. His work involves leading corporate and regulatory air and greenhouse gas emissions reporting, providing strategic advice on emerging environmental issues, and overseeing carbon market optimization decisions. He has developed expertise in evaluating environmental markets for commercialization opportunities, conducting life cycle assessments, and emissions accounting, reporting, and forecasting.

BRETT M. WIENS received the bachelor's degree in geography and the master's degree in geographic information systems from the University of Calgary, Calgary, Alberta, Canada, in 2004 and 2005, respectively. He has worked as a GIS Analyst with Golder Associates Ltd., and Cenovus Energy Inc., Calgary. He is currently a Senior Data Science Advisor with Cenovus Energy. He is a member of the Society of Petroleum Engineers (SPE).



KE DU received the B.S. degree in environmental engineering from Tsinghua University, in 1998, the M.S. degree in atmospheric physics and atmospheric environment from Peking University, in 2001, and the Ph.D. degree in environmental engineering from the University of Illinois at Urbana-Champaign, in 2007. He is currently an Associate Professor with the Department of Mechanical and Manufacturing Engineering, University of Calgary, Canada. His research interests include air emission monitoring technology and instrumentation, air pollution processes and mechanisms, and air pollution control technologies. He is a P.Eng. with the Association of Professional Engineers and Geoscientists of Alberta, Canada.