ORIGINAL PAPER

Sensitivity analysis and ensemble artificial intelligence-based model for short-term prediction of NO$_2$ concentration

V. Nourani[1,2] · **Z. Abdollahi[1]** · **E. Sharghi[1]**

Abstract

In this study, in the first step, three scenarios with different input combinations are created to implement a sensitivity analysis for hourly NO$_2$ prediction in Columbus City, Ohio. Three classes of inputs including concentration-related data (NO$_2$ concentration at previous time steps and NO$_2$ concentration in the suburban monitoring station), meteorology (wind speed, wind direction, and temperature), and traffic-related data (traffic count, hour of the day, and day of the week) are applied to create three scenarios. Also, the support vector regression methodology is employed to perform the sensitivity analysis. Dominant variables determined in the sensitivity analysis are applied as inputs to three models called feed-forward neural network, support vector regression, as well as classification and regression tree. In the last step, ensemble techniques including simple linear averaging, weighted linear averaging, and nonlinear support vector regression ensemble are proposed to improve the performance of sole models. The results indicate that, in the urban area, in addition to NO$_2$ variations in the previous time step, other variables such as hourly traffic count in freeway loop, suburban NO$_2$ concentration, and hour of the day can affect the NO$_2$ concentration. Further, the values of determination coefficient for the individual models, namely classification and regression tree and feed-forward neural network, are 67 and 81% that the ensemble technique as a post-processing approach enhances the performance of them up to 19% and 5% in the verification steps, respectively.

Keywords Air pollution modelling · Traffic-related pollutant · Support vector regression · Feed-forward neural network · Classification and regression tree · Columbus City

Introduction

Air pollution is a serious challenge worldwide, especially in highly populated areas such as metropolises with heavy traffic flows. Due to the development of transportation and urbanization, the number of vehicles has increased tremendously and traffic-related pollution has become one of the major concerns. The role of traffic in producing pollutants such as nitrogen oxides (NO$_x$), carbon monoxide (CO), and aromatic hydrocarbons in urban environments is undeniable, and people living in metropolitan areas are facing progressive health effects (Gilbert et al. 2005). Atmospheric emissions of NO$_2$ are mainly due to combustion processes including vehicle exhaust, coil, oil, and natural gas. Scientific evidence has shown that short-term exposure to NO$_2$ can aggravate asthma symptoms, and in some cases, hospitalization or receiving emergency treatment is necessary (U.S. EPA 2016). NO$_2$ as a gas in the atmosphere can be decomposed into nitric acid effecting both marine and soil environment and also leads to the ozone formation in sunlight. Thus, estimation of the spatiotemporal variations in air pollutants such as NO$_2$ is crucial in determining whether air pollution may cause adverse health outcomes or not. Reliable modelling also provides advanced information at an early stage based on which the government could take measures to control air pollution. On the other hand, Kambezidis et al. (2015) have shown that tropospheric NO$_2$ impacts the incoming solar radiation; they have provided a relationship

✉ V. Nourani
  vnourani@yahoo.com

1 Center of Excellence in Hydroinformatics and Faculty of Civil Engineering, University of Tabriz, Tabriz, Iran

2 Faculty of Civil and Environmental Engineering, Near East University, via Mersin 10, 99138 Nicosia, N Cyprus, Turkey

between the flux of the diffuse solar radiation and NO_2 concentration over the over Athens.

With cities expanding rapidly, estimation of pollutants produced from stationary and mobile sources becomes more complex. Highways and vehicular traffics as line sources of air pollutants are responsible for virtually all of the CO and NO_x emitted to the atmosphere near highways (Hamilton and Harison 1991). Assessing the impact of emitted pollutants' impacts becomes more complex by the fact that air pollutants could be transported far from their sources and are not confined to one location or even one region. Thus, consideration of the leading factors to explain pollutants' variations in a region depends on local meteorology and surrounding traffic patterns.

In an urban environment, determining the efficiency of every parameter contributing to air quality is a key issue in air pollution modelling. Thus, sensitivity analysis might be a major tool for investigating such effects. Such an analysis helps in not only identifying the most important parameters, but also in determining some alternate optimal decision. Sensitivity analysis based on artificial intelligence (AI) is a reliable tool to assess the efficiency of all involved parameters. In this regard, Mehdipour and Memarianfard (2019) performed sensitivity analysis using support vector regression (SVR) to examine the impacts of photochemical precursors and metrological parameters on tropospheric ozone. They found that $PM_{2.5}$, PM_{10}, CO, and NO_2 had great importance in this regard. Radojević et al. (2019) examined the sensitivity of artificial neural network (ANN) to periodic parameters alongside meteorological variables for predicting daily average concentrations of sulphur dioxide (SO_2) and NO_x. They observed that the models based on periodic parameters outperformed other models that use only meteorological variables as inputs. Elangasinghe et al. (2014) analysed the sensitivity of meteorological variables and determined the wind speed and wind direction as the most effective parameters for predicting NO_2 concentration near a major highway in Auckland, New Zealand. Optimization methods contributing to cost-effective and time-efficient models constitute the core aim of researchers when conducting sensitivity analysis. Reporting major variables involved in the air pollution field is an advantage to future researchers intending to simulate pollution trend in municipal areas with distinct geographical and urban road networks.

On the other hand, the application of various spatiotemporal variables (fixed air quality station data, satellite-based information, traffic count, meteorological data, land-used predictors, and periodic variables such as hour of the day) which are accessible and able to explain output variation is a way to develop more accurate air pollution models. Alimissis et al. (2018) applied ANN for the estimation of NO_2 concentrations at each of 13 monitoring sites located in Athens considering a specific site as target and using concentrations at remaining monitoring sites as independent variables. Results showed a wide range of determination coefficients (DCs) from 0.23 to 0.74 at the monitoring sites in the suburban and urban areas. Yeganeh et al. (2018) investigated the application of satellite-based NO_2, traffic, meteorological, and land-used predictors in adaptive neuro-fuzzy interface system (ANFIS) to propose monthly NO_2 predictions. Modelling the NO_2 variation could be conducted for hourly, weekly, and monthly values, but when it is needed to predict its concentration in hourly intervals, some limitations may arise in data accessibility. For instance, satellite-based NO_2 measurements don't cover 24-h records and are limited to a special range of time in a day (Bechle et al. 2013). Moreover, implementation of real hourly traffic as a predictor is a controversial issue in air pollution modelling. That's because the permanent automatic traffic recording stations, which provide hourly intervals, are mostly placed at highways in contrast to short-term traffic counts collected in a large number of road segments (Leduc 2008). Video image detection as non-intrusive method has also been applied to determine hourly traffic flow in multiline intersections (Jamal et al. 2015). Kamińska (2019) applied a random forest partition modelling to predict hourly NO_2 concentration using vehicle count obtained from a video camera at an intersection together with meteorological data. The results showed that the traffic flow had the greatest impact on both upper and lower values of NO_2 concentration. Elangasinghe et al. (2014) employed hour of the day, day of the week, and month of the year for representing NO_2 time variation emission in the ANN model.

AI models have been used in many fields of engineering as well as air pollution modelling (e.g. Agirre-Basurko et al. 2006; Azid et al. 2014; He et al. 2015; Feng et al. 2015; Perez and Gramsch, 2016; Cabaneros et al. 2017; Murillo-Escobar et al. 2019). Machine learning algorithms including ANN (Mishra and Goyal 2015; Bai et al. 2016) and SVR (Osowski and Granty 2007; Moazami et al. 2016) have recently shown reliable abilities in air quality modelling. Linear models such as decision tree and random forest with implementing pre-processing and post-processing approaches have also illustrated fairly successful flexibility for pollutant concentration forecasting (Kamińska 2019; Shang et al. 2019). Although different black-box models have been used in the atmospheric science, these methods may lead to different performances at different situations, and therefore, it seems that combining distinct models outputs by means of ensemble techniques may produce slightly better results. The overall idea of ensemble models is that instead of relying on an individual model or selecting the best model among a number of them, combining AI-based models outputs from linear and/or nonlinear models may capture almost all input information. In a real case, it rarely happens that an environmental time

series is solely linear or nonlinear. Thus, different aspects of fundamental patters can be taken from assembling distinct models. The concept of combining outputs has been discussed in different engineering fields including rainfall runoff models (Shamseldin et al. 1997), seepage analysis (Sharghi et al. 2018), river water quality (Elkiran et al. 2019), and vehicular traffic noise (Nourani et al. 2020). As a novel ensemble technique, one part of the present study has been allocated to the implementation of the ensemble concept in the air pollution field for the prediction of NO_2 concentration.

The main aim of this paper is to analyse NO_2 variations (from 1 January 2019 to 15 March 2019) in the station located almost close to the downtown in Columbus City. Because of the high population density in the urban area, predicting and investigating NO_2 variations in a city are far more important than in other places such as suburban areas. Highways and freeways, which provide access to several suburbs surrounding a city, heavily contribute to air pollution. Although the impact of highways on NO_2 concentration is smaller beyond 100 or 200 m, the number of people living beyond 100 or 200 m from highways may be greater than that of people living in the immediate vicinity of highways (Gilbert et al. 2007). Thus, in this study, the hourly concentration of NO_2 in the suburban station ($C_{s(t)}$) is considered as a secondary input for predicting the hourly concentration of NO_2 in the urban station ($C_{u(t)}$) (as main target). The proposed process can be summarized in three steps: first, the SVR model is applied to perform single and class sensitivity analysis to determine the dominant variables and the important classes of data for predicting $C_{s(t)}$ and $C_{u(t)}$. Three classes of data are considered as inputs including concentration-related data (CR), meteorological data (M), and traffic-related data (TRE) to create three scenarios with different input combinations. In the second step, the SVR model is proposed for $C_{s(t)}$ prediction applying the dominant inputs. Then, three machine learning models, called feed-forward-neural network (FFNN), SVR, and classification and regression tree (CART), are developed for predicting $C_{u(t)}$ using the dominant inputs determined in the sensitivity analysis as well as the values of $C_{s(t)}$ generated from the SVR model. In this case, each of the FFNN, SVR, and CART models is denoted as integrated model because of applying the SVR-generated values of $C_{s(t)}$ instead of the observed values. The FFNN as the most common model among AI models, the SVR as an almost new approach comparing other traditional ANNs, and CART as the linear model were considered for this study. In the last step, three ensembling techniques of simple linear averaging (SA), weighted linear averaging (WA), and nonlinear support vector regression ensemble (SVRE) are implemented on the outputs

of FFNN, SVR, and CART models to enhance the overall performance of the modelling.

Materials and methods

Study area and data

Columbus is the most crowded city in the US State of Ohio. Transportation in this city is based on the interstate highway system, which is a crucial component of the transportation system in the USA. The Beltway as a well-known place on the highway system encircles the city to streamline the inner-city traffic flow. In Columbus City, Interstate 270 is the beltway freeway loop, which provides access to several suburbs surrounding Columbus (Fig. 1). Regarding the air pollution monitoring system, two fixed air quality monitoring stations were considered in this study, one located in the urban area and the other in the vicinity of the beltway. Atmospheric concentrations of nitrogen dioxide (NO_2) are measured indirectly by photometrically measuring the light intensity, at wavelengths greater than 600 nm, resulting from the chemiluminescent reaction of nitric oxide (NO) with carbon monoxide. Figure 1 indicates the locations of air quality stations, the traffic counter, and the weather station in the city. From 1 January 2019 to 15 March 2019, $C_{u(t)}$ and $C_{s(t)}$ were collected from the EPA (https://www.epa.gov/), resulting in 1681 instances. Traffic count in the north part of the freeway loop (TR) and M were also gained from Ohio Department of Transportation (https://www.dot.state.oh.us/pages /home.aspx) and the Ohio State University (https://oardc .osu.edu/), respectively. In addition, parameters such as hour of the day (H) and day of the week (D) were considered as inputs in order to represent the emission rate of NO_2 from industrial and manufactural sources. Emission inventory reports explain all air pollution emissions from sources within a specific area over a specific time interval. In such reports, the North American Industry Classification System (NAICS) is used to describe what kind of economic activity is occurring in the facility. The Columbus Emission inventory point source report is available in by the Ohio Environmental Protection Agency (https://www. epa.ohio.gov/) and among 16 facilities discharging NOx to the atmosphere in Columbus; the first five facilities with the most emissions are presented in Table 1.

NO_x emissions from traffic and point sources must be controlled since this pollutant is one of the main factors participating in tropospheric ozone formation. Tropospheric or ground-level ozone is one of the criteria air pollutants that is not emitted directly into the air but is formed when NO_x and volatile organic compounds react in the presence of sunlight; it can even reach high levels during
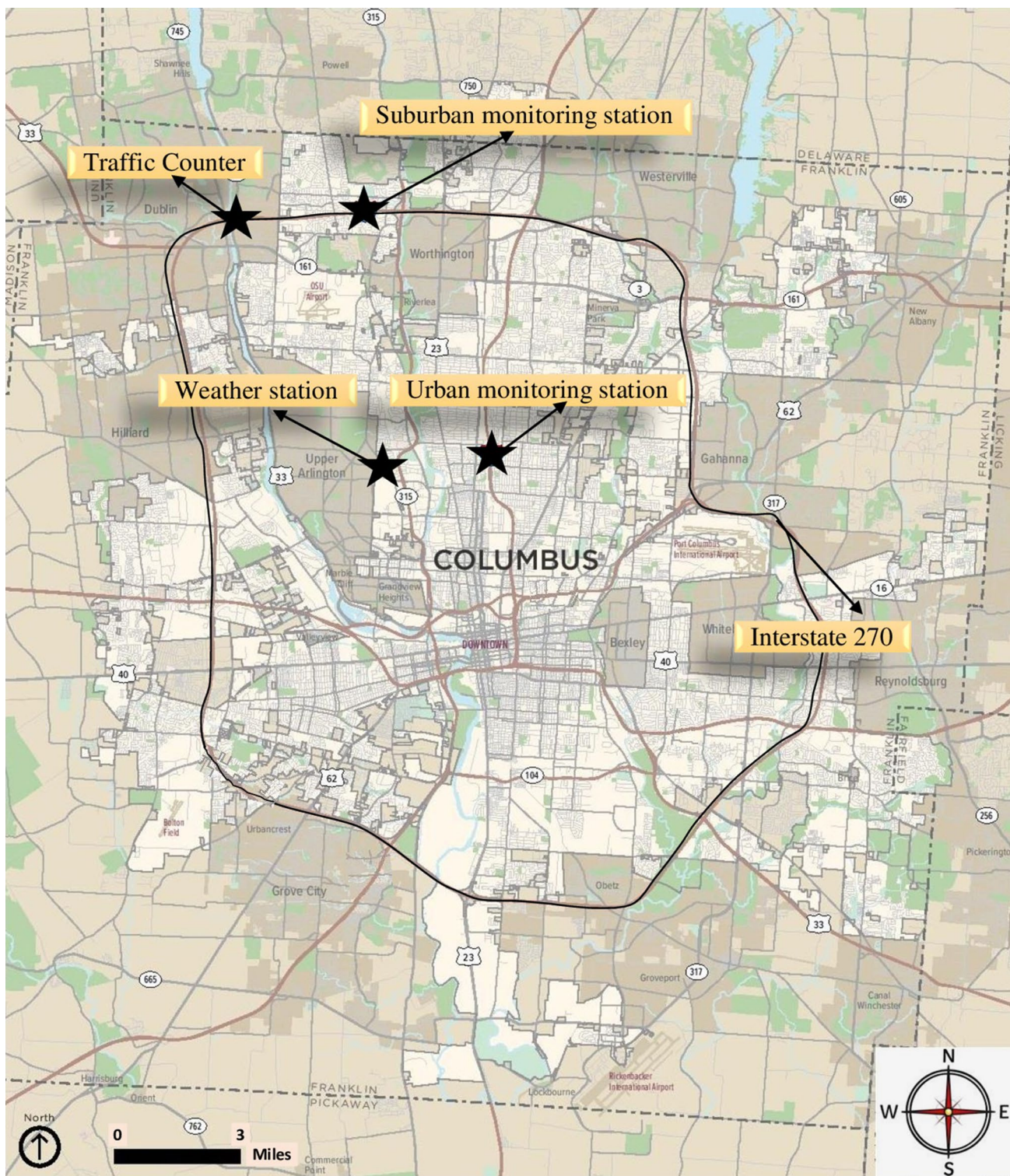
Fig. 1 Study area and data collection sites

colder months. In 3 August 2018, Columbus City was designated as a nonattainment area [any area that does not meet the national primary or secondary ambient air quality standard for a National Ambient Air Quality Standards (NAAQS)] under the 2015 ozone standard (EPA 2018). This city was also classified as marginal area based on the Clean Air Act Amendments where marginal areas have up to three years from designation to attain the NAAQS

Table 1 Five facilities with the most NO$_x$ emissions in Columbus for the year 2018 (https://www.epa.ohio.gov/)

| Facility name	NAICS	Stack/point (Ton)
MPLX terminals LLC—Columbus terminal	Petroleum bulk stations and terminals	160
Anheuser-Busch Columbus Brewery	Breweries	97.44
Ross products—division of abbott	Dry, condensed, and evaporated dairy product manufacturing	22.8136
Columbus Southerly Wastewater Treatment Plant	Sewage treatment facilities	13.92
Georgia Pacific Chemicals LLC	All other basic organic chemical manufacturing	12.279

Table 2 Statistics of the used data

	NO$_2$ concentration (ppb)		Traffic count	Urban weather station	
	$C_{s(t)}$	$C_{u(t)}$	TR (veh/h)	WS (m/s)	T (°C)
Max	48	53	15,218	17	23
Min	1	1	248	0	− 21
Average	11.53	12.08	5879	4.35	− 0.5
Standard deviation	8.87	8.95	4185	2.6	7.1

(EPA 2018). In other words, EPA has set 3-year deadline for Columbus City as a "marginal nonattainment" to come into compliance with Clean Air Act Standards. As such, consideration of measuring the traffic and industrial emissions can be efficient for predicting NO2 concentration as well as tropospheric ozone reduction.

In this study, wind speed (WS), wind direction (WD), temperature (T), NO$_2$ concentration at the suburban station at previous time step ($C_{s(t-1)}$), NO$_2$ concentration at the urban station at previous time step ($C_{u(t-1)}$), $C_{u(t)}$, $C_{s(t)}$, TR, H, and D were used in different steps of modelling in this study. Variables such as $C_{s(t)}$, $C_{s(t-1)}$, $C_{u(t)}$, and $C_{u(t-1)}$ are denoted as CR because of their relation to pollutants' concentration. TR, H and D, WS, WD, and T were also considered as TRE and M, respectively. In the process of constructing the models, collected data were divided into two parts, of which the first 80% applied for training and the rest 20% were used for the model verification purpose. Table 2 summarizes the statistics of the used data.

Table 2 shows that the peak value of $C_{u(t)}$ is higher than $C_{s(t)}$. The maximum and minimum traffic counts reported in Table 2 indicate the number of vehicles this freeway handles ranging between 248 and 15,218 vehicles per hour; nevertheless, the pattern of the traffic may give information about peak hours on the days of a week. Figure 2 compares the temporal variations in the traffic pattern in the eastern and western parts of the beltway with that in the north. The pattern of traffic variation is almost the same on the three sides (north, east, and west) of the beltway in a week (from
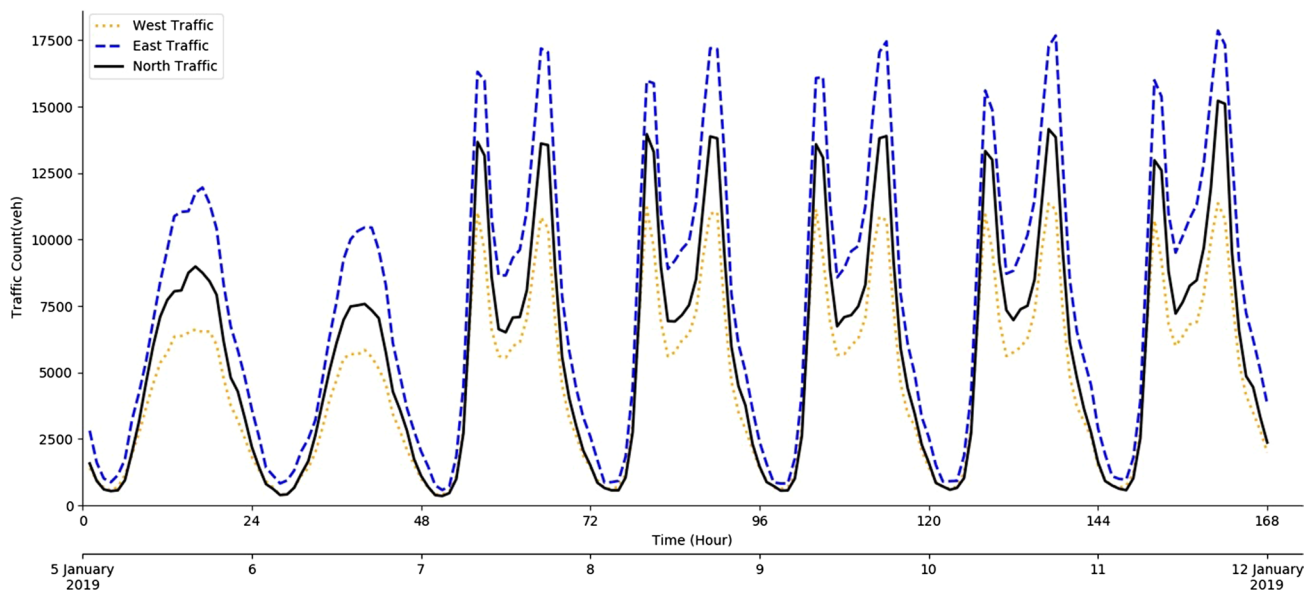
Fig. 2 Recorded traffic counts in the beltway freeway loop in the period 1 January 2019—15 March 2019

5 to 12 in January); this pattern has almost repeated for the other weeks. Further, the wind rose plot using WS and WD gives a concise but information-laden view of how WS and WD are distributed in a specific location. As revealed by the wind rose plot (Fig. 3), the prevailing wind direction from January to March is from west to east in the Columbus weather station. Such information is important to interpret the distribution of pollution over the region.

Proposed methodology

The main aim of this study was to model $C_{u(t)}$ in Columbus City. In the proposed modelling framework of this study, firstly the SVR model was developed and trained to perform left-out sensitivity analysis for both suburban and urban stations. Single and class sensitivity analyses were considered to determine the optimized classes of data and important inputs in the modelling of NO_2 concentration. Note that other machine learning models such as ANNs could be used to perform sensitivity analysis, but the SVR model was applied at this step because of its better performance. In the class sensitivity analysis, three scenarios with different combinations of classes were created to determine the importance of CR, TRE, and M. For the single sensitivity analysis in each scenario, the dominant inputs were determined and the best combination was selected to be used in the modelling. In the second step, $C_{s(t)}$ was predicted by applying the SVR model to historical data. Next, three models of FFNN, SVR, and CART were used to predict $C_{u(t)}$ based on the determined dominant inputs determined from

the sensitivity analysis as well as the generated values of $C_{s(t)}$ by the SVR model as an exogenous parameter. In this regard, each model of FFNN, SVR, and CART was denoted as an integrated model because of applying the SVR-generated values instead of the observed ones of $C_{s(t)}$. Since the SVR model was implemented using historical data from the sub-urban station, this model is also able to model future values or missing real ones consequent on measurements. Hence, the advantage of the integrated model can be attributed to predicting $C_{u(t)}$ from forecasted values of $C_{s(t)}$ using the SVR model. In the last step, three ensemble techniques based on outputs of FFNN, SVR, and CART were formed to improve the overall performance of the single models. Figure 4 presents the schematic of the proposed methodology. According to Fig. 4, initially, 8 inputs ($C_{s(t)}$, $C_{u(t-1)}$, TR, H, D, WS, T, and WD) were fed to the SVR model to perform the sensitivity analysis for determining the important inputs for $C_{u(t)}$ prediction. Since $C_{s(t)}$ was determined as a dominant input, an SVR model was also developed for $C_{s(t)}$ prediction using 3 dominant inputs ($C_{s(t-1)}$, TR, and H) resulting from the sensitivity analysis. Afterwards, the values of $C_{s(t)}$ predicted from the SVR model were applied as input for $C_{u(t)}$ prediction. In the second step, 4 inputs ($C_{u(t-1)}$, TR, H, and generated $C_{s(t)}$) were used as inputs to models of SVR, FFNN, and CART. In the third step, three ensemble techniques (SA, WA, and SVRE) were implemented to combine the outputs of the models of SVR, FFNN, and CART models. Sensitivity analysis and ensembling technique were implemented on inputs and the outputs of diverse models, respectively, to reduce defects that may arise in environmental issues.

Fig. 3 Wind rose plot in the weather station for the period 1 January 2019—15 March 2019
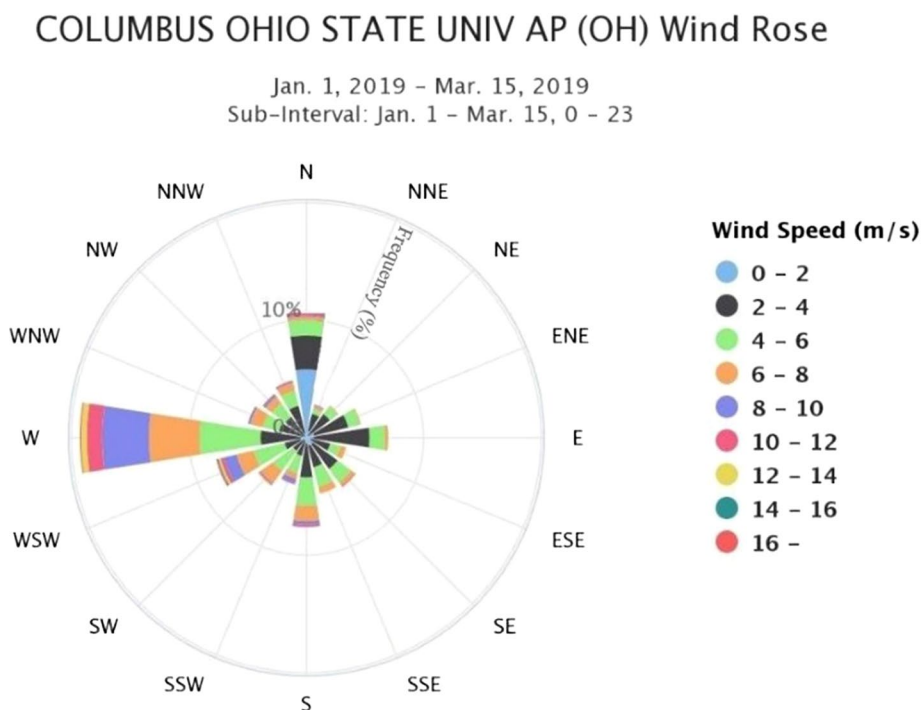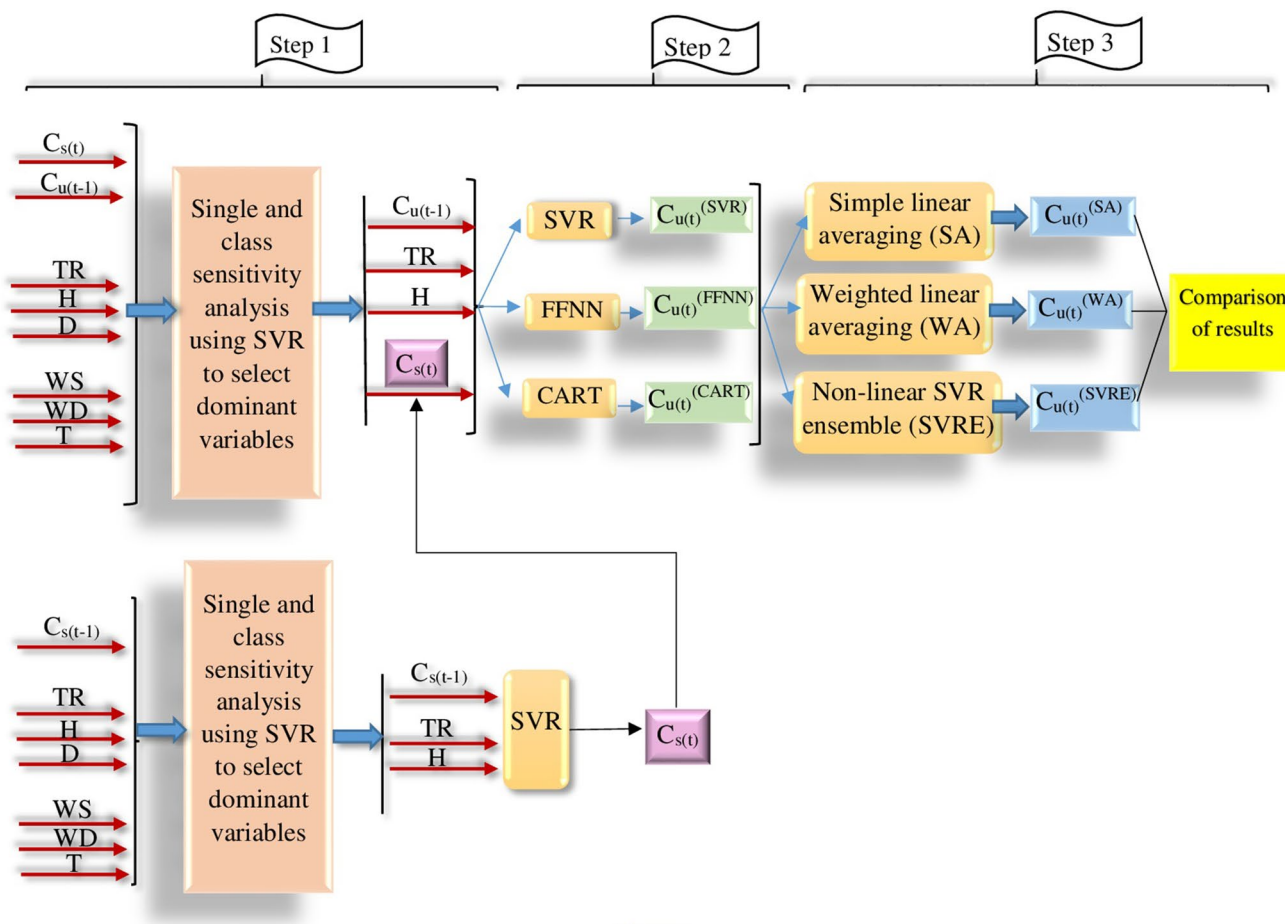
Fig. 4 Schematic of the proposed methodology

For example, there are several factors involved in modelling the concentration of NO_2 that may vary from one region to another; even in one region, the concentration of the pollutant may be more sensitive to some factors. Therefore, performing a sensitivity analysis on inputs is a method to define the dominant variables, which can explain the NO_2 variation regarding the geographical condition and urban road network. On the other hand, by ensembling diverse models, the problem of choosing a suitable model can be handled, because, in real-world cases, it is difficult to define a time series as solely linear or nonlinear. As a result, assigning an AI model to a complex environmental time series may not seem reliable. Previous studies such as Zhang (2003) have proved this point that there is no unique model to define the process perfectly. To investigate this concept in NO_2 time series modelling, linear CART and nonlinear AI models were applied to detect and capture the linear and nonlinear portions of NO_2 time series; the obtained results from ensembling techniques were compared with the individual model outcomes.

Sensitivity analysis

In this study, AI-based (here SVR) left-out method (Nourani et al. 2019) was used for determining every variable efficiency. In the left-out method, one of the variables was left out and the SVR model was trained with the rest of the variables; afterward, the left-out input was switched for every input used in the model. In this way, contributions of all parameters were evaluated and it is clear that the more efficient the variable is, the greater reduction in the model's accuracy occurs. In other words, when the left-out variable is switched for an important input, the model performance abruptly reduces because an important input is extracted and switched for a less important variable (here left-out variable). In addition, in the process of training for a distinct combination of inputs, to find the best fit and tuning the SVR parameters, the grid search approach using cross-validation was applied (Hsu et al. 2003). In this approach, different values of the parameters were examined and the one with the best cross-validation accuracy was selected (Hsu et al. 2003). This method is time-consuming and seems naive,

but it is still more straightforward over several advanced methods; the drawback of being time-consuming can be handled by using a coarse grid to identify "better" region on the grid and then constructing a fine grid on the better region, so using a coarse grid first and then a finer grid search on that region can be used to investigate the best SVR model parameters (Hsu et al. 2003). Other AI models such as ANN and CART could be used for the sensitivity analysis process. If an ANN model was applied instead of the SVR, the problem of tuning the SVR parameters would turn into determining the best architecture for the ANN model. In the present paper, in order to investigate single and class sensitivity analysis, three scenarios were considered based on different classes of data. The main goal of creating three scenarios was to investigate the efficiency of every class of data as well as every single variable in the modelling of NO_2 concentration. Because of the importance of CR such as $C_{s(t)}$ and $C_{u(t-1)}$ in the urban station and $C_{s(t-1)}$ in the suburban station, they were applied as common class of data in all three scenarios, and then, other classes of data were included in each scenario. That way, the importance of every class of data was revealed in the modelling of NO_2 variation. It should be noted that for both urban and suburban stations, the necessity of applying additional NO_2 times series of previous time steps including $C_{u(t-2)}$, $C_{u(t-3)}$, $C_{u(t-4)}$, and $C_{u(t-5)}$ in the urban station and $C_{s(t-2)}$, $C_{s(t-3)}$, $C_{s(t-4)}$, and $C_{s(t-5)}$ in the suburban station was examined and it was concluded that using only $C_{u(t-1)}$ and $C_{s(t-1)}$ as inputs was appropriate for reaching the best performance of the modelling in the urban and suburban stations, respectively.

Scenario 1

In this scenario, 2 classes of data including CR and M were taken into account for hourly NO_2 prediction in both urban and suburban stations as:

$$C_{u(t)} = f\left(C_{u(t-1)}, C_{s(t)}, WS, WD, T\right) \tag{1}$$

$$C_{s(t)} = f\left(C_{s(t-1)}, WS, WD, T\right) \tag{2}$$

where WS, WD, and T represent the wind speed, wind direction, and temperature, respectively; f stands for the predictor model, which can be SVR, FFNN, or CART; $C_{u(t)}$ and $C_{u(t-1)}$ denote the concentration of NO_2 in the urban station in the current and previous time step in the urban station, respectively, and $C_{s(t)}$, $C_{s(t-1)}$, respectively, but in the suburban station.

Scenario 2

Scenario 2 was similar to scenario 1 in terms of CR. This scenario was created by keeping the CR fixed and replacing M with TRE. In other words, for NO_2 concentration prediction in both urban and suburban stations, the CR and TRE classes of data were used as:

$$C_{u(t)} = f\left(C_{u(t-1)}, C_{s(t)}, TR, D, H\right) \tag{3}$$

$$C_{s(t)} = f\left(C_{s(t-1)}, TR, D, H\right) \tag{4}$$

where TR, D, and H present the traffic counts, day of the week, and hour of the day, respectively.

Scenario 3

In the third scenario, which is a combination of scenarios 1 and 2, three classes of data were considered for NO_2 prediction in both urban and suburban stations. Thus, CR, M, and TRE were applied in NO_2 concentration modelling as:

$$C_{u(t)} = f\left(C_{u(t-1)}, C_{s(t)}, WS, WD, T, TR, D, H\right) \tag{5}$$

$$C_{s(t)} = f\left(C_{s(t-1)}, WS, WD, T, TR, D, H\right) \tag{6}$$

Support vector regression (SVR)

SVM was first proposed and developed by Vapnik (1995) based on statistical learning theory and has been prioritized for considering to solve various pattern recognition problems among many available supervised learning methods (Li et al. 2020). SVR is a new and promising approach that employs the structural risk minimization principals. In this approach, instead of minimizing the error between the observed and computed values, the operational risk as the objective function is considered to be minimized. In SVR, first a linear regression is fitted to the data, and then, the outputs go through a nonlinear kernel to catch the nonlinear pattern of the data. The SVR principles for regression are as follows. Given a dataset of N elements $\{(x_i, d_i) i = 1, 2, \ldots N\}$, ($x_i$ is the input vector, d_i is the actual value, and N is the total number of data points); the general SVR function is written as Eq. (7) (Wang et al. 2013):

$$y = f(x) = w\varphi(x_i) + b \tag{7}$$

where $\varphi(x_i)$ represents feature spaces, non-linearly mapped from the input vector x; w and b are the weight vector and

adjustable factor which both can be determined by allocating positive values for the slack parameters of ξ and ξ^* and minimizing the error function [Eq. (8)] (Wang et al. 2013):

$$\text{Minimize}: \frac{1}{2}\parallel w \parallel^2 + C\left(\sum_i^N (\xi_i + \xi_i^*)\right) \tag{8}$$

With the constrains:

$$\begin{cases} w_i\varphi(x_i) + b_i - d_i \leq \varepsilon + \xi_i^* \\ d_i - w_i\varphi(x_i) + b_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \quad i = 1, 2, 3 \ldots N \end{cases}$$

where $\left(1/2\right) \parallel w \parallel^2$ is the weights vector norm and C is referred to the regularized constant determining the trade of the empirical error and the regularized term. ε is called the tube size and is equivalent to the approximation accuracy placed on the training data points. Mentioned optimization problems can be changed to a dual quadratic optimization problem by defining Lagrange multipliers α_i and α_i^*. The vector w in Eq. (7) can be computed after solving the quadratic optimization problem as:

$$w^* = \sum_{i=1}^N (\alpha_i - \alpha_i^*)\varphi(x_i) \tag{9}$$

So the final form of SVR can be expressed as (Wang et al. 2013):

$$f(x, \alpha_i, \alpha_i^*) = \sum_{i=1}^N (\alpha_i - \alpha_i^*)K(x, x_i) + b \tag{10}$$

where α_i and α_i^* are Lagrange multipliers, $K(x, x_i)$ is referred to kernel function, which is capable of nonlinearly mapping into feature space, and b is the bias term. One of the most used kernel functions is the radial basis function (RBF) which is written as follows:

$$K(x_1, x_2) = \exp\left(-\gamma \parallel x_1 - x_2 \parallel^2\right) \quad \gamma > 0 \tag{11}$$
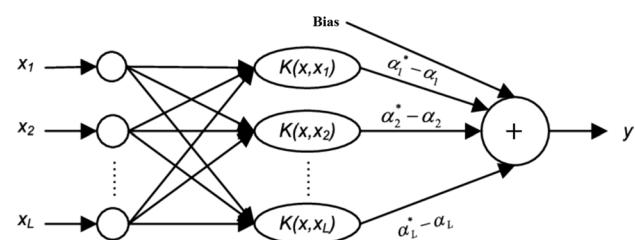
where γ is the kernel parameter.

The generalization capacity of the SVR model is highly dependent on the good tuning of the kernel parameter (γ) in Eq. (11) and tuning parameters C and ε in Eq. (8). A characteristic SVR structure is displayed in Fig. 5. For tuning these parameters in this study, the grid search approach using cross-validation was applied (Hsu et al. 2003).

Feed-forward neural network (FFNN)

ANNs as a black box tool have been widely used in different fields of engineering. Feed-forward neural network (FFNN) as an ANN model is the first and simplest type of neural network in which information moves forward through the input layer, hidden layers, and output layer, sequentially (Fig. 6). Multi-layer feed-forward neural networks, trained with a back-propagation learning algorithm, are the most popular neural networks. The multi-layer neural-network performance can be considered in two modes: training and prediction. Training and test datasets are used for the training and prediction modes. The training mode starts with arbitrary values of the weights and proceeds iteratively. Each iteration of the complete training set is called an epoch. In each epoch, the network adjusts the weights in the direction that reduces the error (back-propagation algorithm). As the iterative process of the adjustment continues, the weights gradually converge to the locally optimal set of values. Many epochs are usually required before training is completed. Researches indicate that a three-layer FFNN, which consists of an input layer, hidden layer, and output layer, has the capability of sufficient performance in the environmental modelling (ASCE 2000; Nourani 2017). The explicit equation to determine the output value of a FFNN is obtained by Eq. (12) (Nourani et al. 2015):

$$\hat{y}_j = f_j\left[\sum_{h=1}^m w_{jh} \times f_h\left(\sum_{i=1}^n w_{hi}x_i + w_{hb}\right) + w_{jb}\right] \tag{12}$$
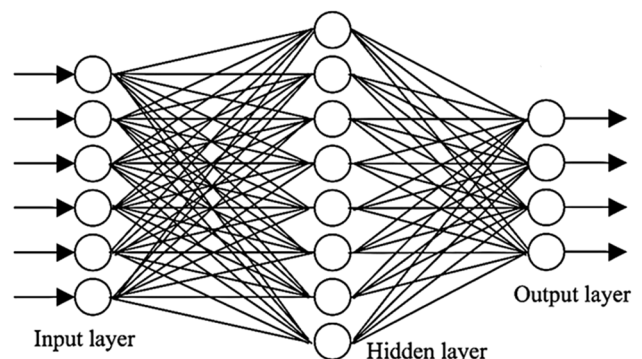

Fig. 5 Structure of SVR

Fig. 6 Structure of FFNN

where i, h, j, b, and w represent the input, hidden, and output layer bias, and the applied weight (or bias), respectively; f_h and f_j stand for the activation function of the hidden and output layers, respectively; x_i, m, n show, respectively, the input layer variable, the number of input, and the number of hidden neurons; and y, $\hat{y}_j$ denote the observed and computed values of the output neuron, respectively. The hidden and target layer weights are different from each other and should be estimated within the training phase.

Classification and regression tree (CART)

Decision tree is one of the non-parametric classification methods which can introduce a pattern classification of observations utilizing a simple technique. Normally, decision tree is drawn from top to the down in which the root is placed at the top. The end of a chain which comprises of root, branch, and node is named as leaf. Each node can be split into two branches. Each node is related to a certain characteristic (input parameter), and branches are described a specific range of input parameters (Liang et al. 2016). Figure 7 schematically shows the structure of a decision tree. The main concept of CART algorithm, developed by Breiman et al. (1984), is to recurrently split the input space into dual subsets until the output becomes more homogenous. Given a dataset of training samples $\{(x_i, y_i) i = 1, 2, \ldots .l)\}$ where $x_i \in R^m$ is the ith input vector and $y_i \in R$ is the corresponding output. CART begins with the root nod, which contains the whole training samples. The next step is to calculate the first split, in which for a regression problem, the split is to minimize the expected sum variances for two resulting subsets (Shang et al. 2019):

$$\min_{j,c} \frac{1}{l} \left(\sum_{k \in S_L} (y_k - \bar{y}_L)^2 + \sum_{k \in S_R} (y_k - \bar{y}_R)^2 \right) \tag{13}$$

$$\begin{cases} \text{s.t} S_L = \{i | x_{ij} \le c, i = 1, \ldots, l\}, \\ \quad S_R = \{i | x_{ij} > c, i = 1, \ldots, l\}, \\ \quad\quad\quad j \in \{1, \ldots, m\} \end{cases}$$

where S_L and S_R are the sets of training indices going to left child node and right child node, $\bar{y}_L$ and $\bar{y}_R$ denote the mean values of the output of samples in the two subsets.

The children of the root node are recursively split in the same manner until some stop criterion is satisfied. By moving from the root to the terminal node (leaf), each sample is assigned to a unique leaf, in which the mean value of samples in a leaf is chosen as the predicted value (Shang et al. 2019).

Efficiency criteria

In this study, a training set was employed to build the predictive model, and a test set was used to examine the trained model. The Determination Coefficient (DC) and root mean square error (RMSE) efficiency criteria were applied in this paper to evaluate the performance of the models as (Nourani 2017). The value of DC represents the percentage of the square of the correlation between the predicted and actual values of the target variable (Armaghani and Asteris 2020). RMSE represents the standard deviation of the fitted error between the predicted and observed values (Zhou et al. 2020). The calculation formulas of the evaluation indicators are presented as follows:

$$DC = 1 - \frac{\sum_{i=1}^{n} \left(C_{N \, obs_i} - C_{N \, com_i} \right)^2}{\sum_{i=1}^{n} \left(C_{N \, obs_i} - \overline{C}_{N \, obs} \right)^2} \tag{14}$$

Fig. 7 Structure of decision tree

= Root Node

= Non-terminal Node

= Terminal Node

= Leaf
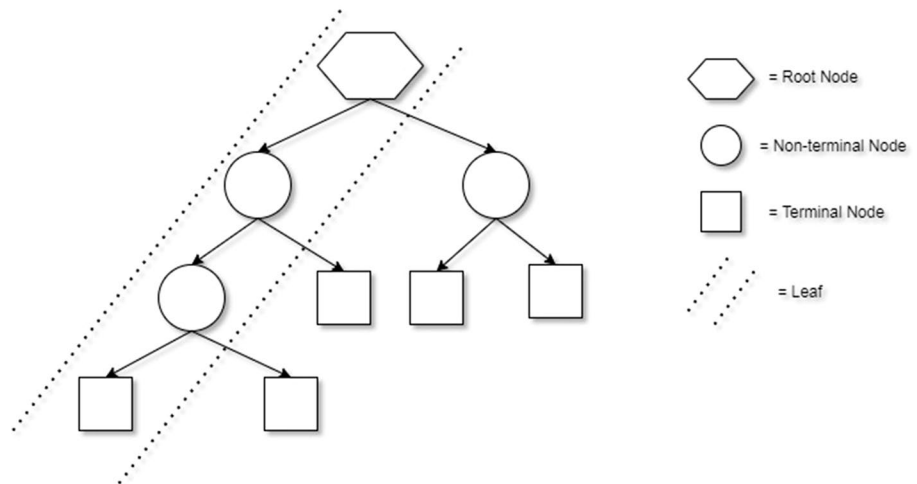
$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(C_{N\,\text{obs}_i} - C_{N\,\text{com}_i})^2}{n}} \qquad (15)$$

where n, $C_{N\,\text{obs}_i}$, $\overline{C}_{N\,\text{obs}}$, and O_{com_i} are the number of data points, the observed NO_2 data, the average value of the observed data, and the calculated values, respectively. DC ranges between $-\infty$ and 1, with perfect score of 1.

Ensembling unit

Ensembling techniques as post-process approaches have shown the ability to improve model's prediction by combining various model outputs. It has been proved that it is less risky to use a combination of relatively simple models than to use a single model, which is more complex and expensive (Makridakis and Winkler 1983). In this paper, three ensembling techniques were applied for combining the outputs of the FFNN, SVR and CART models. The first two linear ensemling techniques including SA and WA were implemented according to Eqs. (16) and (17) (Sharghi et al. 2018).

$$\overline{C_u}(t) = \frac{1}{M}\sum_{i=1}^{M} C_{u_i}(t) \qquad (16)$$

where $C_{u_i}(t)$ is the output of the ith individual model (here, outputs of FFNN, SVR, and CART), $\overline{C_u}(t)$ is the output of the simple linear ensemble model and M is the number of single models (here 3).

$$\overline{C_u}(t) = \sum_{i=1}^{M} w_i C_{u_i}(t) \qquad (17)$$

where w_i is the applied weight on the ith model which can be written as:

$$w_i = \frac{\text{DC}_i}{\sum_{i=1}^{M} \text{DC}_i} \qquad (18)$$

where DC_i is the determination coefficient of the ith individual model.

The third ensembling technique which is nonlinear averaging is implemented by the SVR model using outputs of three models, namely the SVR, FFNN, and CART. It should be noted that the training dataset was used for both computation of W_i in Eq. (18) and training the SVR ensembling technique. In this part, other AI models such as FFNN could be used for ensembling, but the SVR model as an almost new model in machine learning approaches was considered to combine the outputs of three models.

Results and discussion

The results of this study consist of three parts presented in three sections separately as follows:

Results of the sensitivity analysis

At the first step of the modelling, single and class sensitivity analysis were performed based on the SVR model for both suburban and urban stations. Table 3 presents the results of the NO_2 sensitivity analysis for all 3 scenarios in the urban station.

In scenario 1, one of the meteorological variables (T) was left out, where SVR model was trained and verified by the rest of the inputs. According to Table 3, the first row of each scenario is the first step of sensitivity analysis; one of the variables was left out and the SVR model trained by the rest of the inputs. The last row in each scenario is related to the applying of all classes of data based on the specific scenario. The rest of the rows in Table 3 represent the process of sensitivity analysis that the left-out variable switches for each input. For instance, the fourth row in scenario 1 shows the switching of T for $C_{s(t)}$ comparing the first row. So, the percentage of change in DC value in the verification step is 17% (0.84–0.67 = 17%).

By replacing T with WS and WD parameters, no remarkable changes were observed in DC; in contrast, replacing T with $C_{s(t)}$ and $C_{u(t-1)}$ led to an abrupt reduction in the model performance by up to 17% and 7% in the verification step, respectively. As such, $C_{u(t)}$ is more sensitive to the $C_{s(t)}$ compared to $C_{u(t-1)}$. This outcome confirms that the NO_2 time series is not an autoregressive process and applying additional variables in previous time steps such as $C_{u(t-2)}$ may not seem reasonable.

In scenario 2, D was first left out, similar to scenario 1 the SVR model was developed and the related parameters were tuned to perform the sensitivity analysis. The results presented in Table 3 indicated that $C_{s(t)}$ and $C_{u(t-1)}$ are still the most sensitive variables, which can affect the model performance by up to 17% and 11%, respectively. Further, by replacing D with TR, the model performance was reduced by up to 9% in the verification step. This means that $C_{u(t)}$ is sensitive to TR. Freeways with a large volume of vehicles (here 5879 vehicles per hour on average) seem to influence NO_2 variations beyond the adjacent region. Gilbert et al. (2007) reported this issue by implementing land-use regression in 55 locations with different distances from the nearest highway (the maximum distance from the nearest highway was 5.264 km). Their research revealed that by excluding locations less than 200 m, the NO_2 concentration was still significantly associated with traffic count in the nearest highway. They also reported that the upwind/downwind location

Table 3 Results of the NO2 sensitivity analysis for all 3 scenarios in the urban station for modelling $C_{u(t)}$

Row number	Model variables	SVR parameters	DC		RMSE (ppb)	
		ε, c, γ	Training	Verification	Training	Verification
Scenario 1						
1	$C_{s(t)}, C_{u(t-1)}$, WS, WD	0.12-20-0.08	0.89	0.84	4.88	3.6
2	$C_{s(t)}, C_{u(t-1)}$, WS, T	0.13-20-0.06	0.88	0.83	4.95	3.63
3	$C_{s(t)}, C_{u(t-1)}$, T, WD	0.13-20-0.07	0.87	0.83	4.99	3.62
4	T, $C_{u(t-1)}$, WS, WD	0.14-25-0.09	0.77	0.67	6.99	5.25
5	$C_{s(t)}$, T, WS, WD	0.15-25-0.02	0.80	0.77	5.90	4.56
6	$C_{s(t)}, C_{u(t-1)}$, WS, WD, T	0.14-25-0.09	0.8	0.82	5.88	3.7
Scenario 2						
1	$C_{s(t)}, C_{u(t-1)}$, TR, H	0.15-45-0.15	0.92	0.87	4.41	3.38
2	$C_{s(t)}, C_{u(t-1)}$, TR, D	0.13-30-0.07	0.83	0.80	4.81	3.59
3	$C_{s(t)}, C_{u(t-1)}$, D, H	0.12-30-0.25	0.80	0.78	4.89	3.84
4	D, $C_{u(t-1)}$, TR, H	0.11-25-0.06	0.72	0.70	5.99	5.33
5	$C_{s(t)}$, D, TR, H	0.12-25-0.05	0.81	0.76	4.85	4.41
6	$C_{s(t)}, C_{u(t-1)}$, TR, H, D	0.15-30-0.06	0.85	0.82	5.23	3.6
Scenario 3						
1	$C_{s(t)}, C_{u(t-1)}$, TR, D, H, WS, WD	0.12-20-0.05	0.856	0.84	4.89	3.32
2	$C_{s(t)}, C_{u(t-1)}$, TR, D, H, WS, T	0.15-20-0.08	0.864	0.82	4.81	3.44
3	$C_{s(t)}, C_{u(t-1)}$, TR, D, H, T, WD	0.12-10-0.08	0.849	0.83	5.03	3.43
4	$C_{s(t)}, C_{u(t-1)}$, TR, D, T, WS,WD	0.13-20-0.07	0.842	0.84	5.07	3.32
5	$C_{s(t)}, C_{u(t-1)}$, TR, T, H, WS,WD	0.13-20-0.07	0.845	0.83	4.98	3.39
6	$C_{s(t)}, C_{u(t-1)}$, T, D, WS, WD,H	0.11-20-0.07	0.843	0.83	5	3.45
7	$C_{s(t)}$, T, D,H, WS, WD, TR	0.16-35-0.009	0.85	0.81	5.02	3.61
8	T, $C_{u(t-1)}$, D, H, WS, WD, TR	0.14-20-0.07	0.743	0.71	6.58	4.78
9	$C_{s(t)}, C_{u(t-1)}$, TR, D, H, WS, WD, T	0.16-30-0.08	0.85	0.82	5.04	3.58

of sampling sites relative to nearest highways was not determined, and therefore, it was impossible to compare the influence of highway between upwind and downwind locations. In other words, their research was conducted without considering meteorological parameters such as wind speed and wind direction. In the current study, the distance of the urban station from the northern part of the beltway was approximately 13.4 km. Also, to explain this sensitivity, the traffic counts in the eastern and western parts of the beltway were investigated; their distances were about 7.7 km and 10.5 km from the urban station, respectively. Figure 2 displays the time series of the traffic counts in the northern, eastern, and western parts of the beltway. The traffic patterns in the three parts of the beltway are very similar, and by replacing the east traffic counts with northern ones, no changes were observed in the DC value of the SVR model. Hence, regarding this pattern similarity, the sensitivity of the NO2 concentration can be attributed to the eastern traffic counts, which has the lowest distance (~ 7.7 km) from the urban station and with the same traffic pattern of the northern part. In addition, replacing D with H also showed a reduction in the modelling performance by up to 7% in the verification step. The main reason that H was considered as input to the NO2

modelling was that it could implicitly represent the point source emissions in a city. According to Table 1, the petroleum industry and brewery manufactory discharged about 160 and 97.44 tons of NO_x and accounted for approximately 70% of total point source NOx emissions in Columbus. The petroleum industry is located in the western part of the urban station where the prevailing wind direction is from west to east (Fig. 3). The brewery is located in the northern part of the urban station where the second most frequent wind direction is from north. It seems that considering the operating hours of the facilities and their locations can almost explain the model's sensitivity to H. However, this result is for a specific time interval (from 1 January to 15 March) and may show less sensitivity in other seasons of the year; therefore, this issue requires further investigation.

In scenario 3, three classes of data were used altogether to compare the changes caused in every variable sensitivity. It was also examined that whether applying all related variables might improve the model performance in comparison with scenarios 1 and 2. The results in Table 3 indicate that replacing T with $C_{u(t-1)}$ did not lead to a specific change in the model's performance. In other words, by applying both TRE and M, the output ($C_{u(t)}$) was not sensitive to $C_{u(t-1)}$

Table 4 Results of the NO$_2$ sensitivity analysis for all 3 scenarios in the suburban station for modelling C$_{s(t)}$

Model variables	SVR parameters	DC		RMSE (ppb)	
		Training	Verification	Training	Verification
Scenario 1					
C$_{s(t-1)}$, WS, WD	0.13-20-0.07	0.851	0.79	4.34	3.27
Scenario 2					
C$_{s(t-1)}$, TR, H	0.15-20-0.8	0.892	0.835	3.94	3.11
Scenario 3					
C$_{s(t-1)}$, WS, WD, T, TR, H	0.19-25-0.9	0.803	0.786	5.13	4.12

anymore. Thus, TR, WS, WD, T, H, and D could be replaced with C$_{u(t-1)}$, though C$_{s(t)}$ is still the most sensitive variable (similar to scenarios 1 and 2) affecting the model's accuracy by up to 13%.

Overall, in terms of the single sensitivity analysis, it was concluded that the NO$_2$ concentration in the urban station could be sensitive to TR and H. This result reveals that depending on the urban road network and freeways, traffic counts should be considered and investigated in a city for modelling NO$_2$ variations. The same traffic pattern at the three sides of the beltway may also give a clue for future studies, in the case that traffic counts in a freeway are accessible for a limited time interval in a city; other freeways or highways with a similar traffic pattern can be used as a surrogate. In addition, in the NO$_2$ modelling, the role of meteorological variables is so complex that even in one region dominant meteorological parameters may differ from one season to another. This complexity is not limited by the seasons; changing dominant meteorological parameters may also differ for the high and low concentrations of NO$_2$. Kamińska (2019) developed two models for upper and lower values of NO$_2$ concentration and showed that the meteorological parameters influencing upper and lower values of NO$_2$ concentration are significantly different, although the hourly traffic count is the most important variable in both parts of the modelling. In the current study, every meteorological input (WS, WD, and T) did not show specific sensitivity to the NO$_2$ concentration modelling. This result was based on the considered time interval (1 January to 15 March), which may vary in other seasons of the year and requires more

attentions. The last point gained from the single sensitivity analysis could be attributed to the importance of CR. The results for scenarios 1, 2, and 3 revealed that C$_{s(t)}$ is the most dominant parameter in all three scenarios. The sensitivity of C$_{u(t-1)}$ diminished in scenario 3, to the extent that it didn't show specific sensitivity to NO$_2$ variations. On the contrary, by excluding C$_{s(t)}$, the model's accuracy dropped by up to 13%, even when all related classes of data (scenario 3) were used for the NO$_2$ concentration modelling. This may reveal the importance of suburban NO$_2$ variations in the prediction of urban NO$_2$ concentration.

Regarding the class sensitivity analysis presented in Table 3, the results for the class sensitivity analysis are bolded and the best combination of inputs in every scenario is highlighted. It is clear that scenario 3 could not be a proper choice for NO$_2$ prediction among the three scenarios. This is because scenario 3 showed almost the same accuracy in the verification step (DC = 0.82) as scenario 1 (DC = 0.82) and 2 (DC = 0.82) in the NO$_2$ prediction, while using more classes of data is not cost-effective. It was also found that applying all related parameters (scenario 3) may not improve the modelling performance. Among scenarios 1 and 2, it could be seen that TRE was almost as efficient as M class of data when they were accompanied by C$_{u(t-1)}$ and C$_{u(s)}$ (bolded in Table 3). One combination of inputs should be selected for the next step of the NO$_2$ concentration modelling at the urban station. Thus, among different input combinations in scenarios 1 and 2, the one with a better performance in the verification step was selected as input to the NO$_2$ modelling in the next step. The results in Table 3 showed that 87% of the NO$_2$ variation could be explained by the variation in 4

Table 5 Results of the integrated models for the prediction of C$_{u(t)}$

Model variables	Model	Model parameters[a]	DC		RMSE (ppb)	
			Training	Verification	Training	Verification
C$_{s(t)}$, C$_{u(t-1)}$, TR, H	SVR	0.15-45-0.15	0.923	0.870	3.38	4.41
C$_{s(t)}$, C$_{u(t-1)}$, TR, H	FFNN	4-10-1	0.85	0.813	3.69	5.27
C$_{s(t)}$, C$_{u(t-1)}$, TR, H	CART	1–6	0.778	0.672	4.89	6.76

[a]In the SVR model a-b-c denote ε, c, γ. In the FFNN model a–b–c in the structure of FFNN represent the number of input layer, hidden layer and output layer neurons. In the CART model a–b refer to the minimum number of samples at a leaf node and maximum depth of the tree
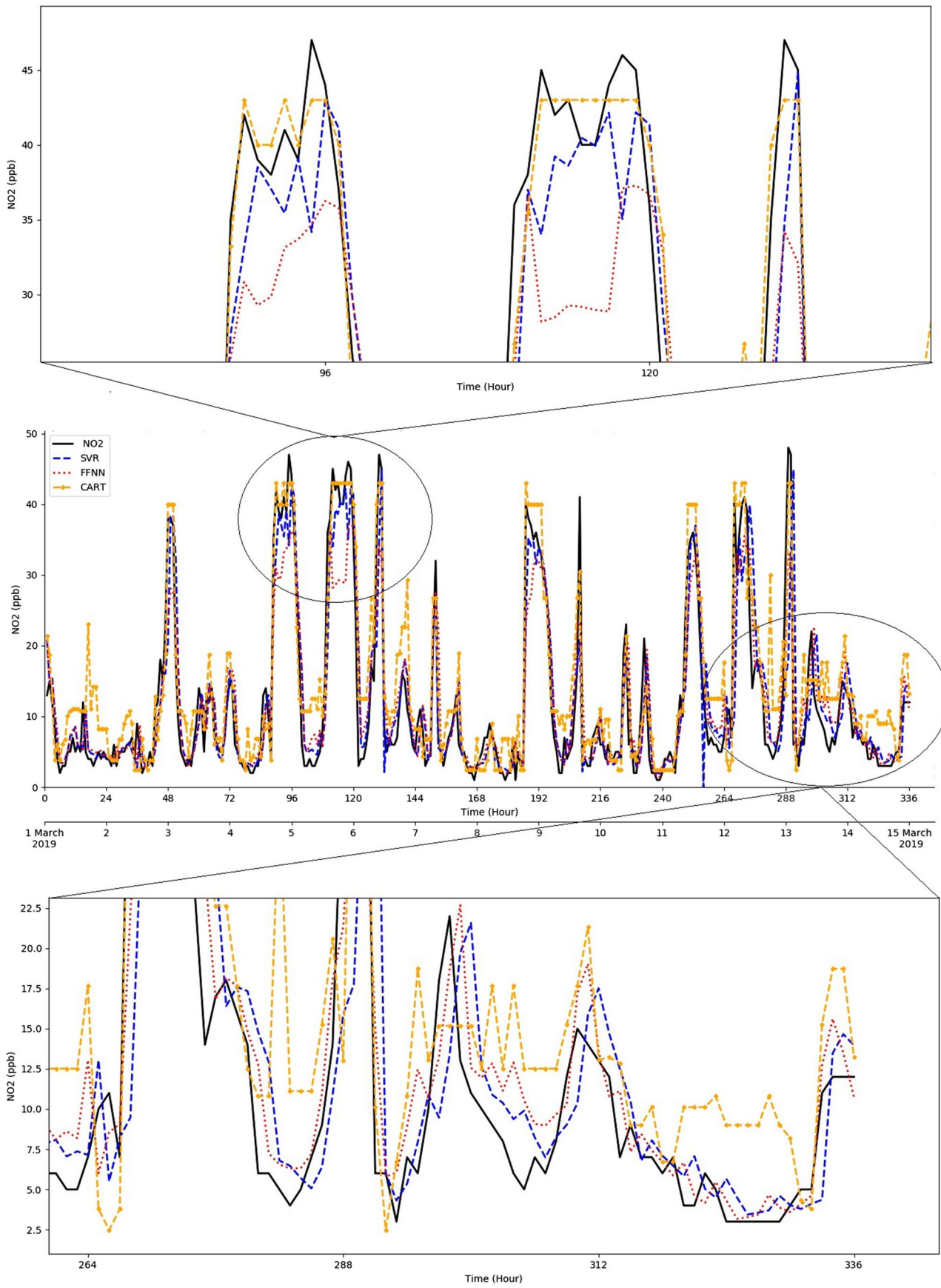
◀**Fig. 8** Estimated NO$_2$ concentration (ppb) from the SVR, FFNN, and CART models

inputs, namely $C_{s(t)}$, $C_{u(t-1)}$, TR, and H. Thus, for the next step of the modelling, they were considered as inputs to the FFNN, SVR, and CART models as:

$$C_{u(t)} = f\left(C_{u(t-1)}, C_{s(t)}, TR, H\right) \qquad (19)$$

where the concentration of NO$_2$ at the urban station ($C_{u(t)}$) could be considered as a function of its concentration in a previous time step ($C_{u(t-1)}$), NO$_2$-concentration in the current time step at suburb station ($C_{s(t)}$), the hourly traffic count in the northern section of the beltway (TR), and the hour of the day (H); f stands for the predictor model, which can the SVR, FFNN, and CART models.

Moreover, sensitivity analysis was also performed in the suburban station to determine the most important inputs to NO$_2$ prediction. The best input combinations of three scenarios for $C_{s(t)}$ prediction are presented in Table 4.

The results in Table 4 show that in the suburban station, the dominant variables contributing to NO$_2$ variation are similar to those for the urban station. Since the suburban station is located in the vicinity of the beltway, it was expected that TR and H were selected as dominant inputs to NO$_2$ modelling. Thus, it could similarly be concluded that applying TR, H, and $C_{s(t-1)}$ could be the best choice for NO$_2$ prediction in the suburban station.

Results of the integrated modelling

At the second step, an integrated model was implemented to predict $C_{u(t)}$. In this model, instead of observed values of $C_{s(t)}$, the generated values from the SVR model were applied for $C_{u(t)}$ prediction. Table 5 presents the results for three models of SVR, FFNN, and CART as integrated models for prediction of the NO$_2$ concentration in the urban station.

For development of the integrated models, the SVR, FFNN, and CART models were trained and evaluated using efficient inputs selected in the previous step. In the SVR case, the model performance is highly depended on the selected parameters; for tuning C, ε (SVR model) and γ (RBF kernel function) grid search method was used (Hsu

et al. 2003). In the FFNN case, considering the tangent sigmoid as the activation function of the hidden and output layers, the FFNN was trained using the scaled conjugate gradient scheme of the back-propagation algorithm (Haykin 1994). In addition, a proper architecture for the network including the number of hidden neurons in the hidden layer and optimal iteration epochs is important to prevent the training process from overfitting. Hence, the range of 1–15 and 500–1000 for the number of neurons in the hidden layer and epoch number were examined, respectively, and the best network was obtained through the trial and error procedure. In the CART case, during the tree-building process, it was difficult to know when to stop the process as different parts of the tree may require markedly different depths (Lewis 2000). Moreover, without defining some stop criteria, the tree-building process is continued until a maximal tree was created which is generally very overfitted (Lewis 2000). Thus, a minimum number of samples at a leaf node (here is 1) and maximum depth of the tree (here is 6) were set to create the best tree via the trial-and-error procedure.

Table 5 compares the results for the integrated models via DC values in the verification steps. The integrated SVR model with DC of 87% signifies that in case the records of the suburban station were missed for any reason in the real time, using generated $C_{s(t)}$ can be reliable enough to be used in prediction of $C_{u(t)}$. In addition, when the SVR model was created and trained using historical data of $C_{s(t)}$, this model can also be used to generate future values of NO$_2$ concentration. That way, the integrated model is capable of applying generated future values of $C_{s(t)}$ to produce future values of $C_{u(t)}$. Thus, the advantage of the integrated model can be revealed when future values are required in the urban station.

Results presented in Table 5 indicate that among various predicting models, SVR and FFNN led to more accurate results than CART. The DC values in the verification step for the SVR, FFNN, and CART are 87, 81, and 67%, respectively. This lower accuracy of CART can be attributed to the linearity of the model and its shortcomings in modelling complex and nonlinear processes such as air pollution. In addition, Fig. 8 reveals each model's advantages and disadvantages. For instance, the FFNN model is not as accurate as the CART and SVR models in the upper values of the NO$_2$ time series (Fig. 8). On the other hand, FFNN and SVR

Table 6 Results of the ensemble models for the prediction of $C_{u(t)}$

Ensemble model	Model parameters[a]	DC		RMSE (ppb)	
		Training	Verification	Training	Verification
Simple linear averaging	–	0.861	0.864	3.54	4.48
Weighted linear averaging	0.36-0.33-0.30	0.873	0.860	3.49	4.55
SVRE	0.013-45-0.3	0.884	0.866	3.23	4.45

[a]Parameters a–b–c in the weighted linear averaging refer to the constants applied to the outputs of the SVR, FFNN and CART models. Parameters a–b–c in the SVRE technique denote ε, c, γ

could explain the process more accurately in the lower values of the NO$_2$ time series comparing to CART. Hence, by combining FFNN, SVR and CART, the performance of the model in upper and lower NO$_2$ values might be improved via the ensemble technique.

Results of the ensemble techniques

In the last step of modelling, three ensemble techniques were established to investigate the ability to fill gaps in the NO$_2$ time series from every single model. To accomplish this, three ensemblig techniques (SA, WA, SVRE) described in Sect. 2 were developed and applied for modelling. Table 6 indicates the results of ensemble techniques in the both calibration and verification steps.

The performance of the ensemble and integrated models can be evaluated by comparing the DC values (see Tables 5 and 6). The results indicate that all ensembling techniques may improve the individual model performance in both calibration and verification steps. In the calibration step, this improvement was up to 11% for the CART model; in the verification step, the ensemble techniques could enhance CART and FFNN predictions by up to 19 and 5%. As described previously, the major goal of ensembling technique is to combine outputs in order to capture patterns
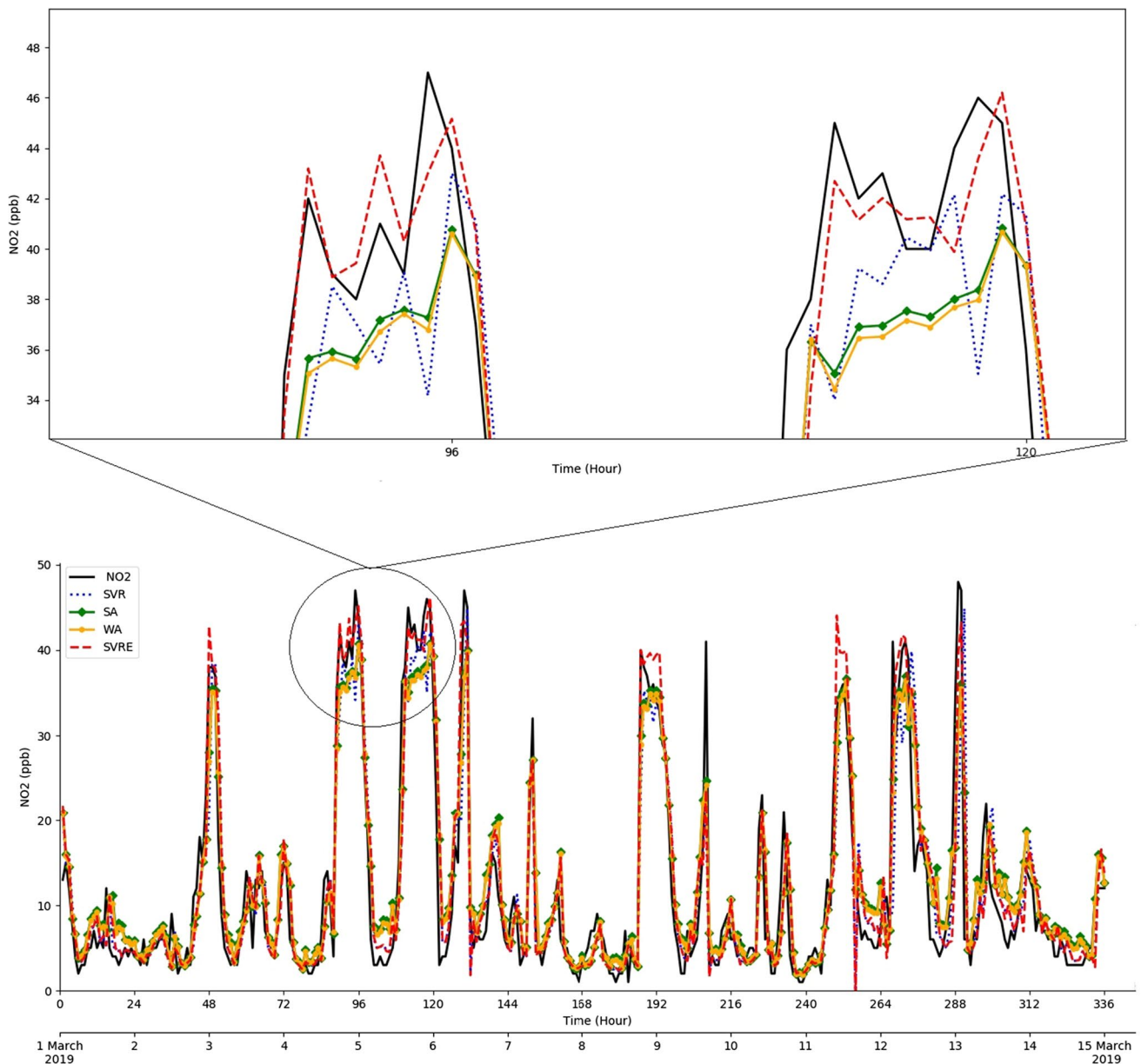
Fig. 9 Comparison of three ensembling techniques with **a** SVR, **b** FFNN, **c** CART
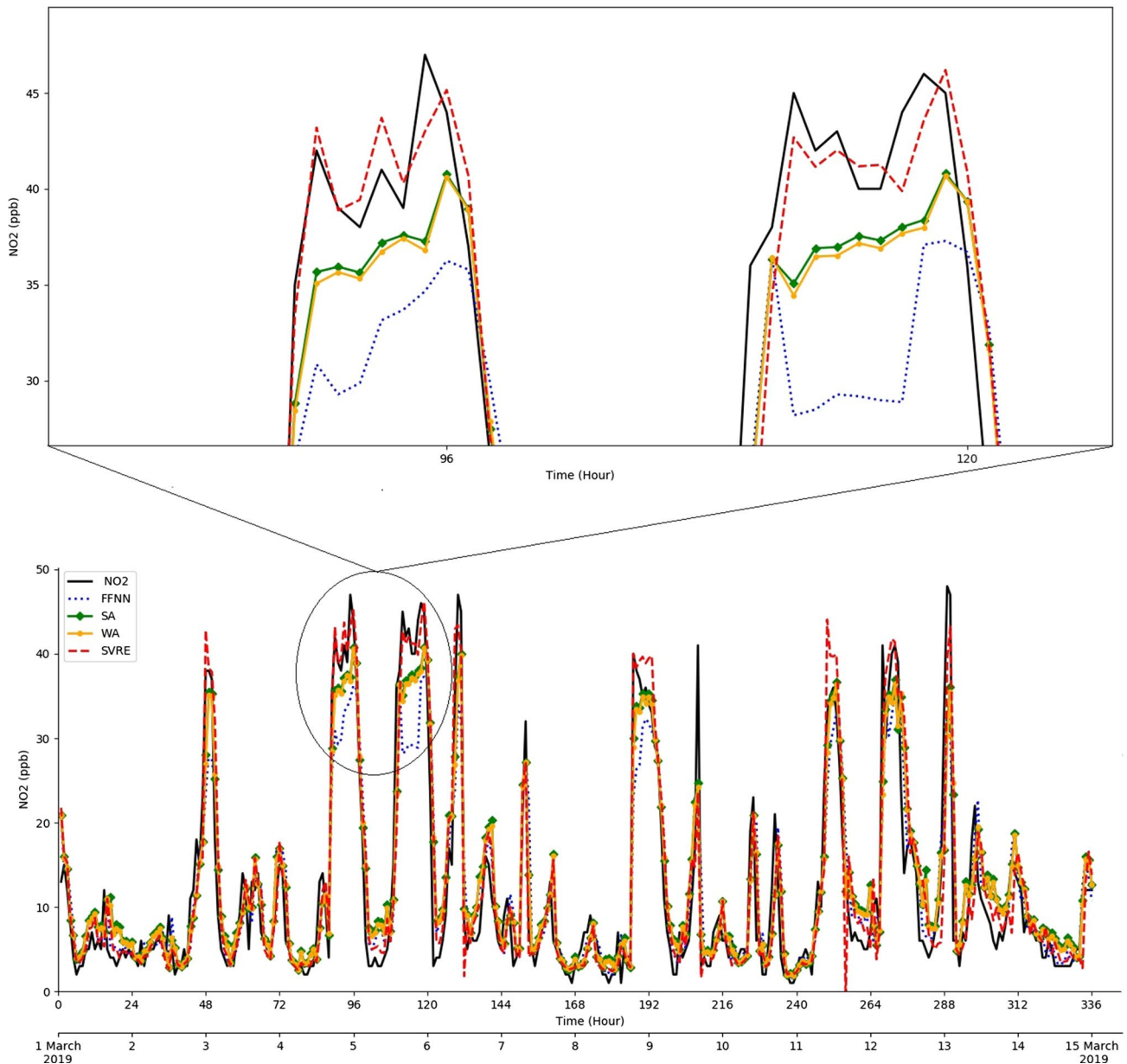
Fig. 9 (continued)

not capable for each single model; this approach can be revealed visually by comparing the time series for both integrated and ensemble results. Figure 9a demonstrates that applying the SVRE technique caused the SVR model to perform almost better in capturing the upper values. Figure 9b shows that SA, WA, and SVRE could also improve the FFNN performance in the upper values. These improvements for the SVR and FFNN models can be attributed to the CART superior performance in the upper values, fact that enhances both FFNN and SVR predictions via the ensemble techniques. On the other hand, in Fig. 9c it can be seen that the better performance of FFNN and SVR in the lower

parts has caused the CART model to overcome its shortcoming in the lower values.

Conclusion

This paper followed three main goals: firstly, single and class sensitivity analyses were performed based on SVR model in order to investigate variables and classes of data which could remarkably influence the NO_2 variations in the suburban and urban environments of Columbus City. Three scenarios based on different classes of data were created to investigate
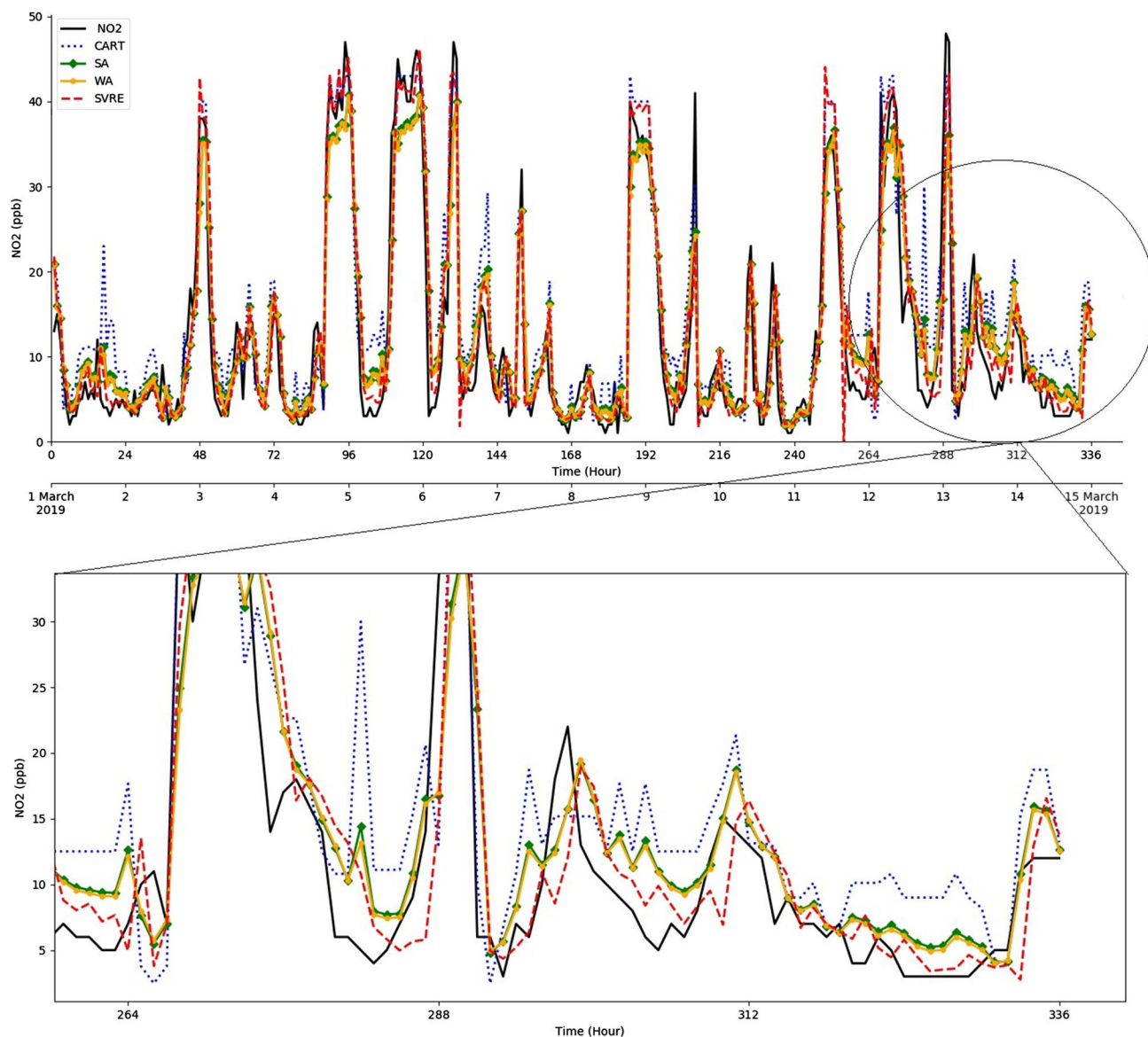
Fig. 9 (continued)

every variable's efficiency and dominant class of data. Secondly, the SVR model was used to predict $C_{s(t)}$, after which the predicted values were applied as one of the inputs for modelling $C_{u(t)}$. Three models (SVR, FFNN, and CART) were used for predicting the $C_{u(t)}$ values. Since generated values of $C_{s(t)}$ were considered as input the SVR, FFNN, and CART models, they were denoted as integrated models. In the last step, three ensemble techniques (SA, WA, and SVRE) were implemented to assess the ability of post-processing techniques in the improvement in the integrated models' performance (SVR, FFNN, and CART). The results of the sensitivity analysis showed that the combination of $C_{s(t-1)}$ TR, and H with DC value of 0.835 is the best choice for $C_{s(t)}$ prediction. In the urban station's modelling, it was

revealed that $C_{s(t)}$ is an important variable, and $C_{u(t)}$ was more sensitive to TR and H. Thus, four variables (TR, H, $C_{s(t)}$ and $C_{u(t-1)}$) were selected as efficient ones. In the second step, three integrated models (SVR, FFNN, and CART) with DC values of 81, 87, and 67% showed a better performance than the CART model. Although CART model, as a linear model, showed a relatively weaker performance, it was able to capture peak values of the NO_2 time series better than the FFNN model. On the other hand, the FFNN and SVR performances were superior to that of CART in order to capture the lower values of the NO_2 time series. Regarding the performances of the integrated models, the SVR model with a DC value of 87% in the verification step indicated the reliability of generated values of $C_{s(t)}$ for application as

an input to $C_{u(t)}$ prediction. In the third step, three ensemble techniques (SA, WA and SVRE) led to the improvement in the CART and FFNN models up to 19 and 5%, respectively.

For future works, it is suggested to apply traffic-related particulate matter (PM) as input to investigate its sensitivity to NO_2 variations. It is also recommended to use other machine learning models for sensitivity analysis and nonlinear ensemble techniques and results were compared with the current study. Plus, the unavailability of hourly traffic count throughout the year in streets around the monitoring stations was the major limitation of this study that is recommended for future work finding an alternative representing this traffic count.

References

Agirre-Basurko E, Ibarra-Berastegi G, Madariaga I (2006) Regression and multilayer perceptron-based models to forecast hourly O_3 and NO_2 levels in the Bilbao area. Environ Model Softw 21(4):430–446. https://doi.org/10.1016/j.envsoft.2004.07.008

Armaghani DJ, Asteris PG (2020) A comparative study of ANN and ANFIS models for the prediction of cement-based mortar materials compressive strength. Neural Comput Appl. https://doi.org/10.1007/s00521-020-05244-4

Alimissis A, Philippopoulos K, Tzanis CG, Deligiorgi D (2018) Spatial estimation of urban air pollution with the use of artificial neural network models. Atmos Environ 191:205–213. https://doi.org/10.1016/j.atmosenv.2018.07.058

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000) Artificial Neural Networks in Hydrology. 2: hydrology applications. J Hydrol Eng 5(2):124–137. https://doi.org/10.1061/(ASCE)1084-0699(2000)5:2(124)

Azid A, Juahir H, Toriman ME, Kamarudin MKA, Saudi ASM, Hasnam CNC, Aziz NAA, Azaman F, Latif MT, Zainuddin SFM, Osman MR, Yamin M (2014) Prediction of the level of air pollution using principal component analysis and artificial neural network techniques: a case study in Malaysia. Water Air Soil Pollut 225(8):1–14. https://doi.org/10.1007/s11270-014-2063-1

Bai Y, Li Y, Wang X, Xie J, Li C (2016) Air pollutants concentrations forecasting using back propagation neural network based on wavelet decomposition with meteorological conditions. Atmos Pollut Res 7(3):557–566. https://doi.org/10.1016/j.apr.2016.01.004

Bechle MJ, Millet DB, Marshall JD (2013) Remote sensing of exposure to NO2: satellite versus ground-based measurement in a large urban area. Atmos Environ 69:345–353. https://doi.org/10.1016/j.atmosenv.2012.11.046

Breiman L, Friedman JH, Olshen R, Stone CJ (1984) Classification and regression trees. Chapman and Hall, New York

Cabaneros SMS, Calautit JKS, Hughes BR (2017) Hybrid artificial neural network models for effective prediction and mitigation of urban roadside NO_2 pollution. Energy Procedia 142:3524–3530. https://doi.org/10.1016/j.egypro.2017.12.240

EPA (2018) Additional air quality designations for the 2015 Ozone National Ambient Air Quality Standards. Federal Regist 83(107):25776–25848. https://www.federalregister.gov/docum ents/2018/06/04/2018-11838/additional-air-quality-designatio ns-for-the-2015-ozone-national-ambient-air-quality-standards

Elangasinghe MA, Singhal N, Dirks KN, Salmond JA (2014) Development of an ANN–based air pollution forecasting system with explicit knowledge through sensitivity analysis. Atmos Pollut Res 5(4):696–708. https://doi.org/10.5094/APR.2014.079

Elkiran G, Nourani V, Abba SI (2019) Multi-step ahead modelling of river water quality parameters using ensemble artificial intelligence-based approach. J Hydrol 577:123962. https://doi.org/10.1016/j.jhydrol.2019.123962

Feng X, Li Q, Zhu Y, Hou J, Jin L, Wang J (2015) Artificial neural networks forecasting of $PM_{2.5}$ pollution using air mass trajectory based geographic model and wavelet transformation. Atmos Environ 107:118–128. https://doi.org/10.1016/j.atmosenv.2015.02.030

Gilbert NL, Goldberg MS, Beckerman B, Brook JR, Jerrett M (2005) Assessing spatial variability of ambient nitrogen dioxide in Montreal, Canada, with a land use regression model. J Air Waste Manag Assoc 55(8):1059–1063. https://doi.org/10.1080/10473 289.2005.10464708

Gilbert NL, Goldberg MS, Brook JR, Jerrett M (2007) The influence of highway traffic on ambient nitrogen dioxide concentrations beyond the immediate vicinity of highways. Atmos Environ 41(12):2670–2673. https://doi.org/10.1016/j.atmosenv.2006.12.007

Hamilton RS, Harrison RM (1991) Highway pollution, Studies in environmental science, 44th edn. Elsevier, Amsterdam

Haykin S (1994) Neural networks: a comprehensive foundation. Macmillan, New York

He HD, Lu WZ, Xue Y (2015) Prediction of particulate matters at urban intersection by using multilayer perceptron model based on principal components. Stoch Env Res Risk Assess 29(8):2107–2114. https://doi.org/10.1007/s00477-014-0989-x

Hsu C-W, Chang C-C, Lin C-J (2003) A practical guide to support vector classification. Technical report, Department of Computer Science and Information Engineering, University of National Taiwan. https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf

Jamal R, Manaa K, Rabee M, Khalaf L (2015) Traffic control by digital imaging cameras. In: Deligiannidis L, Arabnia HR (eds) Emerging trends in image processing, Computer vision and pattern recognition, Morgan Kaufmann, pp 231–247

Kambezidis HD, Melas LD, Kampezidou DH, Psiloglou BE (2015) Effect of tropospheric nitrogen dioxide on incoming solar radiation. J Solar Energy Res Updat 2:14–17 https://www.researchga te.net/profile/Harry_Kambezidis/publication/304251256_Effec t_of_Tropospheric_Nitrogen_Dioxide_on_Incoming_Solar_Radia tion/links/576a95e008aefcf135bd251d.pdf

Kamińska JA (2019) A random forest partition model for predicting NO_2 concentrations from traffic flow and meteorological conditions. Sci Total Environ 651(Part 1):475–483. https://doi.org/10.1016/j.scitotenv.2018.09.196

Leduc G (2008) Road traffic data: collection methods and applications. JRC technical notes, working papers on energy, transport and, climate change, N.1. ftp://ftp.jrc.es/pub/EURdoc/JRC47967.TN.pdf

Lewis RJ (2000) An introduction to classification and regression tree (CART) analysis. In: 2000 annual meeting of the society for academic emergency medicine, San Francisco, California. https://pdfs.semanticscholar.org/6d4a/347b99d056b7b1f28218728f1 b73e64cbbac.pdf

Li E, Zhou J, Shi X, Armaghani DJ, Yu Z, Chen X, Huang P (2020) Developing a hybrid model of salp swarm algorithm-based support vector machine to predict the strength of fiber-reinforced cemented paste backfill. Eng Comput. https://doi.org/10.1007/s00366-020-01014-x

Liang M, Mohamad ET, Faradonbeh RS, Armaghani DJ, Ghoraba S (2016) Rock strength assessment based on regression tree technique. Eng Comput 32:343–354. https://doi.org/10.1007/s00366-015-0429-7

Makridakis S, Winkler RL (1983) Average of forecasts: some empirical results. Manag Sci 29(9):987–996. https://doi.org/10.1287/mnsc.29.9.987

Mehdipour V, Memarianfard M (2019) Ground-level O3 sensitivity analysis using support vector machine with radial basis function. Int J Environ Sci Technol 16(6):2745–2754. https://doi.org/10.1007/s13762-018-1770-3

Mishra D, Goyal P (2015) Development of artificial intelligence based NO2 forecasting models at Taj Mahal, Agra. Atmos Pollut Res 6(1):99–106. https://doi.org/10.5094/APR.2015.012

Moazami S, Noori R, Amiri BJ, Yeganeh B, Partani S, Safavi S (2016) Reliable prediction of carbon monoxide using developed support vector machine. Atmos Pollut Res 7(3):412–418. https://doi.org/10.1016/j.apr.2015.10.022

Murillo-Escobar J, Sepulveda-Suescum JP, Correa MA, Orrego-Metaute D (2019) Forecasting concentration of air pollutants using support vector regression improved with particle swarm optimization: case study in Aburrá Valley, Colombia. Urban Clim 29:100473. https://doi.org/10.1016/j.uclim.2019.100473

Nourani V (2017) An emotional ANN (EANN) approach to modeling rainfall-runoff process. J Hydrol 544:267–277. https://doi.org/10.1016/j.jhydrol.2016.11.033

Nourani V, Alami MT, Vousoughi FD (2015) Wavelet-entropy data pre-processing approach for ANN-based groundwater level modeling. J Hydrol 524:255–269. https://doi.org/10.1016/j.jhydrol.2015.02.048

Nourani V, Elkiran G, Abdullahi J (2019) Multi -station artificial intelligence based ensemble modeling of reference evapotranspiration using pan evaporation measurements. J Hydrol 577:123958. https://doi.org/10.1016/j.jhydrol.2019.123958

Nourani V, Gökçekuş H, Umar IB (2020) Artificial intelligence based ensemble model for prediction of vehicular traffic noise. Environ Res 180:108852. https://doi.org/10.1016/j.envres.2019.108852

Osowski S, Garanty K (2007) Forecasting of the daily meteorological pollution using wavelets and support vector machine. Eng Appl Artif Intell 20(6):745–755. https://doi.org/10.1016/j.engappai.2006.10.008

Perez P, Gramsch E (2016) Forecasting hourly $PM_{2.5}$ in Santiago de Chile with emphasis on night episodes. Atmos Environ 124(Part A):22–27. https://doi.org/10.1016/j.atmosenv.2015.11.016

Radojević D, Antanasijević D, Perić-Grujić A, Ristić M, Pocajt V (2019) The significance of periodic parameters for ANN modeling of daily SO_2 and NO_x concentrations: a case study of Belgrade. Serbia Atmos Pollut Res 10(2):621–628. https://doi.org/10.1016/j.apr.2018.11.004

Shamseldin AY, O'Connor KM, Liang GC (1997) Methods for combining the outputs of different rainfall-runoff models. J Hydrol 197:203–229. https://doi.org/10.1016/S0022-1694(96)03259-3

Shang Z, Deng T, He J, Duan X (2019) A novel model for hourly $PM_{2.5}$ concentration prediction based on CART and EELM. Sci Total Environ 651(Part 2):3043–3052. https://doi.org/10.1016/j.scitotenv.2018.10.193

Sharghi E, Nourani V, Behfar N (2018) Earthfill dam seepage analysis using ensemble artificial intelligence-based modeling. J Hydroinform 20(5):1071–1084. https://doi.org/10.2166/hydro.2018.151

U.S. EPA (2016) Integrated science assessment for oxides of nitrogen. Health criteria. EPA/600/R-15/068. Research Triangle Park. https://ofmpub.epa.gov/eims/eimscomm.getfile?p_download_id=526855

Vapnik V (1995) The nature of statistical learning theory. Springer, Berlin

Wang W-C, Xu D-M, Chau K-W, Chen S (2013) Improved annual rainfall-runoff forecasting using PSO-SVM model based on EEMD. J Hydroinform 15(4):1377–1390. https://doi.org/10.2166/hydro.2013.134

Yeganeh B, Hewson MG, Clifford S, Tavasoli A, Knibbs LD, Morawska L (2018) Estimating the spatiotemporal variation of NO_2 concentration using an adaptive neuro-fuzzy interface system. Environ Model Softw 100:222–235. https://doi.org/10.1016/j.envsoft.2017.11.031

Zhang GP (2003) Time series forecasting using a hybrid ARIMA and neural network model model. Neurocomputing 50:159–175. https://doi.org/10.1016/S0925-2312(01)00702-0

Zhou J, Qiu Y, Zhu S, Armaghani DJ, Khandelwal M, Mohamad ET (2020) Estimating TBM advance rate in hard rock condition using XGBoost and Bayesian optimization. Undergr Space. https://doi.org/10.1016/j.undsp.2020.05.008