International Journal of Pioneering Technology and Engineering

Research Article

# Estimation of $PM_{10}$ and SO2 Substances with Extreme Learning Methods for Analysis of Air Pollution in Amasya Province

**Salih Berkan Aydemir**[a*]

[a] Department of Computer Engineering, Engineering Faculty, Amasya University, Amasya,Türkiye

Article Info

***Article history***
**Received**: 24.09.2022
**Revised**: 15.10.2022
**Accepted**: 13.11.2022

***Keywords:***
*Extreme Learning,*
*Extreme Learning with Adaptive neuro-fuzzy inference System*
*Prediction,*
*Air Pollution*
.

**ABSTRACT**

Air pollution can lead to various diseases. Especially, Sulfur dioxide ($SO_2$) and Particulate Matter ($PM_{10}$) concentrations in the air directly affect air pollution. Therefore, the estimation of $PM_{10}$ and $SO_2$ substances in the air plays an important role in air pollution prediction. In this study, for the data received from the Ministry of Environment and Urbanization air monitoring center, the $PM_{10}$ and $SO_2$ pollutants are estimated based on past data with the extreme-learning-adaptive fuzzy extraction system method. Excessive learning algorithms in prediction problems both work fast and can reach low error values. On the other hand, the adaptive neural-fuzzy inference system offers an effective convergence by handling uncertain and incomplete data. Within the scope of the study, estimated and actual values are visualized with figures. The results of extreme learning algorithms are graphed with the error values they reach. Extreme learning-adaptive neural-fuzzy inference system; It is compared with ELM, KELM, CSELM, CDELM, and SELM algorithms. As a result of the comparison, KELM and ELM-ANFIS in terms of RMSE metric for $PM_{10}$ give the best results close to each other with 0.1151 and 0.1155 values, respectively. For $SO_2$, it has been determined that the ELM-ANFIS showed the best performance with 0.0842 RMSE and 0.8171 $R^2$ metric values.

## 1. Introduction

The high emission values of $SO_2$ (Sulfur dioxide), $NO_X$ (Nitrogen oxide), $NMVOC$ (non-methane volatile organic compounds), $NH_3$ (Ammonia), $CO$ (Cobalt), and $PM_{10}$ (Particulate Matter) on earth are important factors affecting air pollution. In particular, the ratio of $SO_2$and PM in the air adversely affects human health. Amasya is one of the provinces with the highest annual $SO_2$ and $PM_{10}$ averages between 2013 and 2017 [1]. Therefore, estimating harmful substances in Amasya is important for human health. $PM_{10}$ and other particles smaller than 10 micrometers in diameter reach the lungs and cause inflammation. In addition, $PM_{10}$ can cause heart and lung conditions that negatively affect humans. $SO_2$ creates a direct toxic effect. It adversely affects essential respiratory functions. $SO_2$ can indirectly threaten human health if it turns into Sulfuric acid and sulfate form. The main source of $SO_2$ is obtained by burning high sulfur oils, coal, and lignite. It is also produced by melting bronze and bronze with high sulfur content. In addition, the fast-learning structure provided by excessive learning and the success of reaching the smallest error produces better results than other ELM.

In this study, the ELM-ANFIS structure is compared with some other ELM. Within the scope of this study, $SO_2$ and $PM_{10}$ pollutants that adversely affect human health are estimated by ELM-ANFIS. The ELM-ANFIS structure combines the rule mechanism of fuzzy logic with the decision-making mechanism of ELM [2]. In addition, the fast-learning structure provided by excessive learning and the success of reaching the smallest error produces better results than other ELM. In this study, the ELM-ANFIS structure is compared with some other ELMs.

It is calculated by considering the different membership function numbers, the square root of the mean of the squares of the errors (RMSE), and the coefficient of determination (R2). Finally, the RMSE and R2 values of $SO_2$ and $PM_{10}$ pollutants that may be harmful to human health are calculated with extreme learning algorithms. In the estimations made using historical data, the ELM-ANFIS structure reaches lower error values than the other ELM. Different machine-learning mechanisms can be used when estimating $SO_2$ and $PM_{10}$ pollutants. This section is especially included in the literature review section. The adverse effects of air pollutants on human health have increased researchers' predictions of some air pollutants. Particularly, $SO_2$ and $PM_{10}$ pollutants, which affect air pollution, are substances estimated in many studies. In the literature review, predictions of some air pollutants were made with various learning methods. However, although it is known that the province of Amasya has high pollutant rates in terms of air pollution in our country [3], air pollutant concentration estimations have not been made for the province of Amasya using ELM-ANFIS methods. ELM-ANFIS works more stable than other methods thanks to the decision-making mechanism of fuzzy logic, and faster results can be produced with the ELM method. Air pollutant concentration estimations were not made for Amasya province with ELM-ANFIS methods.

***Corresponding author:** Salih Berkan Aydemir
*E-mail address: salih.aydemir@amasya.edu.tr
https://doi.org/10.56158/jpte.2022.33.1.02

Kaplan et al. In this study, $SO_2$ and $PM_{10}$ estimations were made for Kütahya province by using artificial neural networks and Levenberg – Marquardt optimization method [4]. Akbulut and Özcan, on the other hand, selected 31 provinces in their study and made their estimations using artificial neural networks and regression Methods [5].

ELM-ANFIS works more stable than other methods thanks to the decision-making mechanism of fuzzy logic, and faster results can be produced with the ELM method. In the second part of the study, other studies on $SO_2$ and $PM_{10}$ estimation are given. The article's extreme learning structure is mentioned in the third part. In the fourth chapter, the ELM-ANFIS structure is explained in detail. In the fifth chapter, the organization of $SO_2$ and $PM_{10}$ data and the experimental methodology are given. In the sixth section, experimental results and comparisons are detailed. In the last part of the study, the results and estimation comments are explained.

## 2. Literature Review

Estimation of the emission values of dangerous gases in the air can be seen as a frequently studied subject by scientists. Therefore, with machines measuring air quality, this situation was evaluated with principal component and cluster analysis [6]. Based on machine learning techniques and many methods in the literature, $SO_2$ and $PM_{10}$ pollutants have been estimated. Two days' prediction of maximum daily concentrations of $SO_2$, $O_3$, $PM_{10}$, $NO_2$, $CO$ in Palermo, Italy, tested with basic neural network Methods [7]. In another study, a multiple linear regression model with principal component analysis was used for $PM_{10}$ concentration estimation in Penang state in Malaysia [8]. Concentration estimates of air pollutants have been made for many countries in the literature [4, 5, 9-11]. Estimation of $SO_2$ and $PM_{10}$ concentrations has been made for many cities in our country.

Kaplan et al. In this study, $SO_2$ and $PM_{10}$ estimations were made for Kütahya province by using artificial neural networks and Levenberg – Marquardt optimization method [4]. Akbulut and Özcan, on the other hand, selected 31 provinces in their study and made their estimations using artificial neural networks and regression Methods [5]. In another study, Turgut and Temiz estimated the proportion of $PM_{10}$ air pollutants for the province of Ankara using the Box-Jenkins approach. In the study of Güngör and Sevindir, multiple linear regression method $SO_2$ and $PM_{10}$ were estimated for Isparta province. Forecasts are made for the winter period of 2011-2012 [9]. In another study, air pollution modeling in Konya city center was made using fuzzy logic and neural networks. Measurements were made in four different periods by selecting 15 schools in the city center of Konya [10]. Considering all studies, $SO_2$ and $PM_{10}$ substances were estimated to measure the effect of air pollution with extreme learning algorithms. On the other hand, in another study conducted by Ö.Çekim in 2020, it can be seen that Amasya is one of the provinces most affected by air pollution [3]. Therefore, in this study, $SO_2$ and $PM_{10}$ estimations were made for Amasya province with extreme learning algorithms and ELM structure.

Table 1 shows estimation studies for some air pollutants in recent years. Deep learning methods, artificial neural networks and structures combining these methods with optimization algorithms were used in air pollutant forecasting studies in China, Iran, Brazil, and Greece. Tzanis et al. made $PM_{2.5}$ and $PM_{10}$ predictions using daily data from air quality monitors in Attica, Greece, using multiple linear regression, linear mixed model, and forward neural networks [11]. Air pollutant rates were estimated with the test data of 2016. Ventura et al. estimated daily concentrations of $PM_{2.5}$ in rural, urban and industrial areas [12]. In another study, Liu et al. estimated $PM_{2.5}$, $SO_2$, , $NO_2$ and $CO$ air pollutants in Beijing, China with NARX neural networks using wavelet transforms and optimization algorithm [13]. Long-term memory models are also frequently used to predict air pollutants. The most important feature of long-term memory models is evaluating past and current data while predicting future data. In the literature, long-short-term memory networks have been used together with methods such as graph conventional networks [14], convolutional neural networks [15], and particle swarm optimization [16].

**Table 1.** In recent years, literature studies on the prediction of air pollutants

| Author (Year) | Method | Location | Predicted aior pollutants |
|---|---|---|---|
| Tzanis et al. (2019) [11] | Forward neural networks with multiple linear regression and linear mixed model | Attica, Greece | $PM_{2.5}$ and $PM_{10}$ |
| Ventura et al. (2019) [12] | Neural networks and Holt–Winters models | Rio de Janeiro, Brazil | $PM_{2.5}$ |
| Liu et al. (2019) [13] | NARX neural networks with empirical wavelet transform and multi-agent evolutionary genetic algorithm | Beijing, China | $PM_{2.5}$, $SO_2$ , $NO_2$ and $CO$ |
| Qi et al. (2019) [14] | Graph conventional networks and long Short-Term Memory networks | Jing-Jin-Ji, China | $PM_{2.5}$ |
| Qin et al. (2019) [15] | Convolutional neural networks and long Short-Term Memory networks | Shanghai, China | $PM_{2.5}$ |
| Zhu et al. (2019) [17] | Support vector regression model with cuckoo and gray wolf hybrid optimization algorithms | Henan, Hubei and Hunan, China | $SO_2$, $NO_2$ |
| Xu et al. (2020) [18] | deep echo state network | Xiamen, Shenzhen, Kunming, and Lasa, China | $AQI$, $PM_{2.5}$ $PM_{10}$, $SO_2$, $NO_2$, $O_3$ and $CO$ |
| Shams et al. (2021) [19] | Artificial neural networks and multiple linear regression | Tahran, Iranian | $SO_2$ |
| Al-Janabi et al. (2021) [16] | Long Short-Term Memory networks enhanced with particle swarm optimization | Pekin, China | $PM_{2.5}$, $PM_{10}$, $SO_2$, $NO_2$, $O_3$ and $CO$ |
| Mao et al. (2021) [20] | Temporal shift long short-term memory extended model | Jing-Jin-Ji, China | $PM_{2.5}$ |

## 3. Material and Method

All methods used within the scope of the study are structures based on extreme learning. In this study, six different ELM structures were compared. These are the algorithms of ELM, KELM, CSELM, CDELM, SELM and ELM-ANFIS. Since the input parameters are randomly assigned and not subject to any learning algorithm, the results vary depending on the randomness. Different kernel types can be used in the hidden layer to speed up learning and reduce standard deviation. However, KELM structures are not sufficient because they consist of a single hidden layer. The input parameters of CSELM, CDELM, and SELM are created with certain methods from different areas of the solution space, and the learning process of the ELM structure is started. However, these methods are methods that do not use fuzzy logic and are not rule-based. Some of the superior features of extreme learning algorithms compared to neural network structures and regression analysis methods are mentioned below [26].

- Parameters such as learning rate and epoch number do not need to be set manually.
- It can learn quickly.
- Although the input parameters are chosen randomly, only the parameters between the hidden layer and the output are updated because the output parameters are calculated analytically.

ELM-ANFIS provides the expression of extreme learning methods with the help of membership functions. Thus, the real values (crisp values) taken from the outside world are blurred, and fuzzy inference system steps are applied. Some of the advantages of the ELM-ANFIS structure are mentioned below [2].

- It involves the human-like reasoning of fuzzy systems using a language model consisting of fuzzy sets and a set of IF-THEN fuzzy rules.
- Extreme learning can learn quickly.
- Only the coefficients of the zero-order or first-order equation in the output layer are updated.

In this study, the estimation of $PM_{10}$ and $SO_2$ air pollutants are made by ELM methods. Figure 1 shows the general flow of the methods used in the study.

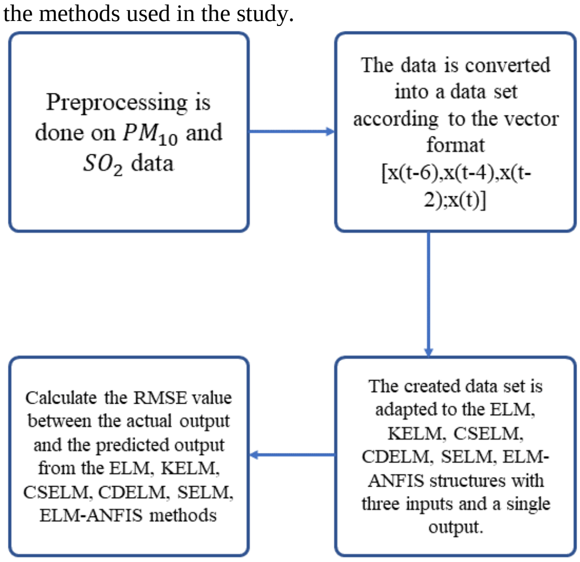


**Figure 1.** General flow of the estimation process

## 4. Extreme Learning Approaches

Extreme learning algorithms are powerful machine learning techniques that work fast and reach the smallest error values. Suggested by Huang et al. in 2006 [21]. Artificial neural network methods work slower since they have feedback structures. On the other hand, both derivative-based and swarm-based optimization algorithms slow down the training of the artificial neural network. However, in overlearning algorithms, the weights between the input and the hidden layer are randomly assigned, and the weights between the hidden layer and the output are determined analytically by the least squares estimator. In this case, the equations obtained from the network output can reach the smallest error with the RMSE. The point to be considered, especially in this part is the number of neurons in the hidden layer. If the number of neurons in the hidden layer is high, overfitting occurs. This indicates that the learned data is now learned by rote and the model is not learning. On the other hand, if the number of neurons in the hidden layer is low, the model learns less than necessary and can reach low error. The structure of a neural network is given in Figure 2.

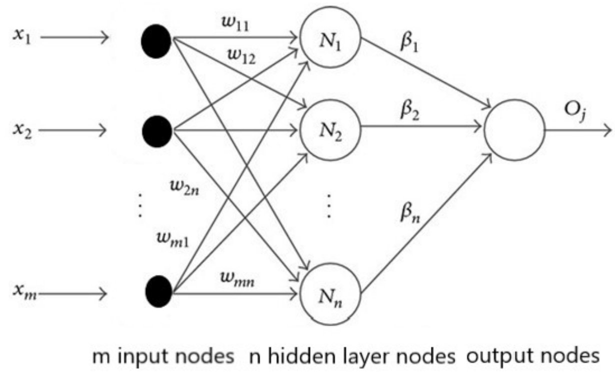


**Figure 2.** Single hidden layer feed forward network structure

To express the extreme learning structure mathematically, suppose our problem consists of N $(x_i, t_i)$ different samples for an arbitrary number N

Here $x_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in \mathbb{R}^n$ and $t_i = [t_{i1}, t_{i2}, \dots, t_{im}]^T \in \mathbb{R}^m$. A single hidden layer extreme learning network defined by $\tilde{N}$ The number of neurons in the hidden layer and the g(x) activation function can be modeled as follows.

$$\sum_{i=1}^{\tilde{N}} \beta_i g(\mathrm{x}_j) = \sum_{i=1}^{\tilde{N}} \beta_i g(w_i \mathrm{x}_j + b_i) = o_j, j = 1,2, \dots N$$

$w_i = [\mathrm{w_{i1}}, \mathrm{w_{i2}}, \dots, \mathrm{w_{in}}]^{\mathrm{T}}$ is the weight vector between the input and the hidden layer. $\beta_i = [\beta_{\mathrm{i1}}, \beta_{\mathrm{i2}}, \dots, \beta_{\mathrm{im}}]^{\mathrm{T}}$ is the weight vector between the hidden layer and the output neurons. $b_i$, i. is the threshold value of the hidden node. On the other hand, if the single-layer network with N samples and $\tilde{\mathrm{N}}$ hidden layer neurons and activation function g(x) reaches zero, it means that it is $\sum_{i=1}^{\tilde{N}} || \mathrm{o_j} - \mathrm{t_j} || = 0$ [21].

$$\sum_{i=1}^{\tilde{N}} \beta_i g(w_i \mathrm{x}_j + b_i) = t_j, j = 1,2, \dots N$$

The mathematical model given in Equation-2 is a linear equation system consisting of N equations. The compact form of Equation-2 is given in Equation-3.

$$H\beta = T$$

The H matrix included in Equation-3 is clearly given in Equation-4.

$$H(w_1, w_2, \dots, w_{\tilde{N}}, b_1, b_2, \dots b_{\tilde{N}}, \mathrm{x}_1, \mathrm{x}_2, \dots, \mathrm{x}_{\tilde{N}}) = \begin{bmatrix} g(w_1 \mathrm{x}_1 + b_1) & \cdots & g(w_{\tilde{N}} \mathrm{x}_N + b_{\tilde{N}}) \\ \vdots & \ddots & \vdots \\ g(w_1 \mathrm{x}_N + b_1) & \cdots & g(w_{\tilde{N}} \mathrm{x}_N + b_{\tilde{N}}) \end{bmatrix} \quad (4)$$

The H matrix is sized as the NxÑ matrix. $\beta$ and T vectors are defined as follows.

$$\beta = \begin{bmatrix} \beta_1^T \\ \beta_2^T \\ \vdots \\ \beta_{\tilde{N}}^T \end{bmatrix}_{\tilde{N}\times m} \quad \text{ve } T = \begin{bmatrix} t_1^T \\ t_2^T \\ \vdots \\ t_{\tilde{N}}^T \end{bmatrix}_{N\times m}$$

The equations mentioned so far are basically like a single hidden layer forward propagation neural network. However, solving the β vector in Equation-3 forms the basis of the extreme learning structure.

The solution of Equation 3 is found by the generalized Moore–Penrose inverse of the matrix [22]. Therefore, leaving β alone in Equation-3 leads to Equation-6.

$$\beta = H^{\dagger}T$$

Here, $H^{\dagger}$ is called the generalized Moore-Penrose inverse of the H matrix. It is also expressed as $H^{\dagger} = (H^T H)^{-1}H^T$. The biggest advantage of the extreme learning structure is that the output parameters are calculated analytically while the input parameters are randomly determined.

## 5. Extreme Learning Approach with Adaptive Neural Fuzzy Inference System

ELM-ANFIS has a successful prediction capability thanks to both the rule structure of fuzzy logic and the convergence feature of neural networks [2]. It can also be seen that the ELM-ANFIS structure gives a lower error value than the neuro-fuzzy inference (ANFIS) system [23]. An example ELM- ANFIS structure with two rules with Sugeno type rules is given in Figure 3. Below is a step-by-step description of the mechanism of the ELM- ANFIS structure.

Fuzzificaiton Layer: A system consisting of two rules is created with membership functions.

IF x, $A_1$ and y, $B_1$ Then $f_1 = p_1x + q_1y + r_1$

IF x, $A_2$ and y, $B_2$ Then $f_2 = p_2x + q_2y + r_2$

Here x and y are data from the outside world. $A_i$ and $B_i$ are linguistic variables. Fuzzy membership functions are also defined in this layer. Fuzzy membership functions are also defined in this layer.

With μ(x) defined as the membership function of x, the membership functions $\mu_{A_i}(x)$ (i=1,2) and $\mu_{B_{i-2}}(x)$ (i=3,4) in the two rule system It is expressed as ,4). In this study, the bell shape membership function is frequently used with the μ(x) membership function. $a_i$, $b_i$ and $c_i$ are expressed as antecedent parameters.

$$\mu(x) = \frac{1}{1 + \left|\frac{x-c_i}{a_i}\right|^{2b_i}} \qquad (7)$$

Multiplicative Layer: In this layer, the product T-norm is generally used. Membership degrees from the previous tier are multiplied by each other based on rules. The values obtained in this layer are also called firing strengths.

$$w_i = \mu_{A_i}(x)\,\mu_{B_i}(y) \qquad (8)$$

Normalization layer: The normalization process is applied to the data received from the second layer.

$$\overline{w_i} = \frac{w_i}{w_1 + w_2} \qquad (9)$$

Defuzzification and output layer: There are output weights in the sharpening layer. The weight parameters . $p_i, q_i$ and $r_i$ are obtained by the least squares estimation method. $p_i, q_i$ and $r_i$ are expressed as output parameters.

$$\overline{w_i}f_i = \overline{w_i}(p_ix + q_iy + r_i) \qquad (10)$$

Finally output, It is expressed as,

$$t = \sum_i \overline{w_i}f_i = \overline{w_1}(p_1x + q_1y + r_1) + \overline{w_2}(p_2x + q_2y + r_2)$$

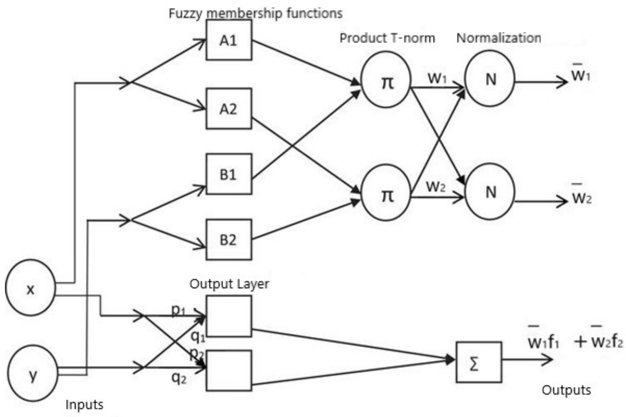


**Figure 3.** ELM-ANFIS structure

Matrix representation of a ELM-ANFIS structure consisting of two rules with antecedent weights $a_i, b_i$ and $c_i$ and output weights $p_i, q_i$ and $r_i$ is denoted by

$$\begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_N \end{pmatrix}_{N\times 1} = \begin{pmatrix} \overline{w}_1x & \overline{w}_1y & \overline{w}_1 & \overline{w}_2x & \overline{w}_2y & \overline{w}_2 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \end{pmatrix}_{N\times 6} \begin{pmatrix} p_1 \\ q_1 \\ r_1 \\ p_2 \\ q_2 \\ r_2 \end{pmatrix}_{6\times 1} \qquad (12)$$

When all parameters are taken into account, the compact form of N linear equations consisting of N data pairs is expressed as,

$$T_{N\times 1} = H_{N\times m^n(n+1)}\beta_{m^n(n+1)\times 1} \qquad (13)$$

Here, m represents the number of membership functions and n represents the size of the inputs. $m^n$ represents the maximum number of rules.

## 6. Experimental Methodology

In this part of the study, the experimental methodology is given. Daily $SO_2$ and $PM_{10}$ concentration values for Amasya province were taken from the website of the Ministry of Environment and Urbanization [1]. In order to reduce the effect of fluctuations on the output of the data, the data set was normalized between 0.1 and 0.9 as in other studies [4, 24].

While making the estimation process, the current data value was tried to be estimated by using the past data. However, since the use of sequential data while using the past data may lead to overfitting, the data set two, four and six days ago and the data from one, three and five days ago were used and the forecast for that day was made. Statistical analysis of the data is given in Table-2.

**Table 2.** Statistical Analysis of Data

| | $PM_{10}$ | $SO_2$ |
|---|---|---|
| Minimum | 2.76 | 0.18 |
| Maximum | 198.54 | 96.22 |
| Mean | 31.61 | 9.38 |
| Standard deviation | 21.04 | 8.53 |

Vector format of the data, t. represented as $[x(t-6), x(t-4), x(t-2); x(t)]$ for the day and x being the data set used.

The data were selected for 2162 days between 22.01.2015 - 21.12.2020. However, since this date range covers a long time, there are missing data due to the failure of the devices to measure the air quality in this interval or the inability to measure due to technical reasons. 95.51% of $SO_2$ data and 94.63% of $PM_{10}$ data are filled. The missing data were filled with the average of other existing data not to affect the estimation process positively or negatively.

When the data of 2162 days is converted to matrix format, a data set of size (694,44) is obtained. The first three columns represent the entries, respectively, and are denoted as$[x(t-6), x(t-4), x(t-2)]$. As the output, the output t th day x(t) is estimated. Of the 694 data, 500 were used for training and 194 for testing of the 694 data, 500 were used for training and the rest for testing. Sample data Some of the $PM_{10}$ data are given in the matrices below. Vector A is taken as a time series flat and resized according to the vector $[x(t-6), x(t-4), x(t-2); x(t)]$. A part of the matrices obtained after dimensioning the vector A is the matrix B. vector $[x(t-6), x(t-4), x(t-2); x(t)]$ in matrix B $[60.66, 75.69, 45.12;\ 39.46]$. This process is applied for 2162 days.

$$A = \begin{pmatrix} 60.66 \\ 74.84 \\ 75.69 \\ 59.60 \\ 45.12 \\ 61.65 \\ 39.46 \\ 32.51 \\ 85.02 \\ 25.60 \end{pmatrix},$$

$$B = \begin{pmatrix} 60.66 & 74.84 & 75.69 & 59.60 \\ 75.69 & 59.60 & 45.12 & 61.65 \\ 45.12 & 61.65 & 39.46 & 32.51 \\ 39.46 & 32.51 & 85.02 & 25.60 \end{pmatrix}$$

Error values, root-square error and coefficient of determination are calculated by Equations-14.1 and 14.2.

$$\text{Root mean square error (RMSE)}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(o_i-t_i)^2} \quad (14.1)$$

$$\text{Coefficient of Determination } (R^2)= 1-\frac{\sum(t_i-o_i)^2}{\sum(t_i-\bar{t})^2} \quad (14.2)$$

$o_i$, is the estimated values, , $t_i$ are the actual values, and $\bar{t}$ is the mean of the data.

## 7. Experimental Results

In this part of the study, the estimation process is carried out using extreme learning algorithms and the ELM-ANFIS method. The number of neurons in the hidden layer is systematically determined as the number of neurons that gives the best value out of 30 runs. Similarly, the results of all constructs are based on the smallest error value obtained after running 30 times. On the other hand, the ELM-ANFIS structure is also run 30 times. The number of neurons in the hidden layer used in each structure was determined as a result of experiments. Algorithms were coded on the R2020b MatLab program and with a computer with i7-4700MQ CPU, 2.40GHz 8.00 GB RAM, 64 bit operating system. As a result of the experimental studies, the error values with the over-learning algorithms and the ELM-ANFIS method are given in Table-3. Considering Table-3, the ELM-ANFIS structure produces the best results. On the other hand, the low number of neurons and having lower error values or coefficient of determination than other methods show that the ELM-ANFIS structure is more effective in regression problems. The kernel parameter of the radial basis function (RBF) is also taken as 20 (Best fit as a result of experiments). In addition, the sigmoid activation function was used in ELM, KELM, CDELM, CSELM, SELM algorithms as in other studies [25, 26]. (14.1)

**Table 3.** $PM_{10}$ and $SO_2$ test prediction result

| Pollutants | $PM_{10}$ | | | $SO_2$ | | |
|---|---|---|---|---|---|---|
| Extreme learning approaches / Error values | RMSE | $R^2$ | Neuron size | RMSE | $R^2$ | Neuron size |
| Kernel Extreme Learning [27] | 0.1206 | 0.7843 | - | 0.0882 | 0.7661 | - |
| Kernel Extreme Learning (Linear kernel) [27] | 0.1255 | 0.8024 | - | 0.0926 | 0.7024 | - |
| Extreme Learning machine (ELM)[21] | 0.1248 | 0.7775 | 20 | 0.0885 | 0.7793 | 20 |
| Constrained Sum Extreme Learning (CSELM)[25] | 0.1211 | 0.7800 | 20 | 0.0871 | 0.7867 | 20 |
| Constrained Difference Extreme Learning (CDELM)[25] | **0.1151** | 0.8157 | 20 | 0.0852 | 0.7962 | 20 |
| Sample Extreme Learning (SELM) [25] | 0.1205 | 0.7839 | 20 | 0.0874 | 0.7848 | 20 |
| Adaptive neural fuzzy inference system with extreme learning (ELM-ANFIS) [2] | **0.1155** | **0.8578** | - | **0.0842** | **0.8171** | - |

**Table 4.** Output parameters in $PM_{10}$ and $SO_2$ with ELM-ANFIS

| | | | | |
|---|---|---|---|---|
| $PM_{10}$ output parameters | -2015.7894 | 203.0948 | -726.4615 | 411.1777 |
| | 1965.1902 | -339.4260 | 700.7440 | -360.0566 |
| | 2144.1616 | -316.0520 | 610.8624 | -374.0642 |
| | -2093.3593 | 444.9455 | -589.1882 | 326.3912 |
| | 1493.31857 | 514.8872 | 1148.356 | -566.9978 |
| | -1462.5331 | -322.4299 | -1090.149 | 500.5160 |
| | -1659.0905 | -346.5404 | -1012.866 | 521.4809 |
| | 1629.2873 | 163.8693 | 962.5334 | -458.3368 |
| $SO_2$ output parameters | -4429.5303 | -5479.9475 | 6517.6665 | 3136.7492 |
| | 7079.0702 | 8163.4660 | -9734.6099 | -4759.0017 |
| | 3081.7959 | 5472.4504 | -6273.0958 | -3941.5605 |
| | -5027.4574 | -8195.0210 | 9395.1755 | 5997.1651 |
| | 4788.9513 | 6287.8344 | -7395.6186 | -3374.7780 |
| | -7612.1112 | -9322.7695 | 11099.042 | 5087.3773 |
| | -3592.1471 | -6412.3219 | 7098.5950 | 4538.5011 |
| | 5794.2282 | 9572.0705 | -10674.612 | -6877.7987 |

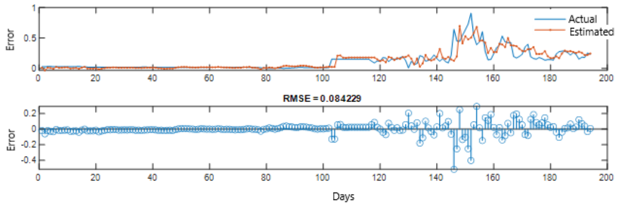


**Figure 4.** $SO_2$ data estimation

## 8. Conclusion

While estimating daily life data such as air pollutants forecast, weather forecast, wind speed forecast, it is important to obtain data and produce results quickly. Because data changes over time. These forecasting operations can be done daily as well as hourly. Using ELM on time-varying data is more advantageous than other machine learning techniques, since ELM produces fast results thanks to its analytical structure. On the other hand, the ELM-ANFIS mechanism includes the human-like reasoning style of fuzzy systems by using a language model with the rules it creates in the data while producing the prediction.

One of the reasons why the first 100 data is concentrated around zero in the estimation for $SO_2$ pollutant is due to the normalization process. However, considering Figure 4, it still becomes difficult to learn the extreme values, and while a steady state appears in the first 100 data of the test, there is a large change in the remaining 94 data. The biggest reason for this is the effect of the COVID 19 pandemic. At first, with the coming of the bans, the closure of the workplaces and factories and the coming of the restrictions, then the short-term removal of the restrictions and the return to the new normal, fluctuations started. 180. After the data, it can be seen that the $SO_2$ pollutant data is again concentrated around zero. This is due to the fact that some bans and restrictions have started again due to the increase in cases in recent days. On the other hand, estimated and actual $PM_{10}$ data are also given in Figure 5.

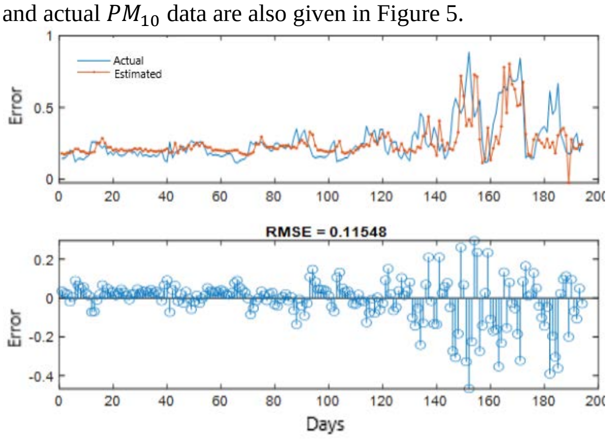


**Figure 5.** $PM_{10}$ data estimation

Similarly, $PM_{10}$ data estimation is made in Figure 5. Likewise, the last 194 days have been selected for the test data. Considering Figure 4 and 5, there are fluctuations between days 140 and 180. This situation shows the ability of the ELM-ANFIS structure to predict the extreme points that the concentration of $PM_{10}$ and $SO_2$ in the air of Amasya province increased during these days.

In the future, air pollutants such as $PM_{2.5}$, $NO_2$, $O_3$ and $CO$ can be estimated with the same methods. Artificial neural networks and differentiation in differentiation methods can be used for the prediction of air pollutants [28]. In addition, by estimating $PM_{2.5}$, $PM_{10}$, $SO_2$, $NO_2$, $O_3$ and $CO$ pollutants with long short-term memory networks, future predictions can be made in Amasya province and some measures can be taken for air pollution. Although ELMs are stable machines, it is very difficult to adjust the input weights. However, various optimization-based ELM structures proposed in the literature can be used for the estimation of air pollutants [29-33]. Thus, the input weights can be learned with optimization methods and the output can be calculated with analytical processes.

**Declaration of conflicting interests**

The authors declared no conflicts of interest with respect to concerning the authorship and/or publication of this article.

**Funding**

The authors received no financial support for the research and/or authorship of this article.

## References

[1]. Web Page: *Hava Kalitesinde PM10 ve SO2 Ortalamaları*, https://cevreselgostergeler.csb.gov.tr/hava-kalitesinde-pm10-ve-so2-ortalamalari-i-85734. Last access date: 23.12.

[2]. Pillai, G., Pushpak, J., and Nisha, M.G. *Extreme learning ANFIS for control applications*. 2014 IEEE Symposium on Computational Intelligence in Control and Automation (CICA): IEEE, 2014.

[3]. Öncel Çekim, H., *Forecasting PM10 concentrations using time series models: a case of the most polluted cities in Turkey*, 2020.

[4]. Kaplan, Y., Saray, U., and Azkeskin, E., *Hava Kirliliğine Neden Olan PM10 ve SO 2 Maddesinin Yapay Sinir Ağı Kullanılarak Tahmininin Yapılması ve Hata Oranının Hesaplanması,* Afyon Kocatepe University Journal of Science & Engineering, 14(2), 2014.

[5]. Akbulut, İ.L. and Özcan, B., *Air Pollution Forecast: A Comparison with Artificial Neural Networks and Regression Methods,* Kocaeli Üniversitesi Fen Bilimleri Dergisi, 3(1): 12-22.

[6]. Pires, J., Sousa, S., Pereira, M., Alvim-Ferraz, M., and Martins, F., *Management of air quality monitoring using principal component and cluster analysis—Part I: SO2 and PM10,* Atmospheric Environment, 42(6): 1249-1260, 2008.

[7]. Brunelli, U., Piazza, V., Pignato, L., Sorbello, F., and Vitabile, S., *Two-days ahead prediction of daily maximum concentrations of SO2, O3, PM10, NO2, CO in the urban area of Palermo, Italy,* Atmospheric Environment, 41(14): 2967-2995, 2007.

[8]. Ul-Saufie, A., Yahya, A., and Ramli, N., *Improving multiple linear regression model using principal component analysis for predicting PM10 concentration in Seberang Prai, Pulau Pinang,* International Journal of Environmental Sciences, 2(2): 403-409, 2011.

[9]. Güngör, A. and Sevindir, H.C., *Isparta İlindeki Atmosferde Bulunan Kükürt dioksit (SO2) ve Partikül Madde (PM) Konsantrasyonunun Çoklu Doğrusal Regresyon Yöntemi İle Modellenmesi,* Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 17(1): 95-108, 2013.

[10]. Kunt, F., *Bulanık mantık ve yapay sinir ağları yöntemleri kullanılarak Konya il merkezi hava kirliliği modellenmesi,* Selçuk Üniversitesi Fen Bilimleri Enstitüsü, 2014.

[11]. Tzanis, C.G., Alimissis, A., Philippopoulos, K., and Deligiorgi, D., *Applying linear and nonlinear models for the estimation of particulate matter variability,* Environmental Pollution, 246: 89-98, 2019.

[12]. Ventura, L.M.B., de Oliveira Pinto, F., Soares, L.M., Luna, A.S., and Gioda, A., *Forecast of daily PM 2.5 concentrations applying artificial neural networks and Holt–Winters models,* Air Quality, Atmosphere & Health, 12(3): 317-325, 2019.

[13]. Liu, H., Wu, H., Lv, X., Ren, Z., Liu, M., Li, Y., and Shi, H., *An intelligent hybrid model for air pollutant concentrations forecasting: Case of Beijing in China,* Sustainable Cities and Society, 47: 101471, 2019.

[14]. Qi, Y., Li, Q., Karimian, H., and Liu, D., *A hybrid model for spatiotemporal forecasting of PM2. 5 based on graph convolutional neural network and long short-term memory,* Science of the Total Environment, 664: 1-10, 2019.

[15]. Qin, D., Yu, J., Zou, G., Yong, R., Zhao, Q., and Zhang, B., *A novel combined prediction scheme based on CNN and LSTM for urban PM 2.5 concentration,* IEEE Access, 7: 20050-20059, 2019.

[16]. Al-Janabi, S., Alkaim, A., Al-Janabi, E., Aljeboree, A., and Mustafa, M., *Intelligent forecaster of concentrations (PM2. 5, PM10, NO2, CO, O3, SO2) caused air pollution (IFCsAP),* Neural Computing and Applications: 1-31, 2021.

[17]. Zhu, S., Qiu, X., Yin, Y., Fang, M., Liu, X., Zhao, X., and Shi, Y., *Two-step-hybrid model based on data preprocessing and intelligent optimization algorithms (CS and GWO) for NO2 and SO2 forecasting,* Atmospheric Pollution Research, 10(4): 1326-1335, 2019.

[18]. Xu, Y., Liu, H., and Duan, Z., *A novel hybrid model for multi-step*

*daily AQI forecasting driven by air pollution big data,* Air Quality, Atmosphere & Health, 13(2): 197-207, 2020.

[19]. Shams, S.R., Jahani, A., Kalantary, S., Moeinaddini, M., and Khorasani, N., *The evaluation on artificial neural networks (ANN) and multiple linear regressions (MLR) models for predicting SO2 concentration,* Urban Climate, 37: 100837, 2021.

[20]. Mao, W., Wang, W., Jiao, L., Zhao, S., and Liu, A., *Modeling air quality prediction using a deep learning approach: Method optimization and evaluation,* Sustainable Cities and Society, 65: 102567, 2021.

[21]. Huang, G.-B., Zhu, Q.-Y., and Siew, C.-K., *Extreme learning machine: theory and applications,* Neurocomputing, 70(1-3): 489-501, 2006.

[22]. Rao, C.R. and Mitra, S.K. *Generalized inverse of a matrix and its applications.* Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Theory of Statistics: The Regents of the University of California, 1972.

[23]. Jagtap, P. and Pillai, G. *Comparison of extreme-ANFIS and ANFIS networks for regression problems.* 2014 IEEE International Advance Computing Conference (IACC): IEEE, 2014.

[24]. Gündoğdu, S., *Comparison of static MLP and dynamic NARX neural networks for forecasting of atmospheric PM10 and SO2 concentrations in an industrial site of Turkey,* Environmental Forensics, 21(3-4): 363-374, 2020.

[25]. Zhu, W., Miao, J., and Qing, L., *Constrained extreme learning machines: A study on classification cases,* arXiv preprint arXiv:1501.06115, 2015.

[26]. Zhu, W., Miao, J., and Qing, L. *Constrained extreme learning machine: a novel highly discriminative random feedforward neural network*. 2014 International Joint Conference on Neural Networks (IJCNN): IEEE, 2014.

[27]. Iosifidis, A., Tefas, A., and Pitas, I., *On the kernel extreme learning machine classifier,* Pattern Recognition Letters, 54: 11-17, 2015.

[28]. Amiri, E., Amiri, S., Raghibnezhad, H., Kazemi, A. *Differentiation in Differentiation Method and Neural Network to Investigate Different Criteria of Economic Productivity,*: International Journal of Pioneering Technology and Engineering, 1(01),1-5, 2022

[29]. Shariati, M., Mafipour, M.S., Ghahremani, B., Azarhomayun, F., Ahmadi, M., Trung, N.T., and Shariati, A., *A novel hybrid extreme learning machine–grey wolf optimizer (ELM-GWO) model to predict compressive strength of concrete with partial replacements for cement,* Engineering with Computers: 1-23, 2020.

[30]. Xu, Y. and Shu, Y. *Evolutionary extreme learning machine–based on particle swarm optimization*. International Symposium on Neural Networks: Springer, 2006.

[31]. Cao, L., Yue, Y., Zhang, Y., and Cai, Y., *Improved crow search algorithm optimized extreme learning machine based on classification algorithm and application,* IEEE Access, 9: 20051-20066, 2021.

[32]. Wei, Y., Lv, H., Chen, M., Wang, M., Heidari, A.A., Chen, H., and Li, C., *Predicting entrepreneurial intention of students: An extreme learning machine with Gaussian barebone Harris hawks optimizer,* IEEE Access, 8: 76841-76855, 2020.

[33]. Du, P., Wang, J., Hao, Y., Niu, T., and Yang, W., *A novel hybrid model based on multi-objective Harris hawks optimization algorithm for daily PM2. 5 and PM10 forecasting,* Applied Soft Computing, 96: 106620, 2020.