PREDICTION OF PARTICULATE MATTER CONTENT PM_{10} WITH ARTIFICIAL NEURAL NETWORK AND MULTIPLE LINEAR REGRESSION

STOYANOV, N.* – PANDELOVA, A. – DZHUDZHEV, B. – GEORGIEV, TZ. – KALAPCHIISKA, J.

Technical University of Sofia, Sofia, Bulgaria
(phone: +359-02-965-2382)

**Corresponding author*
e-mail: n_stoyanov@tu-sofia.bg; phone +359-02-965-23-82

(Received 7th Jul 2023; accepted 29th Aug 2023)

Abstract. Air pollution in industrial areas in Bulgaria are among the most serious problems affecting the environment and human health. Therefore, the accurate prediction of air pollution levels is particularly important. The present study compares the predictive capabilities of two modelling approaches using multiple linear regression and artificial neural networks. The models are based on the relationship between particulate matter concentration with meteorological variables and gaseous pollutants. The comparative statistics show that the complex interactions between the studied input factors are better modelled using the nonlinear approach implemented by multilayer perceptron. The obtained results have very high accuracy and reliability and can be successfully used to predict PM_{10} one day ahead in Sofia, the capital city of Bulgaria.
Keywords: *PM_{10} forecasting, air quality, multilayer perceptron, modeling*

Introduction

Air pollution is one of the most important factors affecting human health and the environment in global aspect (Manisalidis et al., 2020). On one hand, the quality of the atmospheric air is directly related to the climatic features of the geographical regions, and on the other hand to the industrialization and increase of the population (Shahrayni and Sodoui, 2016).

The unfavorable weather conditions, such as low wind speed, temperature inversions and high humidity lead to advantageous conditions for the accumulation of various pollutants in the ground layer of the atmosphere (Abdullah et al., 2020). A very important factor is the topography of the area, which can limit the movement of air masses and retain polluted air. The air quality in populated areas also depends on the sources of emissions related to human activities, such as industry, car traffic, construction objects, etc.

In Bulgaria very often exceedances of the permissible levels of atmospheric air pollution determined by European and national legislative norms are registered (Doncheva and Boneva, 2013). One of the most polluted areas is Sofia, where theparticulate matter concentrations PM_{10} are serious problems to which special attention should be paid. The PM_{10} levels are directly or indirectly related to various meteorological variables and gases (Afzali et al., 2014; Uzunov et al., 2019). These factors determine the retention or distraction of PM_{10} on the one hand, and on the other they affect the chemical transformations in the ground layer of the atmosphere.

Two categories of methods can be used to study the degree of atmospheric air pollution - traditional and intelligent. The traditional methods use analytical dependencies that can be deterministic or statistical. The deterministic models require a detailed picture

http://www.aloki.hu ● ISSN 1589 1623 (Print) ●ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2106_56435655

related to the overall knowledge of atmospheric processes, the chemical composition of harmful emissions, as well as their change intime (Honore et al., 2008). They are obtained after solvingdifferential equations and not requiring large arrays of measurement data. The statistical models, on the other hand, have higher accuracy and can reveal relationships with additional variables (Zafra et al., 2017). One of the main disadvantages of statistical models is that their action is valid only for the specific investigated area.

Intelligent methods, such as artificial neural networks (Vinas et al., 2022), support vector machines (Shaziayani et al., 2022) and fuzzy logic (Alyousifi et al., 2021) are alternative approaches to traditional methods. They provide new advanced capabilities for accurate and reliable predictions in the field of ecological monitoring.

In recent years, artificial neural networks have become an increasingly popular machine-learning technique for modeling atmospheric processes. This is due to the possibilities of these methods for modeling highly non-linear interactions between different variables and the better results compared to statistical techniques (Zhang et al., 2021). Important characteristics of neural networks are the large adaptability of the structural model, the possibilities for generalization and tolerating errors. There is a great variety of structures used, mainly based on multilayer perceptron MLP (He et al., 2015; Abderrahim et al., 2016) or radial basis functions RBF (Wahid et al., 2011; Yadav and Nath, 2018). New algorithms as ensemble machine learning methods (Ejohwomu et al., 2022) which increase the accuracy of the results are also being successfully implemented.

The neural networks for PM_{10} prediction commonly use various meteorological variables as input parameters, such as temperature, humidity, windspeed, atmospheric pressure, as well as concentrations of various gases (Ul-Saufie et al., 2011; de Gennaro et al., 2013). They are most often applied to forecasts of PM_{10} concentrations from few hours to several days ahead. Combined models including a neural network and an ARIMA method are also often implemented (Wongsathan and Seedadan, 2016). The hybrid ARIMA-ANN model has higher accuracy, as well as improved forecasting performance.

An interesting topic is comparison of the predictive models, based on different synthesis techniques. This comparison can show the advantages and disadvantages of individual modeling approaches and give an answer to the question of choosing the most suitable method. The subject of research is usually the analysis of the properties of models built with multiple linear regression, artificial neural networks, or hybrid models (Aslanargun et al., 2007; Abdullah et al., 2019).

The purpose of the present study is to investigate two approaches for forecasting of PM_{10} concentrations one day ahead. The methods of multiple linear regression and artificial neural networks were used for modeling.The construction of the models is based on the relationship of PM_{10} with several meteorological variables and the gases - Carbon Monoxide (CO) and Sulfur Dioxide (SO_2). The software packages Statgraphics and Matlab were used for processing and analyzing the experimental data.

Materials and Methods

Explored area

The subject of the investigation is the capital city of Bulgaria - Sofia, which is among the European cities with the most complex characteristics, in terms of geographical relief, the distance to sea or oceans (influence of winds), temperature inversions and the presence of dense fogs. The area is located in the western part of the territory of Bulgaria,

http://www.aloki.hu ● ISSN 1589 1623 (Print) ●ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2106_56435655

at an altitude of 550 m, surrounded by five mountains. The analyzed data for PM_{10} is taken from Mladost district, located in the south-eastern part of Sofia and forming 10% of the city's territory. It is one of the official sites of the Executive Environment Agency, which is part of the national air quality control system in Bulgaria and reports to the European Commission. The location of the studied area is shown in *Figure 1*.

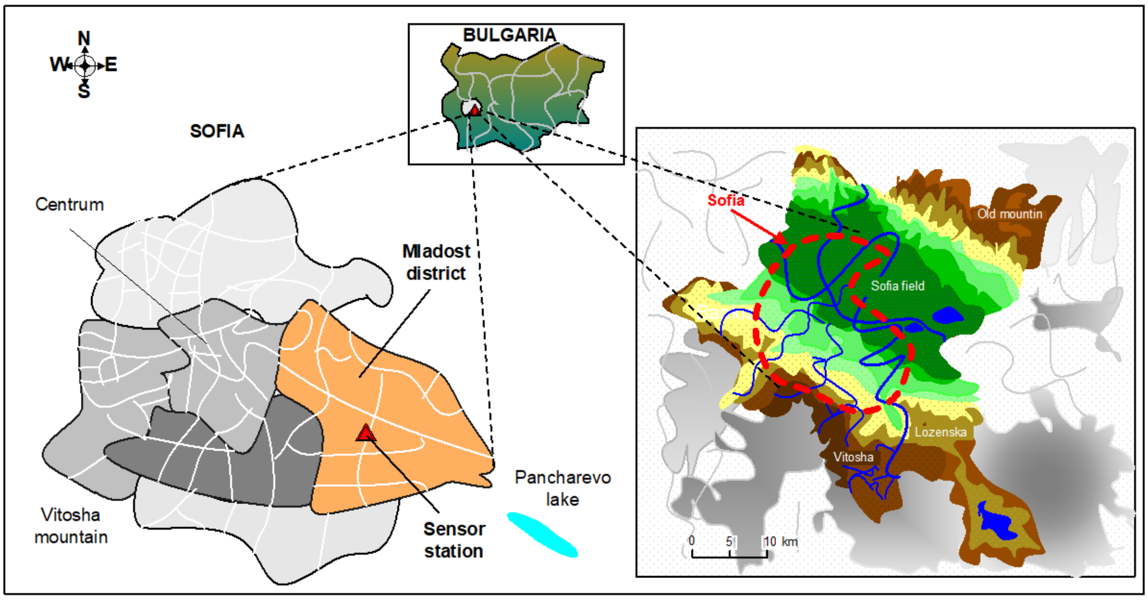

Figure 1. *Location of explored area – Mladost region, Sofia, Bulgaria*

The comparative analysis in this study was made for average daily concentrations of particulate matter with an aerodynamic diameter of less than (≤) 10 µm - PM_{10}. The models are built on the basis of the relationship of PM_{10} with five meteorological variables - air temperature T, solar radiation R, wind speed S, wind direction D and atmospheric pressure P and the gases - Carbon Monoxide (CO) and Sulfur Dioxide (SO_2). The experimental data are taken for a period of one calendar year from 01.01.2017 to 12.31.2017, including daily average values for each variable. The used equipment to collect the data is an automatic sensor station Termo Scientific model SHARP 5030 (Thermo Fisher Scientific Inc.).

Data normalization

The concentration of PM_{10}, meteorological variables and gases that are used to construct the models have different units of measurement. For this reason data normalization is necessary (Abdullah et al., 2019). A min-max technique was used, which generated a new data set with values in the interval [0 ÷ 1] according to the following expression:

$$z_i = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{Eq.1}$$

where z_i is the normalized value of the corresponding x_i.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ●ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2106_56435655

Multiple linear regression method

The MLR method establishes the relationship between several explanatory variables and the responsible variable, by using a linear function (Abdullah et al., 2017). The mathematical equation can be presented in the following generalized form:

$$Y = a_0 + \sum_{i=1}^{n} a_i X_i + \varepsilon \quad \text{(Eq.2)}$$

where *Y* is the responsible variable, X_i are the independent variables, which must be at least two, a_i are regression coefficients, and $\varepsilon$is stochastic error related to the regression method.

Compulsory conditions for the adequacy of MLR models are related to the residuals – they must have a normal distribution with zero mean and constant variance (Abdullah et al., 2017). Synthesis and tests with the MLR model were performed in *Statgtraphics software* (StatPoint Technologies, Inc).

Artificial neural network – Multi-layer perceptron

The artificial neural network is a computer system in which the parallel processing of information is realized, similar to the processes taking place in the human brain. The neural network consists elementary processing units (neurons) connected in a certain way to each other (Feng et al., 2015; Biancofiore et al., 2017). Each neuron has inputs that can receive outer signals or signals from the outputs of other neurons. The inputs have weights that determine the strength of the connection between the neurons. The incoming signals multiplied by the weights are summed, after which the output signal is formed by activation function. The neurons are connected in layers that form the structure of the network. The layers are divided into input, hidden and output, depending on the location of the neurons. The processes in one neuron with number j can be described mathematically according to the following relationship (Abdullah et al., 2019):

$$y'_j = \sum_{i=1}^{n} w_{i,j} u_i + b_j \quad \text{(Eq.3)}$$

where y_j is the output signal, w_{ij} is the input weight u_i, and b_j is the bias of neuron j. The activation function *f* by which the output signal is obtained is usually non-linear:

$$y_j = f(x)\left[\sum_{i=1}^{n} w_{i,j} u_i + b_j\right] \quad \text{(Eq.4)}$$

A Feed-Forward topology based of multilayer perceptron MLP is commonly used to predict PM_{10}, depending on a set of meteorological variables and gases. For the present study, a neural network consistinginput, hidden and output layer is developed, which structure is shown in *Figure 2*.

The set of input data is divided into two parts. The first includes 70% observations and is used to train the network. Avalidation process was performed with the remaining 30%. A Levenberg–Marquardt training algorithm was used to train the network. MLP studies were performed in *Matlab software* (The MathWorks Inc., Natick, MA, USA).

Performance indicators

The evaluation of the models is realized according to several criteria that show the overlapping of the predicted values to the data (Nazif et al., 2018; Dutta and Jinsart,

http://www.aloki.hu ● ISSN 1589 1623 (Print) ●ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2106_56435655

2021). The performance comparison was made by Correlation coefficient R^2, Root Mean Square Error (RMSE), , Mean Absolute Error (MAE) and Index of Agreement (IA). The selected accuracy criteria are as follows:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{n}} \qquad \text{(Eq.5)}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t| \qquad \text{(Eq.6)}$$

$$R^2 = \left(\frac{\sum_{t=1}^{n}(y_t - y_t^{mean})(\hat{y}_t - \hat{y}_t^{mean})}{n.S_{forec}.S_{obs}}\right) \qquad \text{(Eq.7)}$$

$$IA = 1 - \left[\frac{\sum_{i=1}^{n}(y_t - \widehat{y_t})^2}{\sum_{i=1}^{n}(|y_t - \hat{y}_t^{mean}| + |\hat{y}_t - \hat{y}_t^{mean}|)^2}\right] \qquad \text{(Eq.8)}$$

where n is the total number of observations, y_t are the predicted values, $\hat{y}_t$are the observed values, y_t^{mean} – mean value of the predicted values, , $\hat{y}_t^{mean}$ mean value of the observed values, S_{forec} is the standard deviation of the predicted values, S_{obs} is the standard deviation of the observed values.

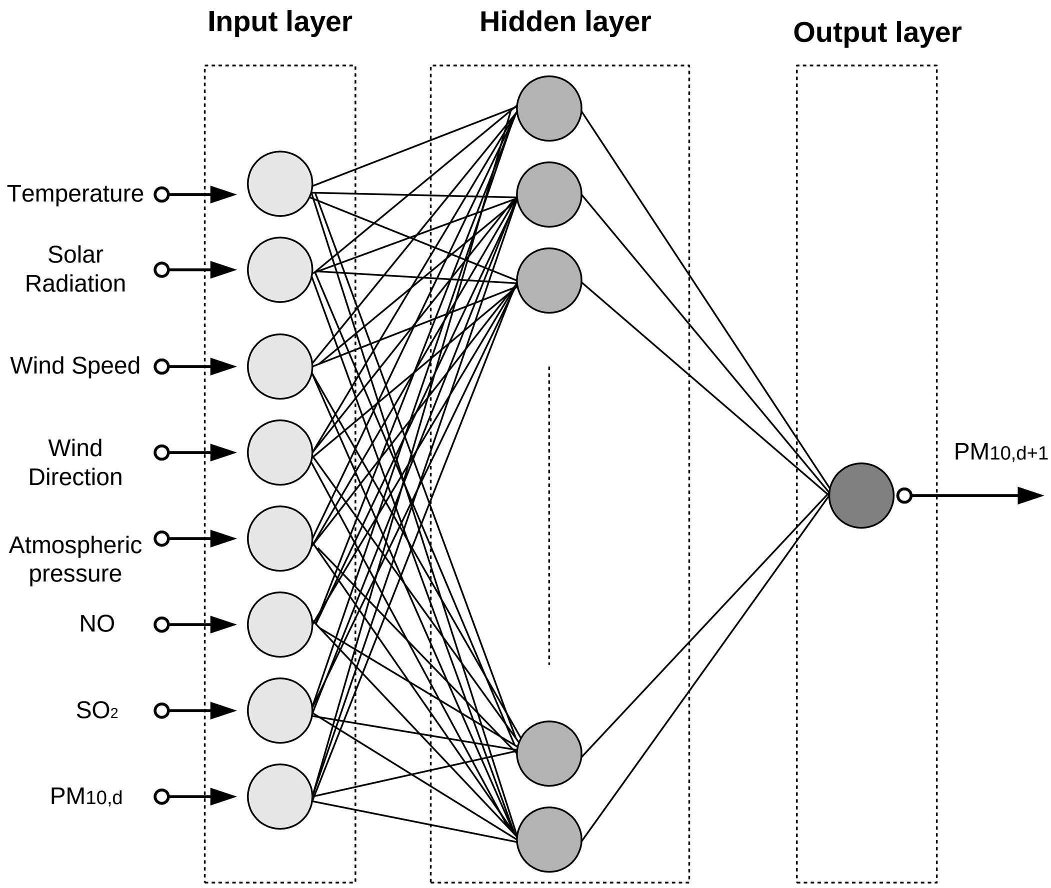

Figure 2. *Schematic diagram of Multi-Layer Perceptron*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ●ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2106_56435655

Results

A comparison of the predictive capabilities of models created using MLR and ANN was performed. The study was carried out on the same input data set and established the relationship of PM_{10} with five meteorological variables and the concentrations of two gases.

Descriptive statistics

The total dataset for the eight variables, covering a period of one year is 2917. Due to measurement equipment problems, there are less than 0.7% missing values and one of the data is out of range. The missing values were filled using linear interpolation (Abdullah et al., 2019). Descriptive statistics of the input data are presented in *Table 1*.

Table 1. *Descriptive statistic of input data*

Statistic	Minimum	Mean	Maximum	Stand. Dev.
T, °C	-12,8	11,73	27,8	8,6435
R, W/m^{-2}	8,7	166,092	379,5	105,597
S, m.s^{-1}	0,57	1,50786	3,15	0,419511
D, degree	51,34	147,992	261,99	52,6419
P, mbar	912,0	922,723	934,0	3,83365
C, µg.m^{-3}	7,89	33,671	153,76	20,7459
CO, mg.m^{-3}	0,05	0,547	3,1	0,457
SO_2, µg.m^{-3}	0	7,241	35,66	5,119

Multiple linear regression model

Spearman's rank correlation analysis was carried out to determine relationships between all variables. From the obtained results, statistically significant non-zero correlations were found, indicating the possibility of searching for a mathematical model. In the second stage, an analysis of the distributions of the variables was performed. After a large number of transformations, normal or admissible distributions were found for all variables.

After establishing the distributions, a search for mathematical model describing the relationship between the independent variable PM_{10} and the meteorological factors and gases started. The used method is stepwise multiple linear regression. Unsatisfactory results, to reach a normal distribution of the residuals, required the use of the Box-Cox procedure for transformation of the dependent variable PM_{10}. After a series of trials, a function was established, that satisfied all the adequacy conditions. The analytical form of this model has the following form:

$$\begin{aligned} PM'_{10,d+1} = {} & -0{,}067 * T_d + 0{,}257 * R_d + 0{,}072 * D_d + 0{,}223 * S_d \\ & + 0{,}269 * P_d + 0{,}259 * SO_{2,d} + 0{,}686 * PM_{10,d} \end{aligned} \quad \text{(Eq.9)}$$

where index *d+1* indicates the next day, and *d* indicates the current day.

The obtained model has the following characteristics - correlation coefficient $R^2 = 0.9658$, Root Mean Square Error RMSE = 0.1042, Mean Absolute Error MAE = 0.0792, Index of Agreement AI = 0,791062 and Durbin-Watson statistic = 1.49.

APPLIED ECOLOGY AND ENVIRONMENTAL RESEARCH 21(6):5643-5655.
http://www.aloki.hu ● ISSN 1589 1623 (Print) ●ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2106_56435655

Various statistical tests were performed to check the normal distribution of the residuals. In *Figures 3 and 4* are presented the distribution of the residuals and plot of the fitting results of the MLR.

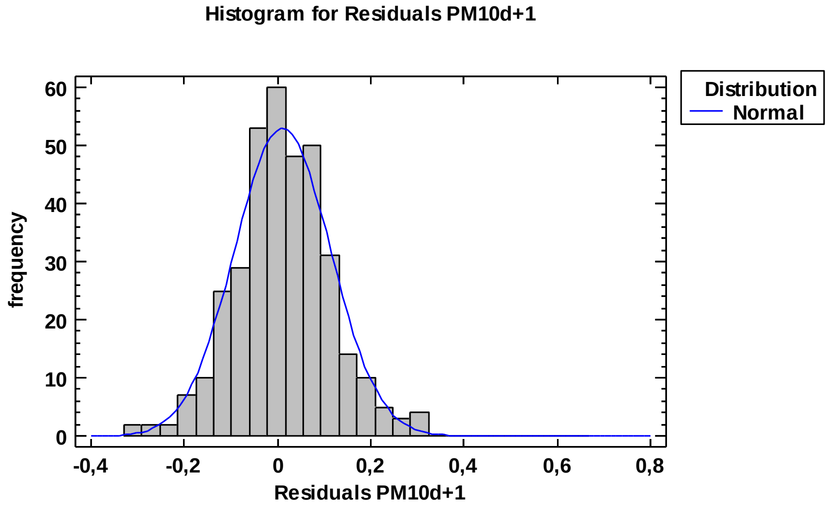

Figure 3. *Histogram of the residuals of MLR*

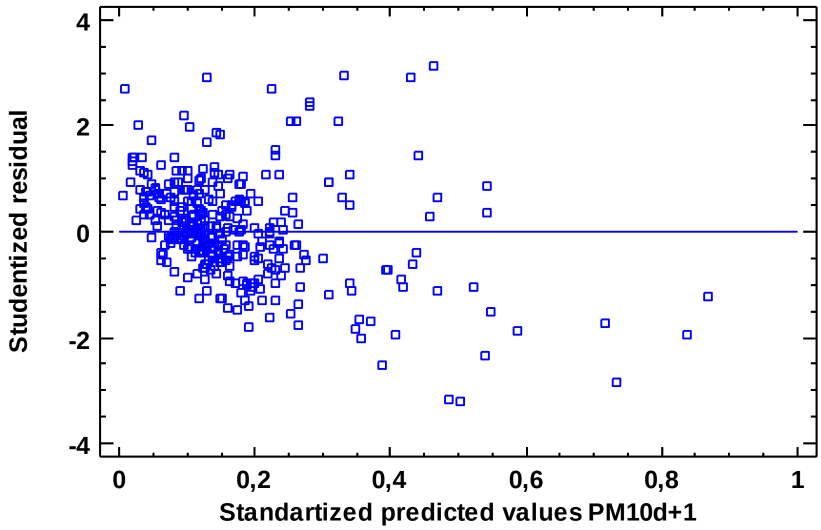

Figure 4. *Fitting results of PM_{10} using MLR*

Table 2 shows the characteristics of the residuals of the regression model. The p-value for all tests is greater than 0.05 that shows the residuals have a normal distribution.

Table 2. *Residuals tests*

Residuals	Statistic	p-value
χ^2	23,7042	0,956
Shapiro-Wilk W	0,992	0,055
Kolmogorov-Smirnov		0,768

http://www.aloki.hu ● ISSN 1589 1623 (Print) ●ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2106_56435655

The plot between real and predicted values is presented in *Figure 5*. The brown line shows the permissible limits of PM_{10} (50 µg/m^3) in the atmospheric air. For the studied period the observed data exceeded the prescribed treshhold value more than 45 times.

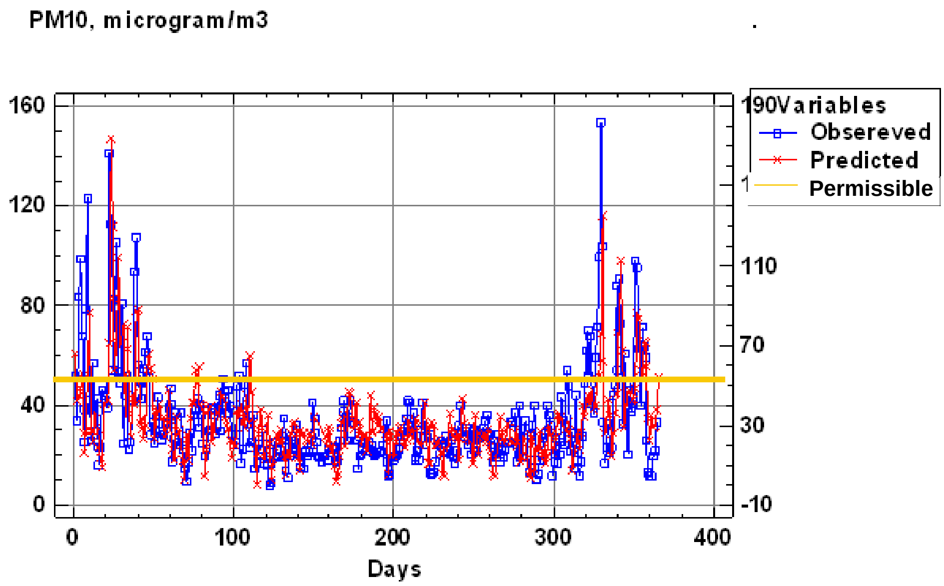

Figure 5. *Plot of predicted versus observed PM_{10} concentrations*

Artificial neural network

The most important matter in creating the neural network is determining the hidden layers and the number of neurons in them. It is interesting to note that more neurons in the hidden layers may lead to over-fitting the output and a smaller number may not model well enough the data (Meerasri and Sothornvit, 2022).

A three-layer neural network with 8-34-1 topology was developed. The structure is a multilayer perceptron MLP, realizing Feed-Forward topologies. The input layer includes eight nodes (neurons) to which the arrays of normalized air temperature, solar radiation, wind speed, wind direction, atmospheric pressure, carbon monoxide (CO), sulfur dioxide (SO_2) and the concentration of $PM_{10,d}$ are fed.

Many experiments have been done to determine the hidden layers and the number of neurons involved in them. The criterion was reaching the maximum value of the correlation coefficient R^2 and minimum values of RMSE and MAE. It was found that the best results of MLP structure are obtained for one hidden layer with an optimal number of neurons 34.

The output layer of the MLP represents one neuron for which a linear transfer function recommended in Abdullah et al. (2019) is chosen. Additional parameters that have been used for the MLR network are a learning rate of 0.05 (Hossain et al., 2013) and a number of epochs, that set after examining the training error and the validation error. Different activation functions in the layers were also tested, taking into account the variants recommended in Voukantsis et al. (2011). The best combination was obtained for tangent-sigmoid '*tansig*' for the hidden layer and log-sigmoid '*logsig*' for the output layer.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ●ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2106_56435655

After simulations with the optimal MLP architecture, the following results were obtained – R^2 = 0.99373; RMSE = 0.091, MAE = 0.0572 and IA = 0,801525. The results of the regression analysis are shown in *Figure 6*.

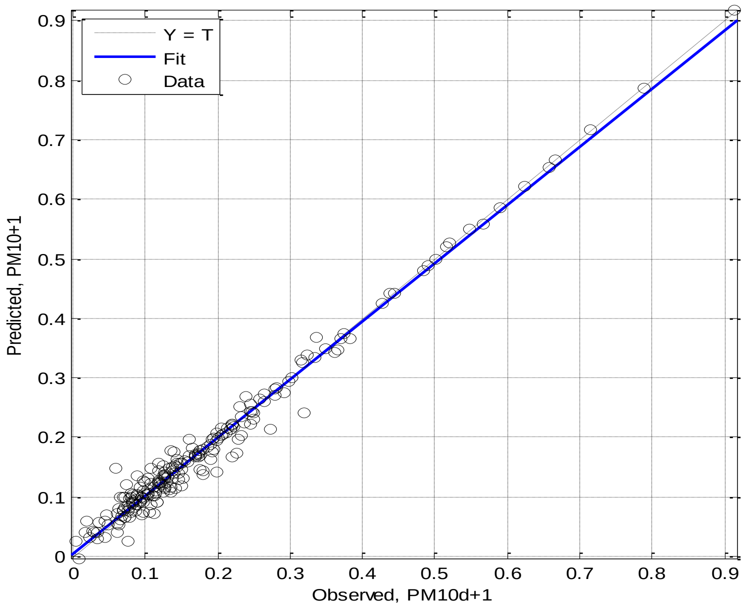

Figure 6. *Scatter plot of forecast PM_{10} concentrations against observed by MLP*

It was found that with the increase the number of neurons above 25 in the hidden layer, the values of the preformance indicatorsreached those of MLR - the coefficient of determination is in the interval 0.95 ÷ 0.97, RMSE - between 0.0967 ÷ 0 .1103 and MAE – 0.0642 ÷ 0.0817. On the other hand, increasing the number of neurons above the optimum does not lead to an increase the characteristics of the neural network. The results of the predicted versus observed values after the MLP simulation are presented in *Figure 7*.

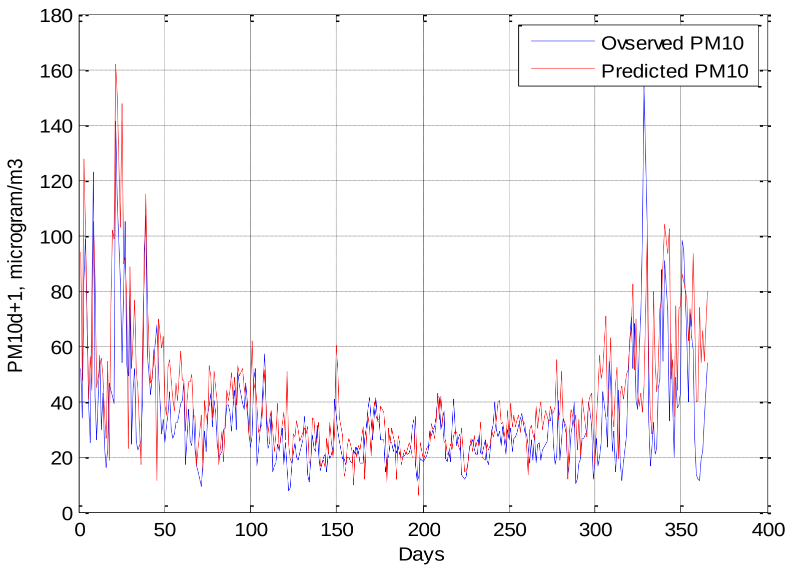

Figure 7. *Plot of predicted PM_{10}concetrations versus observed by MLP*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ●ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2106_56435655

Studies were also performed on the most appropriate selection of activation functions in the MLP layers. The combinations tested for the chosen variants *tansig*, *logsig* and *purelin* are shown in *Table 3*. The best combination determined after comparing the R^2, RMSE and MAE criteria is *tansig* for the hidden layer and *purelin* for the output layer.

Table 3. *Tested variants of activation functions*

Activation function hidden layer	Activation function output layer	R^2	RMSE	MAE
Tansiq	Purelin	0,99373	0,091	0,0572
Logsig	Purelin	0,96523	0,1030	0,0614
Tansig	Tansig	0,97449	0,1037	0,0676
Logsig	Tansig	0,96791	0,0879	0,0637
Tansig	Logsig	0,46299	0,3529	0,3375
Logsig	Logsig	0,32613	0,3534	0,3385

Discussion

A comparative analysis between MLR and MLP methods

The comparison between MLR and MLP is made according to the performance indicators – Correlation Coefficient R^2, Root Mean Square Error RMSE, Mean Absolute Error MSE and Index of Agreement AI. When the coefficient of determination R^2 tend to 1, and the RMSE and MAE are closerto 0, the model has better fitting. *Table 4* presents the values of the compared indicators for the two modeling methods.

Table 4. *Comparison between MLR and MLP*

Method	R^2	RMSE	MAE	IA
LMR	0,9658	0,1042	0,0792	0.791062
ANN - MLP	0,99373	0,091	0,0572	0,801525

Fromtheobtainedresults, itcanbesummarizedthatthecharacteristicsofneuaral network is higher that of linear regression. The coefficient of determination of MLP (0.99373) exceeds that of MLR (0.9658), indicating that the neural network, even a little, better describes the experimental data. On the other hand, the RMSE (0.091) and MAE (0.0572) values for MLP are also in favor of the non-linear method, compared to MLR (0.1042 and 0.0572).

The combinations of different activation functions show that four variants have better indicators than those obtained by MLR. The *logsig-purelin*, *tansig-tansig* and *logsig-tansig* variants are approximately identical and tendto the values of the best *tansig-purelin* combination, without reaching them.

In general, it can be concluded that the nonlinear algorithm with Feed-Forward topologies is a better modeling technique than MLR. The results of all four investigated criteria R^2, RMSE,MAE and IA, although to a small extent, are in favor of the neural network. A peculiarity of the MLP method is that, unfortunately, the initial values of the weights in the neural network structure are set randomly and repetition are needed to obtain the best results.

The obtained results of the comparative analysis are in agreement with those obtained in similar studies (Abdullah et al., 2020; Dutta and Jinsart, 2021). The nonlinear approach is characterized by higher accuracy, as the essential peculiarity is the selection of the hidden layers and the neurons in them. The neural networks with one or two hidden layers usually show good performance, defined by high R^2 and minimal errors. With an appropriate selection of activation functions in the individual layers and number of hidden neurons, the coefficient of determination can reach values around 0.999 (Meerasri and Sothornvit, 2022), which is in accordance with the present results.

Conclusion

The purpose of this study is to investigate the predictive capabilities of two methods for modeling of air pollution concentrations PM_{10} in Sofia,the capital city of Bulgaria. The models use a linear and non-linear approach to predict PM_{10} one day ahead, realized by step-wise multiple linear regression and Feed-Forward topology, respectively.

MLR and MLP were used for modeling the relationship between the concentrations of particulate matter $PM_{10,d+1}$ (one day ahead) with the variables - air temperature, solar radiation, wind speed, wind direction and atmospheric pressure, concentration of Carbon Monoxide (CO), concentration of Sulfur Dioxide (SO_2) and $PM_{10,d}$ (present day). After establishing of the relationships between all the variables, two forecasting models were created. The adequacy of the MLR model was proven by analyzing the behavior of the residuals. A neural network with Feed-Forward topologies consisting of one input, one hidden and one output layer is developed. Studies to determine the most appropriate topology and number of neurons in the hidden layer have been carried out.

A comparative analysis of the developed models was performed, comparing their main performance indicators – correlation coefficient R^2, Root Mean Absolute Error RMSE, Mean Absolute Error MAE and Index of Agreement AI. The better predictive capabilities offered by the non-linear algorithm method were established.

The obtained results show that the constructed models can achieve a high accuracy of one-day-ahead forecasting, based on actual data of meteorological variables and gas components. The models can successfully be used to create a methodology for air pollution control in this area.

Acknowledgements. The authors would like to thank the Research and Development Sector at the Technical University of Sofia for the financial support.

REFERENCES

[1] Abderrahim, H., Chellali, M. R., Hamou, A. (2016): Forecasting PM_{10} in Algiers: efficacy of multilayer perceptron networks. – Environmental Science and Pollution Research 23(2): 1634-1641.

[2] Abdullah, S., Ismail, M., Fong, S. Y. (2017): Multiple linear regression (MLR) models for longterm PM_{10} concentration forecasting during different monsoon seasons. – Journal of Sustainability Science and Management 12(1): 60-69.

[3] Abdullah, S., Ismail, M., Ahmed, A. N., Abdullah, A. M. (2019): Forecasting Particulate Matter Concentration UsingLinear and Non-Linear Approaches for Air QualityDecision Support. – Atmosphere 10: 667.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ●ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2106_56435655

[4] Abdullah, S., Napi, N. N. L. M., Ahmed, A. N., Wan Mansor, W. N., Abu Mansor, A., Ismail, M., Abdullah, A. M., Ahmad Ramly, Z. T. (2020): Development of Multiple Linear Regression for Particulate Matter (PM_{10}) Forecasting during Episodic Transboundary Haze Event in Malaysia. – Atmosphere 11: 289.
[5] Afzali, A., Rashid, M., Sabariah, B., Ramli, M. (2014): PM_{10} Pollution: Its Prediction and Meteorological Influence in PasirGudang, Johor. – 8th International Symposium of the Digital Earth (ISDE8), IOP Conf. Series: Earth and Environmental Science 18.
[6] Alyousifi, Y., Othman, M., Husin, A., Rathnayake, U. (2021): A new hybrid fuzzy time series model with an application to predict PM_{10} concentration. – Ecotoxicology and Environmental Safety 227: 112875.
[7] Aslanargun, A., Mammadov, M., Yazici, B., Yolacan, S. (2007): Comparison of ARIMA, neural networks and hybrid models in time series: tourist arrival forecasting. – Journal of Statistical Computation and Simulation 77(1): 29-53.
[8] Biancofiore, F., Busilacchio, M., Verdecchia, M., Tomassetti, B., Aruffo, E., Bianco, S., Di Tommaso, S., Colangeli, C., Rosatelli, G., Di Carlo, P. (2017): Recursive neural network model for analysis and forecast of PM_{10} and $PM_{2.5}$. – Atmospheric Pollution Research 8: 652-659.
[9] de Gennaro, G., Trizio, L., Di Gilio, A., Pey, J., Pérez, N., Cusack, M., Alastuey, A., Querol, X. (2013): Neural network model for the prediction of PM_{10} daily concentrations in two sites in the Western Mediterranean. – Science of the Total Environment 463-464C(4): 875-883.
[10] Doncheva, M., Boneva, G. (2013): Particulate Matter Air Pollution in Urban Areas in Bulgaria. – Journal of Environmental Protection and Ecology 14(2): 422-429.
[11] Dutta, A., Jinsart, W. (2021): Air Pollution in Indian Cities and Comparison of MLR, ANN and CART Models for Predicting PM_{10} Concentrations in Guwahati, India. – Asian Journal of Atmospheric Environment 15(1): 2020131.
[12] Ejohwomu, O. A., Oshodi, O. S., Oladokun, M., Bukoye, O. T., Emekwuru, N., Sotunbo, A., Adenuga, O. (2022): Modelling and Forecasting Temporal $PM_{2.5}$ Concentration Using Ensemble Machine Learning Methods. – Buildings 12: 46.
[13] Feng, Q. L., 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. – Atmospheric Environment 107: 118-128.
[14] He, H. D., Lu, W. Z., Xue, Y. (2015): Prediction of particulate matters at urban intersection by using multilayer perceptron model based on principal components. – Stochastic Environmental Research and Risk Assessment 29: 2107-2114.
[15] Honore, C., Rouıl, L., Vautard, R., Beekmann, M., Bessagnet, B., Dufour, A., Elichegaray, C., Flaud, J. M., Malherbe, L., Meleux, F., Menut, L., Martin, D., Peuch, A., Peuch, V. H., Poisson, N. (2008): Predictability of European air quality: Assessment of 3 years of operational forecasts and analyses by the PREV'AIR system. – Journal of Geophysical Research 113: d04301. doi:10.1029/2007jd008761.
[16] Hossain, Md. M., Neaupane, K., Tripathi, N. K., Piantanakulchai, M. (2013): Prediction of Groundwater Arsenic Contamination using Geographic Information System and Artificial Neural Network. – Environment Asia 6(1): 38-44.
[17] Manisalidis, I., Stavropoulou, A., Stavropoulos, A., Bezirtzoglu, E. (2020): Environmental and health Impacts of Air pollution: A review. – Frontiers in Public Health 8(14): 1-13.
[18] Meerasri, J., Sothornvit, R. (2022): Artificial neural networks (ANNs) and multiple linear regression (MLR) for prediction of moisture content for coated pineapple cubes. – Case Studies in Thermal Engineering 33: 101942.
[19] Nazif, A., Mohhamed, N. I., Malakahmad, A., Abualqumbos, M. S. (2018): Regression and multivariate models for predicting particulate matter concentration level. – Environmental Science Polution 25(1): 283-289.
[20] Shahrayni, H. T., Sodoui, S. (2016): Statistical Modeling Approaches for PM_{10} Prediction in Urban Areas; A Review of 21st-Century Studies. – Atmosphere 7(2): 15.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ●ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2106_56435655

[21] Shaziayani, W. N., Ahmat, H., Razak, T. R., Abidin, A. W. Z., Warris, S. N., Asmat, A., Noor, N. M., Ul-Saufle, A. Z. (2022): A Novel Hybrid Model Combining the Support Vector Machine (SVM) and Boosted Regression Trees (BRT) Technique in Predicting PM10 Concentration. – Atmosphere 13: 2046.

[22] Ul-Saufie, A. Z., Yahya, A. S., Ramli, N. A., Hamid, H. A. (2011): Comparison Between Multiple Linear Regression And Feed forward Back propagation Neural Network Models For Predicting PM_{10} Concentration Level Based On Gaseous And Meteorological Parameters. – International Journal of Applied Science and Technology 1(4).

[23] Uzunov, M. I., Pilev, D. I., Savova-Videnova, M. (2019): Statistical evaluation and forecasting of the dust particlesconcentrations depending on the meteorological conditions. – Journal of Chemical Technology and Metallurgy 54(1): 149-162.

[24] Vinas, M. J. D., Gerardo, B. D., Medina, R. P. (2022): Forecasting $PM_{2.5}$ and PM_{10} Air Quality Index using Artificial Neural Network. – Journal of Positive School Psychology 6(5): 6863-6871.

[25] Voukantsis, D., Karatzas, K., Kukkonen, J., Räsänen, T., Karppinen, A., Kolehmainen, M. (2011): Intercomparison of air quality data using principal component analysis, and forecasting of PM_{10} and $PM_{2.5}$ concentrations using artificial neural networks, in Thessaloniki and Helsinki. – Science of the Total Environment 409: 1266-1276.

[26] Wahid, H., Ha, Q. P., Duc, H. N. (2011): Computational intelligence estimation of natural background ozone level and its distribution for air quality modelling and emission control. – Proceedings of 28th International Symposium on Automation and Robotics in Construction, ASARC: 1157-1163.

[27] Wongsathan, R., Seedadan, I. (2016): A hybrid ARIMA and Neural Networks model for PM10 pollution estimation: The case of Chiang Mai city moat area. – 2016 International Electrical Engineering Congress iEECON 2016, 2-4 March, Chiang Mai, Thailand, pp. 273-276.

[28] Yadav, V., Nath, S. (2018): Daily Prediction of PM10 using Radial Basis Function and Generalized Regression Neural Network. – 2018 Recent Advances on Engineering, Technology and Computational Sciences (RAETCS): 18045140.

[29] Zafra, C., Angel, Y., Torres, E. (2017): ARIMA analysis of the effect of land surface coverage on PM10 concentrations in a high-altitude megacity. – Atmospheric Pollution Research 8: 660-668.

[30] Zhang, H., Srinivasan, R., Yang, X. (2021): Simulation and Analysis of Indoor Air Quality in Florida Using Time Series Regression (TSR) and Artificial Neural Networks (ANN) Models. – Symmetry 13: 952.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ●ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2106_56435655