Engineering Applications
https://publish.mersin.edu.tr/index.php/enap
e-ISSN 2979-9201

Carbon Monoxide forecasting with artificial neural networks for Konya (Case Study of Meram)

Esra Aslı Çubukçu *[1], **Vahdettin Demir** [1], **Mehmet Faik Sevimli** [1]

[1]KTO Karatay University, Faculty of Engineering and Natural Sciences Faculty, Civil Engineering Department, Konya, Türkiye, cubukcuasli@gmail.com, vahdettin.demir@karatay.edu.tr, mehmet.faik.sevimli@karatay.edu.tr

Cite this study: Çubukçu, E. A., Demir, V., & Sevimli, M. F. (2023). Carbon Monoxide forecasting with artificial neural networks for Konya (Case Study of Meram). Engineering Applications, 2 (1), 69-74

Keywords
Air quality
Forecast
Artificial Neural Network
Konya

Research Article
Received: 12.12.2022
Revised: 13.01.2023
Accepted: 22.01.2023
Published: 03.02.2023

Abstract

The first use of the term air quality problem, which emerged with the industrial revolution, date back to the 18th and 19th centuries. Natural causes such as forest fires and volcanic eruptions caused by air pollution, as well as the effect of increasing human activities on air quality with the industrial revolution, are more than natural effects. Consequences of air pollution; acid rains, climate change, respiratory diseases, occurrence of extreme weather conditions, decrease/increase in the number of species in the ecosystem. Especially in megacities, human health is closely affected due to wrong construction, heavy traffic and population density. For this reason, the preliminary forecast and model of air quality has an important place for possible health problems and global problems. In this study, Carbon Monoxide (CO, μg/m^3) records of Meram district of Konya were modeled with three different Artificial Neural Networks (ANN) methods. These are Multilayer, Radial-Based and Generalized Regression ANN. Input parameters in modeling are air quality parameters such as; PM_{10}, SO_2, NO_2, NO_X and periodicity. CO is the output parameter. CO is quite harmful for human health; It is a colorless, odorless gas and is formed when the carbon in fuels is not fully combusted. When the comparison criteria are examined, it is seen that the best result is the input model of the Multilayer ANN model (RMSE=90.361, MAE=74.206, R^2= 0.824).

1. Introduction

The use of the term air pollution emerged with the industrial revolution. In addition to natural effects, human factor is the main cause of air pollution. The negative effects of air pollution on health are caused by human inhalation, direct exposure of pollutants or as an indirect exposure to air, soil, water, plants, animals, etc. can occur with pollutants accumulating in environmental environments. Today, most of the air pollution is carried out by motor vehicles and industrial facilities. Air quality is very important for the safety and sustainability of human health. Every year, thousands of people die from respiratory diseases caused by air pollution. Especially in big and industrial cities, air pollution causes an increase in health problems in people. Changes in the physical, chemical and biological properties of the air affect natural and artificial non-living beings as well as living things. Primary pollutants and secondary pollutants are the two main categories of air pollutants. Primary air pollutants are emitted directly from the source to the atmosphere Primary air pollutants include Particulate Substances (PM), Oxides of Sulfur (SO_x), Oxides of Nitrogen (NO_x), Hydrocarbons (HC) and Carbon Monoxide (CO). Secondary air pollutants are primary air pollutants formed as a result of chemical and/or physical reactions. Among them, PAN (Peroxy Acetyl Nitrate), Ozone (O_3), Sulfuric Acid (H_2SO_4). When water vapor in clouds condenses in the form of water droplets or snow particles, it combines with nitrogen dioxide as it descends from the atmosphere to form a highly corrosive rain of nitric acid. In addition, NO, NO_2 concentrations vary according to place and time. For people to be ready for and respond to upcoming extreme occurrences, air pollution forecasting is essential [1,2]. Unal et al. [3] modeled the air pollution of Ankara with machine learning. They compared the results as RMSE, MAE and R^2. Tunç and Toros [4] investigated the effect of the Covid-19 pandemic on air pollution in Adana. When the results are examined; While there was a slight decrease in the concentration of environmental pollution both during the pandemic and compared to previous years, they found that control measures during the pandemic did not have a significant effect on the general air quality of Adana province. In their study conducted in 2020, Kara et al. analyzed NO, NO_2, NOx levels in Turkey by dividing them into Black Sea region, Marmara region, Aegean region, Mediterranean region, Central Anatolia region, Eastern regions. Looking at the results; It has been shown that the amount of nitrogen dioxide and its variations (NOx's) have been decreasing over the years in Turkey [5]. Bayati et al. in their study in 2021; In 2019, they aimed to contribute to the improvement of air quality by monitoring the behavior of air pollutants over Van through statistical analysis of the data. Looking at the results, they observed that the wind plays an effective role in the horizontal transport of pollutants in the atmosphere. They revealed that if the wind is calm, the air pollution stays where it is and the precipitation helps the pollutants in the atmosphere to collapse, and because of this feature, precipitation is a cleaner of the atmosphere [6].

In this study, the amount of CO in the air (μg/m^3) of Meram district was estimated according to the daily data of Meram air quality station in the central district of Konya. Three different Artificial Neural Networks (ANNs) were used in the modeling. These are Multilayer ANN (MANN), Radial Based ANN (RBANN) and Generalized ANN (GRANN). Primary pollutant elements as input parameters; PM_{10}, SO_2, NO_2, NO_X and periodicity are selected. The output parameter is CO. CO is quite harmful for human health; It is a colorless, odorless gas and is formed when the carbon in fuels is not fully combusted. Its main source is internal combustion engines (85-95%). Industry, wood burning and forest fires are major sources of CO emissions. All units are μg/m^3. Since the most dangerous parameter is CO, it has been chosen as the output parameter.

2. Material and Method

2.1. Material

The data were obtained from the air quality measurement station in Meram district of Konya, and from the website of the Ministry of Environment, Urbanization and Climate Monitoring [7]. The date of the data was taken daily between 01.11.2020 and 31.10.2021. 80% of the data from the beginning was used in training, and the most recent 20% was used in testing. Table 1 provides descriptive statistical information about the data used.

Table 1. Statistical information of parameters

Statistic	PM_{10} (μg/m^3)	SO_2 (μg/m^3)	NO_2 (μg/m^3)	NO_X (μg/m^3)	CO (μg/m^3)
Average	28.467	12.474	38.759	60.550	908.204
Standard Eror	1.355	0.624	0.692	1.966	39.918
Median Term	20.465	7.620	37.070	50.425	601.305
Standard Deviation	24.769	11.405	12.640	35.938	729.535
Kurtosis	5.143	1.646	-0.223	4.978	5.050
Skewness	2.181	1.430	0.428	1.962	2.268
Range	157.720	62.430	70.200	254.290	3995.870
Max	161.010	63.690	82.420	269.290	4201.670
Min	3.290	1.260	12.220	15.000	205.800
Total	9507.890	4166.230	12945.560	20223.560	303340.240
Number	334.000	334.000	334.000	334.000	334.000

2.2. Method

2.2.1. Multi-Layered ANN

One of the most widely used ANN models, MANN is an input layer, at least one-cell intermediate It consists of a layer and an output layer. The input layer enters the problem to be solved into the ANN. provides. The output layer is the layer where the processed information is exported in the network. between two layers hidden layer can be found. Adding all the inputs and weights by multiplying and passing them through the function. The output value of that neuron is calculated. Since the information flow in the MANN model takes place in the forward direction, should feed. It propagates backwards until the error is minimal. Besides these classical algorithms Newton, which uses numerical optimization methods and provides fast convergence in solutions, Levenberg-Marquardt algorithms have also become frequently preferred [8]. You can find more detailed information about the Levenberg-Marquardt algorithm here [9].

2.2.2. Radyal Based ANN

The concept of Radial-Based ANN (RBANN) was introduced to the literature by Broomhead and Lave in 1988. Radial-based functions and ANN model of neuron cells in the nervous system seen in humans. It has been developed considering the action-reaction situations. RBANN models train their training in multidimensional space. It is possible to see it as a curve fitting approach [10-12]. Thus, the training performance of the RBANN sample, the output vector Finding the closest result to the data in the space and thus turns into an interpolation problem [10]. RBANN structure input layer, hidden It consists of a layer and an output layer. However, unlike other ANNs, it is from the input layer. When passing to the hidden layer, the data is transferred to radial basis activation functions and a nonlinear set. subjected to analysis. The structure between the hidden layer and the output layer is like other ANN types. and the actual training takes place in this layer.

2.2.3. Generalized Regression ANN

The generalized regression neural network proposed by Specht [13] is in the back propagation method. It does not require an iterative training procedure. GRANN used training data It estimates any function between the input and output vectors. As the training set expands estimation error decreases to zero [14]. Regression, x, and training as it is known by definition estimating the most probable value of a dependent variable y based on the independent variable x, given the set it does. The regression method estimates y to minimize the common squared error. GRANN, a It is a method that estimates the joint probability density function of x and y given the training set. Since the probability density function is obtained from the data without making any pre-acceptance, the system is generally ideal [14].

3. Applications

As comparative metrics, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Coefficient of Determination were used. Equation 1-3 describe related equations.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|CO_e - CO_o\right| \tag{1}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(CO_e - CO_o)^2} \tag{2}$$

$$R^2 = \left(\frac{N*\left(\sum CO_o * CO_e\right) - \left(\sum CO_o\right)*\left(\sum CO_e\right)}{\sqrt{\left(N*\sum CO_o^2\right) - \left(\sum CO_e\right)^2 * \left(N*\sum CO_e^2\right) - \left(\sum CO_e\right)^2}}\right)^2 \tag{3}$$

In the equations, “CO_e” and “CO_o” show the estimated and observed elevation values and “N” represents the sample number.

Table 2. Training results

Criteria	Methods	Input Number				
		1	2	3	4	5
MAE (Training)	MANN	181.666	115.479	119.393	115.825	115.213
	RBANN	178.469	111.864	112.055	99.329	423.427
	GRANN	7.411	7.281	104.701	91.942	174.252
RMSE (Training)	MANN	279.622	164.681	165.080	166.536	165.696
	RBANN	272.461	158.695	156.525	166.536	165.696
	GRANN	8.603	8.488	145.860	127.395	221.364
R^2 (Training)	MANN	0.876	0.957	0.957	0.956	0.957
	RBANN	0.883	0.960	0.961	0.970	0.474
	GRANN	0.026	0.060	0.967	0.975	0.934

Table 3. Test results

Criteria	Methods	Input Number				
		1	2	3	4	5
MAE (Test)	MANN	170.781	141.939	176.423	79.309	74.206
	RBANN	169.079	138.491	145.816	112.288	263.051
	GRANN	8.271	8.343	129.755	90.172	137.007
RMSE (Test)	MANN	218.841	191.428	226.142	96.259	90.361
	RBANN	218.461	188.240	188.885	135.072	289.460
	GRANN	9.381	9.465	208.476	131.549	164.048
R^2 (Test)	MANN	0.125	0.379	0.372	0.792	0.824
	RBANN	0.121	0.381	0.326	0.834	0.016
	GRANN	0.067	0.122	0.339	0.657	0.400

When the results are examined; It is seen that the highest coefficient of determination, R^2 is in RBANN with 4 inputs (MAE=112.288, RMSE=135.072, R^2=0.834), but the lowest error values are seen in the MANN method with 5 input (MAE=74.206, RMSE=90.361, R^2=0.824) for RMSE and MAE. It is seen that R^2 is quite low, RMSE and MAE are high for all methods in 1, 2 and 3 inputs. It is seen that the GRANN method gives low R^2 and high error compared to other methods, but it is observed that it gives the best result in 4 inputs (MAE=90.172, RMSE=131.549, R^2=0.657).

The graph of the most successful result of each method is given in Figure1, 2 and 3.

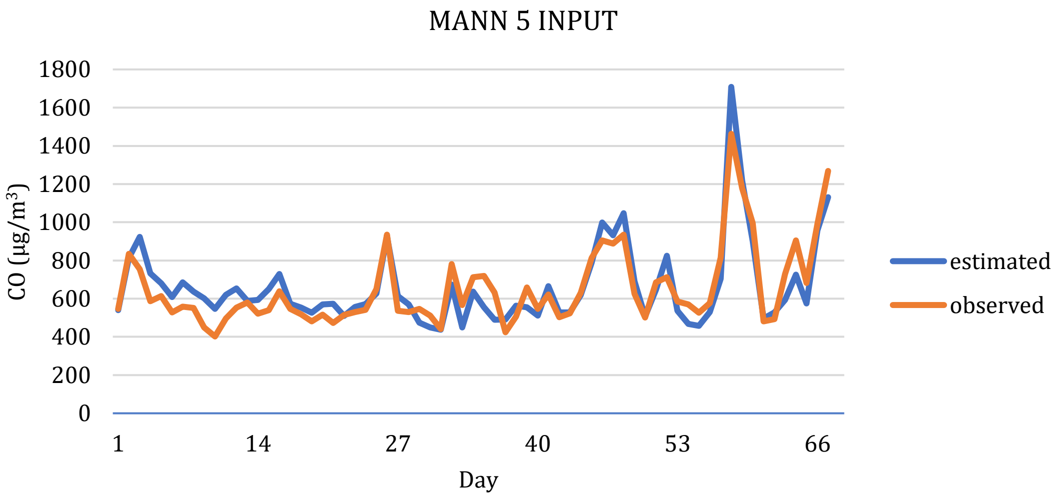

Figure 1. MANN 5 Input test results graph

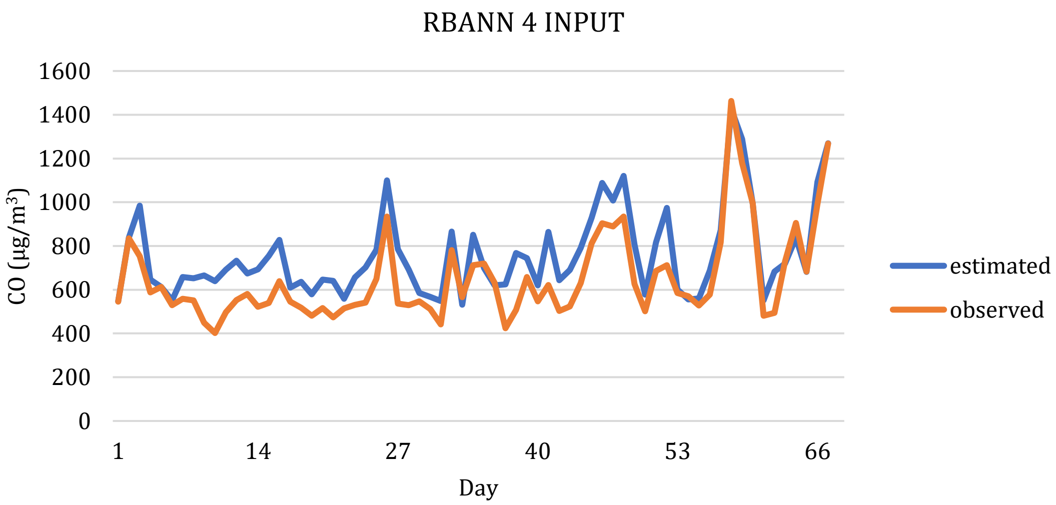

Figure 2. RBANN 4 Input test results graph

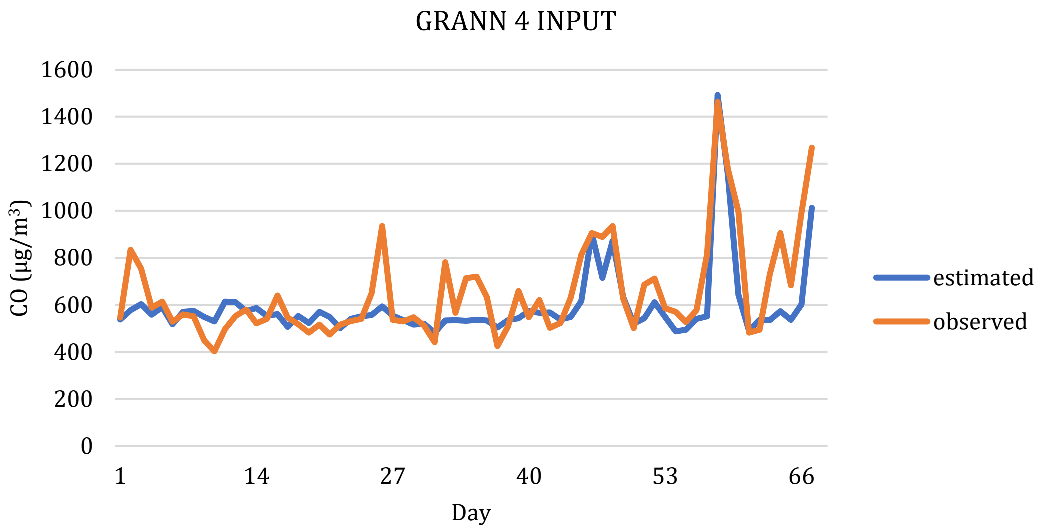

Figure 3. GRANN 4 Input test results graph

When the graphs are examined, estimates closer to the observed values were observed in the MANN method. When the results are compared with the previous study by Çubukcu et al. [15] using fuzzy logic methods (MAE=96.797, RMSE=115.972, R^2=0.824), the MANN method is more successful.

4. Conclusion

In this study, the amount of CO in the air (μg/m³) was estimated for Meram. It was estimated according to the daily data of Meram air quality station in the central district of Konya. In the model, the parameters are primary air polluting gases. The data has been kept up to date for one year. Three different ANNs were used in the modeling. These are MANN, RBANN and GRANN.

Although the lowest RMSE and MAE in the training results are at the 1 and 2 inputs of the GRANN method, it is seen that it gives a very low result in terms of the coefficient of determination and cannot form an accurate model. However, when the 4-input model of GRANN is examined, it is observed that the training result is the best (RMSE=127.395, MAE=91.942, R^2=0.975). Although the training results of RBANN and MANN methods gave better results than GRANN from the 1st input, they could not achieve the best result.

When the test results are compared, it is seen that the best estimation model was created with the MANN method in 5-input (RMSE=90.361, MAE=74.206, R^2=0.824). Then the best result is in the 4-input model by GRANN (RMSE=135.072, MAE=112.288, R^2=0.834).

Periodicity is included in the algorithm as the 5th input in the models, but it has been observed that it does not make a significant improvement in the results.

Acknowledgement

This study was partly presented at the 2nd Advanced Engineering Days [15].

Funding

This research received no external funding.

Author contributions

Esra Aslı Çubukçu: Methodology, Software **Vahdettin Demir:** Writing-Original draft preparation, Editing. **Mehmet Faik Sevimli:** Last Editing.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Mills, I. C., Atkinson, R. W., Kang, S., Walton, H., & Anderson, H. R. (2015). Quantitative Systematic Review Of The Associations Between Short-Term Exposure To Nitrogen Dioxide And Mortality And Hospital Admissions. BMJ Open, 5(5). https://doi.org/10.1136/bmjopen-2014-006946
2. Saraçoğlu, H. (2010). Investigation Of Exhaust Gas Emissions of Ships Calling Izmir Port and Their Environmental Impacts. Master Thesis. Istanbul Technical University, Istanbul, Turkey.
3. Ünal, Z. F., Dinç, U., Özen, C., & Toros, H. (2019). Air Pollution Forecasting for Ankara with Machine Learning Method. Journal of Research in Atmospheric Science, 1(1), 42–48.
4. Tunç, D. Ç., & Toros, H. (2021). The impact of COVID-19 measures on air quality in Turkey. Journal of Research in Atmospheric Science, 2(2), 46–50. https://doi.org/10.1080/15275922.2021.1892876
5. Kara, Y., Karakaya, T., Pirselimoğlu, G., Dursun, Ş., & Toros, H. (2020). Overall Evaluation of NO_X, NO, NO_2 Gasses in Turkey and Their Data Quality Control. Research Article Journal of Research in Atmospheric Science, 2(1), 12–16. http://resatmsci.com/
6. Al-bayati, R. M., Bulut, B., Adeeb, H. Q., & Toros, H. (2021). Air Pollution Data Analysis Over Van City, Turkey. Journal of Research in Atmospheric Science, 3(1), 8–12.
7. https://www.havaizleme.gov.tr
8. Okkan, U., & Mollamahmutoğlu, A. (2010). Modeling Of Daily Flows of Yigitler Stream with Artificial Neural Networks and Regression Analysis. Dumlupınar University Journal of Science Institute, 23, 33–48.
9. Çavuşlu, M. A., Becerikli, Y., & Karakuzu, C. (2012). Hardware Implementation of Neural Network Training with Levenberg-Marquardt Algorithm. Turkish Informatics Foundation Journal of Computer Science and Engineering, 5(5), 1–7
10. Okkan, U., & Dalkılıç, H. Y. (2012). Monthly Runoff Model for Kemer Dam with Radial Based Artificial Neural Networks. IMO Technical Journal, 5957–5966.
11. Partal, T., Kahya, E., & Cığızoğlu, K. (2008). Estimation Of Precipitation Data Using Artificial Neural Networks and Wavelet Transform. ITU Journal of Engineering, 7(3), 73–85.
12. Poggio, T., & Girosi, F. (1990). Regularization Algorithms for Learning That Are Equivalent to Multilayer Networks. Science, 247(4945), 978–982. https://doi.org/10.1126/science.247.4945.978
13. Sürel, A. (2006). The Use of Generalized Regression Neural Network In Water Resources Engineering. Master Thesis. Istanbul Technical University, Istanbul, Turkey.
14. Alp, M., & Cığızoğlu, K. (2004). Modelling Rainfall-Runoff Relation Using Different Artificial Neural Network Methods. ITU Journal of Engineering, 3(1), 80–88.
15. Çubukçu, E. A., Demir, V., & Sevimli, M. F. (2022). Carbon monoxide forecasting with air quality parameters and fuzzy logic for Konya: A case study of Meram. *Advanced Engineering Days (AED)*, *2*, 65-68.