Proceeding Book of ISESER 2021

O 17. FORECASTING AIR POLLUTANT INDEX (API) USING NONLINEAR AUTOREGRESSIVE (NAR) NEURAL NETWORK DURING COVID-19 PANDEMICS IN MALAYSIA

Samsuri Abdullah[1,2*], Nurul Adyani Ghazali[1], Amalina Abu Mansor[3], Ku Mohd Kalkausar Ku Yusof[3], Nazri Che Dom[4], Ali Najah Ahmed[5], Marzuki Ismail[2,3]

[1]*Faculty of Ocean Engineering Technology and Informatics, Universiti Malaysia Terengganu, Kuala Nerus, 21030, Malaysia*
[2]*Institute of Tropical Biodiversity and Sustainable Development, Universiti Malaysia Terengganu, Kuala Nerus, 21030, Malaysia*
[3]*Faculty of Science and Marine Environment, Universiti Malaysia Terengganu, 20130, Kuala Nerus, Terengganu, Malaysia*
[4]*Faculty of Health Sciences, Universiti Teknologi MARA, UiTM Cawangan Selangor, 42300 Puncak Alam, Selangor, Malaysia*
[5]*Institute of Energy Infrastructure (IEI), Department of Civil Engineering, College of Engineering, Universiti Tenaga Nasional (UNITEN), Kajang, Selangor Darul Ehsan 43000, Malaysia*

E-mail: *samsuri@umt.edu.my*

ABSTRACT: Coronavirus Disease 2019 (COVID-19) pandemics have emerged in Malaysia since 18 March 2020, which then the government has announced for Movement Control Order (MCO) as a method to curb the transmission in public. The air quality is expected to be good as most of the operations are closed. Thus, we evaluated and predicted the Air Pollutant Index (API) during the MCO in Malaysia for an overview of the air quality level during the pandemic. As the API is complex in the atmosphere, we used a nonlinear autoregressive (NAR) neural network model for the nonlinear dataset. Urban cities are generally having higher pollutants concentrations along with the urbanization process. High pollutant concentrations led to health problems, especially respiratory illness, either in the short or long term. We used the data from 18 March 2020 (the first day of Movement Control Order, MCO) until 31 December 2020. Results revealed the NAR model executed higher R^2 for Kuala Terengganu (99.23%). The optimum NAR model architectures which are trained using the Levernberg-Marquardt training algorithm is 1:14:1 for Kuala Terengganu. NAR neural network is capable of modeling and forecasting nonlinear time series during the COVID-19 pandemic.

***Keywords**: COVID-19, Malaysia, Nonlinear Autoregressive, Movement Control Order, Levernberg-Marquardt*

INTRODUCTION

Urban air quality has caused many respiratory diseases due to air pollutants interactions with humans (Duan et al. 2020). In line with that, the Air Pollutant Index (API) is used as a communication medium to evaluate the risk, especially towards sensitive groups such as the children and elderly in the ambient air. Urban air quality is associated with respiratory problems such as asthma and bronchiolitis in children (Ortega-Garcia et al. 2020) and the elderly (Karimi and Shokrinezhad, 2021). Malaysian Department of Environment publicly displayed the API for five categories. API of 0-50 is considered good, 51-100 is moderate, 101-200 is unhealthy, 201-300 is very unhealthy, and above 300 is hazardous (DOE, 2021). Malaysia recorded the number of confirmed cases since January 2020, and the cases are escalating until the Malaysian government announced the Movement Control Order (MCO) to be implemented starting 18 March 2020 (Aziz et al. 2020) as an early step to reduce the SARS-CoV-2 infections (Shah et al. 2020). During this time, only essential services are allowed to operate. The reduction of air pollutions has happened chiefly all around the world. Specifically, in Malaysia, Abdullah et al. (2020) found that the decrease in $PM_{2.5}$ is up to 58.4% during MCO (lockdown). Othman and Latif (2021) added that the concentrations of air pollutants are declining during the lockdown period. The world also witnessed the reduction of air pollutants in the atmosphere in Morocco (Khomsi et al. 2020), New York (Perera et al. 2021), and Iraq (Hashim et al. 2021). Machine learning and artificial intelligence had to get significant attention for air quality forecasting. The intelligence concept can capture the nonlinearity and complexity of the air pollutants in the atmosphere and vast data (Liu et al., 2021). Previously, the linear time series model for air pollutants was used to model linear time series (Zhang et al. 2018). API time series is known as the sequence of API on a time basis. On the other hand, nonlinear approaches, such as the nonlinear autoregressive (NAR) neural network, are a powerful tool for time series forecasting. Wu and Lin (2019) and Zhang et al. (2020) proved that neural network models could predict AQI in China. Janarthanan et al. (2021) further verified that the neural network is a compelling model for the prediction of AQI. Currently, Li et al. (2021) have successfully analyzed the time series of AQI before and during COVID-19 lockdown in Shanghai with human activities. The results can help in decision making and forecasting for the guidance of authority. This study aims to establish the NAR neural network models in an urban city during the MCO in Peninsular Malaysia.

MATERIALS AND METHODS

The status of air quality in Malaysia is displayed on an hourly basis by the Malaysian Department of Environment (DOE) via the Air Pollutant Index (API). There are six criteria pollutants measured, including fine particulate matter ($PM_{2.5}$), coarse particulate matter (PM_{10}), sulphur dioxide (SO_2), nitrogen dioxide (NO_2), carbon monoxide (CO), and ground-level ozone (O_3). Before the API execution, the sub-index for each criteria pollutants are calculated, and the maximum sub-index is considered the API, showing the status of air quality at that particular area. The monitoring was under the concession of DOE and Transwater Sdn. Bhd. covers urban, suburban, industrial, and background stations. The stations are known as Air Quality Monitoring Station (AQMS). In this study, we used the AQMS located in the urban area, as urban areas are known as the areas composed of high pollutants concentrations. The selected AQMS is Kuala Terengganu (East region). We collect the API data in this study at the website of DOE (http://apims.doe.gov.my/public_v2/api_table.html). We used the data from 18 March 2020 until 31 December 2020, as 18 March 2020 is the Movement Control Order (MCO) starting day in Malaysia. Artificial Intelligence (AI) is a subset of computer science that tends to develop several computer programs in demonstrating intelligibility. Artificial Neural Network (ANN) is one of the crucial groups under AI, commonly used for pattern recognition and function approximation (Cakir and Sita, 2020). ANN rapidly gives realistic solutions and also serves as a universal data estimation with no prior assumptions are made like traditional statistical prediction techniques (Mao et al., 2021). The generalization ability of ANN makes it very useful in function approximations whereby it able to learn from the previous pattern of information without supplying a mathematical model. The ideas of the development of ANN inspired by the human brain (Alonso-Montesinos et al., 2021). The brain-controlled the nervous system of a human, which is composed of massive neurons. Once a neuron communicates with another neuron, it set up a complex connection. The human brain structure, with its outstanding functions assembled to form mathematical and computational models to represent the real-world scenario. Nonlinear autoregressive (NAR) designed to forecast a time series of the past values. We used API data for the time series analysis. The data set is divided into three parts for training (70%), validation (15%), and testing (15%). We optimize the NAR model using the Levernberg-Marquardt (LM) training algorithm for all study areas. The best models for each study area are selected based on the coefficient of determination (R^2). A value close to one indicates the model is adequate for API forecasting. We used MATLAB 2019b to establish NAR models. Equation 1 shows the NAR model.

$$y(t) = b_0 + b_1 y(t-1) + b_2 y(t-2) + \cdots + b_n y(t-n) \quad (1)$$

Where $y(t)$ is the output at time t, and b is the coefficient.

RESULTS AND DISCUSSION

Air Pollutant Index (API) was calculated based on the Pollution Standard Index (PSI), which the United States Environmental Protection Agency (USEPA) accepted the API at the international level (DOE, 2021). API reading generally represented by the concentration of particulate matter ($PM_{2.5}$), which dominated among the criteria pollutants. API status has five categories, which are good (API:0-50), moderate (51-100), unhealthy (101-200), very unhealthy (201-300), and hazardous (>300). Figure 1 shows the percentage of API indication status from 18 March 2020 until 31 December 2020. API indicators showed that the status was good and moderate during the MCO period. Kuala Terengganu showed a good status with 37.02%, while the rest was moderate level. The good and moderate status mainly due to several prohibitions of movement and gathering activities in Malaysian such as travelling abroad, visitors and tourist (tourism) and educational institutions, government and private agencies (except for essential services) closure indirectly reduce the air pollution (Abdullah et al., 2020; Leal and Hernandes., 2020). The implementation of MCO is changing based on the current situations and confirmed COVID-19 cases. The Malaysian government has implemented several restrictions such as Conditional Movement Control Order (CMCO), Recovery Movement Control Order (RMCO), Enhanced Movement Control Order (EMCO), and Semi-Enhanced Movement Control Order. Some relaxation of restrictions allows the non-essential services to operate; thus, no limitation on human mobility. Emissions of air pollutants in Malaysia have mainly come from stationary and mobile sources, deteriorating the air quality. That is the main reason for the variation of API between good and moderate levels during MCO in Malaysia.

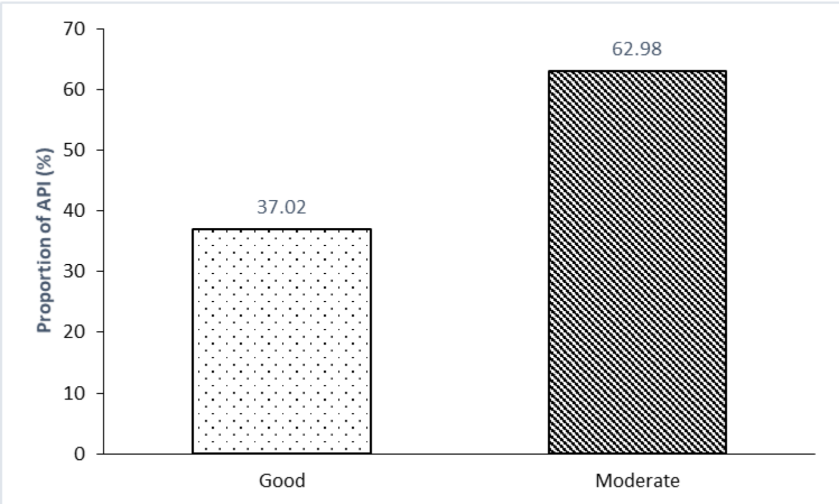

Figure 2. The proportion of Air Pollution Index (API)

We develop the NAR models using trial and error method, added one neuron after another. The number of neurons tested is in the range of 1-15. We trained the NAR model with one hidden layer. Table 1 shows the evaluated results of the NAR model. The ranges of R^2 during training (0.988553 - 0.991080), validation (0.984699 - 0.993351), and testing (0.983171 - 0.992255) for Kuala Terengganu. The bold signifies the best neuron number. The optimum neuron numbers for Kuala Terengganu is 14 (R^2 = 0.99226). Thus, the NAR model explained the variance in the data set with the highest of 99.23% for Kuala Terengganu. Due to the complexity of urban areas, the nonlinearity of the API time-series dataset conclusively is deduced as the high accuracy measure. The NAR model is a compelling model that can capture nonlinear and complex data, as proven in previous studies. The trial and error method in determining the optimum neuron number in the hidden layer is important to avoid any over and less fitting of the NAR model. Optimizing the NAR model via the LM algorithm's adoption helps reduce the bias in the model during training, validation, and testing. LM algorithm is considered the fast convergence that helps train the dataset as it is a combination of gradient descent and the Gauss-Newton method for an optimal solution (Du and Stephanus, 2018).

Conclusively, the overall performance of the NAR model is in Figure 2. Several previous studies on different fields had successfully utilized the NAR model for time series forecasting. Saba and Elsheikh (2020) predicted the COVID-19 outbreak in Egypt and executed it with a high R^2 of 0.999, using the Levernberg-Marquardt training algorithm. The COVID-19 prediction is also established by Khan and Gupta (2020), revealing a high R^2 of 0.97 using the NAR model. Sunayana et al. (2021) successfully developed the NAR model for the monthly prediction of SWM with high accuracy in the waste management field. Figure 6 shows the regression equations for NAR models, and the equation used for time series forecasting is in Equation (2).

Table 2. Coefficient of Determination (R^2) of NAR Model

Neuron Number	Training	Validate	Testing
1	0.97682	0.99176	0.98682
2	0.97909	0.99088	0.98885
3	0.97297	0.99082	0.98881
4	0.97524	0.99082	0.98913
5	0.98002	0.98925	0.98861
6	0.97881	0.98470	0.99040
7	0.97373	0.99289	0.99076
8	0.97127	0.98911	0.99154
9	0.97543	0.98859	0.98847
10	0.97721	0.98921	0.98317
11	0.97599	0.99335	0.99086
12	0.97780	0.99235	0.99212
13	0.97658	0.99233	0.99026
14	0.97119	0.98969	**0.99226**
15	0.97688	0.99034	0.99152

$$Predicted_{S4} = 0.99\ (Observed_{S4}) + 0.36 \quad (2)$$

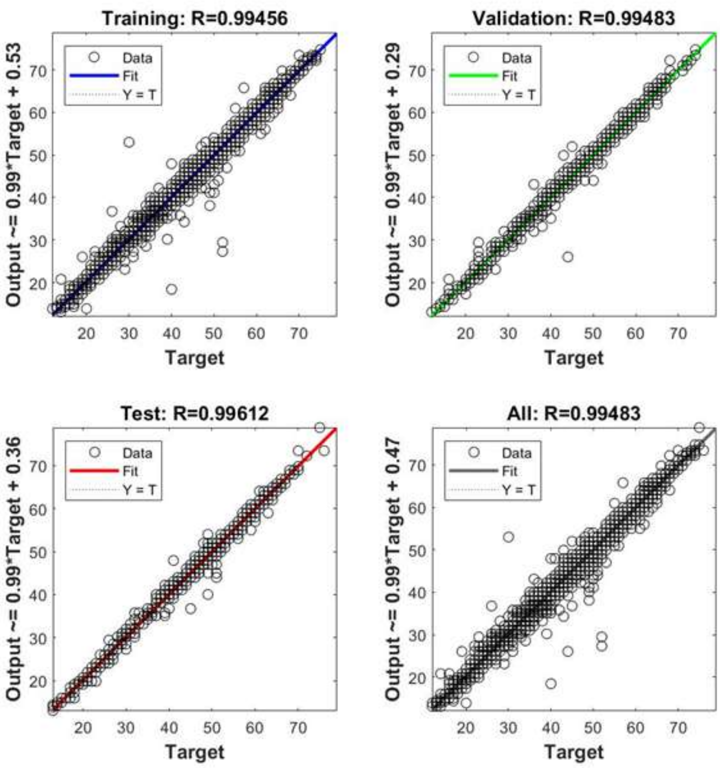

Figure 2. Performance of NAR models during training, validation, and testing

CONCLUSION

This study used one of the time series models of the neural network, the nonlinear autoregressive (NAR), to capture the nonlinear and complexity of Air Pollutant Index (API) data. The NAR models established created explicitly for the API during the COVID-19 pandemic period in Malaysia. The proposed API NAR models showing their superiority by having high accuracy of up to 99% for prediction purposes. An accurate and robust NAR prediction model is necessary to promote public health, especially in urban areas where high pollutants exist. Adopting the NAR models for air quality management is in line with the Sustainable Development Goal-13 on Sustainable Cities and Communities, supporting the aspiration of United Nations 2030.

REFERENCES

Abdullah, S., Mansor, A.A., Napi, N.N.L.M., Mansor, W.N.W., Ahmed, A.N., Ismail, M. and Ramly, Z.T.A. 2020. Air quality status during 2020 Malaysia Movement Control Order (MCO) due to 2019 novel coronavirus (2019-nCoV) pandemic. *Science of the Total Environment.* 729, 139022. https://doi.org/10.1016/j.scitotenv.2020.139022

Alonso-Montesinos, J., Ballestrin, J., Lopez, G., Carra, E., Polo, J., Marzo, A., Barbero, J. and Batlles, F.J. 2021. The use of ANN and conventional solar-plant meteorological variables to estimate atmospheric horizontal extinction. *Journal of Cleaner Production*, 285, 125395. https://doi.org/10.1016/j.jclepro.2020.125395

Ash'aari, Z.H., Aris, A.Z., Ezani, E., Ahmad Kamal, N.I., Jaafar, N., Jahaya, J.N., Manan, S.A. and Umar Saifuddin, M.F. (2020). Spatiotemporal Variations and Contributing Factors of Air Pollutant Concentrations in Malaysia during Movement Control Order due to Pandemic COVID-19. *Aerosol and Air Quality Research,* 20, 2047–2061. https://doi.org/10.4209/aaqr.2020.06.0334

Aziz, N.A., Othman, J., Lugova, H. and Suleiman, A. Malaysia's approach in handling COVID-19 onslaught: Report on the Movement Control Order (MCO) and targeted screening to reduce community infection rate and impact on public health and economy. *Journal of Infection and Public Health,* 13(12), 1823-1829. https://doi.org/10.1016/j.jiph.2020.08.007

Cakir, S. and Sita, M. 2020. Evaluating the performance of ANN in predicting the concentrations of ambient air pollutants in Nicosia. *Atmospheric Pollution Research,* 11(12), 2327-2334. https://doi.org/10.1016/j.apr.2020.06.011

DOE. 2021. *Pengiraan Index Pencemar Udara (IPU*). Retrieved http://apims.doe.gov.my/public_v2/home.html

Du, Y-C. and Stephanus, A. 2018. Levenberg-Marquardt Neural Network Algorithm for Degree of Arteriovenous Fistula Stenosis Classification Using a Dual Optical Photoplethysmography Sensor. *Sensors,* 18(7), 2322. https://doi.org/10.3390/s18072322

Duan, R-R., Hao, K. and Yang, T. 2020. Air pollution and chronic obstructive pulmonary disease. *Chronic Diseases and Translational Medicine,* 6(4), 260-269. https://doi.org/10.1016/j.cdtm.2020.05.004

Hashim, B.M., Al-Naseri, S.K., Al-Maliki, A. and Al-Ansari, N. 2021. Impact of COVID-19 lockdown on NO_2, O_3, $PM_{2.5}$ and PM_{10} concentrations and assessing air quality changes in Baghdad, Iraq. *Science of the Total Environment.* 754, 141978. https://doi.org/10.1016/j.scitotenv.2020.141978

Janarthanan, R., Partheeban, P., Somasundaram, K. and Elamparithi, P.N. 2021. A deep learning approach for prediction of air quality index in a metropolitan city. *Sustainable Cities and Society,* 67, 102720. https://doi.org/10.1016/j.scs.2021.102720

Karimi, B. and Shokrinezhad, B. 2021. Air pollution and the number of daily deaths due to respiratory causes in Tehran. *Atmospheric Environment,* 246, 118161. https://doi.org/10.1016/j.atmosenv.2020.118161

Khomsi, K., Najmi, H., Amghar, H., Chelhaoui, Y. and Souhaili, Z. 2020. COVID-19 national lockdown in morocco: Impacts on air quality and public health. *One Health,* 11, 100200. https://doi.org/10.1016/j.onehlt.2020.100200.

Leal, E.T. and Hernandes, B.E.M. 2020. Association of environmental and meteorological factors on the spread of COVID-19 in Victoria, Mexico, and air quality during the lockdown. *Environmental Research,* 110442. https://doi.org/10.1016/j.envres.2020.110442

Liu, H., Yan, G., Duan, Z. and Chen, C. 2021. Intelligent modeling strategies for forecasting air quality time series: A review. *Applied Soft Computing,* 102, 106957. https://doi.org/10.1016/j.asoc.2020.106957

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. https://doi.org/10.1016/j.scs.2020.102567

Ortega-Garcíaa, J.A., Martínez-Hernández, I., Boldo, E., Cárceles-Álvarezb, A., Solano-Navarro, C., RebecaRamis, R., Aguilar-Ros, E., Sánchez-Solis, M. and López-Hernández, F. 2020. Urban air pollution and hospital admissions for asthma and acute respiratory disease in Murcia city (Spain). *Anales de Pediatría (English Edition),* 93(2), 95-102. https://doi.org/10.1016/j.anpede.2020.01.006

Othman, M. and Latif, M.T. 2021. Air pollution impacts from COVID-19 pandemic control strategies in Malaysia. *Journal of Cleaner Production,* 291, 125992. https://doi.org/10.1016/j.jclepro.2021.125992.

Perera, F., Berberian, A., Cooley, D., Shenaut, E., Olmstead, H., Ross, Z. and Matte, T. 2021. Potential health benefits of sustained air quality improvements in New York City: A simulation based on air pollution levels during the COVID-19 shutdown. *Environmental Research,* 193, 110555. https://doi.org/10.1016/j.envres.2020.110555

Saba, A.I. and Elsheikh, A.H. 2020. Forecasting the prevalence of COVID-19 outbreak in Egypt using nonlinear autoregressive artificial neural networks. *Process Safety and Environmental Protection*, 141, 1-8. https://doi.org/10.1016/j.psep.2020.05.029

Shah, A.U.M., Safri, S.N.A., Thevadas, R., Noordin, M.K., Rahman, A.A., Sekawi, Z., Ideris, A. and Sultan, M.T.H. COVID-19 outbreak in Malaysia: Actions taken by the Malaysian government. *International Journal of Infectious Diseases,* 97, 108-116. https://doi.org/10.1016/j.ijid.2020.05.093

Sunayana., Kumar, S. and Kumar, R. 2021. Forecasting of municipal solid waste generation using nonlinear autoregressive (NAR) neural models. *Waste Management,* 121, 206-214. https://doi.org/10.1016/j.wasman.2020.12.011

Wu, Q. and Lin, H. 2019. Daily urban air quality index forecasting based on variational mode decomposition, sample entropy and LSTM neural network. *Sustainable Cities and Society,* 50, 101657. https://doi.org/10.1016/j.scs.2019.101657

Zhang, K., The, J., Xie, G. and Yu, H. 2020. Multi-step ahead forecasting of regional air quality using spatial-temporal deep neural networks: A case study of Huaihai Economic Zone. *Journal of Cleaner Production*, 277, 123231. https://doi.org/10.1016/j.jclepro.2020.123231

Zhang, L., Lin, J., Qiu, R., Hu, X., Zhang, H., Chen, Q., Tan, H., Lin, D. and Wang, J. 2018. Trend analysis and forecast of $PM_{2.5}$ in Fuzhou, China using the ARIMA model. *Ecological Indicators,* 95, Part 1, 702-710. https://doi.org/10.1016/j.ecolind.2018.08.032