Caspian Journal of Environmental Sciences
Online ISSN: 1735-3866
Print ISSN: 1735-3033

Regression and validation studies of the spread of novel COVID-19 in Iraq using mathematical and dynamic neural networks models: A case of the first six months of 2020

Anees A. Khadom[1*], Ali Khudhair Al-Jiboory[2], Mustafa S. Mahdi[1], Hameed B. Mahood[3]

1. Department of Chemical Engineering, College of Engineering – University of Diyala – Baquba City 32001, Diyala governorate, Iraq
2. Department of Mechanical Engineering, College of Engineering – University of Diyala – Baquba City 32001, Diyala governorate, Iraq
3. Department of Chemical and Process Engineering, Faculty of Engineering and Physical Sciences, University of Surrey, Guildford GU2 7XH, UK

** Corresponding author's E-mail: aneesdr@gmail.com*

ABSTRACT

The dramatic spread of COVID-19 has put the entire world at risk. In this work, the spread of COVID-19 in Iraq has been studied. Due to the increase in the number of positive cases and deaths with this disease, huge pressure acts on the economy and world professionals worldwide. Therefore, building mathematical models to predict the growth of this serious disease is extremely useful. It helps to predict the future numbers of cases in Iraq. Therefore, dynamic neural networks and curve fitting techniques have been developed to construct such a model. Data from the World Health Organization (WHO) are used as a source for mathematical model construction. The period between 25, February to 18, June 2020 was used for regression, validation, and model construction of Dynamic Neural Networks (DNN). Nine samples (19 – 27 June 2020) were used for predicting the future infected and death cases. Descriptive statistical studies showed that the standard deviation varies sharply on June as compared with earlier months of 2020. Three mathematical models are proposed. Linear, polynomials (2^{nd}, 3^{rd}, and 4^{th} orders), and exponential models are used to correlate confirmed infected cases (CIC) and confirmed death cases (CDC) that represent the dependent variables as function of time (independent variable). Nonlinear regression based on least-square method is used to estimate the coefficients of models. Exponential models were the most significant with 0.9964 and 0.9974 correlation coefficients for CIC and CDC, respectively. Validation analysis showed a significant deviation between real and predicted cases using exponential models. However, DNN models showed better response than exponential models. This is evidenced by objective and subjective comparisons. Finally, the CIC and CDC may be increased with time to approach 50000 and 2000 respectively at the end of June 2020.

Keywords: Mathematical modelling, COVID-19, Statistical analysis, Confirmed cases, Neural networks.

INTRODUCTION

Iraq represents one of the countries affected by the spread of COVID-19. Although the spread was very limited at the beginning of 2020, however, recent events showed an elevated level of confirmed infected cases during March, April, May, and June by World Health Organization (WHO 2020). The Iraqi government established very restricted rules for quarantine and control of the people's transportation, which keeps the low level of cases during the early months of 2020. They founded the Higher Committee for Health and National Safety, responsible for the evaluation of the pandemic situation in the country. Unfortunately, the number of confirmed infected cases was increased over time. Regardless of the health, biological issues, and source of COVID-19 that studied elsewhere (Minhas 2020; Acter 2020; Cortegiani 2020; Maliki *et al*. 2021), other branches of knowledge have to

Caspian J. Environ. Sci. Vol. 19 No. 3 pp. 431~440
DOI: 10.22124/CJES.2021.4930
Received: March 26. 2021 Accepted: May 23. 2021
Article type: Research

participate in the war against this serious threat. Zhao *et al.* (2020) used the Maximum-Hasting parameter estimation method and the modified Susceptible Exposed Infectious Recovered model to simulate and predict the epidemic growth in six African countries under different epidemic intervention modeling scenarios. They proposed a series of epidemic control measures. South Africa, Egypt, Algeria are in a severe epidemic situation, while the situation in Nigeria, Senegal, and Kenya are in rapid deterioration leading to more difficulties in preventing the epidemic. Adekunle *et al*. (2020) investigated the spatial density and placed statistical credence to the global deliberation. Spatial variations of clusters that examined the connection between COVID-19 attributable deaths and confirmed cases were constructed. It was found that a ~0.045 COVID-19 deaths as a result of confirmed cases in Africa. Ren *et al*. (2020) studied the communities with confirmed cases from January 21 to February 27 were collected and considered as the specific epidemic data for Beijing, Guangzhou, and Shenzhen. The spatiotemporal variations of the epidemics were evaluated, and then the ecological niche models (ENM) was applied. This model was used to assemble the epidemic data and nine socioeconomic variables for identifying the potential risk of COVID-19 infection zones. The risk zones covered about 75% to 100% of currently infected societies. Bendaif *et al*. (2020) investigated the spread of COVID-19 in Morocco. The data of the Ministry of Health were used as a source for nonlinear regression, residual sum of squares, and Gaussian distribution. Results showed that Moroccan citizens must stay at home as much as possible until June 12, 2020.

Mathematical and statistical analyses have received high attention and they can potentially give a prediction relationship between response (dependent variable) and factors (independent variables). Another important method for establishing input-output relationships is the Artificial Neural Network (ANN, Dua 2011; Rashidi 2012). It denotes one of the artificial intelligence capable of resembling (to some extent) the characteristic of human problem. It solves the problems that is difficult to simulate using expert system analysis techniques and standard software technologies (Moral *et al*. 2008; Fahmi & Cremaschi 2012). Without formal knowledge of an empirical model, ANN can model linear and nonlinear processes in a versatile way. For some applications, this gives artificial neural networks a benefit and advantage over the old-style fitting methods (Khadom *et al*. 2009). Artificial neural network removes the restrictions of traditional methods using input data to exploit the preferred information. In the application of ANN instead of a mathematical equation, sufficient input and output data are required. Data can be used for input and output training and ANN checking. There are several similarities, benefits, and drawbacks between mathematical and ANN modeling available in the literature. Mathematical modeling indicates the simplicity of language, proof, and usage. The type of solutions can be observed and analyzed, and their sensitivity to two or multi-parameter. In our previous work; a mathematical and ANN analyses were applied successfully for industrial and engineering fields (Yaro & Khadom 2010; Khadom *et al.* 2010; Khadom & Yaro 2011; Khadom & Yaro 2014; Hassan *et al.* 2016; Rashid & Khadom 2020).

In the present work, an attempt was made to apply mathematical and ANN models to represent the relationship between the dependent variable (confirmed infected and dead cases), and independent variables (time, t). The study represents the first forecast analysis for the spread of COVID-19 in Iraq using mathematical and ANN analyses. According to our information, this work represents the first research dealing with the spread of COVID-19 in Iraq, which take into account the application of mathematics and ANN.

MATERIALS AND METHODS

Mathematical analysis

Mathematical and statistical analyses were used as a tool to correlate response (dependent variable) and factors (independent variables). Three models were proposed; linear, polynomial (2nd, 3rd, and 4th orders), and exponential models as shown below (Connor 1994):

Linear model (LM)

$$Y = \beta_o + \sum_{i=1}^{n} \beta_i x_i + \epsilon \quad \textbf{(1)}$$

Polynomial model (PM)

$$Y = \beta_o + \sum_{i=1}^{n} \beta_i x_i^n + \epsilon \quad \textbf{(2)}$$

Exponential model (EM)

$$Y = \beta_o exp(\beta_1 x) \quad \textbf{(3)}$$

Caspian J. Environ. Sci. Vol. 19 No. 3 pp. 431~440
DOI: 10.22124/CJES.2021.4930

Received: March 26. 2021 Accepted: May 23. 2021
Article type: Research

where Y is the number of confirmed infected cases and x is the time (days), ε is the standard error, and β_o, β_1, β_2, etc. are constants .The influence of time (from February 25 through June 20, 2020) was studied as a function of confirmed infected cases (CIC) and confirmed death cases (CDC). The data from 21 to 28, June 2020 were used in validation process. The data were obtained from World Health Organization website (WHO 2020), which login on 28, June, 2020. Three mathematical models were proposed (as shown in Eqs. 1 – 3) to represent the CIC data. Linear, polynomial, and exponential models can be expanded and written according to present work notation as follows:

$$CIC = \beta_o + \beta_1\, t \quad \textbf{(1a)}$$

$$CIC = \beta_o + \beta_1 t + \beta_2\, t^2 \quad \textbf{(2a)}$$

$$CIC = \beta_o + \beta_1\, t + \beta_2\, t^2 + \beta_3\, t^3 \quad \textbf{(2b)}$$

$$CIC1 = \beta_o + \beta_1\, t + \beta_2\, t^2 + \beta_3\, t^3 + \beta_4\, t^4 \quad \textbf{(2c)}$$

$$CIC = \beta_o\, exp(\beta_1\, t) \quad \textbf{(3a)}$$

Equations 1a, 2a, 2b, 2c, and 3 represent the linear (LM), 2^{nd}-order polynomial (PM1), 3^{rd}-order polynomial (PM2), 4^{th}-order polynomial (PM3), and exponential (EM), respectively, in terms of present work notation. STATISTICA 7 software was used to estimate the coefficients of models (Eq. 1a – 3). The Levenberg-Marquardt non-linear estimation least squares method was used during estimation method. Maximum number of iteration, convergence criterion and the level of confidence were 1000, 1×10^{-6} and 95% respectively. Similar equations were used to confirm death cases (CDC).

Dynamic Neural Networks (DNN)

Dynamic Neural Networks (DNNs) are networks that contain delays and that operate on a sequence of inputs. These dynamic networks can have purely feedforward connections, or they can also have some feedback (*recurrent*) connections. In other words, they have memory and their response at any given time will depend not only on the current input, but on the history of the input sequence (Jiang & Schotten 2020). Fig. 1 illustrates the general architecture for dynamic neural network that have been used in this work to model the growth of COVID-19 in Iraq. Dynamic neural networks proved to be effective for prediction (Jiang & Schotten 2020) due to automatically extracting relevant features from the training samples. DNN is good at processing data and exhibiting great potential in time-series prediction (Hassan *et al.* 2016) through storing large historical information in its internal state.

RESULTS AND DISCUSSION

Mathematical studies

Descriptive statistical analysis

Table 1 shows the descriptive statistics of CIC for COVID-19 in Iraq during the period from February 25, to June 18, 2020. During the six days of February, only 5 CIC were recorded. The standard deviation value was close to the mean. The number of CIC were increased sharply from 83 (March) to 2437 (June 27). Fig. 2 shows the normal distribution of CIC against the category (frequency of upper limit). It is clear that the category magnitude increased from range 5 during February to range 2600 during June, 2020. Descriptive statistics of confirmed death cases (CDC) were also collected in Table 1. Numbers of CDC were low during February, March, April, and May (average = 5.5). It significantly increased to 122 cases during June. Fig. 3 shows the normal distribution of CDC against the category (frequency of upper limit). Categories of 20 – 23, 30 – 40, and 80 – 90 during June were repeated four times as compared to low categories during earlier months of 2020.

Regression analysis

Mathematical and statistical analyses can be used as a tool for the prediction of CIC as a function of time. Regression of Equations 1a – 3 can be carried out to assess the coefficients of equations. The numerical forms of equations were shown below:

$$CIC = -3588.93 + 127.1\, t \quad \textbf{(1b)}$$

$$CIC = 2794.31 - 214.86t + 3.078\, t^2 \quad \textbf{(2b)}$$

Caspian J. Environ. Sci. Vol. 19 No. 3 pp. 431~440
DOI: 10.22124/CJES.2021.4930

Received: March 26. 2021 Accepted: May 23. 2021
Article type: Research

$$CIC = -2400.52 + 334.67\ t - 9.25\ t^2 + 0.07\ t^3 \quad \textbf{(2c)}$$

$$CIC1 = 361.02 - 316.52\ t + 16.873\ t^2 - 0.291\ t^3 + 0.002t^4 \quad \textbf{(2d)}$$

$$CIC = 7.759\ exp(0.0733\ t) \quad \textbf{(3a)}$$

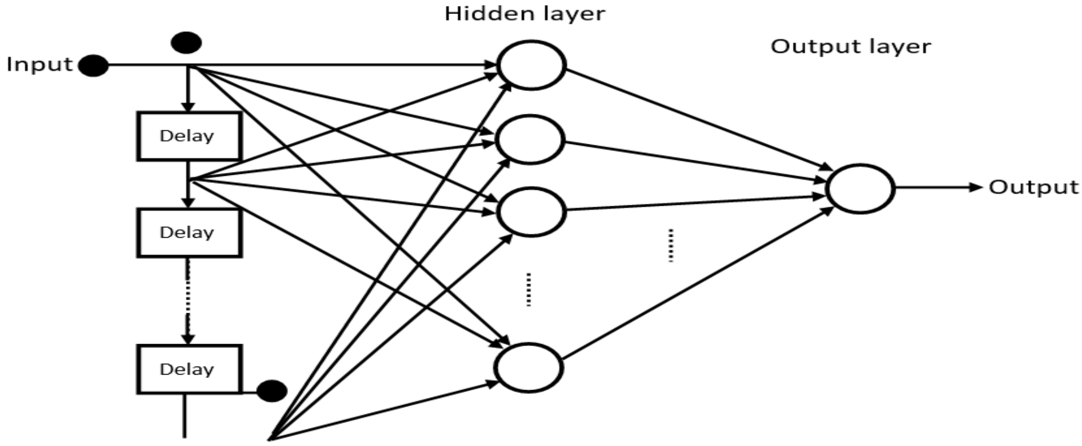

Fig.1: General architecture of dynamic neural network

The correlation coefficients of LM, PM1, PM2, PM3, and EM were 0.7487, 0.9087, 0.9719, 0.9949, and 0.9964, respectively. Fig. 4 illustrates the graphical representation of these models as compared to real data. The best-fitting function was obtained using EM. Similar analysis was repeated to confirmed death cases (CDC). The following equations were obtained:

$$CDC = -82.122 + 3.823\ t \quad \textbf{(2b)}$$

$$CDC = 69.812 - 4.313\ t + 0.073\ t^2 \quad \textbf{(3b)}$$

$$CDC = -78.983 + 11.426\ t - 0.279\ t^2 + 0.0021\ t^3 \quad \textbf{(3c)}$$

$$CDC = 36.8 - 8.618\ t + 0.525\ t^2 - 0.0091\ t^3 + 0.00005\ t^4 \quad \textbf{(3d)}$$

$$CDC = 1.829\ exp(0.053\ t) \quad \textbf{(4a)}$$

The correlation coefficients of LM, PM1, PM2, PM3, and EM were 0.7934, 0.9031, 0.9677, 0.9947, and 0.9974, respectively. As shown in Fig. 5, the best-fitting function was obtained using EM. The CIC and CDC growth were increased over time. Therefore, people's awareness and social distance can be recommended for adaption with COVID-19. Long-term quarantine may lead to many social and psychological problems, in addition to an economical one.

Testing models analysis

The validity of exponential models was investigated from June 19-27, 2020. Figs. 6 and 7 show prediction and residual of CIC and CDC by EM, respectively. The trends of CIC and CDC were identical. Both cases were increased over time. However, the deviation between real data and predicted one was significant. The residual was low, especially, during the days 115, 116, 117, and 118, which equivalent to Wednesday, Thursday, Friday, and Saturday in June, respectively. Within these days, the transportation and movement of people were very limited due to the quarantine and official vacation (Friday and Saturday). In other words, mathematical models were built according to increasing cases over time (exponential growth), while, real cases depend on many other factors, such as State regulations, Ministry of Health activities, etc. On the other side, the residual values were higher during working days (112, 113, 114, and 119) and the predicted values were close to the real one. In general, mathematical models were proven to be powerful in forecasting the trends and behaviour of CIC and

Caspian J. Environ. Sci. Vol. 19 No. 3 pp. 431~440
DOI: 10.22124/CJES.2021.4930

Received: March 26. 2021 Accepted: May 23. 2021
Article type: Research

CDC over time, but it failed in an approximation of CIC and CDC numbers. This may be attributed to many other effective factors which should be taken into consideration.

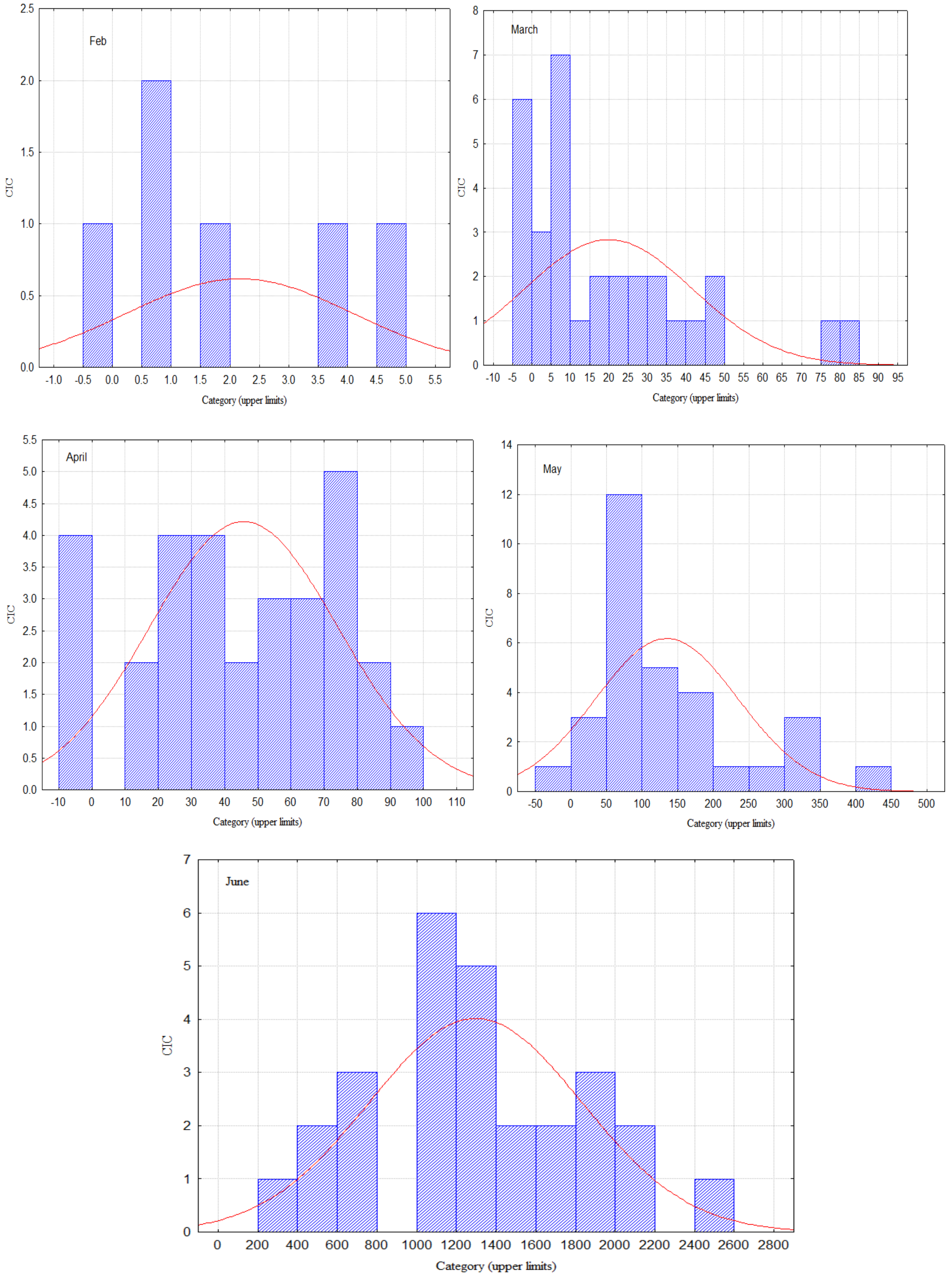

Fig. 2. Normal distribution of CIC during February 25 to June 27, 2020.

Caspian J. Environ. Sci. Vol. 19 No. 3 pp. 431~440
DOI: 10.22124/CJES.2021.4930

Received: March 26. 2021 Accepted: May 23. 2021
Article type: Research

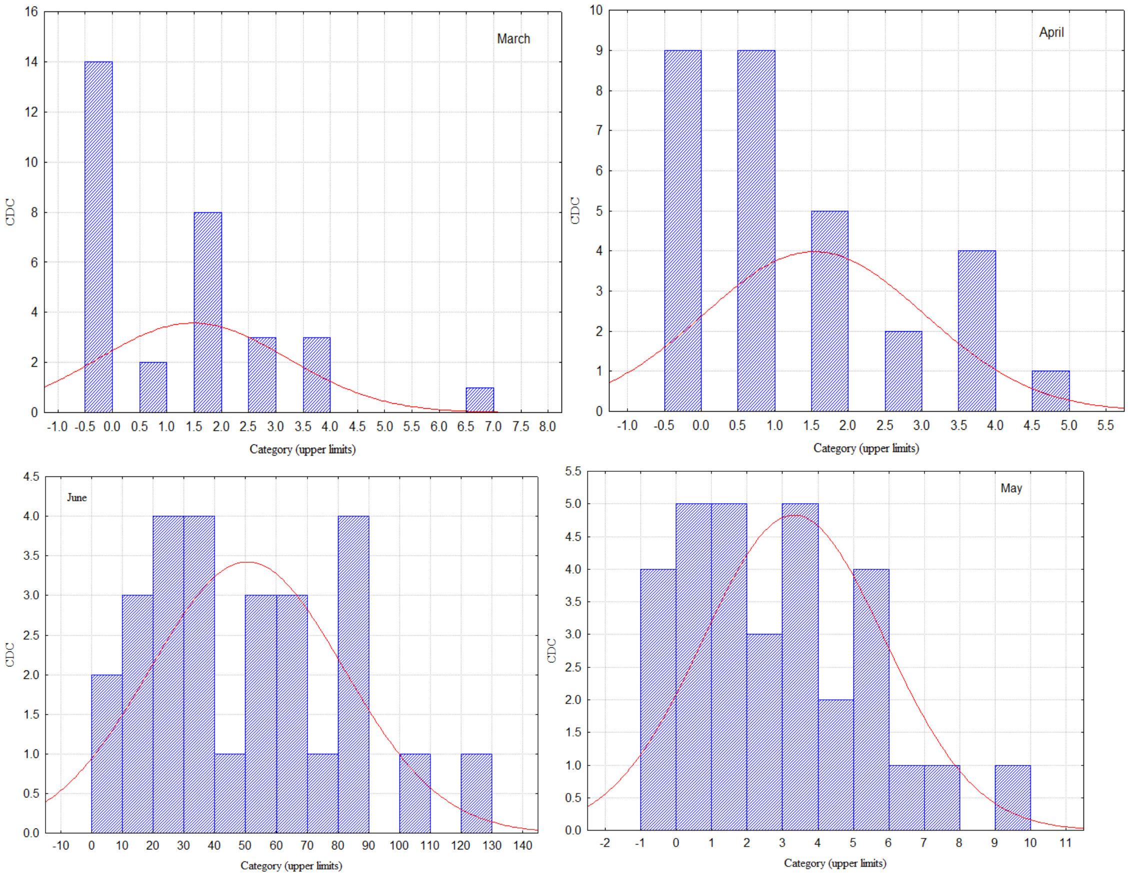

Fig. 3. Normal distribution of CDC during February 25 to June 27, 2020.

Table 1. CIC and CDC descriptive statistics of COVID-19 data in Iraq.

Cases	Months	Days	Mean	Minimum	Maximum	Std. Dev.
	Feb	6	2.1667	0	5	1.9408
	Mar	31	19.9032	0	83	21.8241
Infected	Apr	30	45.7667	0	91	28.3922
	May	31	134.7097	0	416	100.1080
	Jun	27	1296.82	260	2437	536.48
	Feb	6	0	0	0	0
	Mar	31	1.484	0	7	1.729
Death	Apr	30	1.533	0	5	1.502
	May	31	3.322	0	10	2.561
	Jun	27	50.52	10	122	31.456

DNN Model Predictions (CIC)

After finalized training, validation of the proposed model is crucial. Fig. 9 shows the actual confirmed cases (CIC) and network predictions for the entire 124 days. Clearly, the network prediction curve has exponential nature similar to the actual growth. As have been mentioned earlier, the data set have been divided into two sets. The first 115 days are used for training, while the last 9 days are used to validate the ability of the model to do ahead predictions. The vertical dash-dotted line (corresponds to 19-June) indicates the end of training and the start of the ahead predictions. As can be seen from Fig. 9, network predictions are remarkably close to the actual CIC,

Caspian J. Environ. Sci. Vol. 19 No. 3 pp. 431~440
DOI: 10.22124/CJES.2021.4930

Received: March 26. 2021 Accepted: May 23. 2021
Article type: Research

although there are some deviations from the actual cases in the last nine samples. Generally, this relatively small gap is not considered significant as can be noted in June, 24 to 27- predictions.

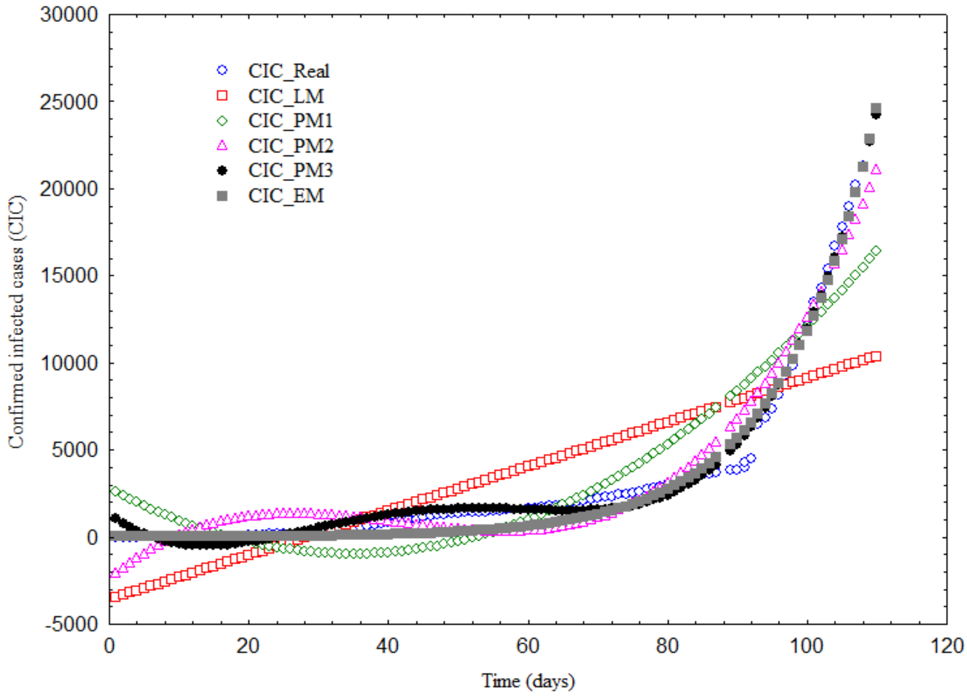

Fig. 4. Mathematical models of CIC as a function of time.

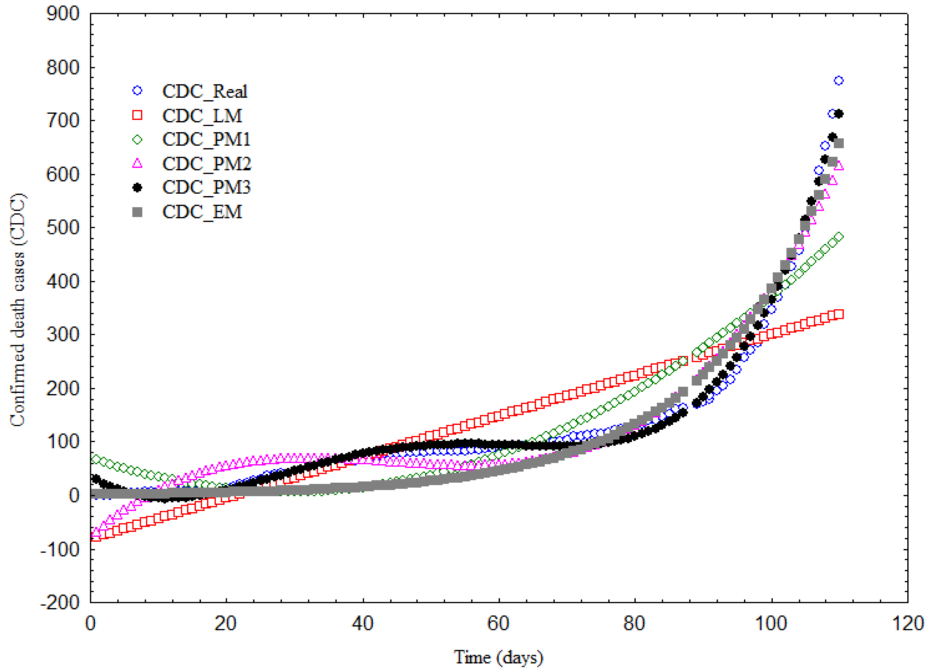

Fig. 5 Mathematical models of CDC as a function of time.

DNN Model Predictions (CDC)

Another dynamic neural network model was developed to predict the growth of the actual confirmed death cases (CDC). Samples (N =119) were used to develop this model from March 1 to June 27, 2020. Similarly, the last nine days were used for validations. After trying many architectures, a network by thirty-five hidden neurons with thirteen delays exhibited good response. The actual CDC and the network predictions are given in Fig. 10. Clearly, the model predictions are quite close to the actual CDC, although there is some deterioration in the response for the validation part. Objective comparisons between DNN and exponential models for CIC and CDC based on the

Caspian J. Environ. Sci. Vol. 19 No. 3 pp. 431~440
DOI: 10.22124/CJES.2021.4930

Received: March 26. 2021 Accepted: May 23. 2021
Article type: Research

root mean squared error for the validation set are illustrated in Table 2. These comparisons indicate that the DNN models are performing better than exponential model in both cases (CIC & CDC).

Table 2. Root mean square errors associated with exponential and dynamic NN model for the two cases CIC and CDC data in Iraq for the nine validation days.

Model	CIC	CDC
Exponential Model	2236.85	428.61
Dynamic NN Model	1978.1	189.99

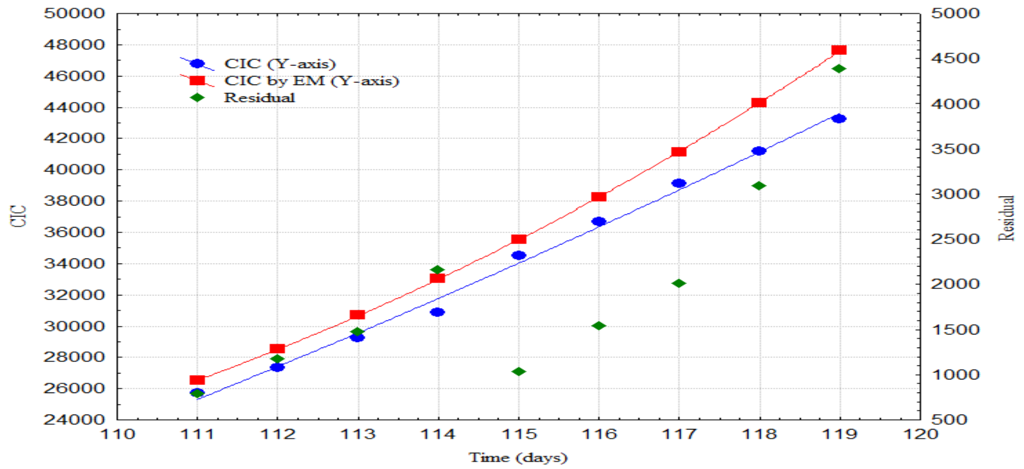

Fig. 6. Prediction of CIC by EM from June 19 - 28, 2020.

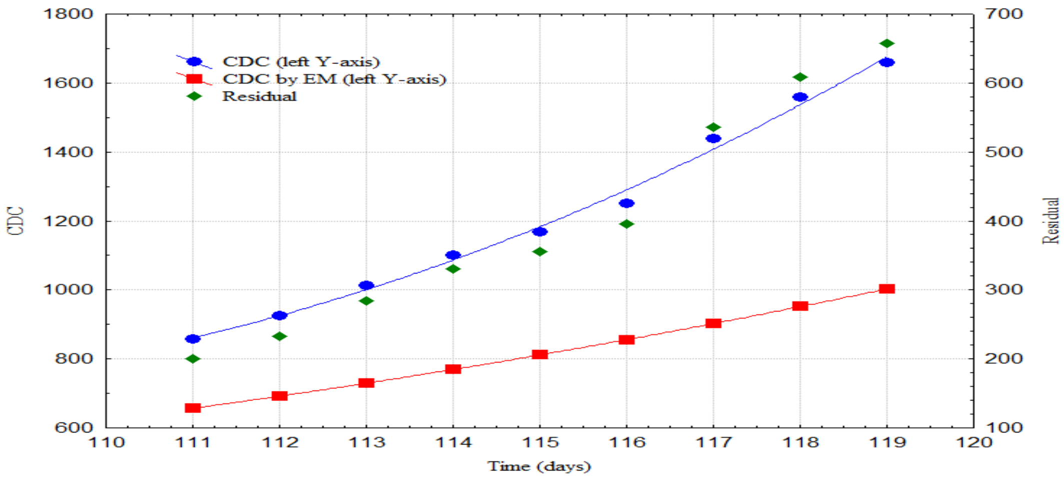

Fig. 7. Prediction of CDC by EM from June 19 - 28, 2020.

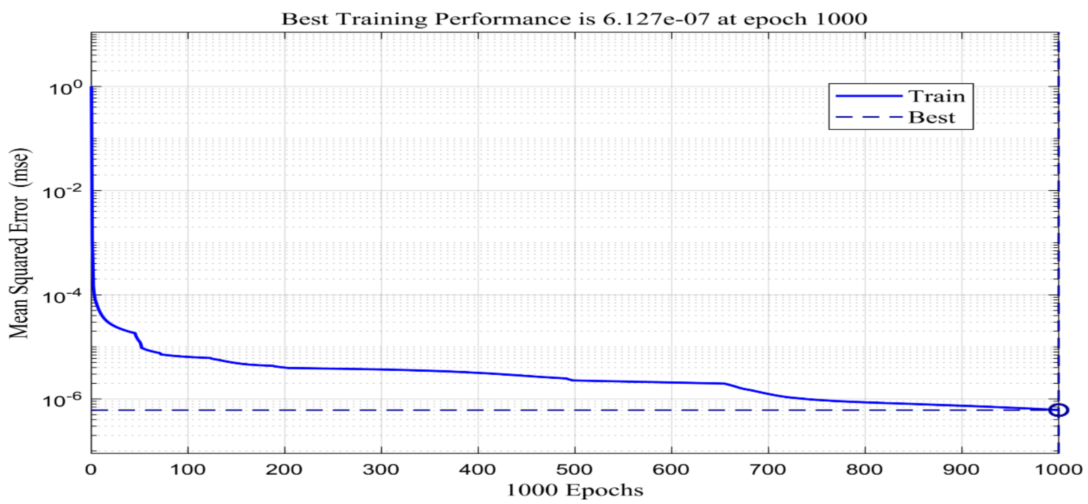

Fig. 8. Training error for the proposed network, 2020.

Caspian J. Environ. Sci. Vol. 19 No. 3 pp. 431~440
DOI: 10.22124/CJES.2021.4930

Received: March 26. 2021 Accepted: May 23. 2021
Article type: Research

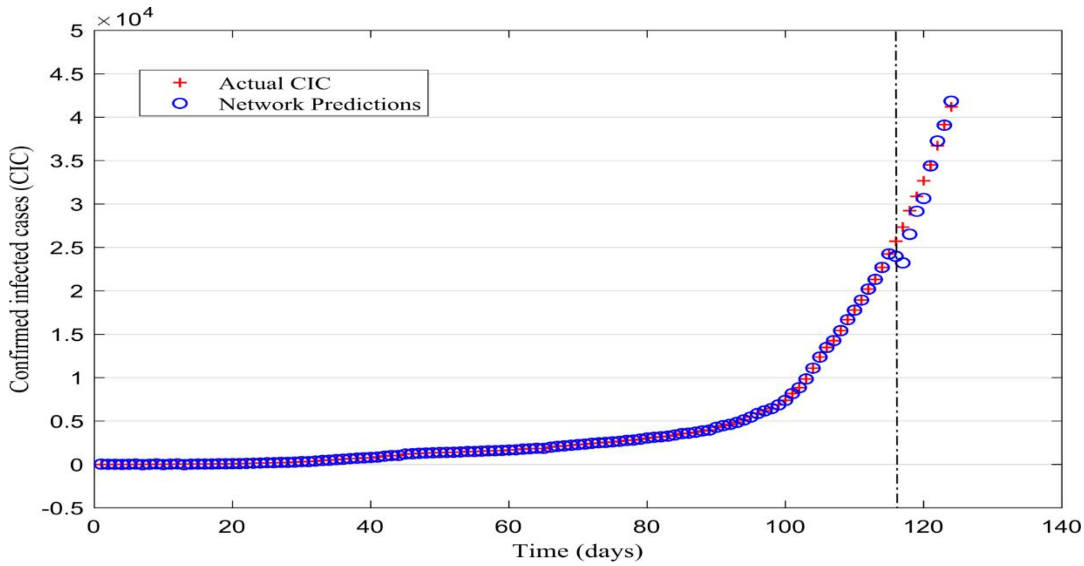

Fig. 9. Prediction of CIC by dynamic NN model from February 25 to June 27, 2020. The last nine days (19 - 27) are used for validation only.

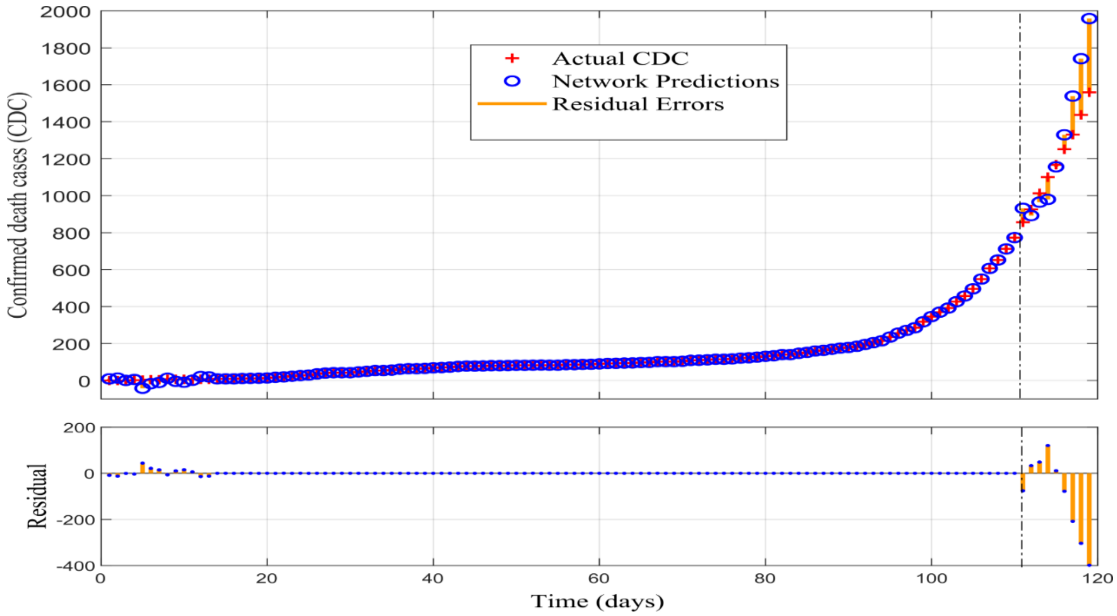

Fig. 10. Prediction of CDC by dynamic NN model from March 1 to June 27, 2020. The last nine days are used for validation only.

CONCLUSION

The spread of COVID-19 in Iraq was studied during the first six months of 2020. Statistical, mathematical, and ANN analyses were used for data fitting and predicting virus growth. Descriptive statistical studies revealed that the standard deviation in June varies sharply as compared to earlier months of 2020. COVID-19 spread in Iraq was studied via mathematical and DNN models. The most important mathematical model was exponential one which used for predicting CIC and CDC with correlation coefficients of 0.9964 and 0.9974, respectively. Validation analysis revealed a large difference using EM. DNN model showed a better response than exponential models.

REFERENCES

Acter, T 2020, Evolution of severe acute respiratory syndrome coronavirus 2(SARS-CoV-2) as coronavirus disease 2019 (COVID-19) pandemic: A global health emergency. *Science of the Total Environment*, 730: 138996.

Caspian J. Environ. Sci. Vol. 19 No. 3 pp. 431~440
DOI: 10.22124/CJES.2021.4930

Received: March 26. 2021 Accepted: May 23. 2021
Article type: Research

Ahmad A, Azid I, Yusof, A & Seetharamu, K 2004, Emission control in palm oil mills using artificial neural network and genetic algorithm. *Computers & Chemical Engineering*, 28: 2709-2715.

Bendaif H, B. Hammouti, I. Stiane, Y. Bendaif, M.A. El Ouadi, El Ouadi, Y 2020, Investigation of spread of novel coronavirus (COVID-19) pandemic in MOROCCO & estimated confinement duration to overcome the danger phase, *Caspian Journal of Environmental Sciences*, 18: 149-156.

Connor, J, Martin R & Atlas, L 1994, Recurrent neural networks and robust time series prediction, *IEEE Transactions on Neural Networks*, 5: 240-254.

Cortegiani, A 2020, A systematic review on the efficacy and safety of chloroquine for the treatment of COVID-19, *Journal of Critical Care*, 57: 279-283.

Dua, V 2011, An Artificial Neural Network approximation based decomposition approach for parameter estimation of system of ordinary differential equations. *Computers & Chemical Engineering*, 35: 545-553.

Fahmi I, Cremaschi, S 2012, Process synthesis of biodiesel production plant using artificial neural networks as the surrogate models. *Comput. Chem. Eng*., 46: 105-123.

Hassan, K, Khadom, A, Kurshed, N 2016, Experimental and Mathematical Studies for Corrosion Reaction of Mild Steel – Sulphuric Acid – Friendly Inhibitor System. *European Journal of Scientific Research*, 139: 163-170.

Hongyan R, Lu Zhao , An Zhang, Liuyi Song , Yilan Liao, Weili Lu , Cheng, C 2020, Early forecasting of the potential risk zones of COVID-19 in China's megacities, *Science of the Total Environment*, 729: 138995.

Ibrahim A, Abayomi Toyin Onanuga, Olanrewaju Olugbenga Akinola, Olakitan, W 2020, Modelling spatial variations of coronavirus disease (COVID-19) in Africa, *Science of the Total Environment*, 729: 138998.

Jiang W, Schotten, H 2020, Deep Learning for Fading Channel Prediction, *IEEE Open Journal of the Communications Society*, 1: 320-332.

Khadom, A, Aprael S Yaro, AS Altaie, AAH, Kadhum 2009, Mathematical modelling of corrosion inhibition behavior of low carbon steel in HCl acid, *Journal of Applied Sciences*, 9:2457-2462.

Khadom, A, Yaro, A 2014, Mass transfer effect on corrosion inhibition process of copper–nickel alloy in hydrochloric acid by Benzotriazole, *Journal of Saudi Chemical Society*, 18: 214-219.

Khadom, A, Yaro, A 2010, Abdul Amir H. Kadhum, application of some basic corrosion equations for copper-nickel alloy in inhibited acid media, *Journal of Tribology and Surface Engineering*, 1: 169-183.

Khadom, A, Yaro, 2011, Modelling of Corrosion Inhibition of Copper–Nickel Alloy in Hydrochloric Acid by Benzotriazole, *Russian Journal of Physical Chemistry A*, 85: 2005-2012.

Maliki, I, Elmsellem, H, Hafez, B, EL, Moussaoui, A, Reda Kachmar, M, Ouahbi, A 2021, The psychological properties of the Arabic BDI-II and the psychological state of the general Moroccan population during the mandatory quarantine due to the COVID-19 pandemic. *Caspian Journal of Environmental Sciences*, 19: 139-150.

Minhas, S 2020, Could India be the origin of next COVID-19 like epidemic? *Science of the Total Environment*, 728: 138918.

Moral, H, Aksoy A, Gokcay, F 2008, Modeling of the activated sludge process by using artificial neural networks with automated architecture screening. *Comput. Chem. Eng*. 32, 2471 – 2478.

Rashid, K.H., Khadom, A.A. 2020, Mathematical Modeling and Electrochemical Behaviour for Corrosion Inhibition of Steel by Kiwi Juice Extract. *J Bio Tribo Corros,* 6:13.

Rashidi, A 2012, A Galvanostatic modelling for preparation of electrodeposited Nanocrystalline coatings by control of current density. *J. Mater. Sci. Technol*. 28: 1071-1076.

World Health Organization (WHO) 2020, website. https://covid19.who.int/region/emro/country/iq. Login on 29, June, 2020.

Yaro, AS, Khadom, AA 2010, Polarisation resistance behaviour of corrosion inhibition of low carbon steel in H3PO4 acid. *International Journal of Surface Science and Engineering*, 4: 429-438.

Zebin Z, Xin, Li, Feng, Liu, Gaofeng, Zhu, Chunfeng, Ma, Liangxu, W 2020, Prediction of the COVID-19 spread in African countries and implications for prevention and control: A case study in South Africa, Egypt, Algeria, Nigeria, Senegal and Kenya, *Science of the Total Environment*, 729:138959.

Bibliographic information of this paper for citing:

A. Khadom, A, Khudhair Al-Jiboory, A, S. Mahdi, M, B. Mahood, H 2021, Regression and validation studies of the spread of novel COVID-19 in Iraq using mathematical and dynamic neural networks models: A case of the first six months of 2020. Caspian Journal of Environmental Sciences, 19: 431-440

Caspian J. Environ. Sci. Vol. 19 No. 3 pp. 431~440
DOI: 10.22124/CJES.2021.4930

Received: March 26. 2021 Accepted: May 23. 2021
Article type: Research