ISSN 1453 – 7303 "HIDRAULICA" (No. 3/2022)
Magazine *of Hydraulics, Pneumatics, Tribology, Ecology, Sensorics, Mechatronics*

Using Hydrological Concepts and an Artificial Neural Network to Model the Rate for COVID-19 Infections versus Deaths

Maritza Liliana ARGANIS JUÁREZ[1,2,*], Margarita PRECIADO JIMÉNEZ[3], Sandra Lizbeth ROSALES SILVESTRE[1]

[1] National Autonomous University of Mexico, Institute of Engineering

[2] National Autonomous University of Mexico, Faculty of Engineering

[3] Mexican Institute of Water Technology, Hydrology Coordination

* Corresponding Author: MArganisJ@iingen.unam.mx

preciado@tlaloc.imta.mx

SRosalesS@iingen.unam.mx

Abstract: *Models were run to reproduce COVID-19 infections versus deaths in Mexico City. The first model was made using rain runoff concept, emulating rain as number of infections reproducing runoff as number of deaths given as of March 2020. The second consisted of using an artificial neural network (ANN) proposed as an initial condition function to be implemented in the model with delay. These models were applied to fit accumulated confirmed case data, obtaining fit corroborated by coefficient of determination, R^2. The R^2 value produced by model was 0.0528 in case of infections comparison vs. official deaths reported by the Ministry of Health, 0.0571 for t case of infections vs. modelling using the HEC-HMS tool, and 0.0937 for case of contagion vs. modelling using ANN.*

Keywords: *Rain-Runoff, Artificial intelligence, infections, deaths, SARS-CoV-2, analogy*

1. Introduction

The COVID-19 pandemic due to the SARSCov-2 virus has persisted for just over two years worldwide. Numerous efforts have been made to obtain models that reproduce contagion behaviour events against deaths in different parts of the world. Basu and Campbell [1] proposed a Long Short-Term Memory (LSTM) based model, used accumulated data on infections and deaths, as a decision support tool for reopening or staggering of activities in a country, utilize data from the Johns Hopkins University repository and made cumulative case predictions for the United States. Dal Molin Ribeiro et al. [2], used regression models to obtain short-term predictions of cumulative covid-19 infections for Brazil.

Understanding population growth phenomena has been a task that over time has provided various challenges to mathematicians, physicists, biologists, medics, economists and many others. From economic areas, where applying growth models to poultry allows making imperative predictions for the profitability of operations, to biological and medical areas, where growth models have been applied to the growth of animals, plants, yeast cells, tumours, and recently to adjust and model COVID-19 pandemic data in order to model phenomena with greater precision, fractional calculus has been implemented in growth models and they were applied to describe cumulative confirmed cases for COVID-19 in Mexico, US and Russia, obtaining an excellent adjustment corroborated by a coefficient determination R^2>0.999 [3].

Since the beginning of pandemic, the effect of COVID-19 has been different throughout the world; however, even with the various measures that each country has taken, the accumulated confirmed cases continue to have a sigmoidal behaviour.

For approximation to description for flow models characteristics, the most conventional hydrological models, it has been divided into three sections that cover (1) the technical and modelling aspects, (2) the enumeration of examples with work done with hydrological models and (3) the description of some widely disseminated models. The models to which reference will be made are those of a semi-distributed and distributed nature, those with the greatest potential and current development [4]. The characteristics of the models are analysed, firstly, taking into account the architecture or way of structuring the modelling process, usually by components of the hydrological cycle, and, secondly, by the modelling approaches. The structure that a model acquires is independent of the representation detail, the number of parameters and the temporal definition used by the hydrological model in question. Currently those of a distributed and/or semi-distributed nature have approached the modelling process by building a modular structure. The conceptualization in components practically forces the models to be built using this modular architecture; that is, dividing the model into different interconnected sub models.

Hydrological simulation models can be of two types: An event model simulates a specific hydrological event: "This downpour would produce this hydrograph." It calculates what part of the precipitation will be net precipitation, and with it calculates the direct runoff that is generated and the rest of the precipitation (abstractions or losses) forgets it. A continuous model attempts to simulate the evolution of the entire hydrological process. It calculates what part of the rainfall is retained on the surface (interception in the vegetation and 'puddles'), what part infiltrates into the ground and what part generates surface runoff. Once the precipitation has passed, it should be considered if the precipitation that was stored in soil evapotranspires or infiltrates into aquifers. Finally, from these they can be lost to a deep circulation (outside the scope of the model) or feed the channels. An event model usually works from a few minutes to several days, while in continuous models' periods from months to several years are common. HEC-HMS was initially a model to simulate specific events, although now it has methods that allow it to be used continuously, as is the case of the application presented in this document.

In this paper, two models will be applied to simulate the process of infections against deaths by COVID-19. The first with a rain runoff model, in the free use program HEC-HMS, calibrating the time of occurrence between infections and deaths (delay time), emulating the rain as infections number to reproducing runoff as deaths number that occurred from March 2020 when the pandemic officially began in Mexico. The second model consisted of using an artificial neural network (ANN) that takes as input data the daily infections number as outputs functions number, preserving the cyclical nature of the phenomenon. The procedure was applied to adjust the data on cases of COVID-19 in Mexico City, obtaining the adjustment.

2. Material and methods

2.1 HEC-HMS

Adjustment parameter used in modelling used by HEC-HMS was implementation of lag effect; global phenomenon was decomposed into its local parts, allowing each outbreak and its different characteristics to be directly compared. The Soil Conservation Service (SCS) dimensionless unit hydrograph procedure is one of best-known methods for deriving synthetic unit hydrographs in use today. References for this method can be found in most hydrology textbooks or manuals. The main reference for this method can be considered as the Soil Conservation Service - National Engineering Manual, Section 4, Hydrology [5].

Dimensionless unit hydrograph used by the SCS was developed by Victor Mockus and was derived based on a large number of unit hydrographs from basins that varied in characteristics such as size and geographic location. Unit hydrographs were averaged and final product was made dimensionless considering the ratios of q/qp (flow rate/peak flow rate) on the ordinate axis and t/tp (time/time to peak) on the abscissa axis, where the units of q and qp are volume/time. This final dimensionless unit hydrograph, which is the result of averaging a large number of individual dimensionless unit hydrographs, has a time to peak located at approximately 20% of its time base and an inflection point at 1.7 times the time to peak final top. The dimensionless unit hydrograph and the cumulative mass curve for dimensionless unit hydrograph are illustrated in Fig. 1.

Curvilinear unit hydrograph can also be represented by an equivalent triangular unit hydrograph. However, time parameter is somewhat difficult to estimate and quite subjective; this parameter has a considerable influence on unit hydrograph values. Underestimating the unit hydrographs "time" will cause peak to occur earlier and higher, while overestimating will result in a later and lower peak. There are several methods to estimate time parameter in UHG. The SCS lag equation is an empirical approach developed by SCS, which estimates lag time directly.

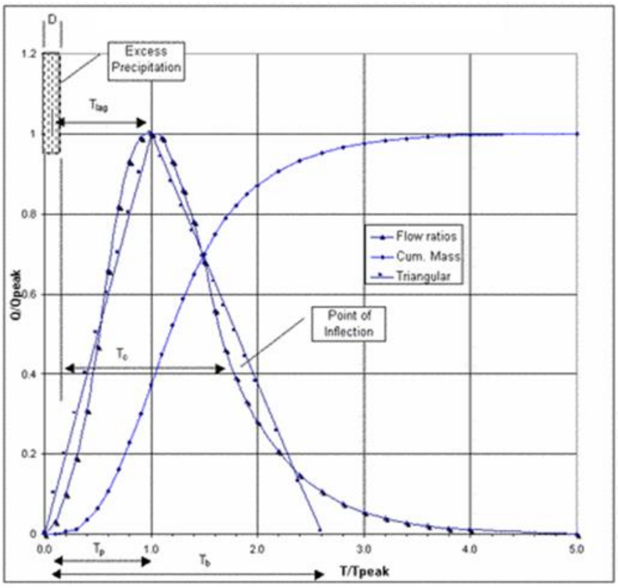

Fig. 1. Dimensionless unit hydrograph of the Soil Conservation Service (SCS)

2.2 ANN

A typical artificial neural network consists of an input (or inputs), a hidden layer of neurons defined by a typically sigmoid function, and a layer with output neurons, traditionally given by a linear function; each layer has weight factors (w) and thresholds or bias (bias, b) and one or more outputs. The neural network goes through the stages of training, validation and testing. The most commonly used algorithm in training is back propagation. In this study, an ANN was used with the configuration given in Fig. 2, from the MATLAB applications [6].

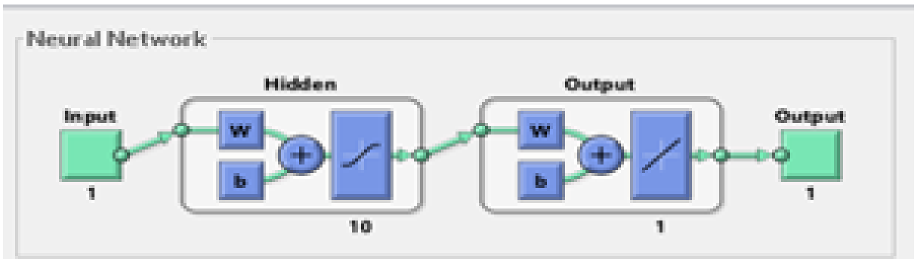

Fig. 2. Configuration of the artificial neural network considered (The MathWorks, 2022)

In this analysis, we propose 10 neurons in hidden layer; a Levenberg-Marquardt backward propagation algorithm for network training, of 646 input data, was considered, too; 646 random samples were considered, occupying 65% for training (Samples that network uses for its training and error is evaluated with these data), 20% for validation (used to measure t network generalization and to stop training when it stops improving said generalization) and 15 % for the test (they do not influence training and give an independent measure of performance for network during and after the training). In this test, the process finished in 12 iterations and in a few seconds.
Additionally, the code of the function created by the neural network can be saved to obtain new deaths from giving the number of cases.

2.3 Delay or lag time

The first case for COVID-19 in Mexico was confirmed by federal government on February 28, 2020, although in its current version, the official database includes a positive case a month before that according to Ministry of Health (SS). One year later, as of mid-February 2021, more than two million infections have been reported, 94.6% of which were confirmed by RT-PCR or antigen

testing. The remaining 5.4% of patients were diagnosed with COVID-19 based on clinical presentation and epidemiological association, in the absence of a valid test result.
In this paper, lag time parameter was proposed where initial function was generalized to model phenomena with recurrent outbreaks. This model was implemented to describe data on confirmed cumulative cases for COVID-19 in Mexico, in the beginning for pandemic until February 2020 to December 25, 2021. Finally, as a result of lag effect, implementation it was possible to break down global phenomenon into its local parts, allowing each outbreak and its respective characteristics to be directly compared.
It is important to note that there is a lag time between date on which a person presents suspected symptoms for COVID-19 and date on which they enter a medical unit to receive care, and therefore be registered into open database. This lag is mainly due to natural development for respiratory disease, which determines t time a person waits to seek medical attention, from time symptoms began. Characterizing lag time is very important to estimate trending the COVID-19 syndrome, since it represents a delay in availability of information that causes an apparent decrease in cases number in the most recent days, which does not correspond with a real decrease in infections. A very important factor in simulating the disease evolution is the time that elapses between infection by virus and symptoms appearance for the disease. Most researches estimated incubation period for COVID-19 range from 1 to 14 days, generally around five days.
Every week a report is issued to the public from the Ministry of Health (SS), in order to provide information on the most up-to-date trend for COVID-19 pandemic by federal entity and by each metropolitan area. These indicators are currently three for ten risk indicators that define the Epidemiological Risk Traffic Light 1 to move towards a new normality, which is a monitoring system for the regulation of public space use of in accordance with the risk for COVID-19 contagion regulated by the SS [7].
Based on lag time analysis, it is possible to determine parameter value such that it corresponds to days that must be discarded at the end of the time series of daily new cases trend for COVID-19 syndrome. Only the delay between the date of symptoms and the date of admission is relevant, which represents the time it takes for a person to seek medical care. In the document entitled Methodology for the analysis of trends in epidemic curves of new cases for COVID-19 syndrome and bed occupancy and deaths registered in the IRAG Network [8], calculation for lag time for Mexico City was 8.85 days, Fig. 3.

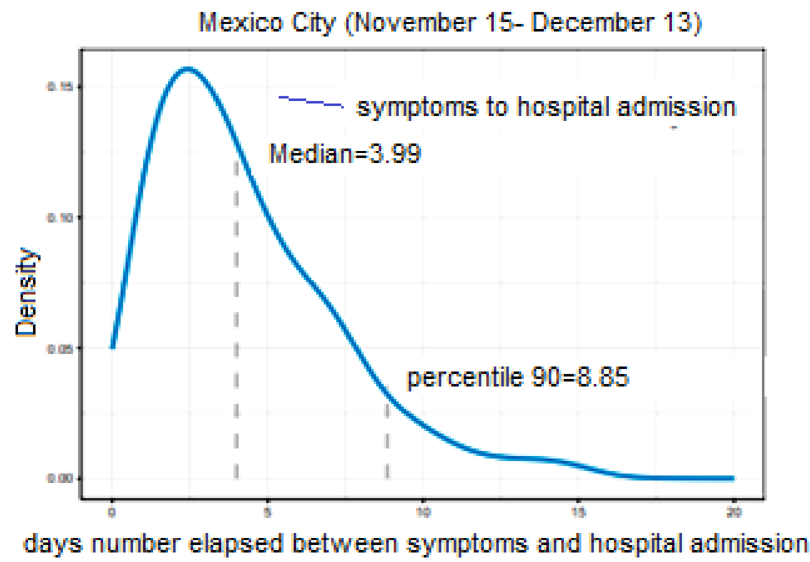

Fig. 3. Estimation between the delay time between infections and deaths in CDMX Source: Methodology for the analysis of trends in epidemic curves of new cases for COVID-19 syndrome and bed occupancy and deaths registered in the IRAG Network

3. Results

To calibrate our model in HEC HMS, calibration of the t_{lag} parameter or lag time was carried out in order to adjust official descent curve with those modeled in HEC-HMS. Delay time used in final modeling was 23,000 minutes. Therefore, lag time used at modeling is 6.38 days.

ANN considered number of infections reported between March 26, 2020 and December 31, 2021 as input variable and deaths data officially reported on those dates in Mexico City as output variable.

The model was applied to adjust accumulated data of confirmed cases in Mexico City, obtaining corroborated adjustment by determination coefficient, R^2, said value gave 0.0528 in case of infections comparison vs. official deaths reported by the Secretary of Health, 0.0571 for case of contagion vs. modeling using HEC-HMS tool and 0.0937 for contagion case vs. modeling using ANN, as shown in figures 4, 5 and 6.

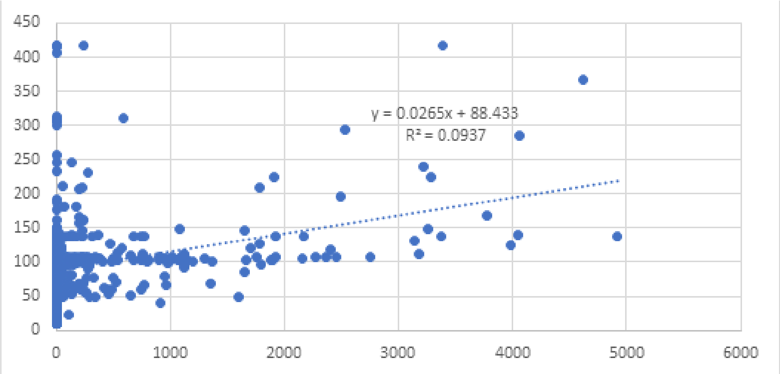

Fig. 4. Infections rate trends (horizontal axis) vs. deaths simulated with ANN (vertical axis)

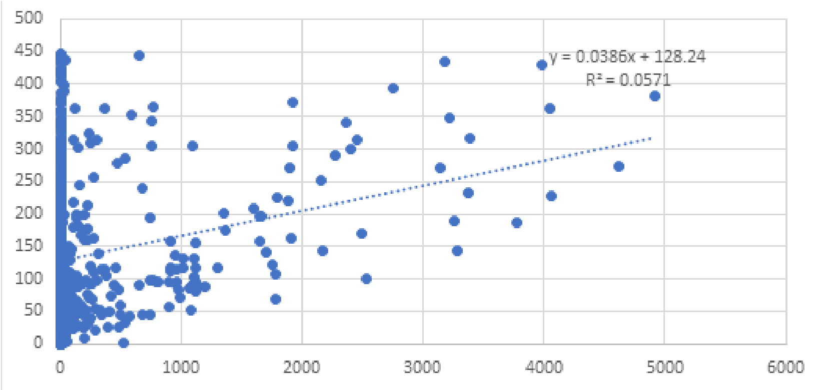

Fig. 5. Infections rate trends (horizontal axis) vs. deaths simulated with HEC-HMS (vertical axis)

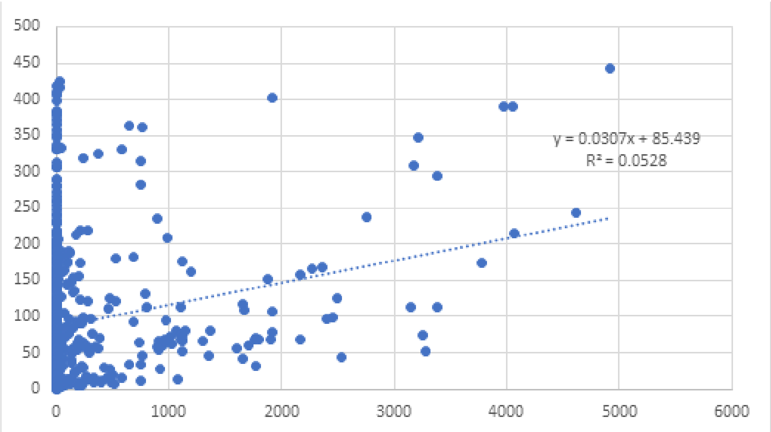

Fig. 6. Infections rate trends (horizontal axis) vs. official deaths (vertical axis)

The vaccination advance for international and national levels was positive altered demographic profile and contagion-hospitalization-death dynamics for pandemic, as one can see in Figure 7. A change in trend is particularly observed at the end of 2021 year. Worth emphasizing, unlike the first two epidemic waves, which are clearly observed in said figure, the growth ratio of the curves of cases, hospitalizations and deaths changed radically in last iteration, which is why the period of time to model it was from the month of February 2020 to December 2021.

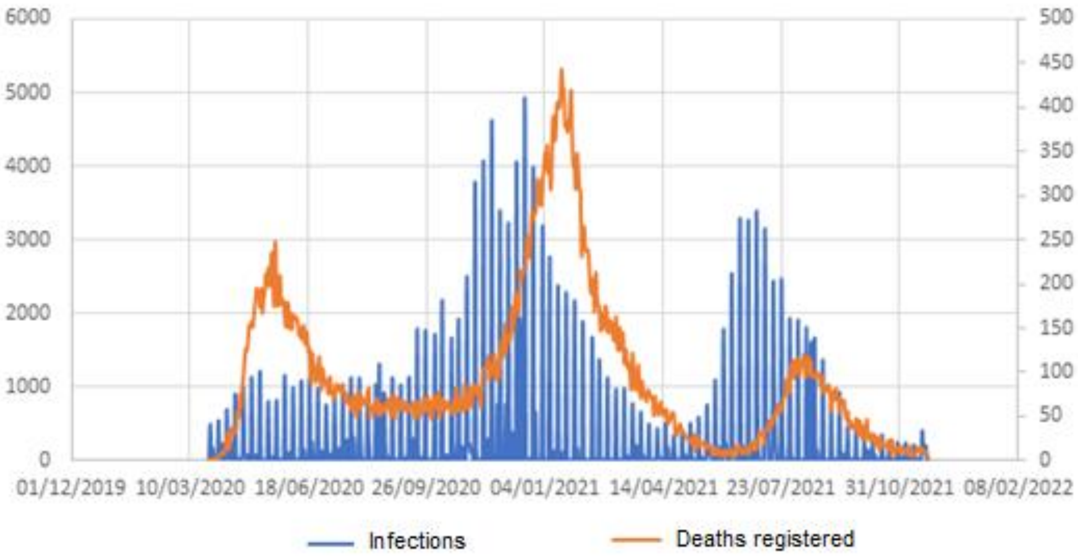

Fig. 7. Infections comparison vs. official registered deaths

Fig. 8 shows the behavior of deaths by considering contagion as an input variable and applying an artificial neural network ANN.

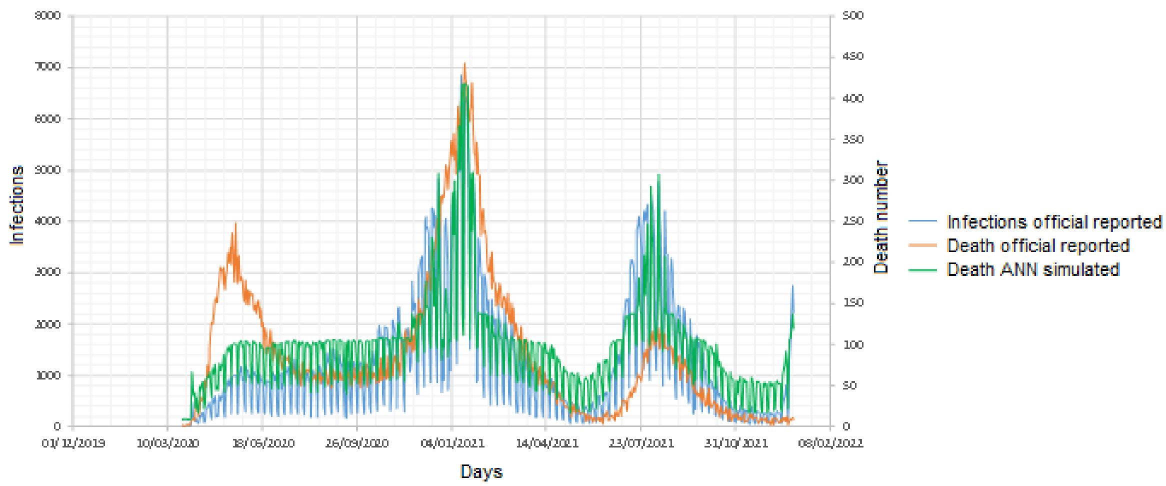

Fig. 8. Infections comparison vs. official and simulated deaths from an ANN

Fig. 9 shows the modeling of the contagion-death process using the HEC-HMS model, which reproduces the behavior of the official contagions reported by the SS, clearly observing the effect of the lag time between the two curves corresponding to the modeled t_{lag} of 6.38 days.

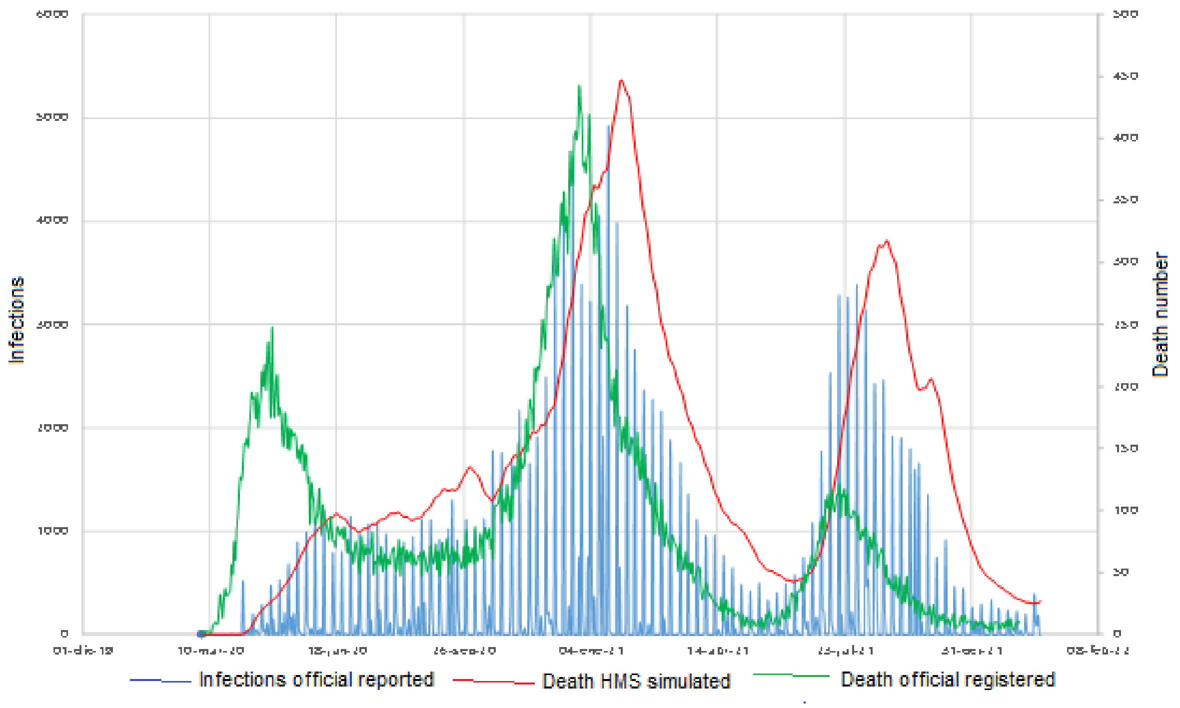

Fig. 9. Infections comparison vs deaths simulated from the HEC-HMS model

Figure 10 shows comparison between official deaths against data modeled from an ANN and what is obtained with the HEC-HMS modeling, observing that in the second wave or peak to HEC-HMS model manages to reproduce the trend and the peak of the officially reported, with an approximate delay of 6 days. On the other hand, the ANN follows the peak times of deaths but, in the third peak, neither of the two models reproduces the number of deaths, but rather the form, which suggests that it is due to the effect of vaccination, which managed to break the upward trend in the number of deaths from COVID-19.

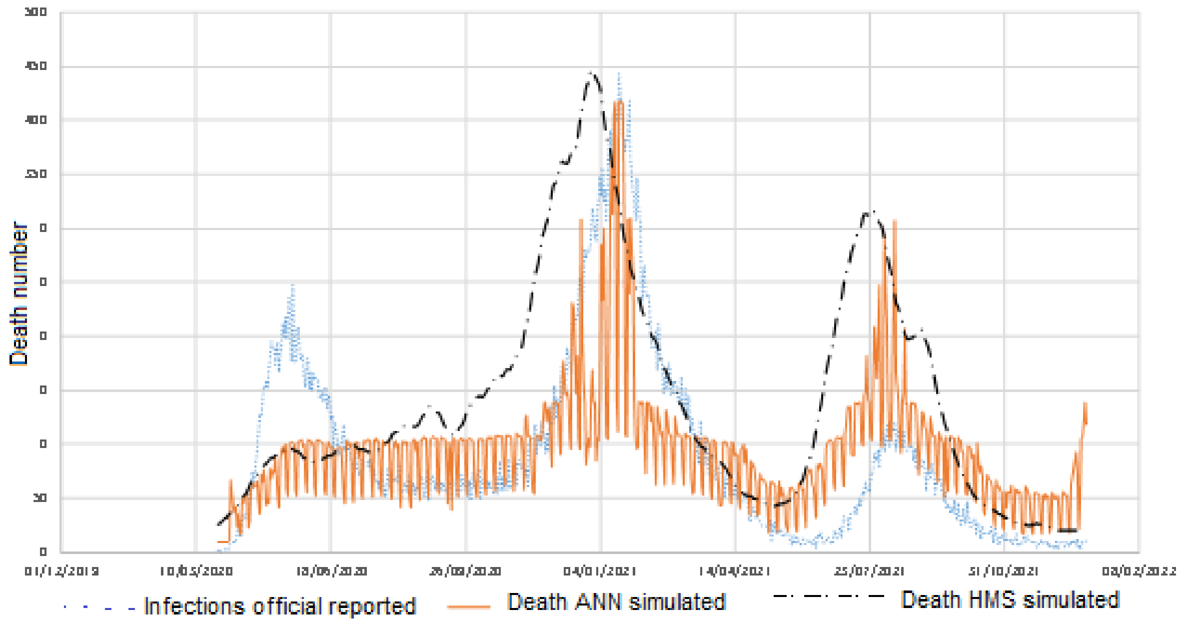

Fig. 10. Comparison of official deaths vs. those modeled by HEC-HMS and ANN

4. Conclusions

Rain runoff models’ theory used in hydrology problems was applied to reproduce behavior in deaths by COVID 19 using official information rate infections, taking data in Mexico City. Additionally, an artificial neural network was used to reproduce contagion-death process, which helped to estimate time in which, on average, the COVID 19 disease can lead to an adverse case fatality situation; tested models adequately reproduced shape of death curve from infection function over time, especially in wave that corresponded to the alpha and beta variant; also, with the HEC-HMS it was estimated that approximate time between contagion and possible death was 6.38 days. In the wave for delta variant, an overestimation was observed given by the HEC HMS models and also by the ANN, attributed to a reduction effect in the lethality of the virus due to the intensification of vaccination campaigns. It should be noted that on the part of the ANN the input and output data are daily data of the reports of contagions and officially occurred deaths, that is to say that the data of the death that occurred on the day it is reported, is attributed to a contagion that occurred days ago, so the model implicitly has an autoregressive component. With the HEC HMS, when the curve of infections calculated against the official infections is drawn, an occurrence of the peak of deaths is observed approximately 6 days later compared to what was officially reported.

Acknowledgments

We thank the Government of Mexico for public databases available on the internet for the data used in this study.
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] Basu, S., and R.H. Campbell. "Going by the numbers: Learning and modeling COVID-19 disease dynamics." *Chaos, Solitons & Fractals* 138 (September 2020): 110140.
[2] Dal Molin Ribeiro, M.H., R. Gomes da Silva, V. Cocco Mariani, and L. Dos Santos Coelho. "Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil." *Chaos, Solitons & Fractals* 135 (June 2020): 109853.
[3] Alcántara-López, F., C. Fuentes, C. Chávez, J. López-Estrada, and F. Brambila-Paz. "Fractional Growth Model with Delay for Recurrent Outbreaks Applied to COVID-19 Data." *Mathematics* 10, no. 5 (March 2022): 825.
[4] Jaiswal, RK, S. Ali, and B. Bharti. “Comparative evaluation of conceptual and physical models of rainfall-runoff.” *Applied Water Science* 10 (January 2020): 48.
[5] SCS. *Soil Conservation Service. Engineers Corp Army Hec-HMS elementary manual*. Spain, Department of Geology - Univ. Salamanca, 1972.
[6] The MathWorks, Inc., 2022. MATLAB. Version R2014a [Software]. https://la.mathworks.com/.
[7] Health Secretary / Secretaría de Salud. "Guideline for the calculation methodology for COVID-19 epidemic risk traffic light" / “Lineamiento para la metodología de cálculo del semáforo de riesgo epidémico COVID-19”. *gob.mx (Gobierno de México)*. Accessed May 20, 2022. https://coronavirus.gob.mx/semaforo/.
[8] Esquivel Vázquez, J., J. Peña Acevedo, D.I. Rodríguez González, I. Sánchez Bravo, and J. L. Salazar Villanueva. *Methodology for the analysis of trends in epidemic curves of new cases for COVID-19 syndrome and bed occupancy and deaths registered in the IRAG Network / Metodología para el análisis de tendencias de curvas epidémicas de casos nuevos de síndrome COVID-19 y de ocupación de camas y defunciones registrados en la Red IRAG*. Mexico, Coordination of Technological Services of CIMAT / Coordinación de Servicios Tecnológicos del CIMAT, 2021.