Proceedings of the 9th World Congress on Electrical Engineering and Computer Systems and Sciences (EECSS'23)
Brunel University, London, United Kingdom – August 03-05, 2023
Paper No. ICBES 155
DOI: 10.11159/icbes23.155

Predicting the COVID-19 Pandemic Spread: An Analysis Using ARX and ARMAX Models

Cristina-Maria Stancioi[1], Vlad Muresan[1], Iulia Clitan[1], Mihail Abrudean[1], Mihaela-Ligia Unguresan[2]
[1]Automation department, Technical University of Cluj-Napoca
28 Memorandumului Street, 400114, Cluj-Napoca, Romania
Stancioi.Ni.Cristina@campus.utcluj.ro; Vlad.Muresan@aut.utcluj.ro; Iulia.Clitan@aut.utcluj.ro; Mihai.Abrudean@aut.utcluj.ro
[2]Physics and Chemistry Department, Technical University of Cluj-Napoca
28 Memorandumului Street, 400114, Cluj-Napoca, Romania
Mihaela.Unguresan@chem.utcluj.ro

Abstract **-** December 2019 will be known as the month that drastically change the people's normal lives. It is the month in which the new coronaviral disease (COVID-19) rise and soon became a global health problem. COVID-19 is believed to be spread solely via the respiratory tract, health officials in the affected countries have implemented a variety of stringent prevention steps, including social isolation, travel restrictions, and varying levels of quarantine. In this paper, the mathematical model of the SARS-CoV-2 virus transmission on the Romanian territory is determined. Estimating the epidemiological trend of the outbreaks is crucial for the allocation of medical resources, regulation of production activities, and even for the national economic development. Thus, it is essential to develop reliable and suitable forecasting tools, based on mathematical prediction models, that can help governments by providing faster decision making, while being cost-effective. Based on the simulation results, the conclusions regarding the spread of the pandemic, the effect of the pandemic on the Romanian population, respectively on the healthcare system are drawn.

Keywords**:** COVID-19, virus, mathematical model, ARMAX, ARX, simulation, prediction.

1. Introduction

The first cases of COVID-19 were identified in 2019, in a city in central China but, in just three months, the outbreak has infected more than 100.000 people, killed more than 4000 people, finally affecting all continents except Antarctica [1]. Since then, the pandemic has grown to affect more than 188 countries on earth [2]. By the end of 2020, more than 43 million people in the world have been affected by it [3] [4].

Furthermore, from the first reported death in China on 11 January 2020, on 28 September 2020 the world reached one million recorded deaths from Covid-19 [1]. Nevertheless, the death rate is considered far higher as many cases have not been reported officially [5]. COVID-19 pandemic manages to trigger a major disruption and has significant repercussions for everyday life and business [6] [7]. The pandemic has had a significant impact on daily life, with towns being closed, schools closing, and people being asked to work from home to slow the spread of infections [8] [9].

Unfortunately, Romania was not bypassed by the Covid-19 pandemic. The first case of Covid-19, in Romania, was reported on the 26th of February 2020, and the national authorities imposed a national lockdown from 16th March to 14th May, that decreased the number of active cases in our country.

On the 17th of March 2020, the first COVID-19 death was recorder. From the 15th of May, our country is maintaining a state of alert, established by the President of Romania. However, with the relaxation of restrictions within Romania's border, a first wave of infections appeared in April 2020. Since the 17th of March until the 9th of June, our country totalled a number of 1341 fatalities, of which 545 women and 796 men. Since then, a new wave of infections followed in September, as the second wave, reaching a much higher infection rate than at the beginning of the pandemic, thus increasing the fatalities.

Romania had more than half a million SARS-CoV-2 infections reported by the end of December 2020. The amount actually exceeds this figure by a lack of testing [10]. Notwithstanding the authorities' prohibitions, patients with COVID-19 have overwhelmed hospitals which have hit their limits on the capacity to provide intensive care. Since the Romanian health system is under enormous strain, the most successful way to do this is a proven approach focused on the science of implementation [11]. In this regard, having an accurate viewpoint of the COVID-19 epidemic by estimation models is critical for the implementation of precise interventions and a risk evaluation study but also to provide an accurate view of the situation.

2. Methods

2.1. Data processing

The preparation of the database, used to establish the structure and structural parameters of the mathematical model was an essential key step since the absence or inconsistency of experimental data would have made it impossible to design a model. These data represent values of certain parameters of the pandemic, respectively values of certain input signals for the designed mathematical model. The complexity of data processing was also given by the need to verify their consistency (quality) by making comparisons between the values provided by several data sources.

In this context, to be able to determine the mathematical model, the obtained result was a project database (containing data relevant to the dynamics of SARS-CoV-2 virus transmission in Romania). This database comprises four components, such as:

- data relevant for the dynamics of virus transmission throughout Romania (number of confirmed cases, number of tests performed, number of cured patients, number of active cases, number of deaths, rate of growth, reproduction number, number of people entering / leaving the country, incidence at 14 days, categories of vulnerable groups, environmental factors, etc.);
- relevant data for the dynamics of virus transmission by counties (number of confirmed cases, number of cured patients, number of active cases, number of deaths, incidence at 14 days, vulnerable groups categories, environmental factors, etc.);
- data on legislative measures adopted by the authorities to reduce the pandemic (measures staggered over time, both at national and local level);
- relevant data for the dynamics of pandemic transmission in countries where a sizable number of Romanian citizens live (all countries in Europe, United States of America, Israel, Canada). Examples of data: number of confirmed cases, number of cured patients, number of patients admitted to ATI, number of active cases, number of deaths, etc.).

All values were centralized daily from the beginning of the pandemic, and also cumulated in the case of signals that allow cumulative expressions. Once the database was structured, it was constantly updated.

The above data were grouped into two time intervals, namely the learning time interval (from the beginning of the pandemic (characterized by the first significant variations that occurred around 15th of March, 2020) and until the 31st of October, 2020), respectively the prediction time interval (corresponding to November 2020; the data related to this time interval were used to test the model determined by testing its prediction capacity using input signals that were not included in the learning procedure).

2.2. ARX and ARMAX models

Modelling is an essential step in the parameter estimation process. Identification methods based on input-output measurements are frequently used in many cases where it is not essential to obtain an in-depth mathematical comprehension of the system under analysis, but it is enough to forecast the system evolution.

The concept of parametric model identification is to derive a mathematical model from measured data. If the initial conditions of the system are known, the model must be capable to determine the output of the system regardless of time t.

The approach used for process modelling in this paper is based on a parametric identification of an ARX (Auto-Regressive with exogenous input) model [12]. The simplicity with which this method may be implemented justifies its selection. The variation of the estimated parameters of the identified model allows us to track the dynamical progression of the process and depict the presence of faults. Using the identification methods described above, we ran several tests based on the ARX model structure, varying the value of parameters involved. The objective was to find the best fit using the minimum amount of parameters possible.

An ARMAX (AutoRegressive Moving Average with eXogenous Inputs) model estimates future values of a dependent variable by merging autoregressive (AR) and moving average (MA) components with exogenous factors [13]. To represent the behaviour of the dependent variable, the AR component takes into account its prior values. It is assumed that the variable's current value is connected to its previous values. The MA component takes into account previous prediction failures in order to capture any remaining patterns in the data. It is assumed that the present value of the variable is connected to previously estimated failures.

3. Obtained results

The best fit was obtained in the case of ARX model as it can be seen in Figure 1. The ARMAX model has also a good fit, described by Figure 2. The major differences are observed in the changing points from the graphics. When the number of new cases increased or decreased, the entire prediction was impacted, but the estimation curve with the best fit followed the real data curve in an almost perfect manner.

The learning time interval was chosen to be equal to 120 days and the prediction time interval was equal to 60 days, as it can be also observed in both figures.

By calibrating the number of coefficients in both ARX and ARMAX models, the best fit was obtained. (94.67%)

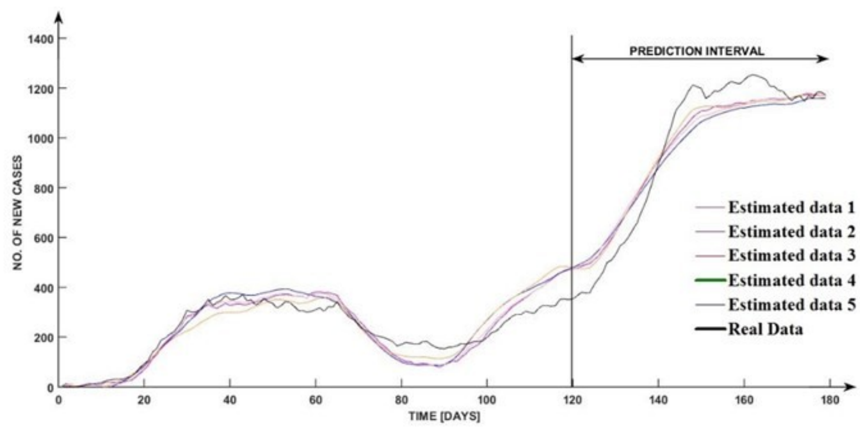

Figure 1 ARX model estimation

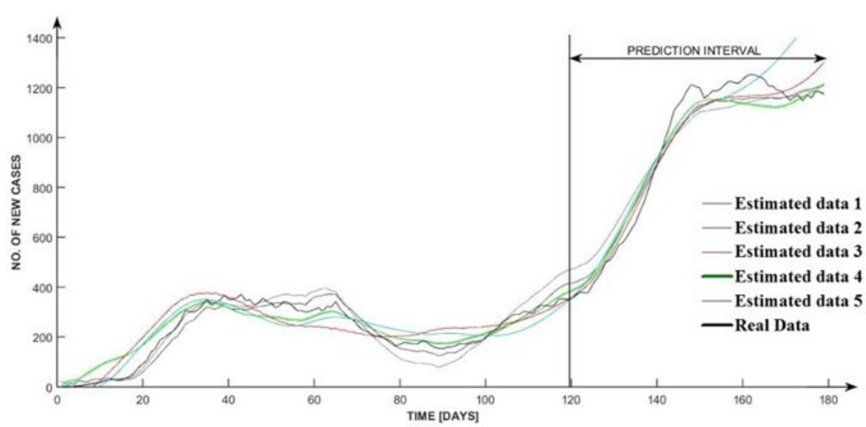

Figure 2 ARMAX model estimation

4. Conclusion

This paper revealed that most developed models and statistical analysis on COVID 19 transmission and progression, so far, concentrate on the evolution of COVID 19 pandemic in one country, which is also the case for our model, that uses time series data collected from Romania. Our study was conducted using publicly available data that did not include personal information, so no ethical review was required. Data on confirmed infected cases were obtained from the daily reports released by the authorities, data collected by different statistical websites such as: covid19.geo-spatial.org , datelazi.ro, that track the evolution of COVID 19 cases in our country.

The determination of the mathematical model of SARS-CoV-2 virus transmission, in relation to time, was based on the processing of the sets of input-output data pairs taken from the project database and their use for determining (learning) the dynamic process of transmission of the SARS-CoV-2 virus over time.

Some other prediction models regard the COVID-19 spread in four other European countries (Spain, Germany, France. Italy) based on the prediction of future COVID 19 daily cases, having for instance a prediction interval of 10 days. Both performed studies, are using ARIMA model (Auto-Regressive Integrated Moving Average). Those studies shown that ARIMA models are suitable for predicting the prevalence of COVID-19 in the future and the results are analyzed in order to understand the trends of the outbreak, in the given countries [14].

It is critical to precisely and particularly anticipate the spread of COVID-19. There is a need to estimate trends in viral transmission using available data. Many governments rely considerably on such forecasts to determine their next steps, whether it's providing medical resources or deciding whether to reduce or increase the number of lockdowns.

References

[1] Van Bavel, J. J., Baicker, K., Boggio, P. S., Capraro, V., Cichocka, A., Cikara, M. and Druckman, J. N. (2020) ‘Using social and behavioural science to support COVID19 pandemic response’, Nature Human Behaviour 4: 460–471.

[2] WHO Coronavirus Disease (COVID-2019) Situation Reports: https://www.who.int/emergencies/diseases/novelcoronavirus-2019/situation-reports/.

[3] Meng Liu, RaphaelThomadsen, Song Yao, Forecasting the spread of COVID 19 under different reopening strategies, Scientifc Reports | (2020) 10:20367 | https://doi.org/10.1038/s41598-020-77292-8.

[4] Wu JT , Leung K , Leung GM, Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study, Lancet. 2020; 395: 689-697.

[5] M. Niazkar and H. R. Niazkar, “COVID-19 outbreak: application of multi-gene genetic programming to country-based prediction models,” Electronic Journal of General Medicine, vol. 17, no. 5, article em247, 2020.

[6] Google COVID-19 community mobility reports. https://www.google.com/covid19/mobility.

[7] Wei Liu , Xiao-Guang Yue and Paul B. Tchounwou, Response to the COVID-19 Epidemic: The Chinese Experience and Implications for Other Countries, Int. J. Environ. Res. Public Health 2020.

[8] OECD - COVID-19: Protecting People and Societies. Available from Paris: https://read.oecd-ilibrary.org/view/?ref=126_126985-nv145m3l96&title=COVID19-Protecting-people-and-societies.

[9] Hamzelou, J., 2020.World in lockdown. New Sci, https://doi.org/10.1016/s0262-4079(20)30611-4.

[10] Alwan NA. Surveillance is underestimating the burden of the COVID-19 pandemic. Lancet. (2020) 396:e24; doi: 10.1016/S0140-6736(20)31823-7.

[11] Bauer MS, Damschroder L, Hagedorn H, Smith J, Kilbourne AM. An introduction to implementation science for the non-specialist. BMC Psychol. (2015) 3:32; doi: 10.1186/s40359-015-0089-9.

[12] Z.M.Yusoff, Z.Muhammad, M.H.F. Rahiman, M.N. Taib, ARX Modeling for Down-Flowing Steam Distillation System, IEEE 8th International Colloquium on Signal Processing and its Application, 2012.

[13] Söderström. Extending the Frisch scheme for errors–in–variables identification to correlated output noise. International Journal of Adaptive Control and Signal Processing, 22:55–73, 2008.

[14] Tahir Mumtaz Awan, Faheem Aslam, Prediction of daily COVID-19 cases in European countries using automatic ARIMA model, Journal of Public Health Research 2020; volume 9:1765, PP 227-233.