# Forecasting the dynamics of COVID-19 Pandemic in Top 15 countries in April 2020: ARIMA Model with Machine Learning Approach

*Pavan Kumar[1], Himangshu Kalita[2], Shashikanta Patairiya[3], Yagya Datt Sharma[4], Chintan Nanda[5], Meenu Rani[6], Jamal Rahmani[7], Akshaya Srikanth Bhagavathula[8*]*

Dr. Pavan Kumar
Department of Horticulture and Forestry,
Rani Lakshmi Bai Central Agricultural University,
Jhansi-284003, India
Email : pawan2607@gmail.com
https://orcid.org/0000-0003-3653-8163
Phone No.: +91-9785879797

Mr. Himangshu Kalita
Haryana Space Applications Centre (HARSAC)
Department of Science & Technology,
CCS HAU Campus, HISAR 125004, India
Email : himangshu.kalita6@gmail.com

Mrs. Shashikanta Patairiya
Anchor Systems Corp.
Reston, VA 20194, USA
Email: shashikantaptairiya28@gmail.com
Orcid ID: 0000-0002-1152-2141

Mr. Yagya Datt Sharma
Hughes SystiqueCorporation,
Germantown (MD), 20876 USA
Email: yagyadatt.sharma@gmail.com
Orcid ID: 0000-0003-4744-8753

Mr.Chintan Nanda
Indian Institute of Remote Sensing,
ISRO, Dehradune, India
Email: chintann549@gmail.com

Mrs. Meenu Rani
Department of Geography
Kumaun University
Nainital, Uttarakhand, India
Email: meenurani06@gmail.com

Dr. Jamal Rahmani,

Department of Community Nutrition,
National Nutrition and Food Technology Research Institute,
Shahid Beheshti University of Medical Sciences, Tehran, Iran
Email: jrahmani@sbmu.ac.ir

**Corresponding author**

Dr. Akshaya Srikanth Bhagavathula, PharmD, PhD student,
Institute of Public Health,
College of Medicine and Health Sciences,
UAE University, Al Ain, UAE
Email: akshaypharmd@gmail.com

Phone: +971543226187

**Abstract**

We here predicted some trajectories of COVID-19 in the coming days (until April 30, 2020) using the most advanced Auto-Regressive Integrated Moving Average Model (ARIMA). Our analysis predicted very frightening outcomes, which defines to worsen the conditions in Iran, entire Europe, especially Italy, Spain, and France. While South Korea, after the initial blast, has come to stability, the same goes for the COVID-19 origin country China with more positive recovery cases and confirm to remain stable. The United States of America (USA) will come as a surprise and going to become the epicenter for new cases during the mid-April 2020. Based on our predictions, public health officials should tailor aggressive interventions to grasp the power exponential growth, and rapid infection control measures at hospital levels are urgently needed to curtail the COVID-19 pandemic.

**Keywords:** *COVID-19; Coronavirus; forecast, machine learning, ARIMA Models, pandemic*

Since January 2020, the novel coronavirus (severe acute respiratory syndrome 2 [SARS-CoV-2]) from Wuhan, China, has continued to spread around the world and turned into Pandemic as novel coronavirus disease (COVID-19) [1-2]. Due to rapid pandemic potential and the absence of vaccines and drugs, the contagious COVID-19 disease devastated the normal life across the globe. Currently, Coronavirus infected more than half a million population, and killed more than 25 thousand people, and also forced more than 3 billion to stay with their homes [3]. Due to a lack of knowledge about this virus, the COVID-19 pandemic placed tremendous strain on everyone around the world. In order to prevent further transmission, strong preventive measures are intensified from the past weeks; however, the number of infected cases are consistently increasing around the world, even after undergoing lockdown.

Mathematical approaches are widely used to infer critical epidemiological transitions and parameters of COVID-19. Methods such as epidemic curve fitting, surveillance data during the early transmission $R_{0,}$ and other epidemic models are frequently applied to generate forecasts of COVID-19 pandemic across the world [4-6].

**Methods**

We used the data of cumulative confirmed death and recovery of COVID-19 cases reported from January 21 until March 26, 2020, that were obtained from John Hopkins Coronavirus resource center (https://coronavirus.jhu.edu/). We analyzed the data using dynamic models to generate 30 days forecasts and to understand the positive effect in the near future as well as projecting trends over trajectories. We used different statistical phenomenological models in the R-language platform to analyze the disease-based trajectories model for prediction purposes. We precisely used four models to analyze the aggregate data set for time series analysis. This includes the

ARIMA model, which is a mass model of two different models, including the AR model (Auto Regression) and the MA (Moving Average) model [7]. This model results over Akaike information criterion (AIC) statistics and coverage of regression analysis.

Another type of COVID-19, like SARS disease (Severe Acute Respiratory Syndrome), is analyzed without breaking the current situation and predicting the future perspective [8]. VAR model (Vector Auto Average) model is used to predict the spatial extinct while using remote sensing data and for the purpose of creation of GIS map of worldwide on three different variables [9]. These three variables in the GIS environment create a map of cumulative confirmed cases country-wise as well as recovered and death map [10]. The use of another statistical analysis is a generalized logistic growth model (GLM), which generally depicted as a scaling parameter of integrating an additional result-oriented value put method [11]. Some epidemic models used in disease epidemic conditions measure oscillates some multiple peak parameters inferred in sub-epidemic and pandemic conditions to determine the projected outcomes [12].

After standardizing all the models, the data of the top 20 countries were included to analyze the forecasting models of differential spatial adjacent, and projected trajectories are analyzed up to April 30, 2020. By using GIS and Remote sensing to determine the epidemic mapping and analyzing the upcoming effects of COVID-19.

**Results and discussion**

An overview of epidemic trends of COVID-19 across 170 countries and territories from January to March 26, 2020, is presented in Figure 1. From these top 20 countries, with a high number of confirmed cases are stratified to include in mathematical mode. The comparison between

cumulative incident cases, mortality, and recovery of COVID-19 information among the top 15 affected countries is shown in Figure 2. Top countries' data of China, Italy, Spain, and Iran showed highly disastrous mortality and badly effected with a vast number of COVID-19 cases. On the other hand, some European countries such as Switzerland, United Kingdom, Netherlands, Austria, Belgium, Norway, and Sweden had huge confirmed cases but negligible number in case of recover and death cases.

Furthermore, data smoothening was applied to stabilize the data by removing changes in the level of a time series, and therefore eliminating (or reducing) trend and seasonality. After this, the forecast prediction model was applied by using AR and MA models to generate plots of the different trends in upcoming days. The outcome of these predictions is presented in Figure 3. Our findings revealed linearity in the current cumulative cases and showed a rapid exponential growth phase in the world span may occur roughly during April 8 to April 30, 2020, when the number of COVID-19 cases may skyrocket near to one million in the USA, 300,000 in Italy, and 250,000 in Germany, 150,000 in the United Kingdom and Iran (120,000). Other countries with a smaller number of cases but showing a sharp upward trend include Switzerland, Austria, and Canada. However, the cases of COVID-19 in China and South Korea remain stable. Public health officials in these countries need to grasp the powerful wave of exponential growth before COVID-19 collapses the entire health system.

The fast raising COVID-19 infection rates may lead to the exponential death toll in the top 15 countries. Our analysis detected that the number of deaths would increase more rapidly in thousands during the mid of April 2020 (Figure 4). The associated mortality rate will be much higher in Italy, and Spain with roughly 30,000 deaths in each country, following France, USA and Iran will record around 10,000 deaths, and also the death in UK, Netherlands, Belgium, and

Switzerland started to incline the death rates rapidly than the rest of current top 15 countries. The recovery rates will stay slow in initial time but then rapidly increase in Italy, Germany, Spain, and Iran by the end of prediction (Figure 5).

Our forecast analysis of COVID-19 dynamics showed a different angle for the whole world, and it looks scarier than imagined. Overall, our analysis confirmed the with an increase in the number of cases, the number of deaths is going to rise exponentially to rising in the future time (Figure 6). Interestingly, the recovery numbers also look promising, going to pack up during mid-April. Thus, it can slow down the surge of COVID-19 pandemic by the end of April.

## Conclusion

Based on our predictions, public health officials should tailor aggressive interventions to grasp the power exponential growth, and rapid infection control measures at hospital levels are urgently needed to curtail the COVID-19 pandemic.

**Conflict of interest:** None

## References

1. Zou L, Ruan F, Huang M, Liang L, Huang H, Hong Z, Yu J, Kang M, Song Y, Xia J, Guo Q. SARS-CoV-2 viral load in upper respiratory specimens of infected patients. New England Journal of Medicine. 2020;382(12):1177-9.

2. Xu XW, Wu XX, Jiang XG, Xu KJ, Ying LJ, Ma CL, Li SB, Wang HY, Zhang S, Gao HN, Sheng JF. Clinical findings in a group of patients infected with the 2019 novel coronavirus (SARS-Cov-2) outside of Wuhan, China: retrospective case series. BMJ. 2020;368.

3. World Health Organization (WHO). Novel coronavirus (COVID-19) situation [Internet]. Geneva: WHO; 2020 [cited 2020 Mar 18]. Available from: https://experience.arcgis.com/experience/685d0ace521648f8a5beeeee1b9125cd.

4. Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, Eggo RM, Sun F, Jit M, Munday JD, Davies N. Early dynamics of transmission and control of COVID-19: a

mathematical modelling study. The Lancet Infectious Diseases. 2020. Doi: 10.1016/S1473-3099(20)30144-4.

5. Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. Journal of travel medicine. 2020 Mar 1.

6. Pirouz B, Shaffiee Haghshenas S, Shaffiee Haghshenas S, Piro P. Investigating a Serious Challenge in the Sustainable Development Process: Analysis of Confirmed cases of COVID-19 (New Type of Coronavirus) Through a Binary Classification Using Artificial Intelligence and Regression Analysis. Sustainability. 2020 Jan;12(6):2427.

7. Chowell, G. Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts. Infect. Dis. Model. 2017, 2, 379–398.

8. Roosa, K.; Lee, Y.; Luo, R.; Kirpich, A.; Rothenberg, R.; Hyman, J.; Yan, P.; Chowell, G. Real-time forecasts of the 2019-nCoV epidemic in China from February 5th to February 24th, 2020. Infect. Dis. Model. 2020.

9. Wang, X. S.; Wu, J.; Yang, Y. Richards model revisited: Validation by and application to infection dynamics. J. Theor. Biol. 2012, 313, 12–19

10. Richards, F. A flexible growth function for empirical use. J. Exp. Bot. 1959, 10, 290–301

11. Roosa, K.; Chowell, G. Assessing parameter identifiability in compartmental dynamic models using a computational approach: Application to infectious disease transmission models. Theor. Biol. Med Model. 2019, 16.

12. Chowell, G.; Tariq, A.; Hyman, J.M. A novel sub-epidemic modeling framework for short-term forecasting epidemic waves. BMC Med. 2019, 17, 164.

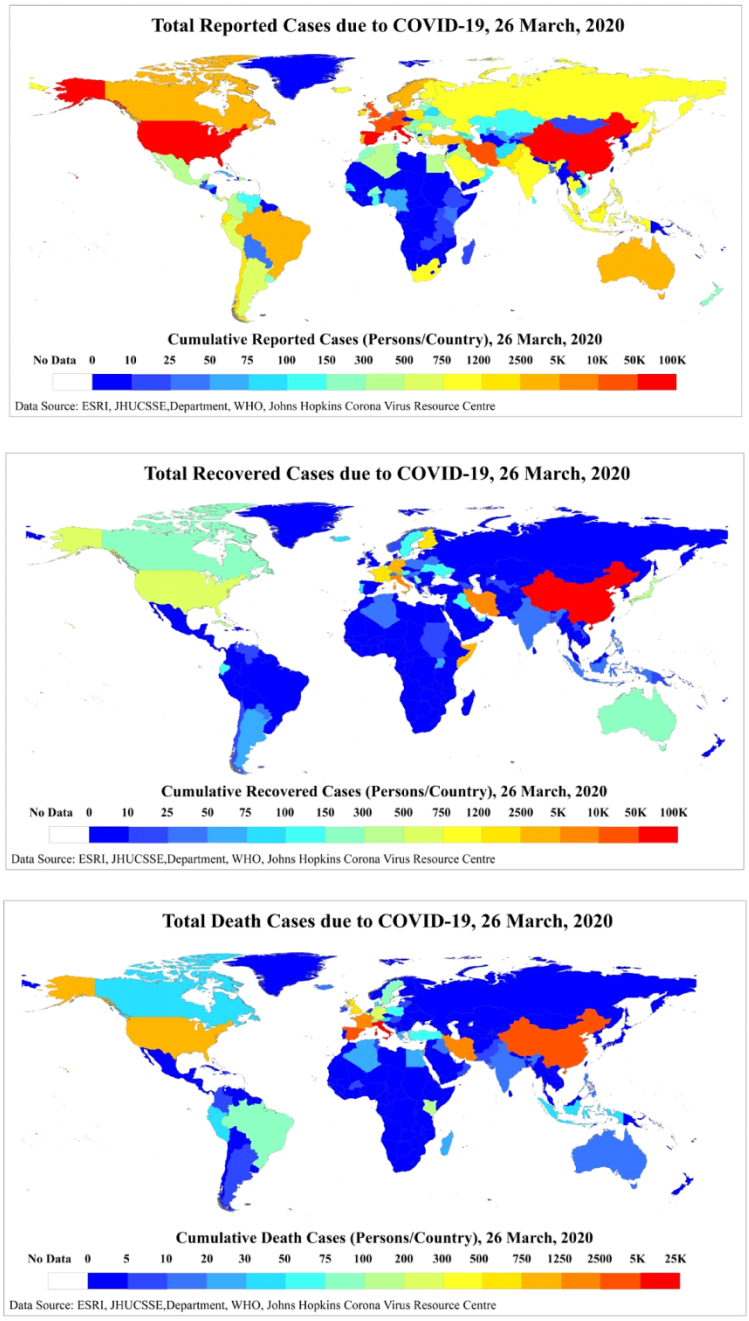


**Fig. 1:** COVID-19 Epidemic Spatial pattern of total confirmed, recovery and death cases from 19 January to 26 March 2020 by Country, Territory, or Conveyance

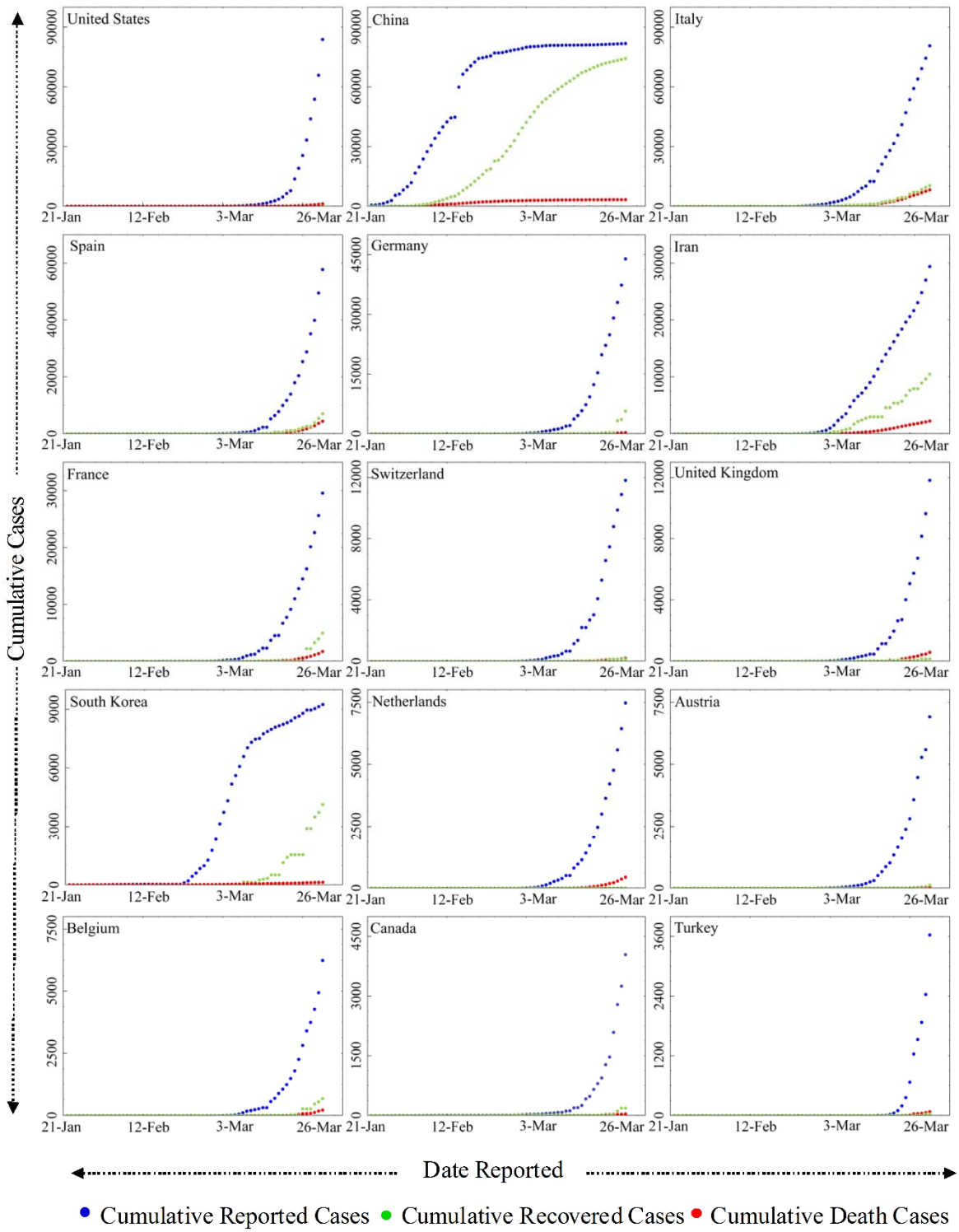


**Fig. 2:** Comparisons between cumulative reported, recovery and death incidence of cases with COVID-19 on the top 15 affected countries

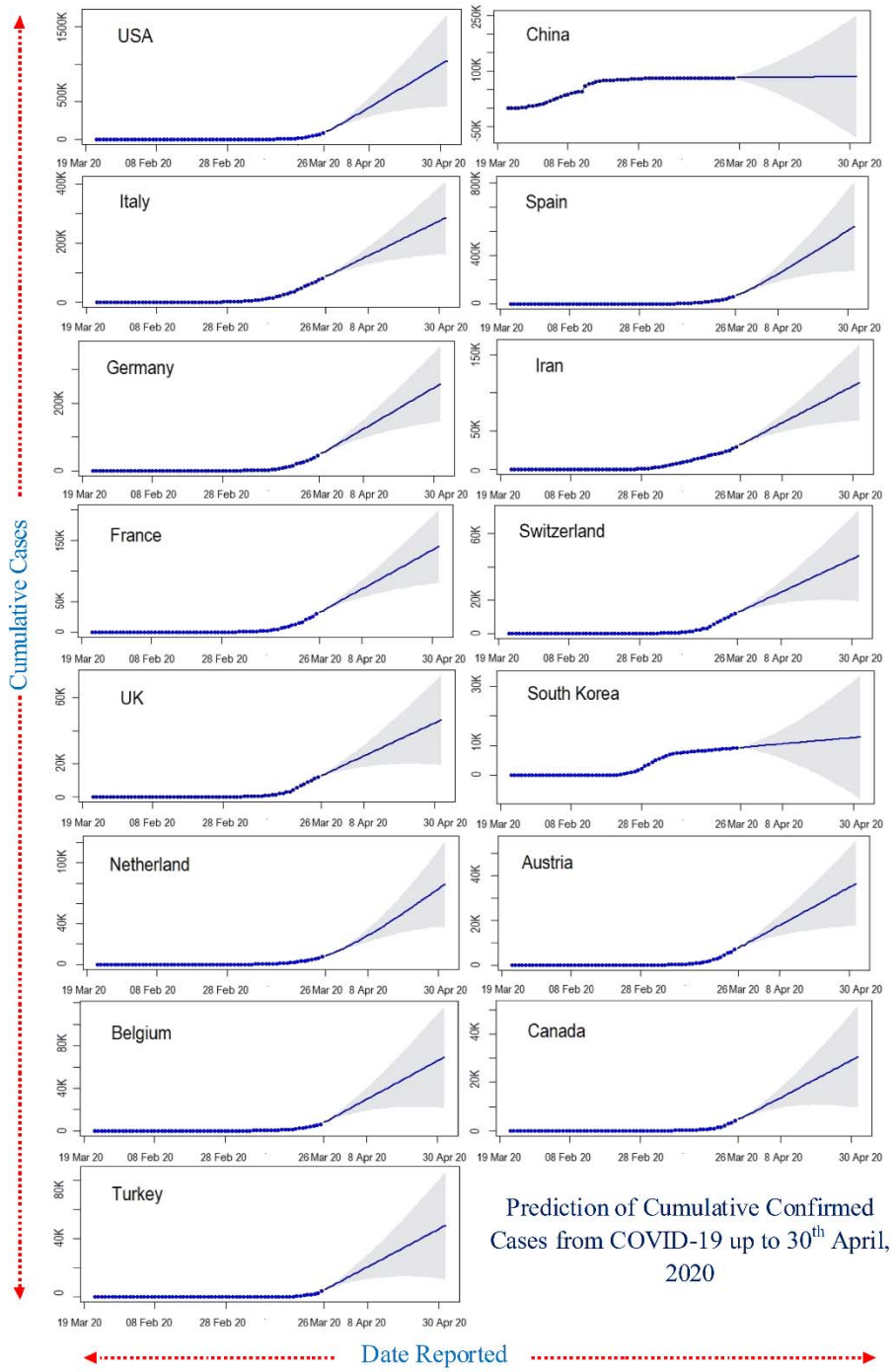


**Fig 3:** Twenty-day ahead ARIMA model forecasts of cumulative confirmed COVID-19 cases in the top 15 affected countries generated on 26 March, 2020. The light black line correspond to the cumulative cases confirmed up until 26 March 2020; the dark blue lines correspond to the mean ARIMA model; the light shadow lines depict the 95% prediction intervals and forecasting periods.

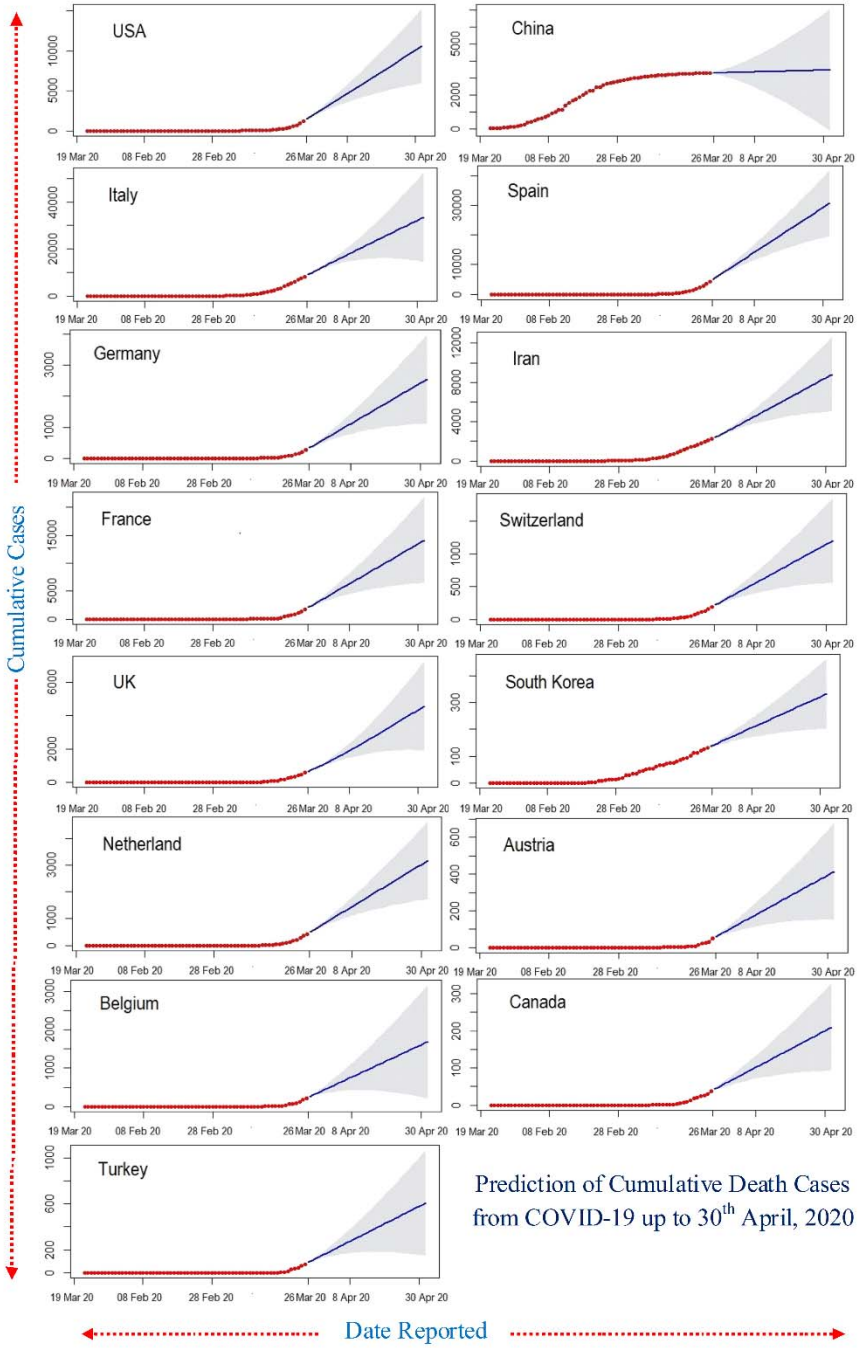


**Fig. 4:** Twenty-day ahead ARIMA model forecasts of cumulative death COVID-19 cases in the top 15 affected countries generated on 26 March, 2020. The light black line correspond to the cumulative cases death up until 26 March 2020; the dark blue lines correspond to the mean

ARIMA model; the light shadow lines depict the 95% prediction intervals and forecasting periods.

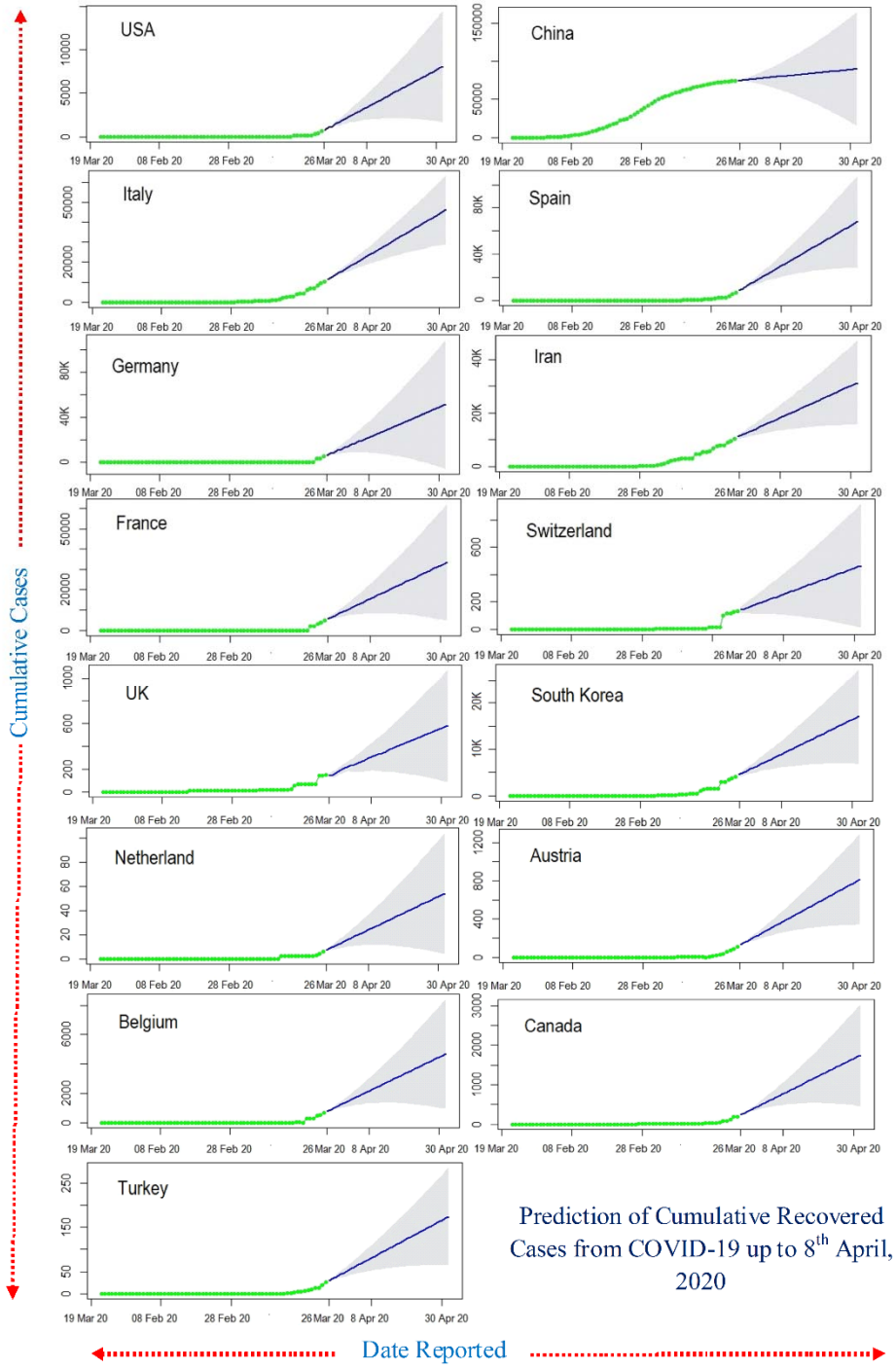


**Fig. 5:** Twenty-day ahead ARIMA model forecasts of cumulative recovered COVID-19 cases in the top 15 affected countries generated on 26 March, 2020. The light black line correspond to the cumulative cases recovered up until 26 March 2020; the dark blue lines correspond to the mean

ARIMA model; the light shadow lines depict the 95% prediction intervals and forecasting periods.

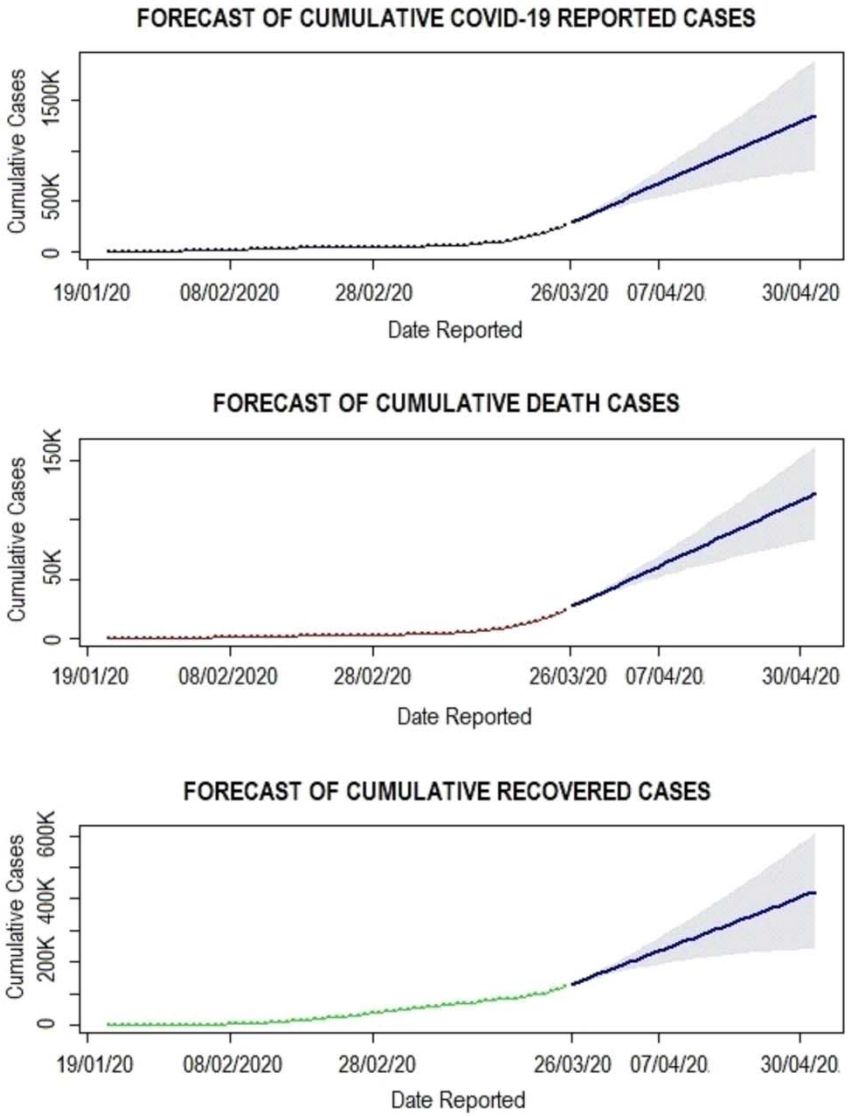


**Fig. 6:** Twenty-day ahead ARIMA model forecasts of cumulative confirmed death and recovered Novel Coronavirus Disease (COVID-19) cases by Country, Territory, or Conveyance generated on 26 March, 2020. The red line correspond to the cumulative confirmed death and recovered up until 26 March 2020; the blue lines correspond to the mean ARIMA model; the light shadow lines depict the 95% prediction intervals and forecasting periods.