Türkiye'de COVID-19 Salgınının Yayılımını Öngörmek İçin Tahmin Modeli
Forecasting Model to Predict the Spreading of the COVID-19 Outbreak in Turkey

Ceyhun Bereketoglu[1*], Nermin Ozcan[1], Tugba Raika Kiran[2], Mehmet Lutfi Yola[3]
[1]Department of Biomedical Engineering, Faculty of Engineering and Natural Sciences, Iskenderun Technical University, Hatay, Turkey
{ ceyhun.bereketoglu, nermin.ozcan }@iste.edu.tr
[2]Department of Medical Biochemistry, Faculty of Medicine, Malatya Turgut Ozal University, Malatya, Turkey
{ raika.kiran } @ozal.edu.tr
[3]Department of Nutrition and Dietetics, Faculty of Health Sciences, Hasan Kalyoncu University, Gaziantep, Turkey
{ mlutfi.yola } @hku.edu.tr

Özetçe*—Bu çalışma, Türkiye için kamuya açık epidemiyolojik verilere dayanarak COVID-19 salgını yayılımı, ölüm oranı ve iyileşme verileri gibi parametlerinin geleceğini tahmin etmeyi amaçlamıştır. İlk olarak Facebook's Prophet, ARIMA ve Decision Tree gibi farklı tahmin yöntemlerini uyguladık. MAPE ve MAE ölçümlerine dayalı olarak, Facebook's Prophet en etkili tahmin modeli çıkmıştır. Daha sonra, Facebook's Prophet kullanarak, salgının Türkiye'deki gelişimi için on beş gün öncesinden bir tahmin modeli oluşturduk. Bildirilen doğrulanmış vakalara dayanarak, simülasyonlar, enfekte olmuş kişilerin toplam sayısının 23 Nisan 2021'e kadar 4328083'e (sırasıyla 3854261 ve 4888611 alt ve üst sınırlar ile) ulaşabileceğini göstermektedir. Simülasyon tahmini, ölü sayısının sırasıyla 34806 ve 36246 alt ve üst sınırlar olmak üzere 35656'ya ulaşabileceğini göstermektedir. Ayrıca bulgularımız, iyileşen vaka oranın %86.38'i aşabileceğini, ancak gelecekteki aktif vakaların mevcut aktif vakalara göre önemli ölçüde artabileceğini de göstermektedir. Bu zaman serisi analizi, yakın gelecekte Türkiye'deki COVID-19 salgınının artış eğiliminde olduğunu göstermektedir. Genel olarak bu çalışma, pandemik yayılmanın simülasyonu ve dolayısıyla COVID-19 salgını için temel müdahalelerin daha ileri düzeyde uygulanması için verimli bir veriye dayalı tahmin modeli analizinin önemini vurgulamaktadır.

Anahtar Kelimeler*—COVID-19, tahmin, Facebook's Prophet, Türk popülasyonu, müdahale.

***Abstract*— This study aimed to forecast the future of the COVID-19 outbreak parameters such as spreading, case fatality, and case recovery values based on the publicly available epidemiological data for Turkey. We first performed different forecasting methods including Facebook's Prophet, ARIMA and Decision Tree. Based on the metrics of MAPE and MAE, Facebook's Prophet has the most effective forecasting model. Then, using Facebook's Prophet, we generated a forecast model for the evolution of the outbreak in Turkey fifteen-days-ahead. Based on the reported confirmed cases, the simulations suggest that the total number of infected people could reach 4328083 (with lower and upper bounds of 3854261 and 4888611, respectively) by April 23, 2021. Simulation forecast shows that death toll could reach 35656 with lower and upper bounds of 34806 and 36246, respectively. Besides, our findings suggest that although more than 86.38% growth in recovered cases might be possible, the future active cases will also significantly increase compared to the current active cases. This time series analysis indicates an increase trend of the COVID-19 outbreak in Turkey in the near future. Altogether, the present study highlights the importance of an efficient data-driven forecast model analysis for the simulation of the pandemic transmission and hence for further implementation of essential interventions for COVID-19 outbreak.**

Keywords*—COVID-19; forecasting, Facebook's Prophet, Turkish populatio, intervention.

I. INTRODUCTION

The coronavirus disease pandemic was started by an outbreak that emerged in December 2019 in Wuhan, China [1]. The source of the pandemic is a virus formerly designated 2019-nCoV, then taxonomically termed as severe acute respiratory syndrome corovirus-2 (SARS-CoV-2) and clinically called coronavirus disease 2019 (COVID-19) [2-4]. On March 11, World Health Organization (WHO) declared a global pandemic for COVID-19. As of April 08, 2021, WHO reported a total of 132.730.691 confirmed cases and 2.880.726 deaths worldwide (https://covid19.who.int/). The COVID-19 has caused a great concern to the health and safety of people all over the world due to its high infection rate and mortality. It is transmitted from person to person by respiratory droplets and causes various symptoms such as fever, cough, shortness of breath and pneumonia particularly in serious cases. According to the Centers for Disease Control and

Prevention (CDC), the COVID-19 has an incubation period range from 2 to 14 days following infection [5], [6]. It has been suggested that older males with comorbidities may be at higher risk for severe illness from COVID-19 [6], [7].

Although there are supportive measures for COVID-19 and currently applied vaccines, there is no specific antiviral treatment for the outbreak [8]. Henceforth, governments have imposed various preventative measures including travel restrictions, quarantines, and event cancellations to stop collapsing of the healthcare systems [9]. The main idea behind this is to flatten the epidemic curve [10]. In Turkey, following the first confirmed case of COVID-19 on March 11, 2020 (fifty days later than many countries), the number of cases in Turkey increased sharply making it one of the top ten countries affected by COVID-19, in less than four weeks. The mortality rate in Turkey is currently lower than that in many of other countries including Italy, France, Spain, UK, and US (https://github.com/CSSEGISandData/COVID-19). Even though the infection and case fatality rates have significantly declined in some countries including Turkey thanks to vaccination and necessary preventions, showing a dim trending in disease prevention and control might result in an a third wave in epidemic [11]. In recent days, the daily cases become higher than recovery cases in Turkey and as a result, the trend in COVID-19 outbreak is now increasing. Besides, the main reasons of the COVID-19 to become a pandemic is due to unawareness of the virus infection rate, inefficient detection, insufficient and delayed preventative and protective measures by policy-makers during the early stages of the epidemic [12]. Taken together, prediction of the virus spread is of great importance for policymakers to understand future development and take effective measures for containing the COVID-19 outbreak.

Several studies have been conducted to make a model for COVID-19 in different countries [1], [2], [4], [13-15]. However, the trends of daily infections and deaths of COVID-19 and prediction of its spreading in Turkey are still poorly studied. Therefore, in the present study, we aimed at generating time series forecasts of the future of the COVID-19 based on three main variables: confirmed cases, deaths and recoveries for Turkey. We simulated a fifteen days ahead of the spreading dynamics of the epidemic using the epidemiological data available in the public domain.

II. METHODS

A. Data

COVID-19 data were retrieved from the github repository (https://github.com/CSSEGISandData/COVID-19) of the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University [16]. The datasets were based on three main variables: confirmed cases, deaths and recoveries. The data refer to daily cumulative cases and cover the period from March 11, 2020 until April 7, 2021.

B. Descriptive analysis of COVID-19 prevalence in Turkey

COVID-19 prevalence was first examined graphically. Then, we calculated mortality and recovery rates. We further analyzed the relation of confirmed cases with the number of active case and recovered people. The main aim of descriptive analysis is to reveal the current state of COVID-19 in Turkey and identify critical points that may become vital in further analysis of the virus transmission.

C. Forecasting models

Autoregressive models (AR), exponential smoothing (ES), moving averages (MA), autoregressive integrated moving average (ARIMA) are fundamental statistical techniques used for time series prediction. Meanwhile, there are also several machine learning techniques including decision trees and other ensemble methods. Besides, Facebook's Prophet (FP) is developed to forecast the behavior of users at particular times of a year in social media and became an effective algorithm which is now used to solve more complex problems [17]. To predict the future of the outbreak, we used some of the most popular regression methods such as DT, ARIMA and FP. The models were trained using the datasets from March 11, 2020 to March 7, 2021. The training data consist of daily confirmed cases, deaths and recoveries for 361 days, while the test data involves 32 days.

DT is an approach constantly used for classification and regression problems. In this algorithm, firstly the feature importance level is determined by the information gain value or the Gini-index [18]. Then, the most significant features are divided into two branches and the limit value is determined. This procedure is repeated and a decision tree is created. Prediction can be done for new values after decision tree is completed. In this work data using DT model was prepared as weekly by time shifting of historical data. The time shifting of model is data preparing phase which maps input feature vectors $x \in \{(t-1), (t-2), \ldots, (t-n)\}$ to output labels $y \in \{t\}$, where X is the feature space and n is days.

ARIMA is a forecasting technique that explains a given time series based on its history, its lags and the lagged forecast errors. It was presented by Box and Jenkins in the 1970s [19]. ARIMA model is characterized as an ARIMA (p, d, q) where p is the AR order, d is the number of differencing required to make the time series stationary, and q is the order of MA [20]. The model for the COVID-19 datasets was computed according to minimum error by grid search technique for p, d, q in range [0,4]. The ARIMA (0, 2, 1) model is used on confirmed cases. Prediction model of deaths and recoveries datasets was determined as ARIMA (3, 2, 3) and (0, 2, 1), respectively.

FP was created by Taylor and Letham to forecasting "at scale" that associates configurable models with analyst-in-the-loop performance analysis. It is coded using 'STAN', an extensive language used by Bayesian modelers, and can also be accessed from programming languages such as Python and R Studio (https://facebook.github.io/prophet). FP model was formed by combining three main model components in the following equation: trend, seasonality, and holidays (Equation 1). In the value of the time series, trend function [g(t)] is nonperiodic changes, seasonality [s(t)] represents periodic changes, and holiday effects [h(t)] represents irregular schedules like holidays. The error term ε_t shows any idiosyncratic changes that are not included by the model [17]. We designed the FP model using trend model components. The change points were determined as growth altering dates and used for forecasting.

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \quad (1)$$

D. Performance evaluation of the forecasting models

In this step, we evaluated the prediction performance of time series models by comparing the predictions of DT, ARIMA and FP models. To detect the prediction performance of models, the mean absolute percentage error (MAPE) and mean absolute error (MAE) were used as the performance metrics. These metrics are the most common measures of forecast accuracy. The formulation of MAPE and MAE can be mathematically expressed as follows:

$$\boldsymbol{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Xt - Xp}{Xt}\right| * 100 \quad (2)$$

$$\boldsymbol{MAE} = \frac{1}{n}\sum_{t=1}^{n}|Xt - Xp| \quad (3)$$

where the Xp is the predicted count model, Xt is the actual test count and n is the number of days passed from the first confirmed case. If the count Xt is 0, a value of 1 is used to avoid infinity state.

E. Forecasting

The models were compared to determine their prediction efficiencies. After the best model was determined, the datasets were not divided to train and test set as was done in Section 2.3. All datasets were separately used as the training phase to estimate the COVID-19 transmission for Turkey. The model was trained using the datasets for 361 days using best prediction model. Then, the model produces fifteen-days-ahead point forecasts and prediction intervals for confirmed cases, death and recoveries.

III. RESULTS

We performed a time series approach based on various prediction models including DT, FP and ARIMA, and evaluated these models using MAPE and MAE metrics. Later on, we built a model using FP method for the prediction of the spreading level of the pandemic by generating fifteen days ahead forecasts of the number of daily confirmed cases, deaths and recoveries for Turkey. The predictions of the future outbreak for Turkey will be useful to take necessary measures such as effective allocation of health care resources, adjusting public health and social interventions by government policymakers. This prediction model performed by data-driven analysis could provide deep insights into the study of early risk assessments for Turkey and other countries.

A. Analysis of COVID-19 for Turkey

We first presented the results for the prevalence of COVID-19 based on total confirmed cases as well as the active cases, recoveries and case fatality (Figure 1). For this purpose, the COVID-19 data were retrieved from the github repository of CSSE at Johns Hopkins University (https://github.com/CSSEGISandData/COVID-19), and the existing data show that the dates of first confirmed case is on March 11, 2020 for Turkey. Since the first case, the infection cases showed a constant icrease until April 21, 2020. However, the cases started to Show dramatic increase after this date. Meanwhile, Turkey have put forward a series of intervention and control measures, including partially lock-down, quarantining all confirmed cases, and also tracking and testing all close contacts. Figure 1 shows that there were 3633925 total cases and 32943 deaths of COVID-19 reported in the Turkey from March 11 to April 7, 2021, with a crude mortality of 0.906%. Besides, as of April 7, 2021, the number of the active cases and recoveries are 406004 and 3194978, respectively. The daily new cases peaked at 5138 cases on April 11, 2020 (mentioned as 1st wave) and 7381 cases on November 24, 2020 (mentioned as 2nd wave), while significantly increase in daily new cases by April 7, 2021 could be an indication of 3rd wave. Meanwhile, the daily deaths peaked at 127 (April 19, 2020), 259 (December 23, 2020) and 276 (April 7, 2021), while the daily recoveries peaked at 5231 on April 29, 2020 and 35511 on December 25, 2020. We also calculated the mortality and recovery rates from March 11, 2020 to the present. Figure 2 shows the change in mortality and recovery rates and the average mortality and recovery rates were found to be 1.996 and 76.534, respectively.

B. Comparison of the DT, ARIMA and FP

To build the forecast model, we first compared the prediction performances of DT, ARIMA and FP models based on two performance metrics, MAPE and MAE. For this, the dataset was divided into 361 and 32 days for training and testing, respectively. Table 1 shows the forecast accuracy of three models. MAE and MAPE measures the mean of the dispersion between predicted and observed value and the model has the least values has the best forecast accuracy. Hence, the performance of the FP method is the best with its lowest size of the error of MAE and MAPE, while DT has the highest errors for all the variable: confirmed cases, deaths and recoveries. Altogether the FP method comprehensively and reasonably estimates the future the number of variables on specific days based on the actual numbers of confirmed cases, deaths and recoveries, hence we decided to use the FP method for further forecasting analysis.

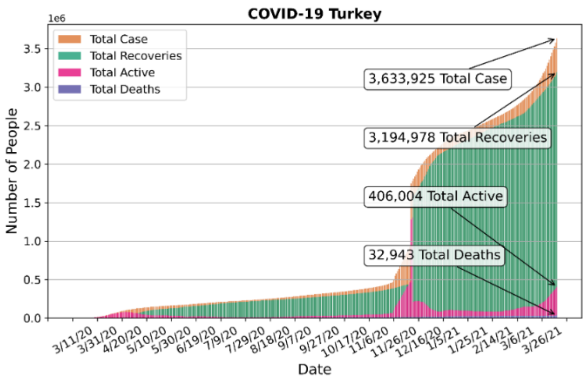

Figure 1. Cumulative distribution of COVID-19 outbreak in Turkey. Orange: total number of confirmed cases, pink: active case, green: recovered and purple: deaths. The number of active cases is calculated by subtracting the total number of deaths and total number of recovered people from the total number of infected people, and those still carrying the COVID-19 virus.

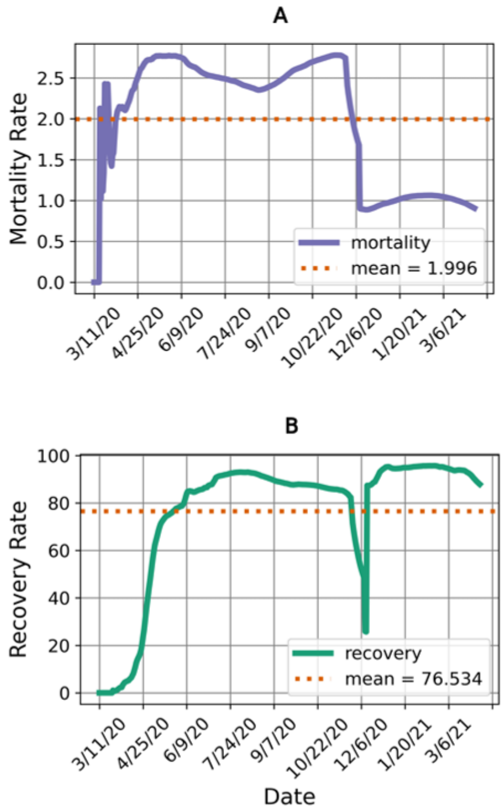

Figure 2. A) Mortality rate, B) Recovery rate. The percentage of total deaths in the total number of cases indicates the mortality rate, while the percentage of total recoveries in the total number of cases shows recovery rate. The orange dashed line represents the average value for each ratio.

Data	Method	MAE	MAPE
Case	FP	301728.48	9.19
	ARIMA	3048418.04	98.02
	DT	1895280.90	62.06
Death	FP	2495.25	8.31
	ARIMA	30463.52	99.76
	DT	14498.18	47.63
Recovery	FP	221402.88	7.55
	ARIMA	2803482.43	97.72
	DT	1346328.78	46.17

Table 1. Error rates of FP, ARIMA and DT methods for case, death and recovery values

C. FP forecasting model results

In this step, the FP model was trained using the dataset for 393 days of confirmed cases, deaths and recoveries. According to the comparison between the actual and predicted data (Figure 3), we can see that the curve of model simulation of the number of confirmed cases, the number of death people and the number of recovered cases matches the curve of actual data very well. As can be seen in Figure 3A and 3C, there are gaps through the curves due to difference between the actual numbers of cases anounced by the Turkish Health Ministry. Before November 25, 2020, people only with symptoms were accepted as cases, while later on, all people with PCR positive were defined as cases by the Turkish Health Ministry (https://covid19.saglik.gov.tr/ (Accessed: 2021-05-24)). Then, we generated fifteen-days-ahead point forecasts and prediction intervals for all the variables (Figure 4). The results of fifteen-day predictions for confirmed, death and recovered cases are shown in Table S1. The forecasts suggest that the total number of infected people could reach 4328083 with lower and upper bounds of 3854261 and 4888611, respectively by April 23, 2021. Besides, the simulations shows that death toll could reach 35656 with lower and upper bounds of 34806 and 36246, respectively. According to the curve of simulated recovered number, we infer that by April 23, 2021, 86,38% (87.65% upper and 85.52% lower bounds) of the total cases will be basically treated. Taken together, it can be inferred that our model simulates the actual data well, and it can be updated anytime since it is a data-driven model.

IV. DISCUSSION AND CONCLUSION

The COVID-19 outbreak globally demonstrates an unprecedented challenge to the world with a significant impact on public health systems, human lives, and world economies [21]. There is currently no specific antiviral treatment for the pandemic [8]. Some authorities have stated that this novel coronavirus might infect up to 70% of the

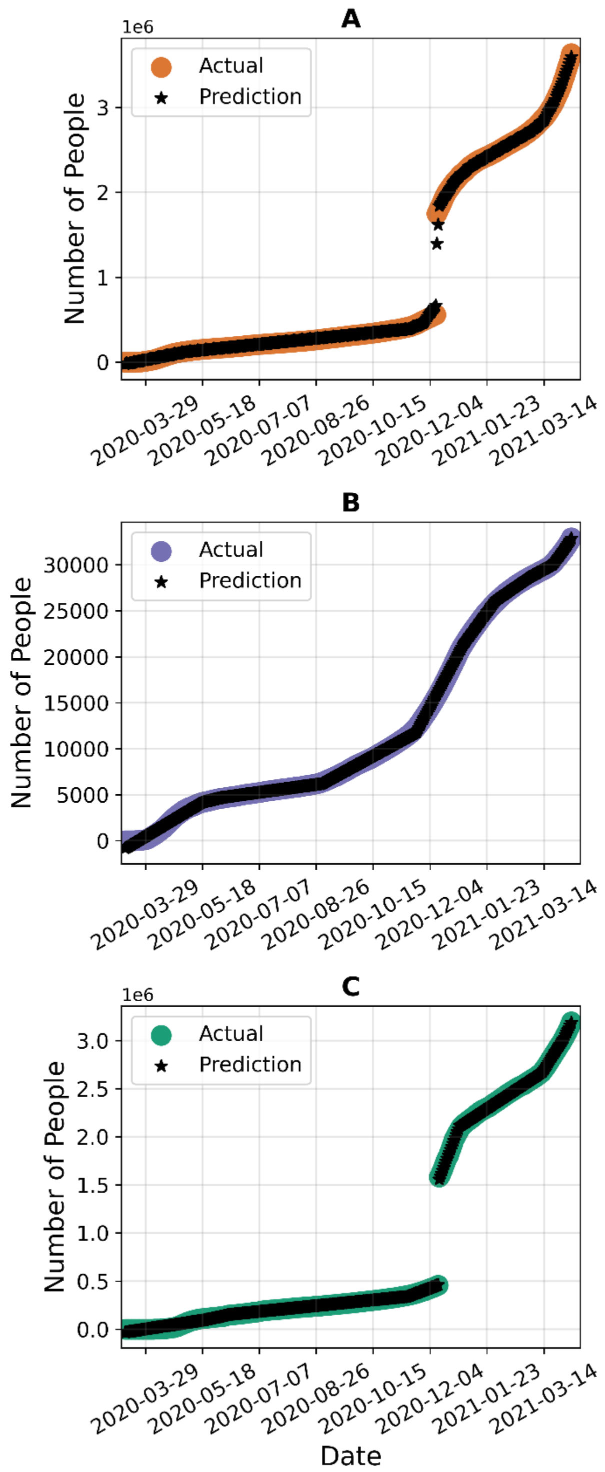

Figure 3. Training model based on the actual values and FP predicted values. A) The number of confirmed cases (orange), B) The number of deaths (purple), C) The number of recovered cases (green). The chart has points representing the actuals values and stars (*) representing the predicted values for each day.

global population in the coming year (https://www.mediaite.com/news/harvard-professor-sounds-alarm-on-likely-coronavirus-pandemic-40-to-70-of-world-could-be-infected-this-year/ (Accessed: 2021-04-07); https://www.bbc.com/news/world-us-canada-51835856 (Accessed: 2021-04-07)). It has also been indicated that the virus is spreading in nonlinear way with the existence of fat-tailed process owing to an increased connectivity globally [22]. In the meantime, there are also opposite opinions arguing that people are overlying the situation and leading people into unnecessary panic. Although some countries have currently taken the outbreak under control, COVID-19 is still spreading all over the world. Hence, showing a weak reaction in disease prevention and control could cause a second wave in pandemic [11]. Therefore, forecast modeling could be an important tool for predicting the future and transmission dynamics of such pandemics. It can provide foresights to the policy-makers in decision-making processes, particularly in public health interventions, global supply chains and economy. Besides, the policy-makers can count the worst-case-scenarios and act accordingly [23].

Although the first confirmed case was detected almost 50 days later than many other countries, the number of cases increased dramatically and Turkey is now among the top 5 countries that have high confirmed cases. We indicate that this could be because of the week preventative measures applied by Turkish Health Ministry, in the recent weeks. We also suggest that comprehensive availability of medical devices, health professionals, and an increase in vaccination in all over Turkey might be effective in lowering of pandemic rates in the near future. Taken together, all these points make Turkey a critical country to figure out the transmission dynamics of COVID-19 outbreak. Therefore, in this study, we intended to contribute to community research with forecasting time-series model that attempts to simulate and predict the spread of COVID-19 in Turkey. Fifteen-days-ahead forecast was provided to estimate three main variables: confirmed cases, deaths and recovered cases. Based on the results of our time series forecast model, we present a list of suggestions below:

- The proposed FP model is a univariate time series analysis and shows a high prediction accuracy. Besides, it is based on a data-driven data as the actual data and predictions can be regularly updated.
- We generated the forecast model using "change point" module of the FP method. This module allows us to find critical points (dramatic changes) in datasets and to form the predictions based on them. Hence, the accuracy of the model has been significantly enhanced. Besides, this may enable one to analyze the impact of measures taken in the country on the spread, particularly the critical dates of interventions such as quarantine, travel restrictions, and performing distance education. Regarding our case, we need further analysis

and detailed information about the incubation time of the virus in Turkey.

The proposed model matches with the actual data very well and is capable of accurately estimating the number of future new cases, deaths and recovered cases. This time series data-driven prediction of the pandemic variables could provide the public, society and policy-makers with more accurate transmission information. Hence, this model can be used as an early warning system and to implement the essential interventions against the COVID-19 outbreak.

- The Turkish government is gradually loosening the restrictions on COVID-19 outbreak. Based on the model estimations, the number of confirmed cases, deaths and recovered cases will be 4328083, 35656, and 3738954, respectively. In addition, the future active cases could significantly increase compared to the current active cases. This suggest that the pandemic will continue to accelerate during the next fifteen days. These findings may raise public concern and panic. However, the future of COVID-19 transmission will still depend on the government regulations and individual citizens in obeying them, particularly in maintaining self-isolation and social distance.
- The proposed model does not include the impact of any government regulations or normalization policies such as travel restrictions and return-to-work order. Such effects should be further investigated to figure out their possible influence on number of cases, second peak etc.

In conclusion, the present approach is easily interpretable, can be incorporated with new data and further updated for new forecasts. With its high prediction accuracy, the model provides an effective future estimations of the main variable of the COVID-19 pandemic which can help the public and policy-makers for the planning of health care systems and effective measures.

AUTHORS CONTRIBUTIONS

Ceyhun Bereketoglu conceived, designed and performed the analysis, and wrote the paper; Nermin Ozcan collected the data, performed the analysis, and wrote the paper; Tugba Raika Kiran performed the analysis and wrote the paper; Mehmet Lutfi Yola performed the analysis and wrote the paper.

ACKNOWLEDGEMENT

This research was supported by the Iskenderun Technical University Scientific Research Commission through project 2020 YP-04.

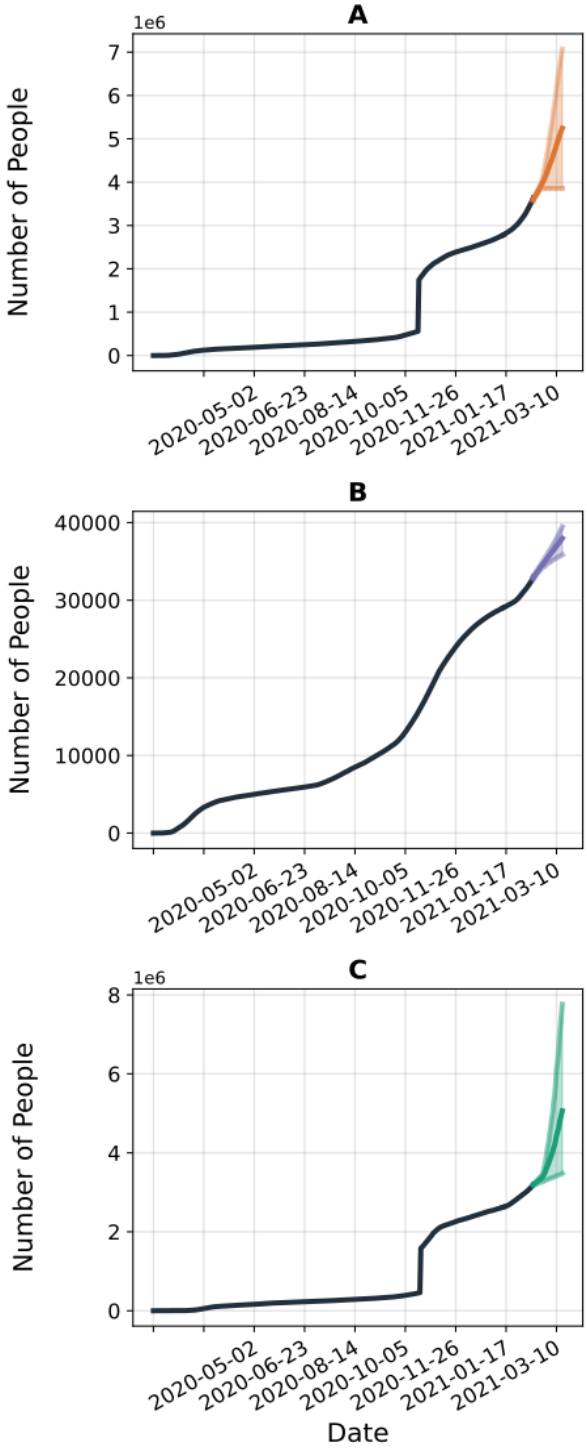

Figure 4. Simulated prediction of COVID-19 spreading trends in Turkey. A) The number of total cases, B) The number of total deaths, C) The number of total recovered cases. The FP model generates fifteen-days-ahead forecasting. The black line show the change in the real values for each variables, while the shaded colored areas indicate the lower and upper boundaries of the forecast model.

REFERENCES

[1] Zhao S, Chen H. Modeling the epidemic dynamics and control of COVID-19 outbreak in China. Quantitative Biology. 2020:1-9.

[2] Vega DI. Lockdown, one, two, none, or smart. Modeling containing covid-19 infection. A conceptual model. Science of The Total Environment. 2020:138917.

[3] 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.

[4] Song P, Karako T. COVID-19: Real-time dissemination of scientific information to fight a public health emergency of international concern. Bioscience trends. 2020.

[5] Patel A, Jernigan DB. Initial public health response and interim clinical guidance for the 2019 novel coronavirus outbreak—United States, December 31, 2019–February 4, 2020. Morbidity and Mortality Weekly Report. 2020; 69:140.

[6] Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. The Lancet. 2020; 395:507-13.

[7] Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The lancet. 2020; 395:497-506.

[8] Tu Y-F, Chien C-S, Yarmishyn AA, Lin Y-Y, Luo Y-H, Lin Y-T, et al. A Review of SARS-CoV-2 and the Ongoing Clinical Trials. International journal of molecular sciences. 2020; 21:2657.

[9] Zhu W, Li X, Wu Y, Xu C, Li L, Yang J, et al. Community quarantine strategy against coronavirus disease 2019 in Anhui: an evaluation based on trauma center patients. International Journal of Infectious Diseases. 2020.

[10] Tobías A. Evaluation of the lockdowns for the SARS-CoV-2 epidemic in Italy and Spain after one month follow up. Science of The Total Environment. 2020:138539.

[11] Newton PN, Bond KC, Adeyeye M, Antignac M, Ashenef A, Awab GR, et al. COVID-19 and risks to the supply and quality of tests, drugs, and vaccines. The Lancet Global Health. 2020.

[12] Lazzerini M, Barbi E, Apicella A, Marchetti F, Cardinale F, Trobia G. Delayed access or provision of care in Italy resulting from fear of COVID-19. The Lancet Child & Adolescent Health. 2020.

[13] Xu J, Cheng Y, Yuan X, Li WV, Zhang L. Trends and prediction in daily incidence of novel coronavirus infection in China, Hubei Province and Wuhan City: an application of Farr's law. American Journal of Translational Research. 2020; 12:1355.

[14] 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. The Lancet. 2020; 395:689-97.

[15] Chakraborty T, Ghosh I. Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: A data-driven analysis. Chaos, Solitons & Fractals. 2020:109850.

[16] Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. The Lancet infectious diseases. 2020.

[17] Taylor SJ, Letham B. Forecasting at scale. The American Statistician. 2018; 72:37-45.

[18] Liu H, Cocea M. Induction of classification rules by gini-index based rule generation. Information Sciences. 2018; 436:227-46.

[19] Box GE, Jenkins GM, Reinsel GC, Ljung GM. Time series analysis: forecasting and control: John Wiley & Sons; 2015.

[20] Shadab A, Said S, Ahmad S. Box–Jenkins multiplicative ARIMA modeling for prediction of solar radiation: a case study. International Journal of Energy and Water Resources. 2019; 3:305-18.

[21] Roda WC, Varughese MB, Han D, Li MY. Why is it difficult to accurately predict the COVID-19 epidemic? Infectious Disease Modelling. 2020.

[22] Norman J, Bar-Yam Y, Taleb NN. Systemic Risk of Pandemic via Novel Pathogens—Coronavirus: A Note. New England Complex Systems Institute (January 26, 2020). 2020.

[23] Petropoulos F, Makridakis S. Forecasting the novel coronavirus COVID-19. PloS one. 2020; 15:e0231236.

[24] Wu X, Nethery RC, Sabath BM, Braun D, Dominici F. Exposure to air pollution and COVID-19 mortality in the United States. medRxiv. 2020.

[25] Martelletti L, Martelletti P. Air pollution and the novel Covid-19 disease: a putative disease risk factor. SN Comprehensive Clinical Medicine. 2020:1-5.

Supplementary File

	Cases			Deaths			Recoveries		
date	***lower***	***mean***	***upper***	***lower***	***mean***	***upper***	***lower***	***mean***	***upper***
2021-04-08	3611530	3644024	3676880	32841	33064	33292	3189070	3208602	3224771
2021-04-09	3644613	3678803	3711707	33024	33244	33462	3194455	3227708	3250764
2021-04-10	3677534	3716135	3754537	33224	33438	33657	3200450	3247465	3278168
2021-04-11	3715677	3755400	3794854	33396	33625	33858	3214955	3268867	3303079
2021-04-12	3754177	3795508	3836966	33503	33784	34037	3229162	3290970	3329773
2021-04-13	3786245	3835673	3885160	33708	33982	34240	3234125	3309749	3355609
2021-04-14	3824189	3878379	3934420	33827	34155	34434	3240280	3330387	3385186
2021-04-15	3858200	3923283	3987343	33969	34330	34627	3253979	3353101	3413837
2021-04-16	3850947	3956977	4059252	34129	34514	34830	3260264	3374200	3444906
2021-04-17	3853793	3998830	4145719	34219	34675	35035	3274691	3404955	3496875
2021-04-18	3850156	4045317	4249482	34297	34825	35222	3275086	3446549	3596470
2021-04-19	3850043	4102964	4382825	34335	34967	35419	3290038	3493275	3696289
2021-04-20	3851202	4149886	4482198	34460	35133	35604	3308152	3553120	3832172
2021-04-21	3859890	4207642	4605863	34641	35335	35824	3319290	3625619	4012350
2021-04-22	3851101	4269222	4751616	34724	35491	36022	3324076	3685991	4162887
2021-04-23	3854261	4328083	4888611	34806	35656	36246	3334783	3738954	4285127

Table S1. Prediction results of FP model fifteen-days-ahead for total cases, deaths and recoveries