DATA ANALYTICS AND SIR MODELING
OF COVID-19 IN BULGARIA

Miroslava Ivanova[1], Lilko Dospatliev[2] [§]

[1]Trakia University
Department of Informatics and Mathematics
Stara Zagora, 6000, BULGARIA
[2]Trakia University
Department of Pharmacology, Animal Physiology
and Physiological Chemistry
Stara Zagora, 6000, BULGARIA

Abstract: The Novel Human Coronavirus (SARS-CoV-2) also well-known as COVID-19 is the greatest public health challenge of the 21st century. The aims of this article were to provide data analytic of the COVID-19 cases in Bulgaria and to modeling their spread using SIR (Susceptible-Infected-Recovered) model. We used the covid19.analytics package from the statistical software R in extracting time-series data of COVID-19 in Bulgaria (dating March 8, 2020 to November 24, 2020), in presenting the data analysis as well as forecast the maximum number of infected people, peak time and the basic reproduction number based on SIR model. As per SIR model, the maximum number of infected people is reached after 61 days of starting of COVID-19 pandemic in Bulgaria (around May 13, 2020) with about $1.298446e^{+05}$ infections. The basic reproduction number R_0 was found to 1.46, which means that on average an infectious individual infects 1.46 susceptible individuals during his infection period. We believe that performed data analytics of COVID-19 cases in Bulgaria and the obtained results of the SIR model will help government of Bulgaria when restricting the spread of the virus.

[§]Correspondence author

1. Introduction

On December 31, 2019, a cluster of cases of pneumonia of unknown cause was detected in Wuhan City, Hubei Province of China (see [43]). It was revealed in trace-back investigations that many of the patients who were infected by the virus had some link to seafood and wet animal market located in Wuhan; which highlighted the possibility of the virus spreading from animal-to-human (see [11]). However, the virus continued to spread through large chunks of the population, it was proved that many of the patients had no exposure to animal markets, which brought up the unpleasant possibility of the occurrence of person-to-person spread (see [36]). This virus was given the name Novel Human Coronavirus (SARS-CoV-2) or known as COVID-19 (see [20]). Within days the virus causes complications in the respiratory tract of the infected person and causing an increase of austere fatal deaths (see [6]).

On January 30, 2020, the World Health Organization (WHO) acknowledged COVID-19 as a pandemic due to its faster spread (see [3], [30]). Despite the on-going efforts of governments, the virus has spread almost everywhere in the world (see [34]).

Social distancing, washing hands frequently, avoiding touching the mouth, nose, and face etc. are important preventive measures for COVID-19 are suggested (see [41]).

The virus was confirmed to have spread to Bulgaria when the country's first cases, a 27-year-old man from the northern city of Pleven and a 75-year-old woman from the central city Gabrovo, were confirmed on March 8, 2020. Neither of the two had traveled to areas with known coronavirus cases. The man tested positive for the virus after being hospitalized for a respiratory infection, and authorities announced plans to test several people who were in contact with the two individuals (see [9]). Two other samples in Pleven and Gabrovo were positive on March 8, 2020 (see [13]). Patient zero remains unknown (see [16]). Now the virus has been spread throughout the country. Bulgaria is in third place in the European Union-European Economic Area in COVID-19 mortality and in 13th place in morbidity, on the basis of 14-day statistics, Chief State Health Inspector Associate Professor Angel Kunchev told the Health Ministry's weekly briefing on November 19, 2020. As per our analysis date (Neveber 24, 2020) there are 124966 total confirm cases, with 38226 recoveries and 3069 deaths in the country (see [18]). Administration and health officials are facing lot of problem to accommodate patients of COVID-19. So, development of some prediction tools to know expected number of cases in upcoming time is the need of hour for future preparations (see [37], [39]).

Since the beginning of the COVID-19 epidemic, there has been various mathematical and statistical modelling that have predicted the global and national epidemic with varying degrees of accuracy and reliability (see [1], [2], [4], [10], [21], [25], [27], [28], [29], [32], [35], [40]). The accuracy of prediction and its uncertainty depend on the assumptions, availability and quality of data (see [31]). The results can vary significantly if there is difference in the assumptions, and values of input parameters. Moreover, an epidemic may not always behave in the same manner as pathogens are likely to behave differently over time (see [42]).

The aims of this article were to provide data analytic of the COVID-19 cases in Bulgaria and to modeling their spread using SIR (Susceptible-Infected-Recovered) model.

2. Materials and methods

2.1. Epidemiological data

In this study we used time series data on the daily progression of the COVID-19 pandemic in Bulgaria from the John Hopkin's University (JHU) (see [14]). We employed the current Bulgarian population based on World meter elaboration for the latest United Nations data (November 24, 2020) as the initial uninfected population ($N = 6927712$ (see [19])) source. We generated a vector with the daily cumulative incidence for Bulgaria, from March 8, 2020 (when our daily incidence data started), through November 24, 2020.

2.2. SIR model

The SIR (Susceptible-Infected-Recovered) model has been extensively utilized by scientists and government policymakers to predict the spread of many diseases. In this study we used a simple SIR model also known as compartmental model or stochastic epidemic model (see [22], [24], [35], [38]) to predict and model the spread of Coronavirus in Bulgaria. SIR model divides the population into three compartments Susceptible (S) (not infected), Infected (I) and Recovered (R) (that is, vaccinated or recovered with immunity).

The SIR model without demographics is given by the following non linear

system of ordinary differential equations (see [12], [24], [38]):

$$\frac{dS(t)}{dt} = -\beta S(t)I(t),$$
$$\frac{dI(t)}{dt} = \beta S(t)I(t) - \gamma I(t),$$
$$\frac{dR(t)}{dt} = \gamma I(t).$$

The function $S(t)$ represents the number of susceptible cases at time t. The function $I(t)$ represents the number of infected cases at time t, and the function $R(t)$ is the compartment of removed cases from the disease, where:

$$\text{Removed cases} = \text{Death cases} + \text{Recovered cases}.$$

The parameters β and γ are called the contact rate and the removed rate, respectively. The initial condition of this model $S(0)$, $I(0)$, $R(0)$ must satisfy the condition:

$$S(0) + I(0) + R(0) = N,$$

where N is the population size (see [12]).

The model can be written also as follows:

$$\frac{dS(t)}{dt} = -\beta \frac{S(t)}{N} I(t),$$
$$\frac{dI(t)}{dt} = \beta S(t)\frac{I(t)}{N} - \gamma I(t), \tag{1}$$
$$\frac{dR(t)}{dt} = \gamma I(t),$$

with $\frac{S(t)+I(t)+R(t)}{N} = 1$ and $S(0) > 0$, $I(0) > 0$, $R(0) = 0$ (see [23]).

An important epidemiological parameter in modelling of infectious diseases is the basic reproduction number denoted by R_0 (see [5]). The basic reproduction number for a pandemic in each region is determined based on the "expected number of secondary cases produced by a single (typical) infection in a completely susceptible population" is given by (see [22]):

$$R_0 = \frac{\beta}{\gamma}. \tag{2}$$

If R_0 is greater than one suggests rapid transmission of the infectious disease whereas if the value of R_0 is less than one, it indicates a decreasing transmission (25) (see [26]). If R_0 is equal to one, it means that each existing infection causes one new infection. For COVID-19, the World Health Organization (WHO) suggests an R_0 ranging between 1.4 to 2.5 (see [26]). Some of the other estimated range of values of R_0 for different countries are: 2.6 to 3.2 for South Korea and 2.6 to 3.3 for Italy (see [45]); and 2.24 to 3.58 for China (see [44]). In [26] is made a meta-analysis of 12 studies which estimated R_0 for COVID-19 in various parts of the world and it is suggested the mean R_0 value of 3.28.

2.3. Covid19.analytics package

Covid19.analytics package (see [15]) is an R package developed and maintained by Marcelo Ponce. The package was released on 3rd May, 2020 and is available on the CRAN Repository as well as on Github. The package uses live data from the John Hopkin's University (JHU) (see [14]) to perform several analytics such as time series, regression, etc. It also uses the SIR model for the prediction of the pandemic.

All data of the reported incidences of coronavirus in Bulgaria (dating March 8, 2020 to November 24, 2020) were extracted from the covid19.data function from covid19.analytics package in the statistical software R program version 3.6.3.

The analysis and visualizations functions from covid19.analytics package used in this study were tots.per.location, growth.rate and totals.plt.

The generate.SIR.model function (see [17]) in covid19.analytics package was used for implements the SIR model from (1) using the actual data from the reported cases. The function identify data points where the onset of the epidemic began and consider the following data points to generate proper guesses for the two parameters describing the SIR ODE system, i.e. β and γ. It does this by minimizing the residual sum of squares (RSS) assuming one single explanatory variable, i.e. the sum of the squared differences between the number of infected cases $I(t)$ and the quantity predicted by the model $\widetilde{I}(t)$,

$$RSS(\beta, \gamma) = \sum_t \left(I(t) - \widetilde{I}(t) \right)^2.$$

The ODE given by (1) is solved numerically using the ode function from the deSolve and the minimization is tackled using the optim function from base R. After the solution for (1) is found, the function provide details about the solution, as well as, a plot of the quantities $S(t)$, $I(t)$, $R(t)$. The generate.SIR.model function also estimates the value of the basic reproduction number R_0, defined by (2) (see [33]).

3. Results and discussion

A plot analysis of Bulgaria showing the relationship between the number of cases and date from which the first case was identified is presented in Figure 1. It can be observed the steady rise of the number of cases started from the end of July.
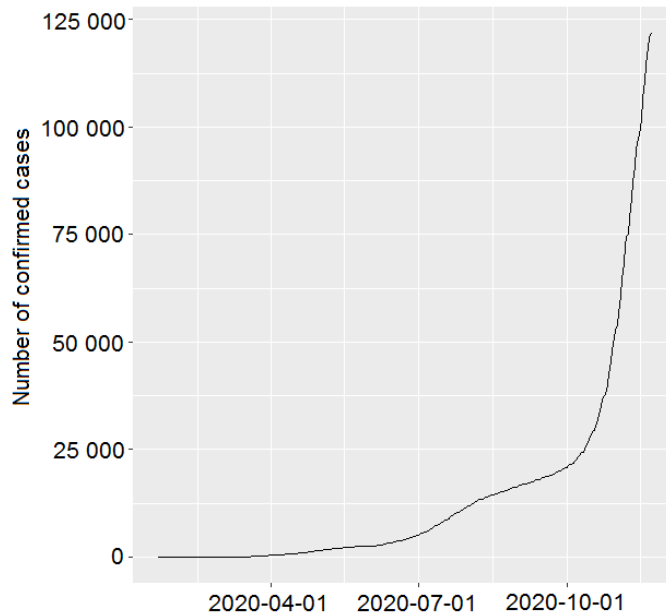
Figure 1: Number of confirmed COVID-19 cases in Bulgaria

The government of Bulgaria has declared a nationwide influenza epidemic and took instant and decisive measures from March 6 to 11. These included mandatory suspension of all educational activities, ban on access to all public places such as theatres and cinemas, and suspension of all types of public events. Hospitals were ordered to postpone elective surgeries and ban patient visits (see [7]). The Minister of Health of the Republic of Bulgaria, Kiril Ananiev has introduced a temporary ban on entry into the territory of the Republic of Bulgaria from April 6, 2020 (see [8]).

These decisive measures led to a reduction in the spread of the infection. There was a weak first wave with a peak around April 20, 2020, when the new cases reached 91 per a day. There was a clear decrease in morbidity over the next month and a half, of which, due to the implemented measures. After restrictions were lifted and borders were opened, cases began to rise rapidly and right now Bulgaria is in "second" wave. The Bulgarian Health Ministry reported a record high of 4828 new COVID-19 cases in a 24-hour period on November 11, 2020. The percentage of positive samples was 41.58%.

Figure 2 A) presents a plot diagram of the number of cases in logarithm and number of days. It is observed a sharp rise in the number of cases from day 45. Day 100 presented a steady rise in the number of cases as compared
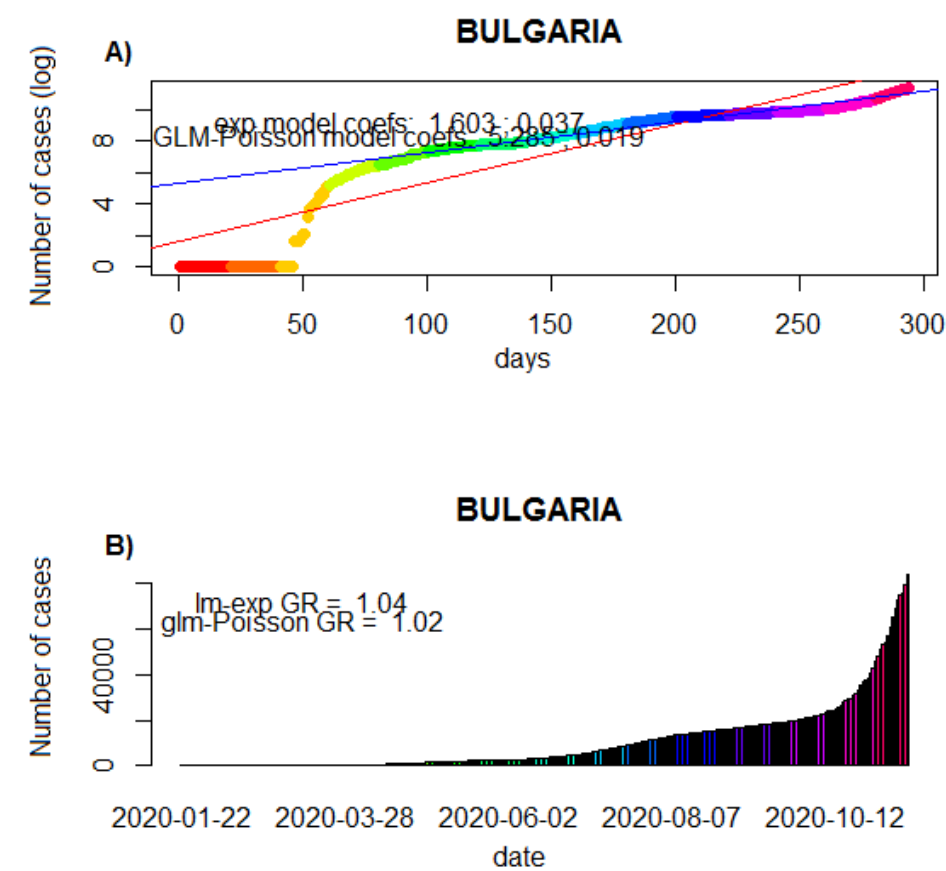
Figure 2: Number of cases (in log) and the growth rate of COVID-19 in Bulgaria

to day 45 increment. Figure 2 B) gives a bar diagram of steady increment of the virus in Bulgaria. It is observed from the plot that there was increased COVID-19 positive cases from Juny 2, 2020. Cases reached more than 80000 as from November 11, 2020 which showed peak value.

Figure 3 A) depicts graphically the number of changes in confirmed cases to the time of the month. It is observed from the graphic, there was a steady difference in changes to number of confirmed cases from the end of January to beginning of March when the first cases were detected. As cases gradually increased from middle of March, the difference remained almost same till beginning of June when there was a sharp reduction in the difference in the number of cases. The cases picked up again towards the second half of June and con-
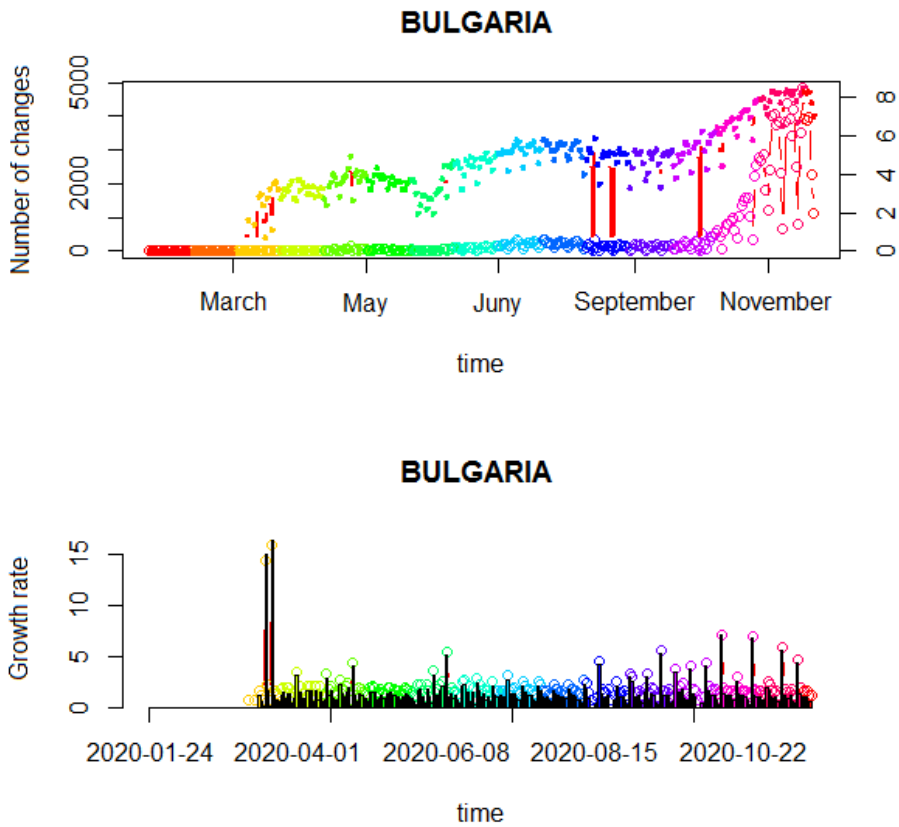
Figure 3: Number of changes (difference in the number of cases) and
the growth rate of COVID-19 in Bulgaria

tinually widened. Figure 3 B) gives a bar chart of the growth rate of the virus
in Bulgaria. It is observed that growth rate peaked on April 1, 2020. 10th of
October also witnessed peak but not as much as the previous date stated.

Figure 4 shows log-scale representation of the global number of confirmed,
active, recovered and deaths of COVID-19 cases vs. the reported cases for
Bulgaria. It is observed that the growth of the epidemic in Bulgaria with the
rest of the world is the same pattern, only that the values are less than that of
the rest of the world.

Figure 5 present the total number of confirmed, recovered, deaths and active
cases in Bulgaria. It is observed from the graph the number of death cases is
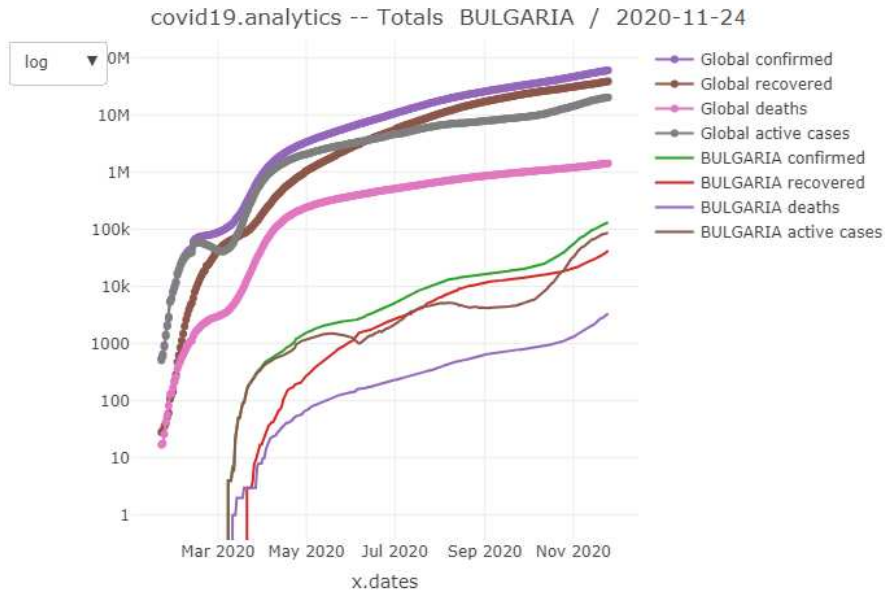
Figure 4: Log-scale representation of the total number of cases vs. the reported COVID-19 cases for Bulgaria

significantly low as compared to the confirmed and recovered cases although the number of death cases is gradually increasing but kept as the barest minimum. Confirmed cases continually increases rapidly as against the recovered, deaths and active cases.

Figure 6 shows the result of the prediction of the SIR model on the COVID-19 cases in Bulgaria implemented by generate.SIR.model function from covid19. analytics package. The first two graphs shows the model accuracy for the infected cases based on the first 25 days of data. The right graph of them is log scale in which the y-axis is logarithm. It is also called log-linear plot or semi-log plot. The dotted line in the plot represents the real data, while the straight line represents the model. The second two graphs show the summary of those who are susceptible to the virus, infected and recovered. The maximum number of infected people is reached after 61 days of starting of this pandemic (around 13 May 2020) with about $1.298446e^{+05}$ infections. Recovery rate is also increasing with the time. It is also observed that the blue line still hight, showing that there is still high risk of susceptibility. A large number of the population is still at the risk of infection if urgent measures are not taken, such as partial lockdown of the country.

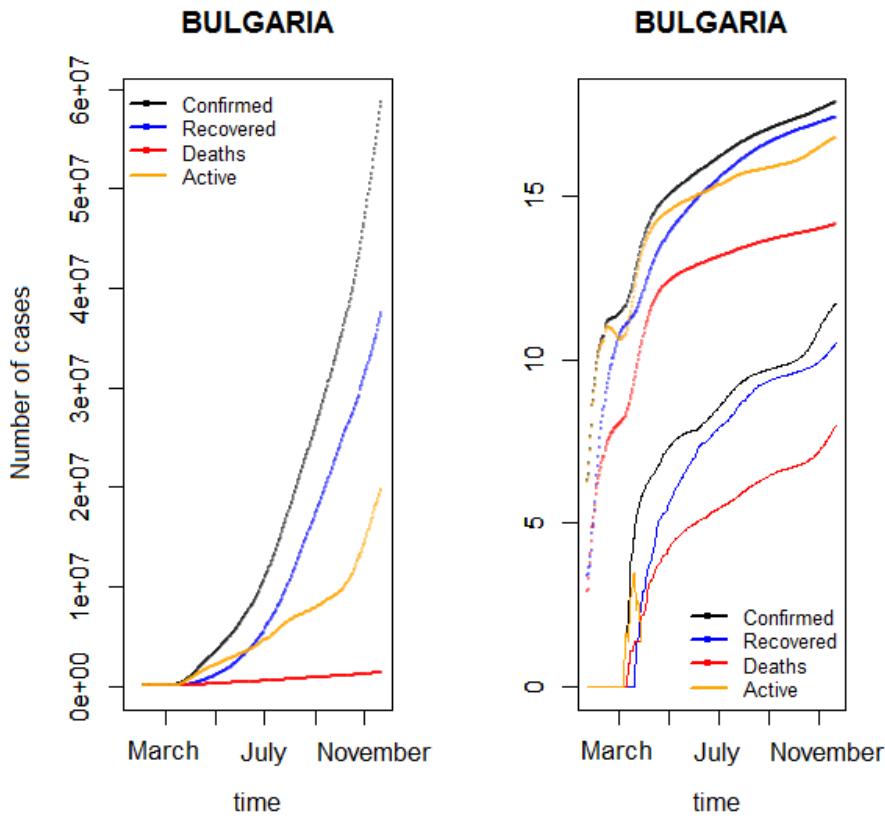After running the SIR model on the available Bulgaria data and in addition

Figure 5: Relationship between the number of confirmed, recovered, death and active COVID-19 cases in Bulgaria with respect to time

to Figure 6 the generate.SIR.model function from covid19.analytics package gives the results for the beta value, the gamma value and the basic reproduction number R_0 which are 0.59, 0.41 and 1.46, respectively. From the obtained value for R_0 it follows that on average an infectious individual infects 1.46 susceptible individuals during his infection period. This obtained value for the basic reproduction number R_0 belong the range between 1.4 to 2.5 suggested of the World Health Organization (WHO) (see [26]).
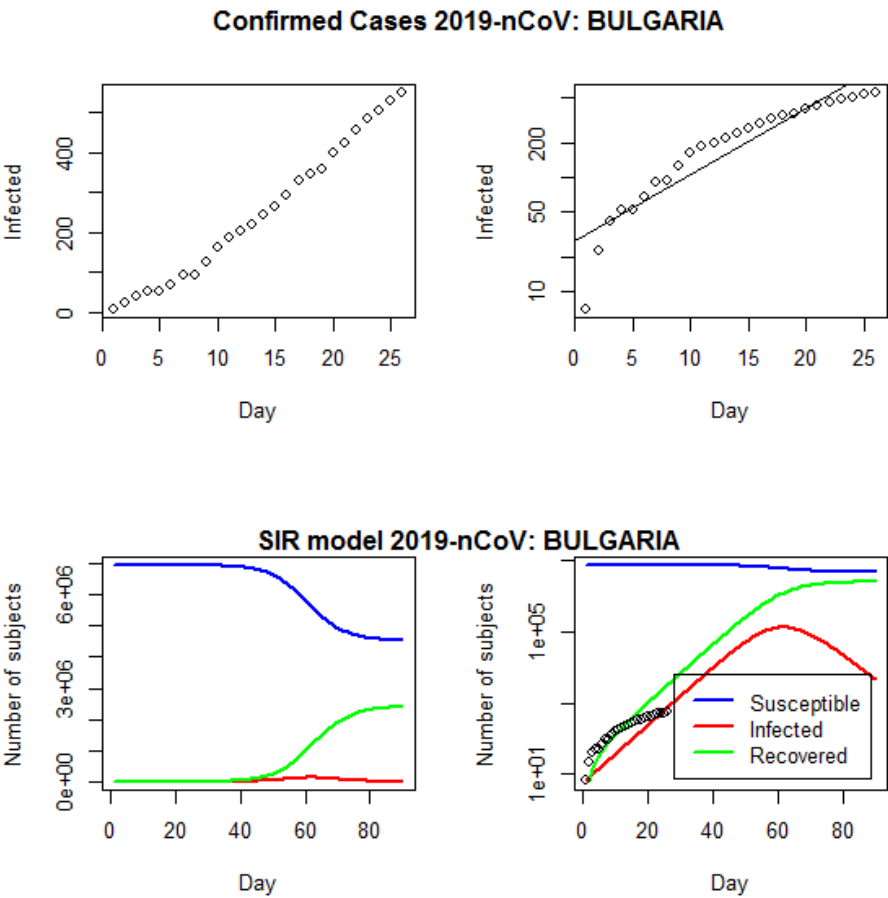
Confirmed Cases 2019-nCoV: BULGARIA

SIR model 2019-nCoV: BULGARIA

Figure 6: SIR model prediction for COVID-19 cases in Bulgaria

4. Conclusion

Predicting the COVID-19 is still a challenge for epidemiologists and data researchers because not much information has been gathered that will of a certainty give the full characteristics of the disease. In this paper, we present data analysis and visualization of COVID-19 cases in Bulgaria based on the number of recovered, deaths, active and confirmed cases. SIR model was used in simulating the susceptible, infected and recovered COVID-19 cases in Bulgaria and the obtained results of the model were presented via graphs. Due to unreported cases, we may have deficiencies of inaccurate data but nevertheless we

believe that this analysis and visualizations will help government of Bulgaria when restricting the spread of the virus.

References

[1] L. Abadias, G. Estrada-Rodriguez and E. Estrada, Fractional-order Susceptible-Infected model: definition and applications to the study of COVID-19 main protease, *Fractional Calculus and Applied Analysis*, **23**, No 3 (2020), 635-655; doi: 10.1515/fca-2020-0033.

[2] A.M. Almeshal, A.I. Almazrouee, M.R. Alenizi et al., Forecasting the spread of COVID-19 in Kuwait using compartmental and logistic regression models, *Applied Sciences*, **10**, No 10 (2020), # 3402; doi: 10.3390/app10103402.

[3] T.M. Alon, M. Doepke, J. Olmstead-Rumsey et al., The impact of COVID-19 on gender equality, *National Bureau of Economic Research* (2020); doi: 10.3386/w26947.

[4] N. Anand, A. Sabarinath, S. Geetha et al., Predicting the spread of COVID-19 using SIR model augmented to incorporate quarantine and testing, *Trans. Indian National Academy of Engin.*, **5** (2020), 141-148.

[5] L.M. Bettencourt and R.M. Ribeiro, Real time Bayesian estimation of the epidemic potential of emerging infectious diseases, *PLoS One*, **3**, No 5 (2008), Art. e2185.

[6] M. Blofield, B. Hoffmann and M. Llanos, Assessing the political and social impact of the COVID-19 crisis in Latin America, *GIGA Focus Latein Amerika*, **3** (2020); Available from URL: `https://www.ssoar.info/ssoar/bitstream/handle/document/67260/ssoar-2020-blofield_et_al-Assessing_the_Political_and_Social.pdf?sequence=1&isAllowed=y&lnkname=ssoar-2020-blofield_et_al-Assessing_the_Political_and_Social.pdf`.

[7] Bulgaria, Ministry for Health, Order No RD-01-114/05.03.2020; Available from URL: `https://www.mh.government.bg/media/filer_public/2020/03/05/rd-01-114_u7ofbPI.pdf` (in Bulgarian).

[8] Bulgaria, Ministry for Health. Order No RD-01-183/06.04.2020; Available from URL: `https://www.mh.government.bg/media/filer_public/2020/04/07/rd-01-183.pdf` (in Bulgarian).

[9] Bulgaria, Moldova Report First Coronavirus Cases, Radio Free Europe/Radio Liberty, 8 March 2020; Available from URL: https://www.rferl.org/a/bulgaria-confirms-coronavirus-cases-amid-flu-outbreak/30475539.html.

[10] Y. Fang, Y. Nie and M. Penny, Transmission dynamics of the COVID-19 outbreak and effectiveness of government interventions: A data-driven analysis, *J. of Medical Virology*, **92**, No 6 (2020), 645-659.

[11] R.T. Gandhi, J.B. Lynch and C. del Rio, Mild or moderate COVID-19, *The New England J. of Medicine*, **383** (2020), 1757-1766.

[12] T. Harko, F.S.N. Lobo and M.K. Mak, Exact analytical solutions of the Susceptible-Infected- Recovered (SIR) epidemic model and of the SIR model with equal deaths and births, *Applied Mathematics and Computation*, **236** (2014), 184-194.

[13] https://botevgrad.com/news/78275/Prof-Todor-Kantardjiev-Ot-Pleven-imame-dve-polojitelni-probi-za-koronavirus-ot-Gabrovo-sashto/ (in Bulgarian).

[14] https://github.com/CSSEGISandData/COVID-19.

[15] https://CRAN.R-project.org/package=covid19.analytic.

[16] https://www.actualno.com/healthy/oficialno-ima-chetiri-potvyrdeni-sluchaja-na-koronavirus-v-bylgarija-news_1440246.html (in Bulgarian).

[17] https://www.rdocumentation.org/packages/covid19.analytics/versions/2.0/topics/generate.SIR.model.

[18] https://www.worldometers.info/coronavirus/country/bulgaria/.

[19] https://www.worldometers.info/world-population/bulgaria-population.

[20] C. Huang, Y. Wang, X. Li et al., Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China, *Lancet*, **395**, No 10223 (2020), 497-506.

[21] N. Hussain and L. Baoming, Using R-studio to examine the COVID-19 patients in Pakistan implementation of SIR Model on cases, *Intern. J. of Scientific Research in Multidisciplinary Studies*, **6**, No 8 (2020), 54-59.

[22] J.H. Jones, Notes on R0. Califonia: Department of Anthropological Sciences. Department of Anthropological Sciences Stanford University, (2007), Available from URL: `http://people.math.gatech.edu/~dmargalit7/classes/math1553Spring2020/JonesR0.pdf`.

[23] M. Keeling and P. Rohani, *Modeling Infectious Diseases in Humans and Animals*, Princeton University Press, Princeton, NJ, USA (2008).

[24] W.O. Kermack and A.G. McKendrick, A contribution to the mathematical theory of epidemics, *Proc. Royal Soc. of London, Ser. A*, **115** (1927), 700-721.

[25] O. Kounchev, G. Simeonov and Z. Kuncheva, The TVBG-SEIR spline model for analysis of COVID-19 spread, and a Tool for prediction scenarios, *arXiv Preprint*: arXiv (2020), 2004.11338, 1-21.

[26] Y. Liu, A.A. Gayle, A. Wilder-Smith and J. Rocklöv, The reproductive number of COVID-19 is higher compared to SARS coronavirus, *J. of Travel Medicine*, **27**, No 2 (2020), 1-4.

[27] B. Malavika, S. Marimuthu, M. Joy et al., Forecasting COVID-19 epidemic in India and high incidence states using SIR and logistic growth models, *Clinical Epidemiology and Global Health* (2020); doi: 10.1016/j.cegh.2020.06.006.

[28] S. Margenov, N. Popivanov, I. Ugrinova et al., Mathematical and computer modeling of COVID-19 transmission dynamics in Bulgaria by time-depended inverse SEIR model, *arXiv Preprint*: arXiv (2020), 2008.10360, 1-16.

[29] T. Marinov and R. Marinova, Dynamics of COVID-19 using inverse problem for coefficient identification in SIR epidemic models, *Chaos, Solitons & Fractals: X*, **5** (2020); doi: 10.1016/j.csfx.2020.100041.

[30] W.J. McKibbin and R. Fernando, The global macroeconomic impacts of COVID-19: Seven scenarios, *CAMA* (2020); doi: 10.2139/ssrn.3547729.

[31] K.R. Moran, G. Fairchild, N. Generous et al., Epidemic forecasting is messier than weather forecasting: the role of human behavior and internet data streams in epidemic forecast, *J. Infectious Diseases*, **214**, No 4 (2016), S404-S408.

[32] U. Nguemdjo, F. Meno, A. Dongfack et al., Simulating the progression of the COVID-19 disease in Cameroon using SIR models, *Plos One* (2020); doi: 10.1371/journal.pone.0237832.

[33] M. Ponce and A. Sandhel, covid19.analytics: An R Package to Obtain, Analyze and Visualize Data from the Corona Virus Disease Pandemic, *arXiv Preprint*: arXiv (2020); 2009.01091, 1-47.

[34] A. Rajesh, H. Pai, V. Roy et al., CoVID-19 prediction for India from the existing data and SIR(D) model study, *arXiv Preprint*: MedRxiv (2020); doi: 10.1101/2020.05.05.20085902.

[35] W.C. Roda, M.B. Varughese, D. Han et al., Why is it difficult to accurately predict the COVID-19 epidemic? *Infectious Disease Modelling*, **5** (2020), 271-281.

[36] P. Sun and K. Li, An SEIR model for assessment of current COVID-19 pandemic situation in the UK, *arXiv Preprint*: MedRxiv (2020); doi: 10.1101/2020.04.12.20062588.

[37] A. Tobías, Evaluation of the lockdowns for the SARS-CoV-2 epidemic in Italy and Spain after one month follow up, *Science of the Total Environment*, **725**, No 138539 (2020); doi: 10.1016/j.scitotenv.2020.138539.

[38] J. Tolles and T.B. Luong, Modeling epidemics with compartmental models, *JAMA Guide to Statistics and Methods*, **323**, No 24 (2020), 2515-2516.

[39] D. Wang, B. Hu, C. Hu et al., Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China, *JAMA*, **323**, No 11 (2020), 1061-1069.

[40] J. Wangping, H. Ke, S. Yang et al., Extended SIR prediction of the epidemics trend of COVID-19 in Italy and compared with Hunan, China, *arXiv preprint*: medRxiv (2020); doi: 10.1101/2020.03.18.20038570.

[41] WHO, World Health Organization Coronavirus disease (COVID-19) advice for the public, (2020); Available from URL: `https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public`.

[42] WHO, World Health Organization Managing epidemics: Key facts about major deadly diseases, (2018); Available from URL: `https://www.who.int/emergencies/diseases/managingepidemics-interactive.pdf`.

[43] WHO, World Health Organization Pneumonia of unknown cause - China (2020); Available from URL: `https://www.who.int/csr/don/05-january-2020-pneumonia-of-unkown-cause-china/en/`.

[44] S. Zhao, Q. Lin, J. Ran et al., Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak, *Intern. J. of Infectious Diseases*, **92** (2020), 214-217.

[45] Z. Zhuang, S. Zhao, Q. Lin et al., Preliminary estimating the reproduction number of the coronavirus disease (COVID-19) outbreak in Republic of Korea and Italy by 5 March 2020, *Intern. J. of Infectious Diseases*, **95** (2020), 308-310.