






Changes in Patterns of Infectivity and Mortality with SARS-CoV-2 Omicron Variant in Bulgaria

Latchezar P. Tomov¹ , Hristina Batselova^{2,3} , and Tsvetelina Velikova⁴ 

¹ Department of Informatics, New Bulgarian University, 1618 Sofia, Bulgaria
lptomov@nbu.bg

² Department of Epidemiology and Disaster Medicine, Medical University, Plovdiv, Bulgaria
dr_batselova@abv.bg

³ Department of Epidemiology and Disaster Medicine, University Hospital “St George”,
6000 Plovdiv, Bulgaria

⁴ Medical Faculty, Sofia University St. Kliment Ohridski, 1407 Sofia, Bulgaria
tsvelikova@medfac.mu-sofia.bg

Abstract. We use our previously developed models (branching processes and time series analysis) to track the changes in the spread of the new SARS-COV-2 variant “Omicron” across age groups. We track changes in transmission rates, lethality, and spread between age groups. Omicron was inferred to be less lethal in other studies. We test that by comparing the prediction from our model, which was trained with the data with the previous variants – wild type, alpha and delta variants. We automatically predict new cases by combining the branching process and change point analysis – a reactive approach for precise short-term prediction - the software is installed on the Avitohol supercomputer. We use time series analysis (regression with Arima errors) to predict deaths and to test hypotheses related to the patterns of spread across age groups and the factors that influence it.

Keywords: Branching processes · Arima · Covid-19 · School closures · Lethality

1 Introduction

1.1 Modeling Tools for Covid-19

The classical approach in epidemiology uses deterministic modeling with nonlinear differential equations to test a different hypothesis for certain epidemic after it ends. Usually, it is based on evidence gathered during the different waves related to the incubation period of the pathogen, the basic reproduction number R_0 , the mode of transmission, et cetera [1]. With these estimated characteristics, the classical model, such as S-I-R (susceptible-infected-recovered) and its various modifications, such as S-E-I-R (susceptible-exposed-infected-recovered), are set up. Then some simulations are made to compare real-world data with the (usually very noisy). Why simulations? Because these equations are nonlinear and analytical solutions do not exist. In 2014 an exact so-called solution for a simple particular case of the S-I-R model was developed. Still, it

involves an integral that can only be solved numerically [2]. These compartmental models have their merit, even though within the deterministic framework, the extinction of a pathogen does not exist. However, they are good as a tool for analysis but have limitations in prediction [3]. Unfortunately, COVID-19 still hasn't ended, and it is not predictable when, how, or if it will end, so post hoc analysis is impossible. New approaches are needed to forecast an ongoing global epidemic in which the main characteristics of the pathogen change rapidly while one dominant variant is replaced by another, and there are no periods with zero new cases, with weak seasonality. Especially a pandemic for which various measures have been taken at various times in various countries to contain cases and, via them – deaths. A new approach, a reactive approach, is needed. We need to be able to forecast for a short-term period without being able to assess precisely the characteristics of tens of different viral variants or be able to predict what the governments and the citizen will do in response to a wave. Thus, we implemented two parallel approaches – one for predicting new cases [4] and one for deaths [5]. The first approach uses a branching process to model the spread and change-point analysis to recalibrate the model automatically from the data with changepoint analysis [6] by estimating the parameters of the whole probability distribution and only then producing a point estimate. The point estimate is made by testing 10 000 different scenarios on a supercomputer in R and calculating the expectation. This is a stochastic approach, which is more suitable for populations that are not very large since the deterministic models are, in a sense, the limit of the stochastics for population tending to infinity. Different stochastic models, such as those described in [7], alleviate some of the difficulties but still lack the ability to adapt to an ongoing epidemic. The second approach uses time series analysis in the form of ARIMA with external regressors with `auto.arima()` in R [8] to infer the deaths from the new daily cases by age groups with different lags. This model is suitable when the errors of estimation are autocorrelated, which is the case in pandemics with multiple waves. It was created in 2021 for the delta wave and shows two major findings – the exponential distribution of the risk per age up to 70 years and the children (0–19 years) are a major factor in the spread – a key regressor despite their low personal risk.

1.2 Modeling Waves in Bulgaria Prior to the SARS-CoV-2 Omicron Variant

Summary of previous research. Prior to Omicron, Bulgaria had three major waves – the wild type from September 2020 to January 2021, the alpha variant from February 2021 to May 2021 and Delta from July 2021 to January 2022. We estimated the case fatality rate and found that alpha was not significantly more lethal but infected older people and caused more deaths. Delta was substantially more lethal, with 18 000 deaths officially only for Delta and 12 000 from the wild type and the alpha combined. [9]. We were able to estimate the case fatality rate by age groups thanks to access to data for deaths per age group. Our previous paper [5] used a more sophisticated model to predict the total number of weekly deaths from new cases by age group since no data for deaths

per age group was accessible at the time. It was an ARIMA model, with added new cases by different age groups and the variants as predictors (the index of the variant as a categorical predictor). It was fitted for the period 6.6.2020–4.11.2021. It worked well for all variants up to and including the Delta variant, which suggests their common nature with the same age groups, and their lags suggest the same mode of transmission and delay between infection and symptom onsets. So, when a novel variant that is significantly different appears, than our model should give predictions for deaths that are incorrect, and the previous regressors would no longer be the optimal ones, with different optimal lags and relative weight of different age groups. In this paper, we extend this model with one more month of Delta wave, which doesn't change its structure or efficiency and try to predict the mixed Delta-Omicron wave after 31.12.2021 that happened in Bulgaria.

2 Modeling the Mixed Omicron-Delta Wave

2.1 Predicting New Daily Cases

We used our mixed model with a branching process and changepoint analysis [4]. We have to note that this model is successfully applied to over 30 different countries from April 2020 up to July 2022, including for Omicron or even for complex mixed waves such as our wave from January 2022 up to March 2022. Results are shown in Figs. 1, 2,

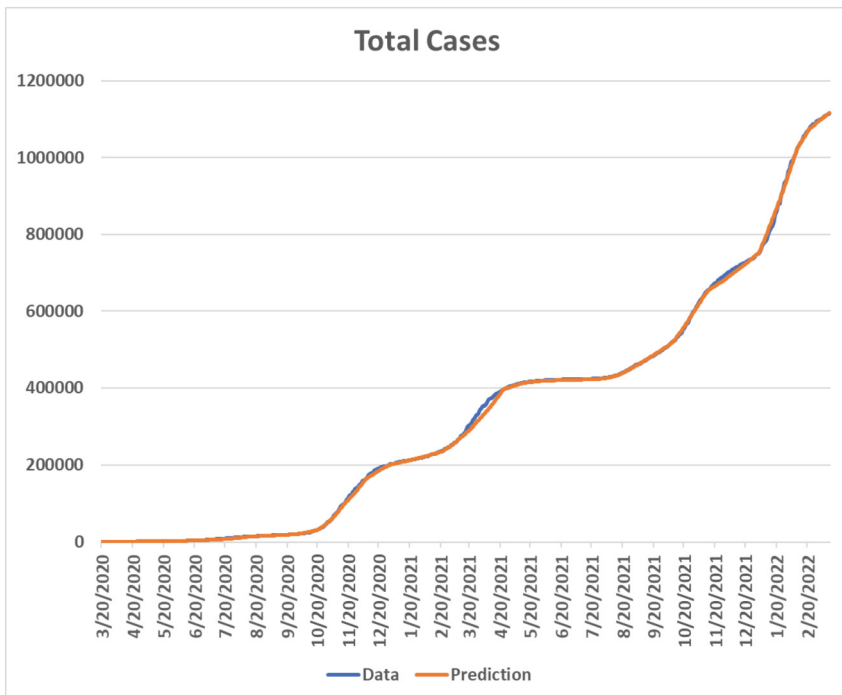


Fig. 1. Prediction for accumulated cases up to 21th of February 2022.

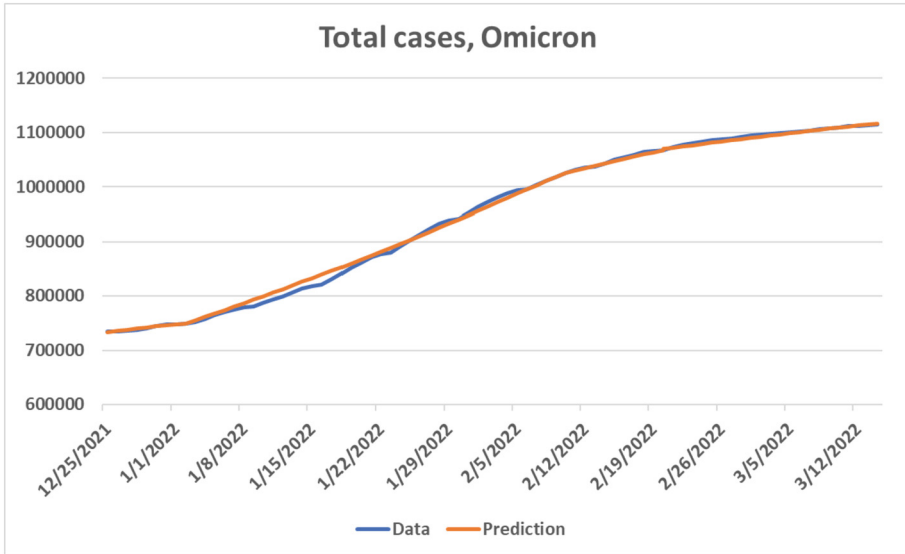


Fig. 2. Prediction for accumulated cases for the mixed Delta-Omicron wave up to 20th of February 2022.

3 and 4. Our model predicts the total or accumulated cases, and the new daily cases are the first finite difference. In Figs. 3 and 4, you can see the discontinuities that reflect the automatically detected regime change. You can also see the significantly higher wave in terms of new daily cases with the appearance of Omicron. Unfortunately, we cannot say how much higher from the new cases are because the testing varies by waves, so further research is needed.

2.2 Predicting Deaths

Here we use our second approach by extending the model [5]. Our data for new daily cases by age group and deaths and comes from the open government portal [10]. It starts on 6.6.2020. The end is 14.3.2022 for this model. We use the data up to four weeks before the end of the pure Delta wave – up to 4.12.2021. We predict the period 4.12–31.12 as a test of the model. Then we try prediction with this model for the mixed wave that started on 1.01.2022. We compare this prediction with the actual official deaths from Omicron to see how well our model is still adequate and if Omicron is similar to all previous variants in characteristics that our models captured very well. We aggregate cases by age group weekly to filter out the noise in the data. Age groups are 0–19, 20–29, 30–39, 40–49, 50–59, 60–69, 70–79, 80–89, and 90+ years old. There is a categorical predictor for the variants with four values. The four variants are Wild type, Alpha, Delta, and Omicron (values 0, 1, 2, 3). A limitation to this model is that we assign value 3 (Omicron) to a wave that starts as a mixture of two variants, and Omicron becomes dominant along the wave. We need further research to predict such complex waves with two different variants.

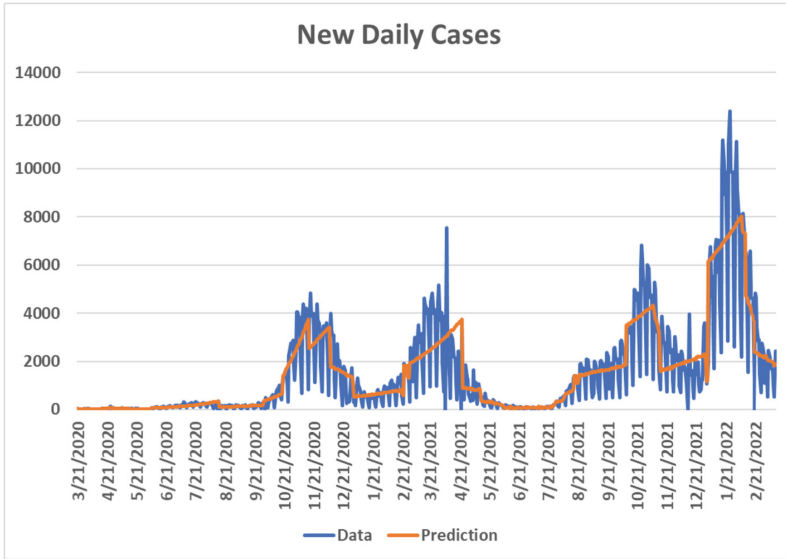


Fig. 3. A derived prediction for new daily cases up to 14th of March 2022.

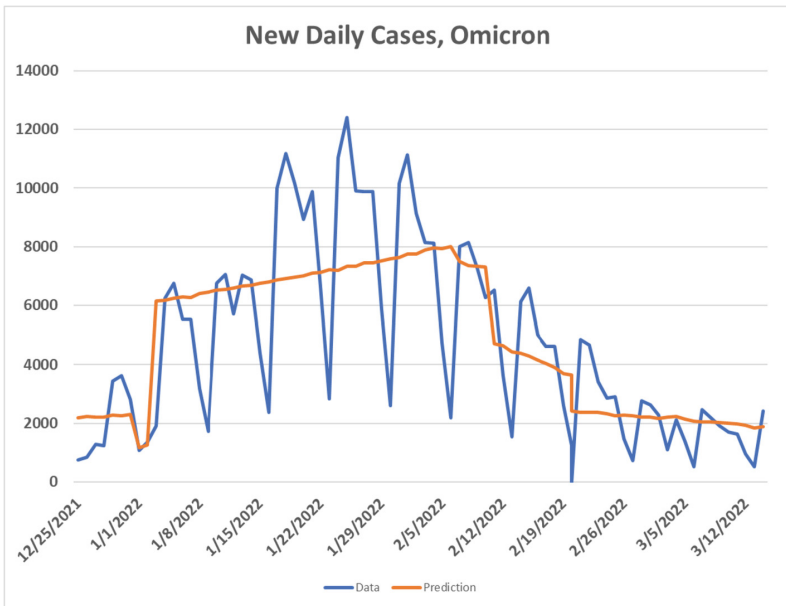


Fig. 4. A derived prediction for new daily cases for the mixed Delta-Omicron wave up to 14th of March 2022.

In building our model, we follow the same steps as described in our previous publication. [5]. We examine the stationarity with different tests. According to the Augmented

Dickey-Fuller test, the observed series are stationary after second-order differencing. So, our ARIMA model necessarily must include second-order differencing to avoid spurious correlations. We made a principal component analysis, shown in Fig. 5. As in our previous publication, we can see a high level of clustering of age groups into two main directions – 20–49 and 60+, with the 0–19 visibly very much apart and 60–90+. The reason is that, just as in previous waves, the pandemic starts from the children who infect their parents. Our question is why this age group significantly influences overall deaths despite having the lowest possible risk and why we have more considerable lags in our models. People in the age group usually have no children left in the household – 50–59, which in Bulgaria means neither parents nor grandparents are visibly alone with a lower correlation. After various tests we made, we extracted predictor by aggregating different age groups – 0–19 was left that way, while 20–29, 20–30 and 40–49 (the parent group) was aggregated. We also aggregated the “active grandparent” group by adding the somewhat less correlated 50–59 into it – 50–69. By “active grandparent,” we mean the usual age at which grandparents can care for children in Bulgaria. However, these categories are mostly based on anecdotal evidence since such research does not exist in Bulgaria. However, this is not relevant to the procedure of building the model per se but only to the interpretation of the results, which may change in the future with new demographic data for the period of the pandemic. Older people above 70 are grouped into one category.

The correlations between the selected optimal regressors after second order difference are generally acceptable, with high value for the parent and grandparent groups, which is further evidence of the role of children (no lag between them also). Here L1 means a lag of one week, L2 – two weeks, etc. (Table 1).

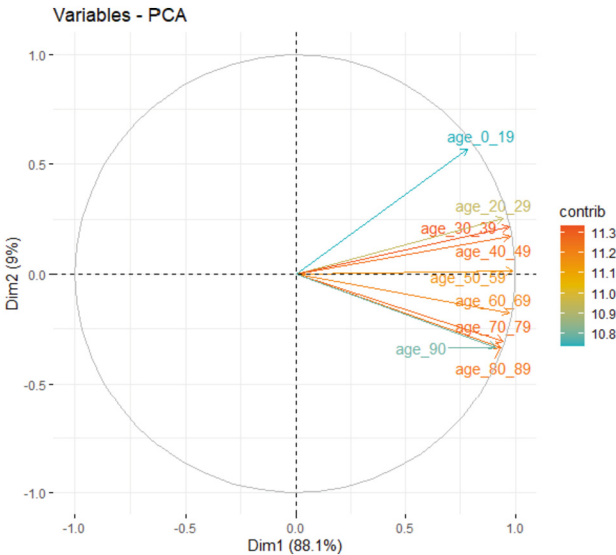


Fig. 5. Principal component analysis of new weekly cases by age groups.

The best model, selected by us and `auto.arima()`, is ARIMA with second-order differencing and second-order autoregressive terms or ARIMA (2, 2, 0). Its predictive ability is shown in Fig. 6, and its forecasting ability – is in Fig. 7. The parameters of the model and the characteristics of the fit, such as the R squared $-R^2$, the standard error of estimation of the coefficients, the mean squared error (MSE), mean absolute percentage error (MAPE) are shown in Table 2. We can see the high value of R^2 , which shows that the model explains more than 97% of the variation and that almost all standard errors of coefficient estimations are at an acceptable level (less than half of the value of the coefficient). There are some interesting things to note about these coefficients. First, the age group 0–19 has a very high coefficient value, noting a high contribution to the deaths, on the same level as the 70 + group. The lowest-risk group is as important as the highest-risk one in predicting deaths for the best model! Our interpretation is in the influence children have on the infection of the parents and grandparents as the main drivers of the pandemic, which is supported for other countries in the latest research [11]. Second, the 70+ group has the lowest or the best ratio of the standard error to coefficient value, which means we are most certainly in that value. This is logical as this is the highest-risk group. The negative coefficients of the autoregressive predictors with one- and two-week lag and the 20–49 group with three weeks lag reflect the nature of the pandemic wave, captured best by compartmental models such as SIR – the idea of the peak of the wave, caused by reaching enough share of people to slow transmission with temporary herd immunity.

Table 1. Correlations between the variables

Variables	0–19 L2	20–49 L1	20–49 L3	50–69 L1	70-plus L0
0–19 L2	1	0.27	–0.36	0.22	0.31
20–49 L1	0.27	1	–0.17	0.87	0.17
20–49 L3	–0.36	–0.17	1	–0.09	–0.15
50–69 L1	0.22	0.87	–0.09	1	0.20
70 + L0	0.31	0.17	–0.15	0.20	1

We tested this model for the rest of the pure Delta wave, and we consider this to be a relatively good forecast, as shown in Fig. 7, including with confidence intervals of 95%.

Predicting SARS-CoV-2 Omicron-Related Deaths

In the case of Bulgaria, we don't have a pure Omicron wave but a mixture of Delta and Omicron variants. This wave, as seen in Fig. 4, was around 10 weeks in length but with very high and early peak around the fourth week, indicating significantly higher effective reproductive and basic reproductive ratios for this variant mixture, mostly due to Omicron. Using our model, we tried to forecast up to the 14th of March, and as it can be seen, the actual deaths were 53% of the predicted ones – something that most probably own to the Omicron. The model underpredicted Delta deaths before Omicron

Table 2. Regression models with ARIMA (2, 2, 0) errors

Model Summary		
	Optimal Model	
Coefficient	Estimate	Standard error
AR1	-0.8119	0.1093
AR2	-0.3899	0.1103
Age 0–19 Lag 2	0.0801	0.0448
Age 20–49 Lag 1	-0.0380	0.0139
Age 20–49 Lag 3	0.0194	0.0054
Age 50–69 Lag 1	0.0803	0.0200
Age 70-plus Lag 0	0.0364	0.0233
Variants ^a	0.7161	34.6957
R²:	0.972	
RMSE:	55.78	
Bias:	-0.97	
MAPE:	17.42	

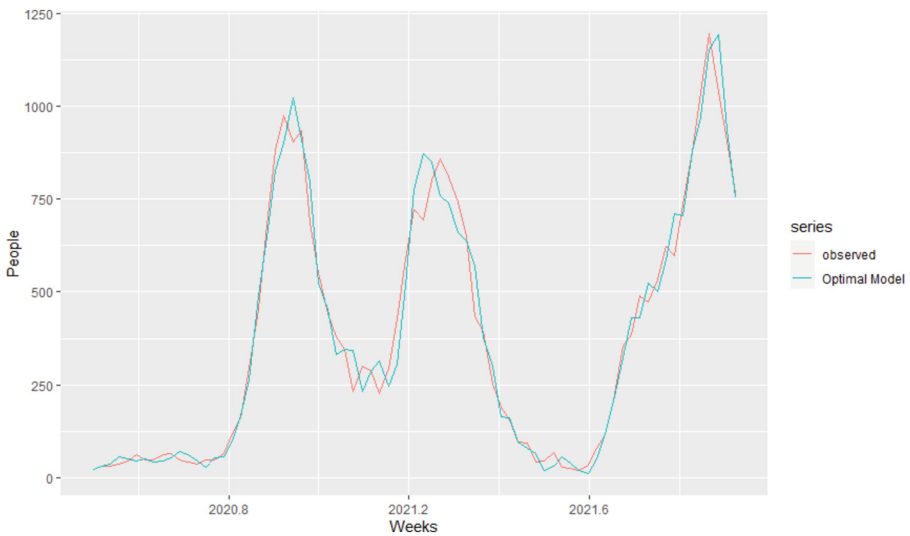


Fig. 6. ARIMA (2, 2, 0) with external regressors for 6.6.2020–4.12.2021.

became dominant in the third week of January 2022. A forecast gets less and less reliable with the extension of the forecasting period, and the uncertainty of how much “milder” Omicron was in that wave is not very low. Still, we consider this enough evidence to

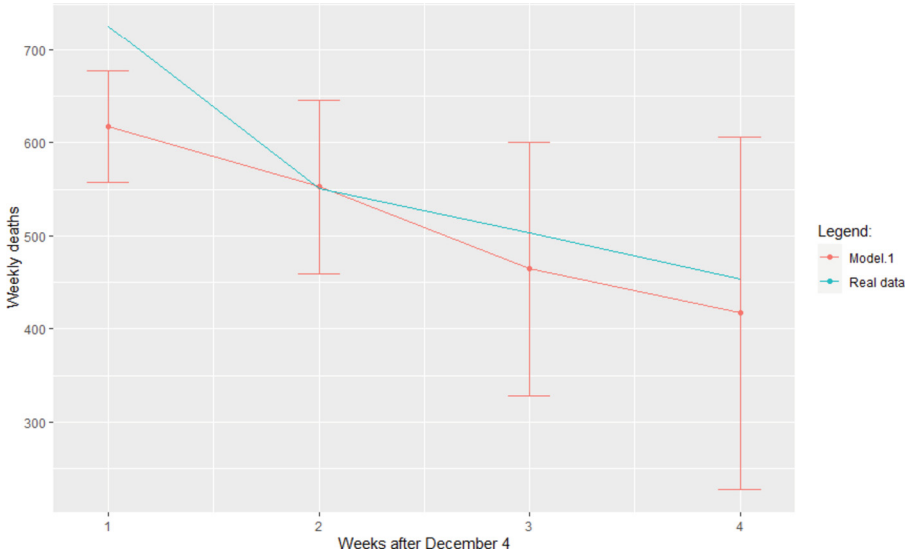


Fig. 7. Prediction for Delta – 4.12.-31.12.2021.

support the more straightforward binary statement that a significant difference exists between deaths from an Omicron and a Delta wave (Fig. 8).

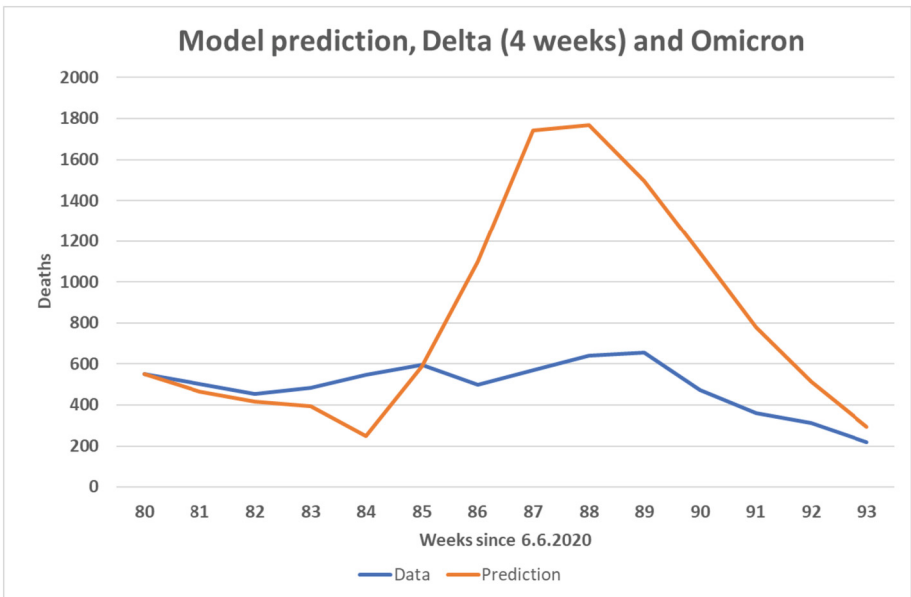


Fig. 8. Prediction for Delta and Omicron –2021/12/4–2022/3/14

3 Conclusion

In this paper, we tested several hypotheses. The first one is that our model from 2021 [5] is suitable for prediction and forecasting for the whole Delta wave in Bulgaria and will remain relatively unchanged when fitted with more data. Our research supports this hypothesis. Another hypothesis is that in Bulgaria, children play a central role in transmission. The evidence here seems to support it due to the surprisingly large weight of the new cases in 0–19 groups in predicting deaths and the fact that they have the largest lag, which was visible just from plotting the recent cases with a moving average with 7 days period as we did in [5]. The third hypothesis was that Omicron is significantly different in its transmissibility and lethality from previous variants in Bulgaria, and the evidence here supports it. For more precise estimations of exactly how much more transmissible and less lethal Omicron was, we will have to use much more sophisticated approaches due to the mixed nature of the wave from January to March 2022 and the various differences in testing volume by time and place. The approach with time-series in predicting deaths here showed a certain limitation – there is autoregressive dependence between cases in one wave and previous waves, and such also exists for mortality. The huge difference in deaths in Bulgaria could also be because, before Omicron, almost 30 000 people died from COVID-19 as an official figure. This is a very high share of the total population of 6.45 million people [12], which limits further deaths since there is an upper limit on the percentage of people that a virus can kill in a limited time frame. An Omicron wave with much lower previous waves would result in many more deaths. Other factors should be taken into account, such as the different vaccination status of the population in various stages of the pandemic and the uncertainty of the data to the anecdotal evidence for fake certificates as mass practice [13]. Nevertheless, this approach and model were suitable to show us some meaningful evidence supporting all three hypotheses. They can be a valuable tool in predicting ongoing pandemics and studying their different aspects as a supplement to stochastic compartmental models and the reactive modeling approach.

Acknowledgment. This study is financed by the European Union-NextGenerationEU, through the National Recovery and Resilience Plan of the Republic of Bulgaria, project № BG-RRP-2.004–0008-C01.

References

1. Brauer, F.: Compartmental models in epidemiology. *Math. Epidemiol.* **1945**, 19–79 (2008). https://doi.org/10.1007/978-3-540-78911-6_2, PMID: PMC7122373
2. Harko, T., Lobo, F., Mak, M.K.: Exact analytical solutions of the Susceptible-Infected-Recovered (SIR) epidemic model and of the SIR model with equal death and birth rates. *Appl. Math. Comput.* **236**, 184–194 (2014). <https://doi.org/10.1016/j.amc.2014.03.030>
3. Melikechi, O., Young, A.L., Tang, T., et al.: Limits of epidemic prediction using SIR models. *J. Math. Biol.* **85**, 36 (2022). <https://doi.org/10.1007/s00285-022-01804-5>
4. Tchordadjieff, A., Tomov, L.P., Velez, V., Dezhov, G., Manev, V., Mayster, P.: On regime changes of COVID-19 outbreak. *J. Appl. Stat.* (2023). <https://doi.org/10.1080/02664763.2023.2177625>

5. Tomov, L., Angelov, S., Tchorbadjieff, A.: Age-specific mortality risk from COVID-19 in Bulgaria. In: Zlateva, T., Goleva, R. (eds.) *Computer Science and Education in Computer Science*, New Bulgarian University, Sofia, pp. 1–24 (2021)
6. Chen, J., Gupta, A.: On change point detection and estimation. *Commun. Stat. Simul. Comput.* **30**, 665–697 (2013)
7. Mamis, K., Farazmand, M.: Stochastic compartmental models of the COVID-19 pandemic must have temporally correlated uncertainties. *Proc. R. Soc. A*.4792022056820220568 (2023). <https://doi.org/10.1098/rspa.2022.0568>
8. Hyndman, R.J., Khandakar, Y.: Automatic time series forecasting: the forecast package for R. *J. Stat. Softw.* **26**(3) (2008)
9. Tomov, L., Batselova, H., Velikova, T.V.: Estimating COVID case fatality rate in Bulgaria for 2020–2021. In: Zlateva, T., Goleva, R. (eds.) *Computer Science and Education in Computer Science. CSECS 2022. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, vol 450. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-17292-2_9
10. Bulgarian open data portal. <https://data.egov.bg/data/resourceView/e59f95dd-afde-43af-83c8-ea2916badd19>. Accessed 10 June 2023
11. Tseng, Y., Olson, K.L., Bloch, D., Mandl, K.D.: Smart Thermometer-based participatory surveillance to discern the role of children in household viral transmission during the COVID-19 pandemic. *JAMA Netw. Open* **6**(6), e2316190 (2023). <https://doi.org/10.1001/jamanetwo rkopen.2023.16190>
12. <https://www.bta.bg/en/news/economy/448417-bulgarian-population-numbers-6-447-710-as-at-december-2022-shrinks-by-34-774-pe>. Accessed 10 June 2023
13. <https://www.rferl.org/a/fake-covid-certificates-romania-bulgaria/31507555.html>. Accessed 10 June 2023