Transactions on Materials, Biotechnology and Life Sciences
ISSN: 2960-1932, eISSN: 2960-2319 | Volume 3
BLSME 2024

Research of the Spread of Covid-19 in World Wide and Particular Countries

Liqiao Zhu [1, *], Yifeng Peng [2], Lancong Xie [3], Yiwen Chen [4]

[1] YALI High School International Department, Changsha, China

[2] PIMA community college, 85704, Tucson, U.S.

[3] Wuhan Britain-China School, Wuhan, 430000, China

[4] Nanjing Normal University Yancheng Experimental School, Nanjing, 224001, China

* Corresponding Author Email: 1812231102@mail.sit.edu.cn

Abstract. To understand the spread of COVID-19, the data sets from Google and GitHub are obtained by using Python and R, and ARIMA model is established for prediction. By visualising the data sets, it can be found that the total trend of the COVID-19 is bimodal and reaches its peaks during the winter and spring time in 2022 and 2023, and this may due to the virus activity and government management. When taking into prediction, ARIMA (3,1,3) is the best predict model to forecast the overall spread trend of COVID-19 as the data is not normal distributed since it keeps fluctuating in the P-P plot, and it can be considered as time series, showing a seasonal pattern. This conclusion is drawn by calculating the values of AIC and BIC to see which model has smallest. At the same time, the ACF and PACF index could determine the p and q value in the model. The model can help to predict the spread of similar diseases.

Keywords: COVID-19; time series; ARIMA; Python; spread trend.

1. Introduction

The COVID-19 pandemic, an unprecedented global phenomenon, has wrought profound changes across the globe, sparing no region from its far-reaching impact. In the wake of soaring infection rates, overwhelmed healthcare systems, and an agonizing toll of lives lost, the urgency of unravelling its mysteries has become glaringly evident. The investigation into the origins and transmission of COVID-19 is no mere scientific pursuit; it stands as a pressing global imperative, poised to shape the trajectory of public health, international cooperation, and the destiny of humanity itself [1].

COVID-19's rapid and extensive dissemination, and the most efficacious measures to curtail its transmission loomed large [2]. Having no effective treatment curing diseases and facing the rapid-spread virus, extreme means of management was adopted; For example, Huhan in China was placed under lockdown to control the spread of COVID-19. The government imposed a strict quarantine on the city and restricted all non-essential travel, thus impose negative effect on the public's daily life. If the number of future infected individuals can be predicted as soon as possible, after the discovery of cases, it is crucial for public health management and the rational allocation of medical resources [3].

Furthermore, the investigation into the spread of COVID-19 has yielded crucial revelations concerning transmission dynamics. It has illuminated the virus's diverse modes of dissemination, its capacity for asymptomatic propagation, and the effectiveness of varied public health measures. This newfound wisdom has empowered governments and healthcare systems with the requisite discernment to enact judicious decisions regarding lockdowns, mask mandates, vaccination initiatives, and travel restrictions. In essence, it has shepherded our collective response to the pandemic, becoming a lodestar that has guided people toward the salvation of innumerable lives [4].

Equally pivotal in this endeavour is the role of international collaboration. The pandemic's global scope has mandated cooperation between nations, scientists, and international organizations. This

cooperative spirit transcends the simple exchange of scientific data; it stands as a testament to humanity's capacity to unify when confronted with a shared peril. Initiatives such as COVAX have witnessed nations collaborating to secure equitable access to vaccines, recognizing the imperative of global immunization in quelling the pandemic and staving off the spectre of future contagions [5].

The investigation into the spread of COVID-19 has further exposed the vulnerabilities within our public health systems, underscoring the urgent need for enhanced preparedness. As the virus traversed international boundaries with unprecedented alacrity, it laid bare deficiencies in surveillance, early warning systems, and response infrastructure [6]. Governments and organizations worldwide now find themselves engaged in a critical revaluation and fortification of their pandemic preparedness strategies, assiduously striving to ensure a more robust response to the exigencies of forthcoming health crises.

In our study, we collect data sets from Google and GitHub, and use ARIMA as a predict model. The model can forecast 3 weeks ahead from when the data is collected. The model can help the government to action earlier and allocate medical resources.

2. Predication and Analyzation of Virus Spread

2.1. Data sources

The datasets of patients are all from Centre for Disease Control and Prevention of America. In this paper, the time line is from January 11th of 2020 to September 30th of 2023(Figure 1). The figure is drawn using data from such research dataset.

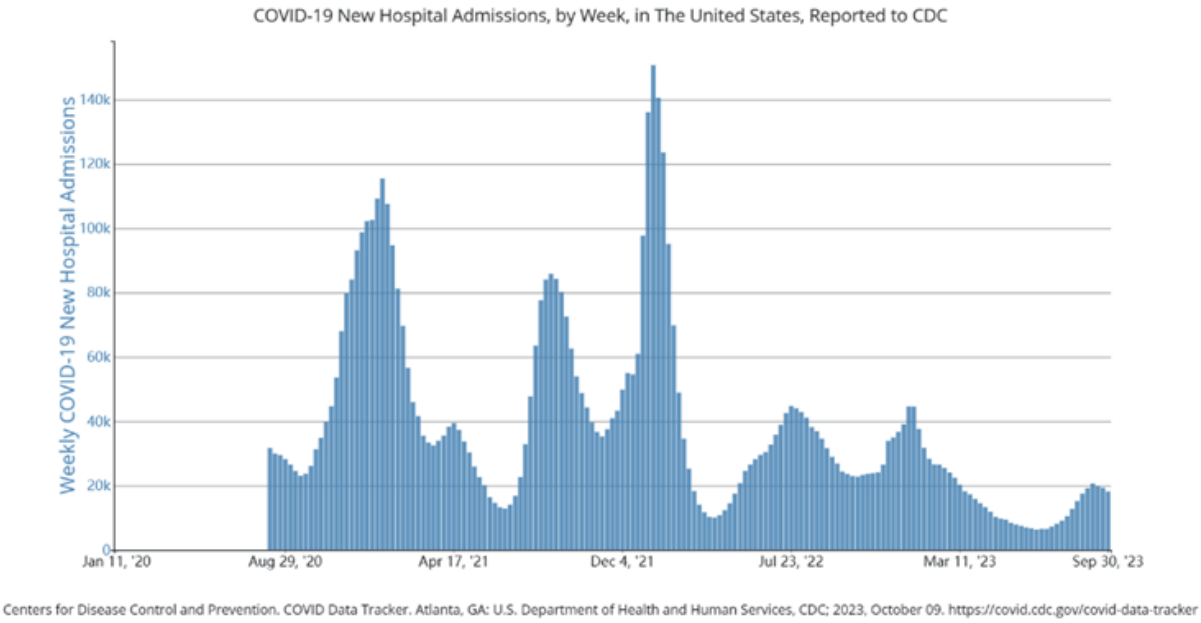

Figure 1. COVID-19 New Hospital Admissions, by Week, in The United States

Google Community mobility report shows the shift of the population mobility intensity [7]. The data is from Feb 17, 2020 to Oct 15, 2022. The categories include in the data is Grocery and Pharmacy stores, including grocery markets, farmers' markets, and pharmacies; Parks, including national parks, public beaches and gardens; Transit Stations, including subways, bus station and train stations. Work place, including companies, offices, etc. Residential areas, including residential communities, apartments, and so on. Retail and Recreation, including restaurants, shopping centres, and cinemas (figure 2).

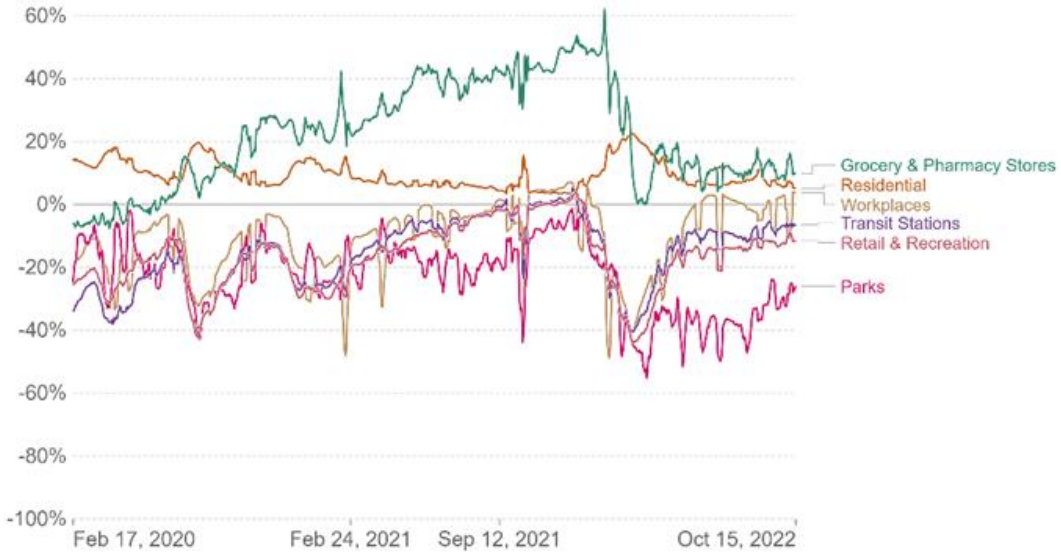

Figure 2. Google Community Mobility Report during COVID-19

2.2. Research protocol

The Python library was developed by Google in collaboration with the Brain Team. It contains the exact syntax, semantics, and tokens of Python. Researchers also use this Python library to solve complex computations in Mathematics and Physics.

The Python library consists of more than 200 core modules. Matplotlib is responsible for plotting numerical data. That's why it is used in data analysis. It is also an open-source library and plots high-defined figures like pie charts, graphs, etc. Pandas is an important library for data scientists. It is an open-source machine learning library that provides flexible high-level data structures and a variety of analysis tools. Pandas support operations like Sorting, Re-indexing and Conversion of data, etc. NumPy is a library that supports large matrices and multi-dimensional data. Array interface is one of the key features of this library [8].

SciPy is an open-source library used for high-level scientific computations. It works with NumPy to handle complex computations. While NumPy allows sorting and indexing of array data, the numerical data code is stored in SciPy. Scikit-learn supports variously supervised and unsupervised algorithms like linear regression, classification, clustering, etc. PyGame is used for developing video games using computer graphics and audio libraries along with Python programming language [9].

PyTorch is the largest machine learning library that optimizes tensor computations. It has rich APls to perform tensor computations with strong GPU acceleration. PyBrain stands for Python Based Reinforcement Learning, Artificial intelligence, and Neural Networks library. It provides fast and easy-to-use algorithms for machine learning tasks [10].

We imported the math library and used one of its methods, i.e., sqrt (square root) without writing the actual code to calculate the square root of a number and use charts to see the number more clearly. That's how a library makes the programmers' job easier. We just needed to fill the data on the x and y axes and then we get the results.

2.3. The overall trend of COVID-19 world wide

The increase population of patients is calculated through using the update data subtract the former day's data. The plot (Figure 3) shows the overall trend of the Spread of COVID-19. The x-axis measures the date- from 2020 to June in 2023 and the y-axis measures the data about patients in the unit of per million. It's easy to see in the plot, there is a significant growth during the Spring of 2022 and the time period between the end of 2022 and the beginning of 2023. The bump during 2022 is

mainly because the variant of Omicron, having stronger infectivity than the former one [2, 3]. Also, some countries give up testing nucleic acid, so maybe there is data missing.

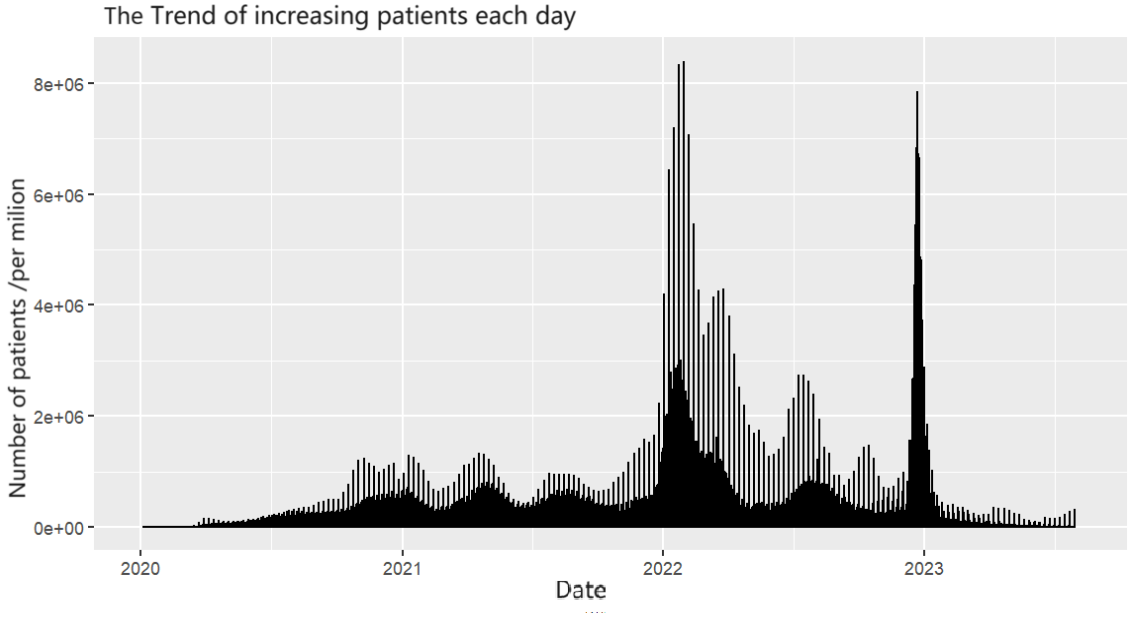

Figure 3. Daily new case over the world.

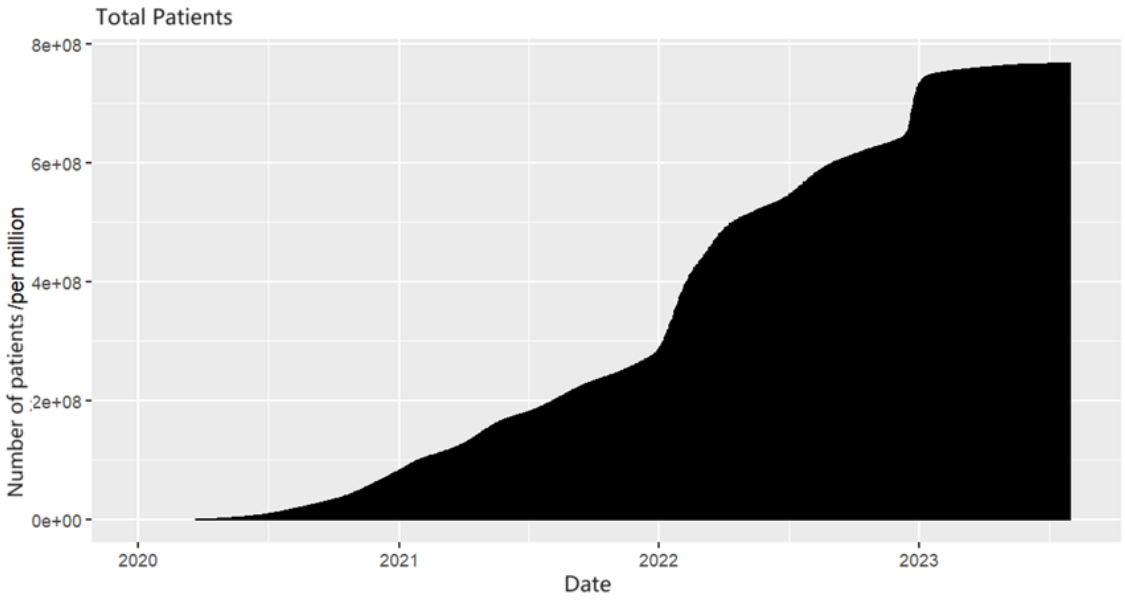

Figure 4. Cumulative confirmed case over the world.

In Figure 4, the x-axis measures the data, and the y-axis measures the number of patients in million. Plot Fig 3 has the same trend in Figure 3. The cliff in the plot fit the increasement in plot Figure 3, too. The total number of patients almost reaches 8 million all over the world, and the steepest increasement happened during about February in 2022 and the beginning of 2023.

2.4. The analyse of the overall COVID-19 data in the US

According to the data download from the Center for Disease Control and Prevention of America, we can get the COVID-19 data of weekly hospital admissions. The data is about the number of increase cases each week. By putting them into Python, we can easily get the basic data (table 1):

Table 1. Descriptive statistics.

Sample size	Minimum	Maximum	Mean	SD*	Median
165	6319.00	150674.000	38686.285	29267.451	31292.000

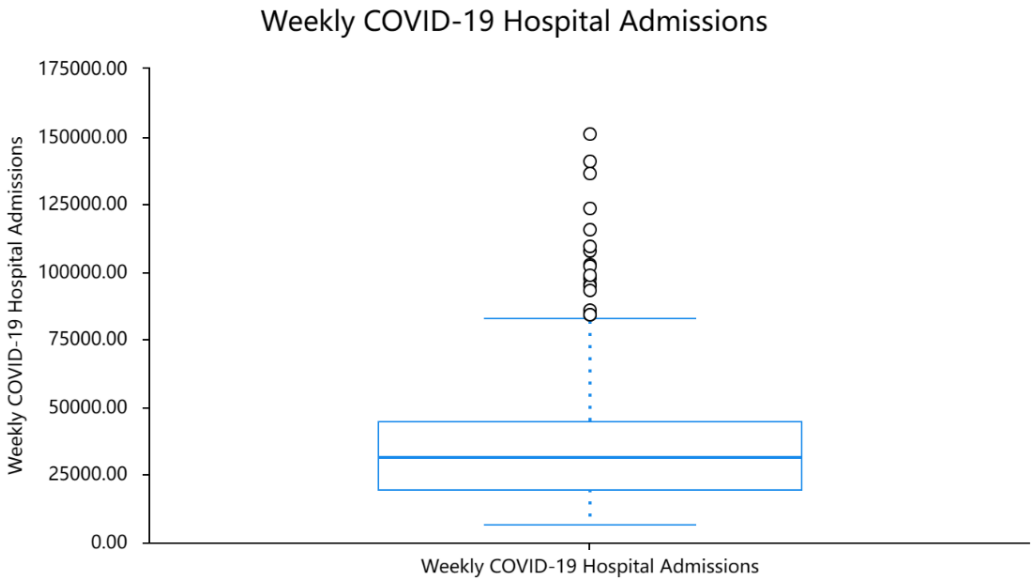

Figure 5. Box plot of the data

Having the above data, we draw the plot (Figure 5) above. By counting the Q1 and Q3 in the data, we can calculate if there is an outlier, which means, if there is a special point. The lower outlier is Q1-1.5IQR, and the lower outlier is Q3+1.5IQR. There are 18 outliers among all data, all happen during the bumps in the former discussion. This can also be a kind of evidence of the weird data happened during 2022 and 2023.

3. Predict model for COVID-19- take the US as an example

We try to predict the future data of COVID-19, and we take the data from the US to set a predict model. The first step is to see whether it is normal distribution or not. We use the P-P plot to see its trend. In such plot, if the data is normal distribution, then the accumulated proportion of data is close to the accumulated proportion of normal distribution. Also, if it is normal distributed, then line is likely to be a diagonal.

It's clear to see that the blue dash line just keeps come across and back then black line, with much difference. So, the data is not normal distributed (figure 6).

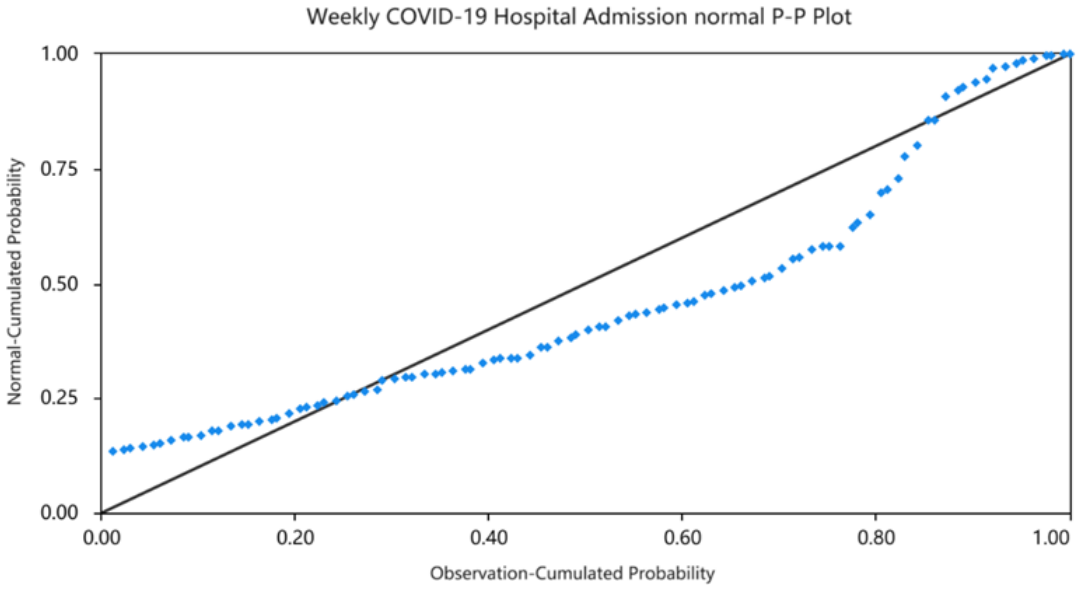

Figure 6. P-P plot of the US data

Considering that the data reaches its two peaks during the winter, it may be a regular pattern hiding in the data. We then consider it as time series, Using ACF and PACF plot to make a predict model.

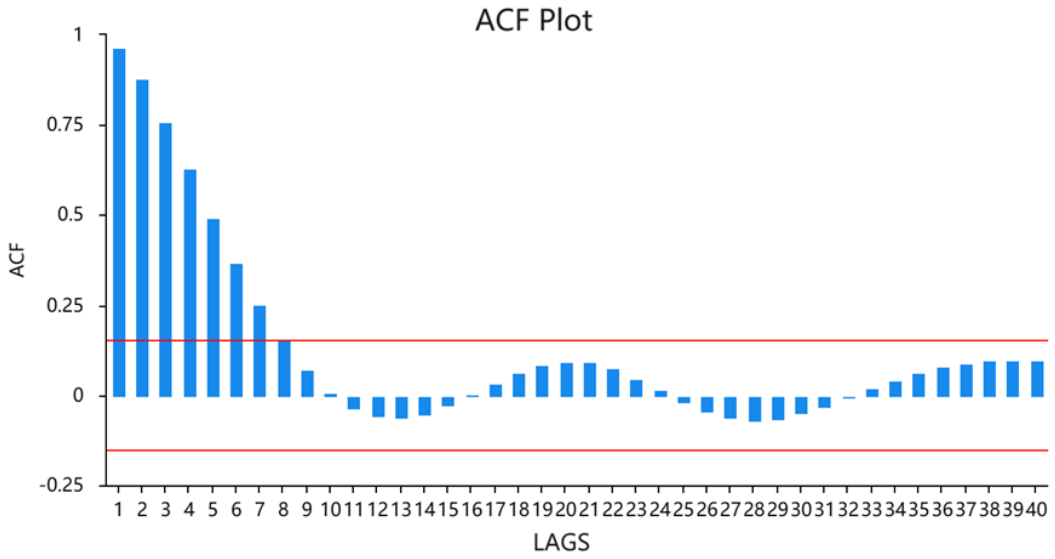

Figure 7. ACF plot of the US data

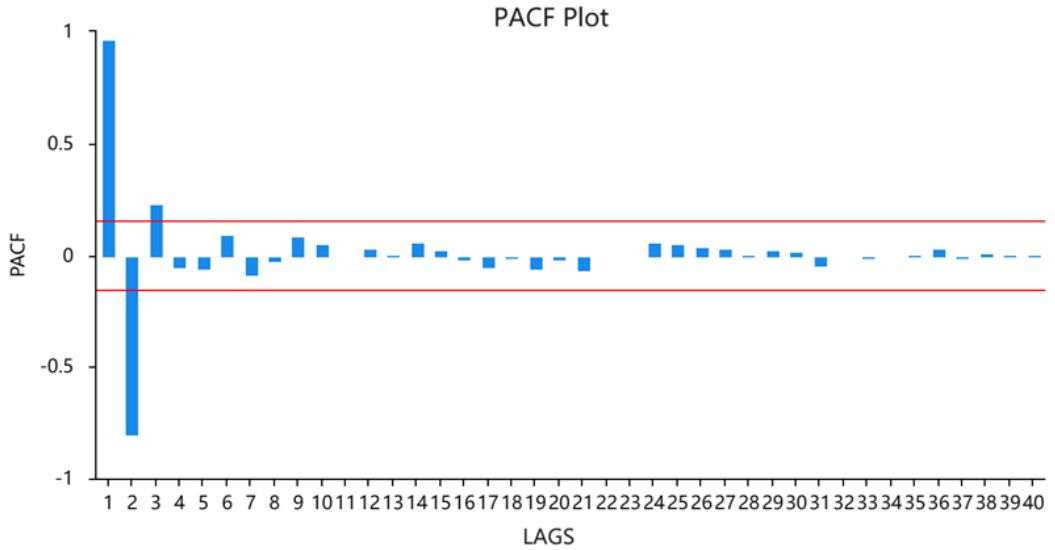

Figure 8. PACF plot of the US data

The ACF plot and the PACF plot is used to predict the autoregression degree p and moving average degree q (figure 7, 8). If ACF plot all becomes 0 after q, the PACF plot is not truncation, Then the ARIMA model can be simply written as MA(q). If ACF is not truncation but PACF does, then the model is AR(p). If both are not truncate, then we need to try the ARIMA model. If both are truncate, then this is a white noise, shouldn't get into consideration (table 2).

Table 2. ACF and PACF results.

LAGS	AC	PAC	Q
1	0.966	0.966	156.703
2	0.879	-0.800	287.250
3	0.761	0.230	385.642
4	0.629	-0.052	453.273
5	0.494	-0.059	495.359
6	0.367	0.090	518.763
7	0.253	-0.086	529.964
8	0.154	-0.026	534.237

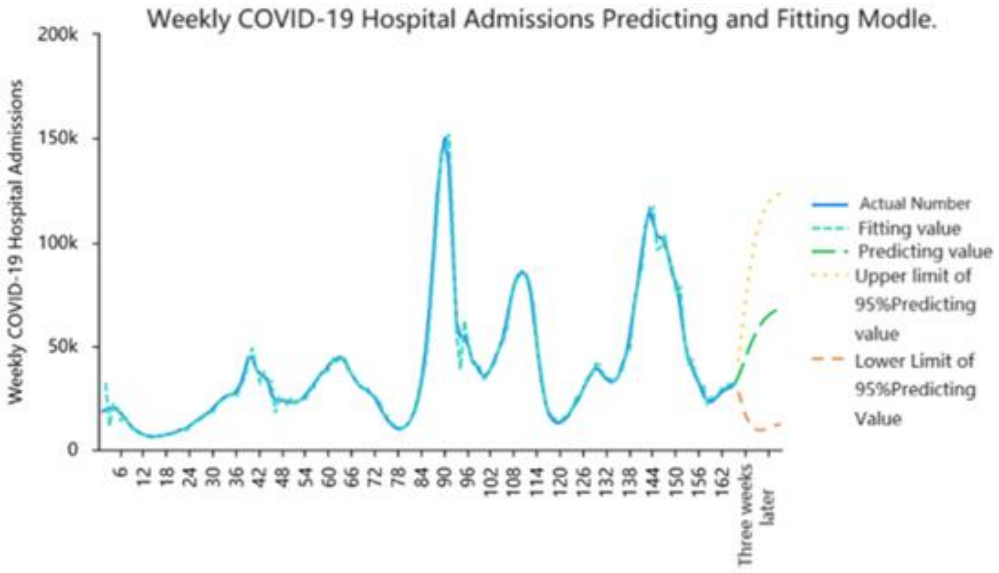

Figure 9. Model fitting results.

We've also made a prediction of the data (figure 9 and table 3). After training the model, it has a well fit of the former data. According to the ACF and the PACF plot, and take into consider the AIC value, the best model we've found is ARIMA (3,1,3), its function is:

$$y(t) = 37.109 + 1.166 * y(t-1) - 0.043 * y(t-2) - 0.236 * y(t-3) + 0.074 * \varepsilon(t-1) - 0.595 * \varepsilon(t-2) - 0.479 * \varepsilon(t-3) \quad (1)$$

Table 3. ARIMA (3,1,3) model results

Term	Sign	Coefficient	SE	z value	p value	95% CI
Constant term	c	37.109	26.755	1.387	0.165	-15.331 ~ 89.548
AR	α1	1.166	0.204	5.709	0.000	0.766 ~ 1.566
	α2	-0.043	0.347	-0.123	0.902	-0.723 ~ 0.638
	α3	-0.236	0.171	-1.383	0.167	-0.571 ~ 0.098
MA	β1	0.074	0.245	0.302	0.763	-0.405 ~ 0.553
	β2	-0.595	0.136	-4.390	0.000	-0.861 ~ -0.329
	β3	-0.479	0.096	-5.004	0.000	-0.666 ~ -0.291

AIC value: 3171.553, BIC value: 3196.352

4. Conclusion

We analyzed the characteristic of the spread of COVID-19 worldwide and in mainland America. Downloading data from Centre for Disease Control and Prevention of America, we used time series to build up a prediction model. Here are the conclusions.

First, the total trend of the spread of COVID-19 all over the world bimodal. It reaches the peak during the winter of 2022 and 2023 because of the change in the government's management. Calculating the outliers, it's easy to see that all the 18 outliers happened in the time range when the total cases have an unexpected increasement during 2022 and 2023. To conclude, COVID-19 spreads the fastest during winter and spring time due to the change of government's management and virus' activity during winter time.

Second, the number of cases isn't normal distributed, and it reaches the peak in winter for twice, so maybe the data is time series. ACF plot, PACF plot all imply that the ARIMA (3,1,3) is the best

predict model, since it has the smallest AIC and BIC value. The function line fits well with the actual number, but because of the quick-shifted world spreading trend, there is still some data points can't be well predicted. The main point is, ARIMA (3,1,3) is the best predict model to predict the total world COVID-19 patients' number.

Authors Contribution

Conceptualization, Liqiao Zhu; data curation, Liqiao Zhu; formal analysis, Liqiao Zhu and Lancong Xie; investigation, Liqiao Zhu and Lancong Xie; methodology, Liqiao Zhu; project administration, Liqiao Zhu and Yifeng Peng; recourses, Yifeng Peng; software, Liqiao Zhu; supervision, Liqiao Zhu and Yifeng Peng; validation, Liqiao Zhu; visualization, Liqiao Zhu and Lancong Xie; writing—original draft, Liqiao Zhu, Yifeng Peng, Lancong Xie, and Yiwen Chen; writing—review and editing, Liqiao Zhu and Yiwen Chen.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] Wang C H, et al. 2022 New progress of COVID-19 mutants and their effects on vaccine immune protection. *Journal of Hainan Medical College*, 28 (6), 401 - 405.

[2] Bian L, Liu J, Gao F, et al. 2022 Research progress on vaccine efficacy against SARS-CoV-2 variants of concern. *Hum Vaccin Immunother*, 18 (5), 2057161.

[3] Saadoon I and Hussein K L 2021 COVID-19 Infection Among Hemodialysis Patients in Tikrit City. *Working paper*.

[4] Mirzaei R, Attar A, Papizadeh S, et al. 2021 The emerging role of probiotics as a mitigation strategy against coronavirus disease 2019 (COVID-19). *Archives of virology*, 166 (7), 1819 - 1840.

[5] Zhu M C, Bin S and Sun X X 2023. Construction and research of an infectious disease model based on the transmission characteristics of COVID-19. *Complex Systems and Complexity Science*, 20 (2), 29 – 37.

[6] Brazer S D, Bauer S C and Lavigne A L 2023 School and district structure adaptations to the COVID-19 super-stressor. *Journal of Educational Administration*, 61 (3), 205 - 221.

[7] Ghislandi S, Muttarak R, Sauerberg M, et al. 2022 Human costs of the first wave of the COVID-19 pandemic in the major epicentres in Italy. *Vienna Yearbook of Population Research*.

[8] Flore J, Hendry N A and Gaylor A 2023 Creative arts workers during the Covid-19 pandemic: Social imaginaries in lockdown. *Journal of Sociology*, 59 (1), 197 - 214.

[9] Khadhraoui M, Bellaaj H, Ammar M B, et al. 2022 Survey of BERT-Base Models for Scientific Text Classification: COVID-19 Case Study. *Applied Sciences*, 12 (6), 2891.

[10] Mgbere O, Nwabuko O, Olateju O, et al. 2023 EPH189 Assessment of Burden of COVID-19 Infection and the Dynamics in African Union Member States. *Value in Health*, 26.