Optimistic Prediction Model For the COVID-19 Coronavirus Pandemic based on the Reported Data Analysis

Khalid S. Aloufi

Department of Computer Engineering
College of Computer Science and Engineering
Taibah University
Medina

Abstract

The world is facing new challenges every day; however, with the spread of the pandemic around the world, this new challenge is different. The pandemic is increasing and concentrating various challenges simultaneously. Although different sectors are facing consequences, the most important sectors, that is, health and economy are the most affected. When the pandemic began, it was not known how long it would last, which complicated health and economic planning. Therefore, it is important for decision makers and the public to know the predictions and expectations of the future of these challenges. In this work, the current situation is analyzed. Then, an expectation model is developed based on the statistics of the pandemic using a growth rate model based on an exponential and logarithmic rate of increase. Based on the available open data about the pandemic spread, the model can successfully predict future expectations, including the duration and maximum number of cases of the pandemic. The model uses the equilibrium point as the day the cases decrease. The model can be used for planning and the development of strategies to overcome these challenges.

Keywords: IR, COVID-19 Coronavirus Pandemic

1. Introduction

In different times in history, humanity has faced great health challenges that affect life in general. Different pandemics have spread around the world in history, such as smallpox, the Spanish flu,tuberculosis and plague [1][2][3] [4] [5] [6]. The effects of pandemics on the health sector are a major challenge that, consequently, affect economy and social life. Humans have daily activities that change substantially during a pandemic. The daily life tasks are thus done with caution until the pandemic ends. However, that can only be done when it is known when the pandemic will end. At that time, people will be able to return to normal daily activities. However, it is convenient to predict the time when life will return to normal. In late 2019, the

---

*Corresponding author
Email address: `koufi@taibahu.edu.sa` (Khalid S. Aloufi)

corona virus, named COVID-19, spread around the world. All aspects of life have suddenly become difficult.

10 Travel between countries have suddenly stopped. Governments around the world are experiencing challenges in providing the daily requirements of food and other consumed goods. Also, governments are having difficulty in stabilizing the economy by paying part of the salaries of employees of the private sector. Some travelers with transit flight are stuck living in an airport because the country they are flying through requires a visa and their continuing flight was canceled. Some people have gone to grocery stores and been unable to obtain

15 required food.

Therefore, planning for the requirements of people in such circumstances is very important. The pandemic requires an expectation model to plan trade and travel activity to decrease the amount of challenges people are facing. Moreover, governments and private sectors require an expectation model to determine the actions that must be taken. For example, a factory needs to know how long it is expected that the factory must

20 work in pandemic conditions. Also, a medicine company will need to predict the production level required for a specific medicine. Coronavirus statistics are gathered from around the world and announced publicity. The world, for the first time, is collecting the data daily, which is helping the world to work together to face the pandemic [7] [8] [9]. The power of open data has been demonstrated during this pandemic [10]. This work obtains the daily statistics and develops a prediction model of the COVID-19 pandemic. Then, the

25 predictions can be used by decision makers to follow and develop plans to address the pandemic. An example trajectory of cases based on daily records of an example territory is shown in figure 1. The green line is the actual current cases, and potential future cases are shown in red. The figure shows different numbers of cases over time with different increase ratios. Also, the figure shows different maximum numbers of cases. Each expected scenario has a different peak number of cases and a different time to reach that peak. Therefore, a

30 model is required to measure the expected peak and the number of days to reach that peak. In this work, based on the actual statistics of each country, a model is developed to predict the number of cases any country will have and the number of days to reach that number.

2. COVID-19 and Open Data

Different organizations around the world are reporting daily or even instant data about COVID-19. The

35 statistics of coronavirus gathered from around the world are announced publicity. The world for the first time is collecting data daily, which is helping the world to work together to face the pandemic [7] [8] [9]. Open data is one of the initiatives of the world wide web consortium. Open data is the publication of data from different sectors, governments or private firms that is made public for general use, which increases the services provided [10]. The Open Data Institute (ODI), one of the initiatives supporting the idea of open

40 data, started in 2012 [11]. Different governments are publishing open data on the web. During the COVID-19 pandemic, health authorities around the world and the World Health Organization (WHO), with continuing
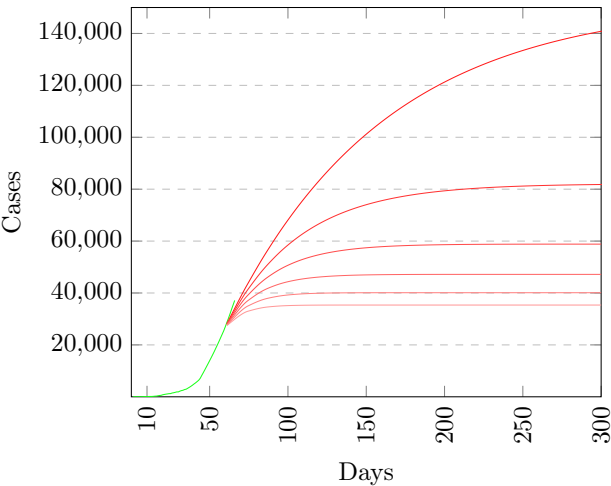
Figure 1: The trajectory of cases of a territory

arrangements with governments and different health organizations, are collecting data daily or even hourly. Data sharing increases the data collection accuracy and the efficiency of the different parties working together to fight the pandemic. In this work, the open data published by different countries is used and analyzed by

45 different organizations. Also, the data are used by researchers in different fields. One of the benefits of the COVID-19 data is that data from the whole world are available. As mentioned earlier, different sectors are collecting the data and publishing it as well. These sources include world meter and the European Centre for Disease Prevention and Control [7] [9].

3. Proposed model

50 In this work, a model is developed using the open data published about COVID-19. After retrieving the data, the data is subject to computational processing. For this reason, the data are available in a format available for direct processing [12]. The main source file is in MS XLSX format. Java is used to process the data. Different packages are available with Java to process data [13] [14]. The data are then transformed to a text file, where the data can be processed to obtain plots using the PDF-LATEX engine.

55 Other methods are possible depending on the requirements. In the collected data, the currently known information about the cases in any country is the number of daily cases and reporting day. Also, the number of death cases is reported, as well as number of recovery cases. From this information, the developed model can predict the situation of the cases in a country and the expected day a steady situation will occur, given the assumption that the cases have started to decrease based on the recovery ratios reported from around the

60 world. Consequently, a mathematical prediction model is presented to predict the short-range and long-range scenarios of the COVID-19 coronavirus pandemic. Then, the predicted data are saved in a text file, where the data can be processed again using the same methods. The model is used to generate a visualization of

the data in text format, where the data can then be processed and presented in any required format. The model is developed using a logistic function as a process that starts with an exponential growth followed

65   by logarithmic growth defined by the function in equation 1. With the following equation, the data can be processed to assess the correlation of the data and the degree of increase or decrease. The covariance of the data and the correlation are computed using equation 2 3. Using 1, the predicted trajectory of the pandemic for each country or the whole world can be computed [15][16][17]. For equation 1, the parameters are e, x_0, k, L. e is the natural logarithm base, known as Euler's number. x_0 is the x value of the sigmoid

70   midpoint, calculated as $SteadyDay/2$ of each country, which is computed using equations 2 and 3. L is the curve's maximum value, and this parameter is $SteadyNumber$, computed by equation 4. k is the logistic growth rate or steepness of the curve.[1] Algorithm 1 presents the algorithm used to generate the predicted trajectory of a country. Then, the algorithm is used to determine the steady date and steady number of cases for each country or the whole world. **Definition**: Let $X = (X_t : t = 1, 2, 3, ...)$ be a wide-sense stationary

75   stochastic process with mean $E[X_t]$ and variance $E[(X_t - E[X_t])^2]$.

$$f(x) = L/(1 + e^{-k(x-x_0)}) \tag{1}$$

$$cov(E[X_1], E[X_2]) =$$
$$\sum_{i=1}^{N} ((X_{1i} - E[X_1])(X_{2i} - E[X_2])) \tag{2}$$

$$corr(E[X_1], E[X_2]) =$$
$$\frac{cov(E[X_1], E[X_2])}{E[(X - E[X_1])]E[(X - E[X_2])]} \tag{3}$$
$$, where -1 < corr(E[X_1], E[X_2]) < 1$$

$$p(X) =$$
$$E[X] + E[(X - E[X])] \tag{4}$$
$$-(E[X] + E[(X - E[X])]) * .33;$$

The situation of the pandemic is either stable, increasing, highly increasing, decreasing or highly decreasing. The situation is stable if either there were no cases reported or there are no more new cases. Each country is in one of the situations shown in table 1.

The pandemic preliminary data of different countries are used in the expectation model. For each country,

80   the daily reported data are used to generate data specifying the Country name, Daily Mean, Max in a Day, Situation, SteadyDay and SteadyNumber. Country is the country name listed with its associated data. Daily Mean is the mean number of daily reported cases by different organizations. Max in a Day is the value of the maximum number of reported cases in a day during the data collection period.

Data: [12]

Result: The COVID-19 prediction data of a country

initialization;

obtain the list of reported data. the sample is the last *sampleSize* of the data in the list.

while *corr > 1* **do**

    for $i = 0$; $i < sampleSize$; $i = i + 1$ **do**

        read sample;

        set Regression;

    end

    compute Corr using equation 3.

    compute predicted values using equation 4

    add the predicted value to the list.

end

Algorithm 1: Prediction Data Setup Algorithms

Table 1: situation

4. Results and analysis

Figures 2,3 and 4 show the reported daily cases of Saudi Arabia. The number of cases is increasing, with a mean of 1452.14 daily reported cases. The first case was reported on 03-03-2020. In the first plot, the daily total reported cases are shown as a blue line. Th red line shows the expected trajectory of the number of cases and the time to reach the steady state. The second plot shows the reported cases on a daily basis. The left plot shows the daily reported death cases. Based on the reported data from each country, the model predicts a different trajectory. The model predicts future cases based on the trajectory reported cases.

For COVID-19 around the world, the total number of cases is decreasing, with a mean of 79314.57 daily reported cases. The first case was reported on 31-12-2019. Figure **??** shows the trajectory of cases and the expectation of when the maximum number of cases will be reach. The expectation is that the total number of cases will reach 6000000 approximately 250 days after the first case was reported. However, the pandemic
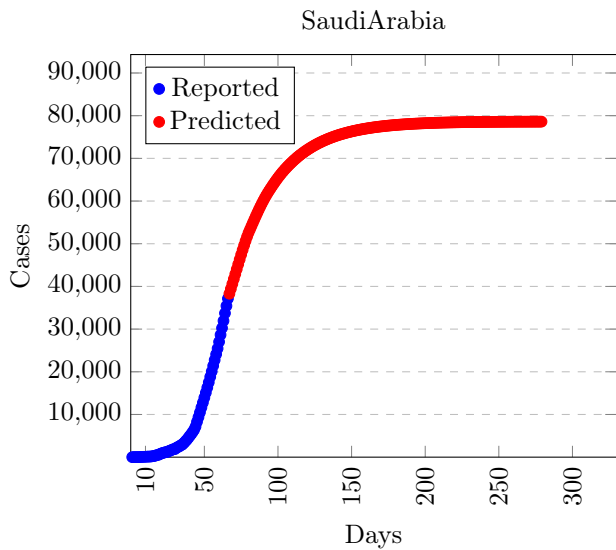
Figure 2: Saudi Arabia reported and predicted
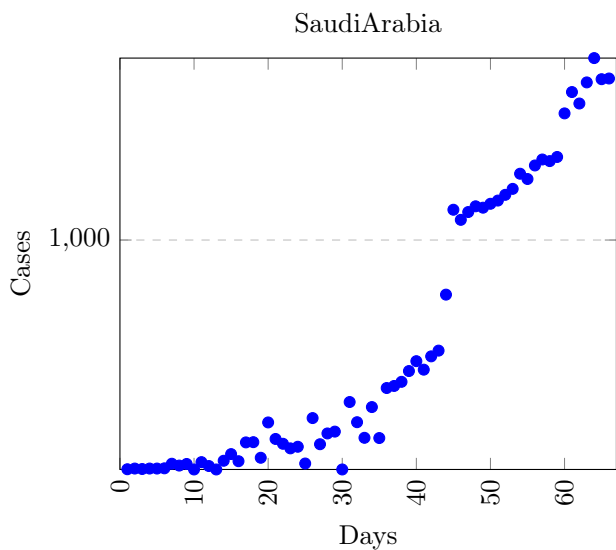
Figure 3: Saudi Arabia daily cases

will continue for as long as 400 days according to the model. Figure 9 shows the daily reported cases from around the world. The figure shows that the range of daily cases is between 60000 and 100000.

Figure 7 shows the plots of all countries from around the world. The blue line shows the reported data. Each country has a different reporting date since the pandemic expected reached countries on different dates. The red line shows the growth in expected cases. Figure 5 shows the maximum and mean number of cases reported per day in increasing order. Figure 6 shows the expected steady day of each country. The model is used to simulate the pandemic trajectory using the equation presented earlier 4. With the appropriate
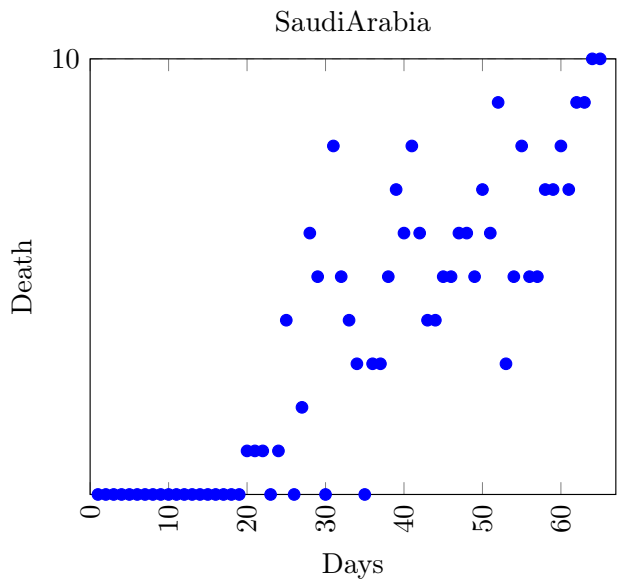
Figure 4: Saudi Arabia reported death cases

parameters, the model can predict the growth rate of cases. For example, for a country where the expected number of cases will reach 50000 and the steady day midpoint is 100, the trajectory is shown in figure 10. Figure 12 shows the initial dates that countries first reported cases. Most countries reported their first case

105 between 29-2-2020 and 9-4-2020, approximately three month after the pandemic started. Figure 13 shows the percentage of countries with decreasing or increasing numbers of cases according to the reported statistics. Figure 11 shows the pandemic expected growth rate based on the expected maximum number of 6 million cases and the midpoint of 150 days.
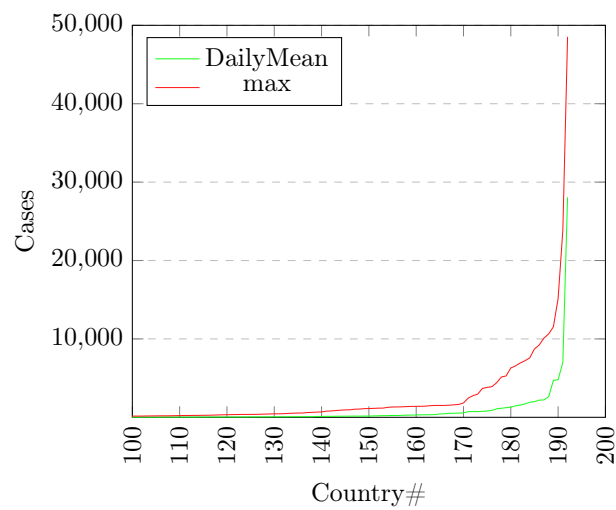
Figure 5: Maximum and mean number of cases for countries
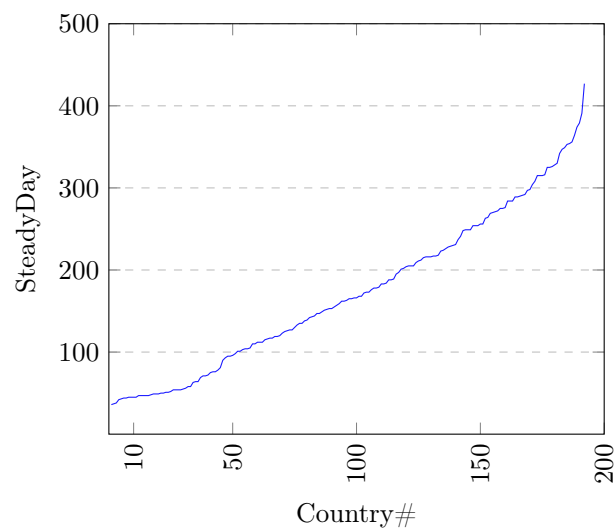
Figure 6: Steady day of countries
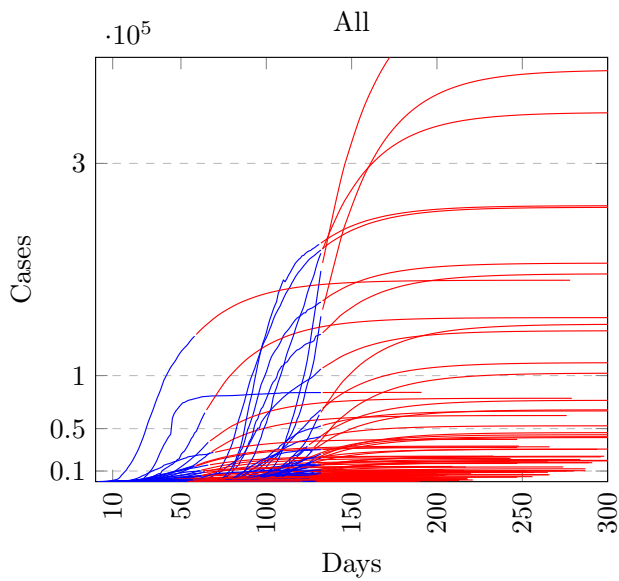
Figure 7: All countries reported and predicted cases
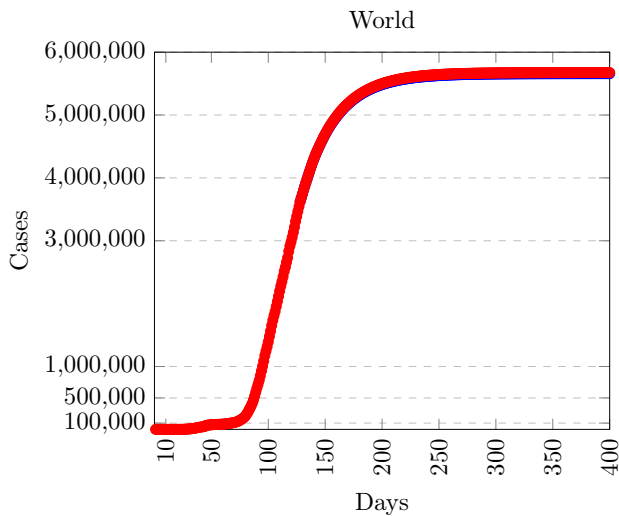
Figure 8: World cases reported and expected
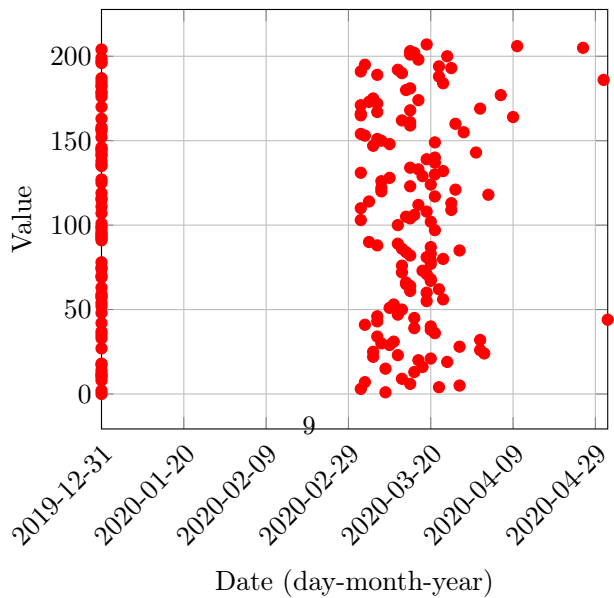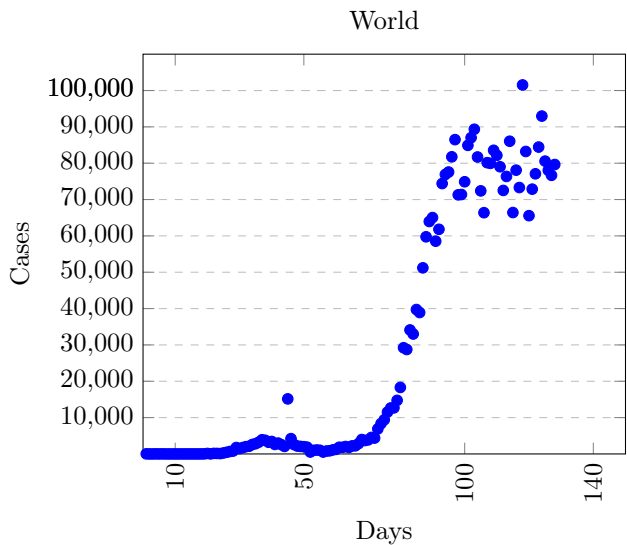
Figure 9: Steady day
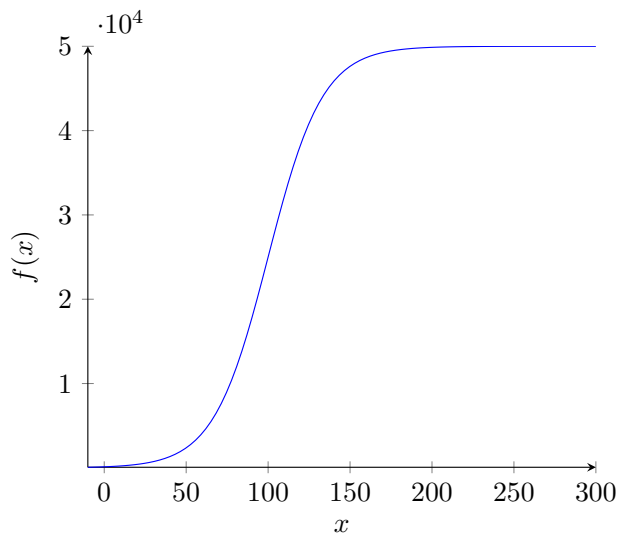
Figure 10: Growth Rate of 50000 cases

5. Recommendations

110   This research recommends using the expectation in planning associated with the pandemic. The model can be used to predict the expected number of cases and in how many days the maximum will be reached. The model can also be used to predict the maximum and minimum numbers of cases and in how many days those numbers will be reached. The model can be used as an optimistic model or pessimists model: both scenarios are important. The model can be used to determine the domain and range of expectation. The

115   domain is the pandemic and its consequences and the associated challenges related to the economy, education
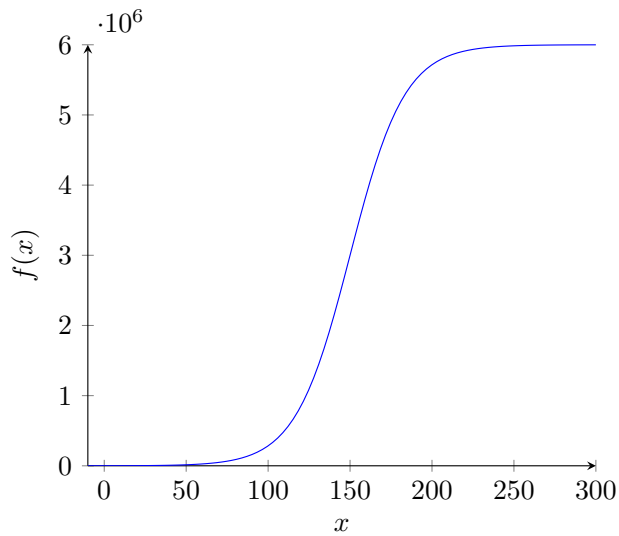
Figure 11: Growth rate to reach 6000000 cases
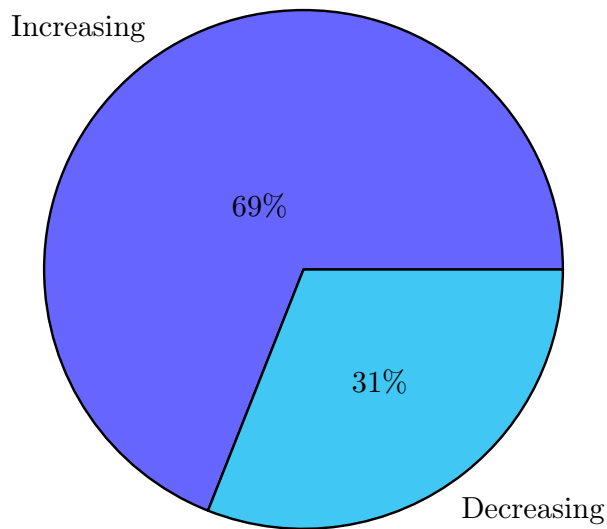
Figure 13: Increasing and decreasing %

and social activities. The range of the time and places the pandemic and the duration of the pandemic cal also be determined. Future work should consider analyzing the data after the pandemic is over to develop mathematical models to help fight such pandemics in a timely manner with more advanced regulations.

6. Conclusion

120   In this work, the data published about the pandemic have demonstrated the power of open data. The data can be used and processed to show the different affects of the pandemic. In this work, the data are

processed and analyzed to show the percentage of countries with decreasing or increasing numbers of cases. Additionally, the mean, variance, covariance and correlation of the data are used to illustrate the pandemic scenario in different countries. Then, the data are used to predict the future state. The data have shown that the growth rate scenario is the best mathematical model to represent the data. Finally, the developed model can successfully represent the pandemic trajectory in a territory.

References

[1] Wikipedia contributors, Smallpox — Wikipedia, the free encyclopedia, `https://en.wikipedia.org/w/index.php?title=Smallpox&oldid=955501571`, [Online; accessed 9-May-2020] (2020).

[2] Wikipedia contributors, Tuberculosis — Wikipedia, the free encyclopedia, `https://en.wikipedia.org/w/index.php?title=Tuberculosis&oldid=955501145`, [Online; accessed 9-May-2020] (2020).

[3] Wikipedia contributors, Plague (disease) — Wikipedia, the free encyclopedia, `https://en.wikipedia.org/w/index.php?title=Plague_(disease)&oldid=952549003`, [Online; accessed 9-May-2020] (2020).

[4] Wikipedia contributors, Spanish flu — Wikipedia, the free encyclopedia, `https://en.wikipedia.org/w/index.php?title=Spanish_flu&oldid=955737989`, [Online; accessed 9-May-2020] (2020).

[5] Wikipedia contributors, Coronavirus disease 2019 — Wikipedia, the free encyclopedia, `https://en.wikipedia.org/w/index.php?title=Coronavirus_disease_2019&oldid=955735118`, [Online; accessed 9-May-2020] (2020).

[6] Wikipedia contributors, Pandemic — Wikipedia, the free encyclopedia, [Online; accessed 9-May-2020] (2020).
URL `https://en.wikipedia.org/w/index.php?title=Pandemic&oldid=955705211`

[7] Worldometers.info, Covid-19 coronavirus pandemic, `https://www.worldometers.info/coronavirus/`, place of publication: Dover, Delaware, U.S.A. (5 May, 2020).

[8] E. O.-O. Max Roser, Hannah Ritchie, J. Hasell, Coronavirus pandemic (covid-19), Our World in DataHttps://ourworldindata.org/coronavirus.

[9] E. C. for Disease Prevention, Control, the geographic distribution of covid-19 cases worldwide, https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide (4 May, 2020).

[10] Wikipedia contributors, Open data — Wikipedia, the free encyclopedia, `https://en.wikipedia.org/w/index.php?title=Open_data&oldid=953018662`, [Online; accessed 10-May-2020] (2020).

[11] Wikipedia contributors, Open data institute — Wikipedia, the free encyclopedia, `https://en.wikipedia.org/w/index.php?title=Open_Data_Institute&oldid=953330894`, [Online; accessed 10-May-2020] (2020).

[12] K. Aloufi, Covid-19, Mendeley Data, v1`doi:10.17632/zk32frw6p5.1`.

[13] B. D. Hahn, K. M. Malan, Acknowledgements, in: B. D. Hahn, K. M. Malan (Eds.), Essential Java for Scientists and Engineers, Butterworth-Heinemann, Oxford, 2002, p. xvi. `doi:https://doi.org/10.1016/B978-075065991-8/50001-0`.
URL `http://www.sciencedirect.com/science/article/pii/B9780750659918500010`

[14] Preface, in: B. D. Hahn, K. M. Malan (Eds.), Essential Java for Scientists and Engineers, Butterworth-Heinemann, Oxford, 2002, pp. xiii – xv. `doi:https://doi.org/10.1016/B978-075065991-8/50000-9`.
URL `http://www.sciencedirect.com/science/article/pii/B9780750659918500009`

[15] Wikipedia contributors, Logistic function — Wikipedia, the free encyclopedia, `https://en.wikipedia.org/w/index.php?title=Logistic_function&oldid=955315778`, [Online; accessed 10-May-2020] (2020).

[16] J. Peng, K. Lee, G. Ingersoll, An introduction to logistic regression analysis and reporting, Journal of Educational Research - J EDUC RES 96 (2002) 3–14. `doi:10.1080/00220670209598786`.

[17] G. Rzadkowski, I. Głażewska, K. Sawińska, Logistic function as a tool of planning, Foundations of Management 6. `doi:10.1515/fman-2015-0004`.