Modeling the Corona Virus Outbreak for Cities, States, and Countries: A Phase Space Approach to Modeling and Interpreting Pandemics

Dr. Joseph Molitoris, Ph.D.

Alexander von Humboldt Foundation Fellow

2 Mitchell Terrace, West Long Branch, NJ 07764

Abstract

The coronavirus outbreak allows for several modeling approaches. Simple models reproduce fairly well the case numbers of the coronavirus pandemic in specific regions. An extended model that includes recovery and death also reproduces the death case numbers. These models may further be used to predict future cases and updated as data becomes available in a region (city, state, country). Since the models are not complicated, they can be used by decision makers to guide their way through the exigencies of this or other pandemics. The models can be run in real time by an analyst to guide the decisions of government officials. Data is shown and compared favorably to the models for several states, countries, and the World. Correlations are typically 98 to 99 percent. A novel phase space approach is introduced; this allows tracking of the pandemic towards its peak and then mathematical extinction.

The Corona Virus outbreak allows for several possible modeling approaches. A number of simple models reproduce fairly well the case numbers (total confirmed cases) of the Corona Virus pandemic in specific regions. An extended model that includes recovery and death also reproduces the death case numbers. These models may further be used to predict future cases and updated as data becomes available in a region (city, state, country). Since the models are not complicated, they can be used by decision makers to guide their way through the exigencies of this or other pandemics. I have run the models on a personal computer with Excel and a MacBook Pro with Numbers and Mathematica applications. The models can be run in real time by an analyst to guide the decisions of government officials. Data is shown and compared favorably to the models for: NYC, New York, NJ, PA, CA, USA, Italy, France, Germany, Spain, China, Japan, South Korea, and the World as a whole. Correlations between model predictions and real data are typically 98 to 99 percent. A novel phase space approach based on available data is introduced; this approach allows tracking of the pandemic towards its peak and then to mathematical extinction.

Since I spent much of my life as a college professor, I have been interested in the pandemic from an academic point of view. The Corona Virus outbreak allows for a number of possible applications to teaching [1] as well as student research [2]. In mathematics or computer science courses, the topic of analytic models and computer modeling is relevant. In biology or earth science classes, the topic of disease and its impact on population is important. In calculus or physics, we often explore exponential growth and computer models. Models can also be used by decision makers to guide their way through the exigencies of this or other pandemics. Decisions

about social distancing, travel restrictions, and the imminent opening of states and countries for normal business are influenced by both actual data and models.

Initially I tried a simple Poisson model [3], but decided that it does not describe the data [4-7] well enough. This is probably because the Poisson probability distribution $p(n)$ only has one parameter λ [3].

Thus I investigated the use of the Binomial distribution which along with a number of trials n has a parameter p that may be thought of as an infection probability for a susceptible population. If the growth curve of the region being modeled indicates more than one outbreak, then more than one binomial distribution helps to describe the cumulative number of cases. See Figure 1 for a comparison of the World Health Organization China data [4] to the Binomial distribution model.

Second, I looked at a simple logistic equation model which can describe all or part of the data for a region. A long start up phase can make it difficult to nonlinearly fit the two parameters S and r . Nevertheless, in many cases this model works very well. See Figure 2 for a comparison of the World Health Organization Germany data [4] to the logistic equation model. The correlation coefficient r -squared is 0.996.

Third, a basic time stepped simulation [1] can be used to describe the data in a region. Again, there are two parameters S and r . S the “susceptibles” is the long term population that becomes infected; r is the growth rate or transmission coefficient. The differential equation that the

computer simulation solves is: $dy/dt = r y (1-y/S)$. The precision of the simulation can be verified by comparison to both exponential growth and logistic model growth, respectively. With a time step of 0.1 day in the simulation, the percent error versus the logistic equation is typically less than 0.1 %.; in all the calculations shown here 0.01 day is used. See Figure 3 for comparison of Centers for Disease Control and Prevention USA data [5] during the pivotal month of March 2020 to the time stepped simulation model. State and city data is also available online [6,7].

A fourth and final model is a bit more sophisticated. The single differential equation is changed to a coupled set of four equations:

$$(1) \quad dy/dt = r y (1-y/S-R/S-D/S) - b y - d y$$

$$(2) \quad dx/dt = -r y (1-y/S-R/S-D/S)$$

$$(3) \quad dR/dt = b y$$

$$(4) \quad dD/dt = d y$$

where y is the number of Infected, x is the number of Susceptibles, R is the Recovered, and D is the deaths. Note that since $x + y + R + D = S$ a constant, the sum of the above differential equations is zero. The parameters b , d , r , and S must be extracted from experimental data of the pandemic and can be adapted to the situation as the disease spreads. The initial conditions that permit the solution of the differential equations using a fixed time step procedure are that at $t=0$, we have $y = 1$, $R = D = 0$, and $x = S-1$. The basics of this type of modeling were developed over 100 years ago [8] and are still in use today in a variety of contexts.

I have then compared the Corona Virus data for several countries (England, France, South Korea, etc.), select states (NJ, PA, NY), and cities (New York City and Los Angeles) to the three models described above. In all cases, fairly good representation of the data is seen. The data may be found by analyst, citizen or teacher online [4-7] either as the pandemic progresses (as an at home lesson or summer project) or from archived data once the crisis is over. County, state, or federal politicians can have an analyst download data in real time and follow the data with the model.

See Figure 4 for a comparison of the World Health Organization World data [4] to my corona differential equation model (equations 1 through 4 programmed in Mathematica). The representation cannot be perfect since early on China drives the World numbers and currently the USA and Europe drive the growth. The China data and model results are shown in Figure 5. Figure 6 is a comparison of the Centers for Disease Control and Prevention USA data [5, 6] to my Corona differential equation model. The representation here is pretty good: the correlation coefficient is 0.999. Figures 7 and 8 show model results compared to data for the states NJ and NY, respectively.

In ONLY the case of NJ and Figure 7, I have shown the impact of limiting the rate of change r based on the fact that for some regions like NJ, NY, and NYC there is a time when the epidemic is spreading in a linear fashion before it again decreases according to the differential equation model. This limitation of r , which might be applied by an analyst or government decision maker (for more realism in certain regions), increases the correlations of the cumulative cases and

deaths data with the model from the typical 0.98 to 0.99 range to a higher 0.999 to 0.9999 spread.

The model also appears to do a reasonable job of describing the infection in New York City, the focal point in the USA. See Figure 9. The data situation in NYC is currently a bit ambiguous with recent reporting of deaths listing both probable and confirmed cases. The last three data points on the Deaths plot in Figure 9 include those larger total numbers, which line up with the model predictions pretty well (correlation is 0.998).

South Korea, Japan, and Spain are shown in Figures 10 through 12. France, Germany, Italy, the United Kingdom, California, and Pennsylvania are shown in Figures 13 through 18.

Another way of interpreting the data or model of the pandemic in a region is via phase space. Looking at the coupled differential equation model, all of the quantities y , D , R , x as well as their derivatives may exhibit interesting behaviour. Figure 19 shows a plot of dy/dt versus y for four regions: China, France, Germany, and NJ. These orbits in phase space define the scope of the pandemic and allow visualization of both disease spread as well as progress towards the end of the event. Unfortunately, not all of the variables may be readily available in reported statistics.

Since the total number of cases (cumulative) as well as the number of new cases - and the corresponding death data - are readily available, consider plotting new deaths versus new cases. Examples are shown in Figures 20 - 22 for China, Korea, and France. The data may either be

viewed as dots (see Figure 23) or as connected points which amounts to a random walk in a nearly contained region of phase space. The trajectory of the random walk (see Figures 20 to 22) or the smoother model data is from pandemic start (1 or a few new cases and new deaths) to finish (0 or few new cases and new deaths). This type of phase space region display allows for the possible near containment of the data into a phase space area where the path to mathematical extinction (pandemic end) can be better seen. The new case data, however, can be somewhat noisy due to small numbers, fluctuations, and reporting issues. In order to improve that data for visualization purposes, a smoothing over several days may be used. Alternatively, the new case and new death data can each be fit by a normal distribution (see Figure 23 Germany). Sample results are shown in the figures.

Another function which fits the coronavirus case and death data fairly well is the hyperbolic tangent: $Y = 1/2 S (1 + \tanh r(x-m))$. This is not surprising since under certain conditions the hyperbolic tangent and logistic function are equivalent. By straightforward differentiation and the use of the Pythagorean identity for hyperbolic tangent, one obtains

$$(5) \frac{dY}{dt} = 2 r Y (1 - Y/S)$$

which is analogous to equation (1) above. This may be coupled with the below three equations to obtain an alternate differential equation model.

$$(6) \frac{dx}{dt} = -2 r Y (1 - Y/S)$$

$$(7) \frac{dR}{dt} = b y$$

$$(8) \frac{dD}{dt} = d y .$$

Since the cumulative number of cases $Y = y + R + D$, the first equation (5) of this alternate model may also be rewritten as $dy/dt = 2 r Y (1 - Y/S) - by - dy$ which replaces (1) and makes it clear that the condition $y + x + R + D = S$ is satisfied.

The alternate coupled differential equation model is compared to the data (online such as at [5,7]) in Figure 24. The data as of May 5 show 70271 confirmed cases and the model gives 68228 or within 2.9%. The data also shows 4212 deaths and the model gives 4149 which is within 1.5%. If one wishes to follow the course of the pandemic in phase space, that is shown in Figure 25.

The coronavirus outbreak allows for several possible modeling approaches. A number of simple models reproduce fairly well the case numbers (total confirmed cases) of the coronavirus pandemic in specific regions. An extended model that includes recovery and death also reproduces the death case numbers. These models may further be used to predict future cases and updated as data becomes available in a region (city, state, country). Since the models are not complicated, they can be used by decision makers to guide their way through the exigencies of this or other pandemics. I have run the models on a personal computer with Excel and a MacBook Pro with Numbers and Mathematica [9] applications. A novel phase space approach for visualizing model and data is introduced. This approach complements other [10] modeling approaches. Further work is in progress.

References

1. Elementary and Advanced Computer Projects, J. Molitoris, J. Coll. Sci. Teach., Sept. (1992) 44.
 2. Proceedings of the Capstone Project Program, J. Molitoris, NJ: Docspics, 2016.
 3. Probability Models and Applications, I. Olkin, L. Gleser, C. Derman, NY: Macmillan, 1980.
 4. World Health Organization:
<https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>
 5. Centers for Disease Control and Prevention:
<https://www.cdc.gov/coronavirus/2019-ncov/cases-in-us.html>
 6. NJ Department of Health:
https://www.nj.gov/health/cd/topics/covid2019_dashboard.shtml and
<https://www.nj.com/coronavirus/2020/03/see-the-spread-of-nj-coronavirus-in-charts-map-saturday-march-21-2020.html>
 7. NYC Health:
<https://www1.nyc.gov/assets/doh/downloads/pdf/imm/covid-19-daily-data-summary.pdf>
- See also NY state Department of Health, Massachusetts dashboard (mass.gov/covid-19), and others online.
8. Kermack, W. and McKendrick, A., Proc. of the Royal Society, Vol. 115A (1927) 700-721.
 9. The Mathematica Book, S. Wolfram, IL: Wolfram Research, 1998-2020.
 10. See papers at <https://wwwnc.cdc.gov/eid/spotlight/coronavirus> such as Mizumoto, K. and Chowell, G., Emerg. Inf. Diseases, Vol. 26 No. 6 (June 2020) eid26.06.200333.

Biography

Dr. Joseph Molitoris received his Ph.D. at Michigan State University in nuclear modeling and has served as a Professor at Muhlenberg College and a Fellow of the Alexander Von Humboldt Foundation and Istituto Nazionale di Fisica Nuclear, (INFN) as well as with the U.S. Army and Navy for program assessments and evaluations. Dr. Molitoris is the author of five books and over fifty publications.

Figures

Figure 1: A comparison of the World Health Organization China data [4] to the Binomial distribution model.

Figure 2: A comparison of the World Health Organization Germany data [4] to the logistic equation model.

Figure 3: Comparison of Centers for Disease Control and Prevention USA data [5] during the pivotal month of March 2020 to the time stepped simulation model

Figure 4: A comparison of the World Health Organization World data [4] to my Corona differential equation model (equations 1 through 4 programmed in Mathematica).

Figure 5: A comparison of the World Health Organization World China data [4] to my Corona differential equation model.

Number

Day

- Recovered
- Model (Deaths)
- Model (Cases)
- Data (Cases)
- Susceptible
- Infected
- Data (Deaths)

Figure 6 is a comparison of the Centers for Disease Control and Prevention USA data [5, 6] to my Corona differential equation model.

Figure 7 shows model results compared to data for the state of NJ.

Figure 8 shows model results compared to data for the state of NY.

Figure 9. New York City.

Figure 10: South Korea.

Figure 11: Japan.

Figure 12: Spain.

Figure 13: Italy.

Figure 14: Germany.

Figure 15: France.

Figure 16: United Kingdom.

Figure 17: Pennsylvania.

Figure 18: California.

Figure 19: Phase space approach to viewing the data or model. China, France, Germany, and NJ are shown.

Figure 20: China phase space plot.

New Deaths

Figure 21: South Korea phase space plot.

New Deaths

Figure 22: France phase space plot.

New Deaths

Figure 23: Germany phase space plot with a normal distribution data fit.

New Deaths

Model

Data (Normal Fit)

Data

Figure 24: Massachusetts daily data for confirmed cases and deaths compared to model results.

Figure 25: Massachusetts phase space plot with data and model compared.

New Deaths