Shirish M. Chitanvis*

Dynamical model for social distancing in the U.S. during the COVID-19 epidemic

https://doi.org/10.1515/cmb-2020-0107
Received June 11, 2020; accepted October 28, 2020

Abstract: Background Social distancing has led to a "flattening of the curve" in many states across the U.S. This is part of a novel, massive, global social experiment which has served to mitigate the COVID-19 pandemic in the absence of a vaccine or effective anti-viral drugs. Hence it is important to be able to forecast hospitalizations reasonably accurately.
Methods We propose on phenomenological grounds a random walk/generalized diffusion equation which incorporates the effect of social distancing to describe the temporal evolution of the probability of having a given number of hospitalizations. The probability density function is log-normal in the number of hospitalizations, which is useful in describing pandemics where the number of hospitalizations is very high.
Findings We used this insight and data to make forecasts for states using Monte Carlo methods. Back testing validates our approach, which yields good results about a week into the future. States are beginning to reopen at the time of submission of this paper and our forecasts indicate possible precursors of increased hospitalizations. However, the trends we forecast for hospitalizations as well as infections thus far show moderate growth.
Additionally we studied the reproducibility *Ro* in New York (Italian strain) and California (Wuhan strain). We find that even if there is a difference in the transmission of the two strains, social distancing has been able to control the progression of COVID 19.
Funding None.

Keywords: IR, random walk, log-normal

MSC: 60, 62, 65

1 Introduction

One goal while tracking the COVID-19 epidemic is to make forecasts of hospitalizations [1], as it indicates the expected loads on local health-care systems. This calls for caution [2] as the uncertainty in the forecast has to be estimated for a novel virus. The uncertainty can be reduced by using what has been learned from the history of hospitalizations. This is especially true for a virgin virus like COVID-19. Hospitalization data have tended to be far more stable than testing data. It has been reported that the IHME model used such an approach to estimate the maximum hospitalizations in NY state within a factor of two or so [3]. Our focus on hospitalizations is complementary to comprehensive epidemiological models such as the one developed by the Covid-19 response team at Imperial College [4].

A straightforward way to make a forecast is to use extrapolation of previous data. One then needs a model of some ilk in order to quantify the uncertainty in the forecast. Uncertainty arises because the transmission of the disease is a probabilistic process which depends on the distance of closest approach, virus load, time of contact, susceptibility of the target etc. Some models use epidemiological knowledge from previous occurrences of a disease to inform their predictions. In the case of COVID-19 which is a novel virus attacking humans, it is not entirely clear that assumptions from other epidemics apply [5]. Indeed, it would be desir-

***Corresponding Author: Shirish M. Chitanvis:** On change of station at Department of Physics and Nuclear Engineering, U.S. Military Academy at West Point, N.Y., E-mail: shirish.chitanvis@westpoint.edu

able to have a model for the evolution of hospitalizations which depends only on the history of the current epidemic.

Added value of this study

We developed a random walk/diffusion equation model for the estimation of uncertainty as we extrapolate in time.

We will show via back-testing that our approach yields useful results about a week into the future. We show formally that our approach gives results with narrower uncertainty bands that do standard distributions. Back-testing shows that the uncertainty bands do provide useful forecasting bounds.

2 Methods

A chain reaction model describes an exponential growth in the number of *active* infections (or hospitalizations) $I(t)$ for the virus:

$$\frac{dI(t)}{dt} = \alpha(t) I(t) \tag{1}$$

where $\alpha/\ln(2)$ is the inverse doubling time. α can be reduced via social distancing and is related to the reproducibility Ro:

$$\alpha(t) = \frac{Ro(t) - 1}{\tau_{incubation}} \tag{2}$$

where $\tau_{incubation}$ is the incubation time for a disease. The incubation time for COVID-19 has been quoted to range from a few days to 14 days.

The formal solution to Eqn. 1 is:

$$I(t) = I_0 \ \exp\left(\int_0^t \alpha(t')dt'\right) \tag{3}$$

where $I_0 \equiv I(t = 0)$.

This chain reaction is essentially a probabilistic process, where the chances of transmission between an infected person and a target could depend on the distance of closest approach, virus load or duration of contact, susceptibility of the target etc. The question we would like to address is whether one can obtain a fundamental model to describe the probabilistic progress of the disease.

Towards that end let us focus on the exponent in Eqn. 3. The number of hospitalizations I at a given time t empirically displays an exponential growth in which the exponent can change in time to reflect the effect of social distancing (see e.g. result for NY state in Fig. 1b, which displays a “square root x" behavior near the beginning of the data):

$$\ln\left(\frac{I(t)}{I_0}\right) = [\tau(t)]^{1/2} \tag{4}$$

where $[\tau(t)]^{1/2}$ is a general function which describes the flattening effects of social distancing, or lack thereof.

Using data, we can obtain $[\tau(t)]^{1/2}$ up to some time $t \leq t_0$. We expect that for a short period of time determined by changes in social distancing behavior for example, we can extrapolate $[\tau(t)]^{1/2}$ for $t > t_0$. The disease progresses exponentially, and hence small changes at a given instant in time can have a large effect on later results. This is the reason why we do not expect forecasting to hold beyond a short period.

If $[\tau(t)]^{1/2} \sim t^{1/2}$, the process is analogous to the process of diffusion where the root mean square distance ξ_{rms} traveled by a particle is proportional to $t^{1/2}$. In our case, $\xi \equiv \ln\left(\frac{I(t)}{I_0}\right)$. And we know that the probability $P(\xi, t)$ of finding a particle diffusing in one dimension a distance ξ after a time t is described by:

$$\frac{\partial P}{\partial t} - D\frac{\partial^2 P}{\partial \xi^2} = 0 \tag{5}$$

where the diffusion constant is D.

Indeed, we found empirically that $[\tau(t)]^{1/2} \sim t^p$ gave a crude fit to data, with $p < 0.5$ for several states we considered. In our case $t \to \tau(t)$ and $x \to \xi \equiv \ln(I/I_0)$, so that the general master equation we seek follows by inspection:

$$\begin{aligned} \frac{\partial P}{\partial \tau} - \frac{\partial^2 P}{\partial \xi^2} &= \mathcal{S}(\xi, \tau) \\ \xi &\equiv \ln\left(\frac{I}{I_0}\right) \end{aligned} \tag{6}$$

where $\tau \equiv \tau(t)$, $P(\xi, \tau)$ is the probability that ξ will have a certain value at a "time" $\tau \equiv \tau(t)$ $[dimensionless]$, while $\mathcal{S}(\xi, \tau)$ is an arbitrary source function. The diffusion constant is taken to be unity without loss of generality. Once we have written down Eqn. 6, it is obvious that the model holds for any function $\tau(t)$.

If $\mathcal{S}(\xi, \tau)$ is a Dirac delta function at the $\xi = \xi_0$, $\tau = \tau_0$, the normalized solution is:

$$P(\xi, \tau) = \frac{1}{\sqrt{2\pi(\tau - \tau_0)}} \exp\left[-\frac{(\xi - \xi_0)^2}{2(\tau - \tau_0)}\right] \tag{7}$$

And since ξ is logarithmic in I, it follows that the natural logarithm of the number of hospitalizations will governed by a normal distribution. In this sense it is different than the log-normal distribution function, which contains a factor which is the inverse of the random variable in its definition. So $P(\xi = \ln(I/I_0), \tau) \to P(\xi, \tau)/I$. The extra factor is simply due to a change in variables, viz. $d \ln I = dI/I$.

The fact that we took a Dirac delta function source at $\xi = \xi_0$ and $\tau = \tau_0$ simply indicates that we start off at the end of our time-series with a known number of hospitalizations on a given date.

Our PDF is right-skewed compared to the normal distribution when plotted as a function of I. The shape is similar to the one in Fig. 1b for the number of current hospitalizations in NY state. We have used Eqn. 7 to make probabilistic predictions of how the number of infections will evolve in time. The variance is given by $2D(\tau(t) - \tau(0))$, and the mean is ξ_0. Here the diffusion constant $D = 1$ $[dimensionless]$.

2.1 Data extraction

We used COVID19 state-wide hospitalization data from the Covid Tracking Project: https://covidtracking.com/api/v1/states/daily.csv. The COVID Tracking Project is a volunteer organization which started from the *The Atlantic* magazine and dedicated to collecting and publishing daily data required to understand the COVID-19 outbreak in the United States. It was started in early March of 2020 by Robinson Meyer and Alex Madrigal at *The Atlantic* who then began a collaboration with Jeff Hammerbacher at *Related Sciences*, along with Erin Kissane, a managing editor at the magazine.

Our interest in this database arose from the fact that this was the first database we found on the web which published daily hospitalization data in all fifty states. Hospitalization data have tended to be far more stable than testing data. These data were not easy to find in early March 2020, when only testing data were prominently available. And with testing for COVID-19 understandably inadequate early in the pandemic in the U.S., it was hard to determine accurately the progress of the disease, given large daily fluctuations.

The Project is supported by the Chan Zuckerberg Initiative, The Rockefeller Foundation, Emerson Collective, Robert Wood Johnson Foundation and the Patrick J McGovern Foundation.

We are grateful to this project for providing these data openly.

Informed consent: Not applicable.

Ethical approval: The conducted research is not related to either human or animal use.

2.2 Data Analysis

A simple IR (Infections (hospitalizations), Recovered cases) model [6], a subset of the SEIR model [7], was used to obtain Ro from this data.

$$\begin{aligned} \frac{dI}{dt} &= \alpha' I - \frac{dR}{dt} \\ \frac{dR}{dt} &= \gamma I \\ \alpha &\equiv (\alpha' - \gamma) \equiv \frac{Ro - 1}{\tau_{incubation}} \end{aligned} \tag{8}$$

We applied Eqn. 8 to analyze the evolution of cases as well as hospitalizations.

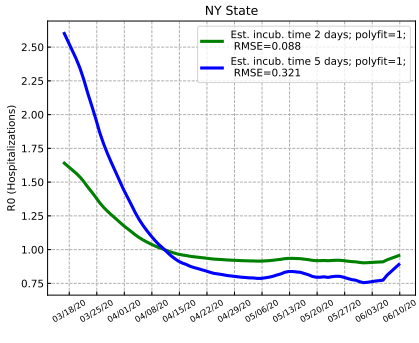
By dealing directly with a population of hospitalized patients, we are not required to track susceptible and exposed persons. We find from data that the recovery rate $\gamma \approx 0.1[1/Day]$, where $dR/dt = \gamma I$. We used Python to perform all the analysis presented in this paper.

The model for extrapolation we have proposed is not unique. There are multiple methods to extrapolate based on previous data. In the case of the novel virus COVID-19 it may be useful to have a diverse set of prediction models to understand the effect of social distancing. The standard probability distribution function used in epidemiology [8, 9] is the Erlang distribution $E_r(k, \mu)$, which is related to the gamma distribution. There are qualitative differences between the standard model and ours. The normal distribution allows the independent specification of the mean and the variance. The Erlang distribution is such that the ratio of the square of the mean ($k\mu$) and the variance ($k\mu^2$) is k. As such the Erlang distribution will have a large variance if the mean is large, but not necessarily in our log-normal distribution, where the ratio is $\xi_0^2/2\ (\tau(t) - \tau(0))$. Our model yields variances that are smaller than those obtained from the Erlang distribution. Back testing results for NY state are included in Appendix A which validate our model.

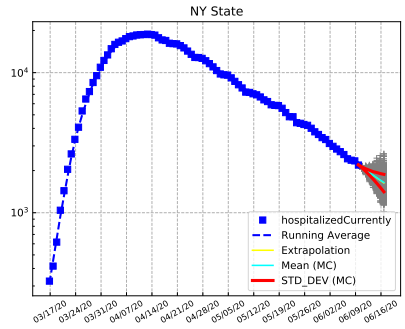
Note that recovery data is unavailable in the dataset we used for some states such as Ohio. In these states we can only estimate α' for infections defined in Eqn. 8.

3 Results

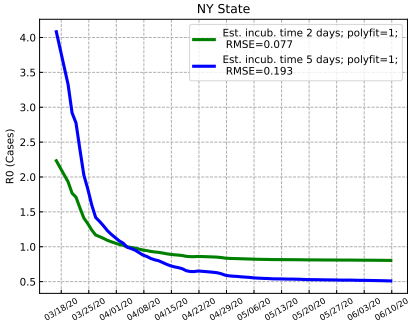
We display in this section exemplary results for two of the many states we have studied.



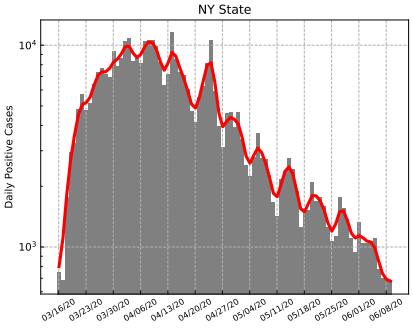
(a) Reproducibility via an IR model for hospitalizations. This measure is a leading indicator of an apex in (b).



(b) Current hospitalizations in the state.

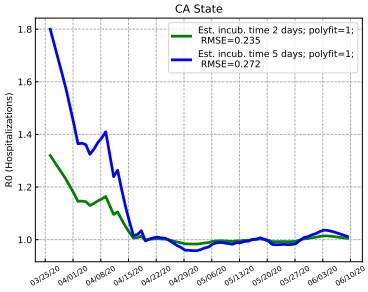


(c) Reproducibility via an IR model for the number of COVID-19 cases.

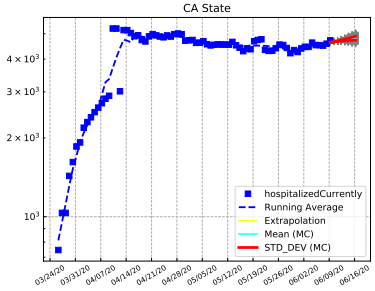


(d) Changes in daily cases.

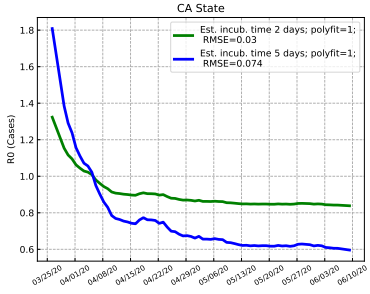
Figure 1



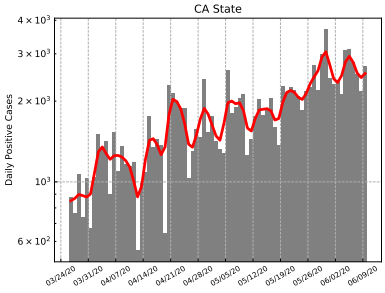
(a) Reproducibility via an IR model for hospitalizations. Data fluctuations around the middle of April (see Fig. 2b) causes the global RMS error to be high. This measure is a leading indicator of an apex in (b).



(b) Current hospitalizations in the state. Notice the slow drop off, as the Ro stays close to unity.



(c) Reproducibility via an IR model for the number of COVID-19 cases.



(d) Changes in daily cases.

Figure 2

4 Discussion

The state of the COVID-19 epidemic is currently fluid. At the time of submission of this paper, states around the U.S. are beginning to reopen, Memorial Day is past, and racial protests have erupted around the country. In addition to hospitalizations, we have studied trends in the number of infections [4] as well. We observed intra-week systematic waves in hospitalizations for every state we have studied.

States were chosen to represent different regions of the U.S. We see that some states like NY have bent the curve over, others like CA have simply flattened the curve, while data in yet other states remains noisy. Our log-normal distribution function is used to forecast hospitalizations a week into the future.

We have encapsulated the current results for states in a color-coded, self-explanatory table (Table 1). Overall, we do not notice a tremendous surge across the country yet. But the tide of states in “red" is rising. However, infections remain “subcritical" in many of the states we studied. On the other hand, several states show a sustained increasing trend in the number of infections, which should eventually affect hospitalizations.

Table 1: The table lists the Ro values for the number of infections (CASES) and hospitalizations (HOSPITLZ.) respectively. The direction of an arrow indicates the trend in the respective Ro. A red arrow indicates a value of $Ro > 1$. A green arrow indicates a value of $Ro < 1$. A yellow cell color indicates that even if $Ro < 1$, the associated error implies the value could be greater than unity. The cell color is green otherwise. Notice that the range of $Ro(HOSPITLZ.)$ is extremely narrow.

State	Ro(HOSPITLZ.)	Ro(CASES.)
WI	0.92↘	0.87↘
WA	0.94↗	0.83↗
IA	0.95↘	0.84↘
NY	0.96↘	0.8↘
IL	0.96↘	0.81↘
WV	0.97↗	0.84↘
MN	0.97↘	0.93↘
OH	0.97↘	0.82↘
VA	0.99↘	0.82↘
KY	1.0↘	0.83↘
MS	1.0↗	1.0↗
CA	1.0↗	0.84↘
MI	1.01↘	0.79↘
AZ	1.01↗	0.9↘
GA	1.01↘	0.82↘
ND	1.02↗	0.93↘
OR	1.02↘	0.87↘
NC	1.03↘	0.91↘
NM	1.03↗	0.83↘
SD	1.03↗	0.91↘
SC	1.04↘	0.95↗
TX	1.04↗	0.93↘
AR	1.04↗	1.02↘

The important point to keep in mind is that the health care system must not be allowed to approach saturation to prevent disastrous situations from developing in a state. This has to be balanced against a difficult decision of what constitutes an “acceptable" casualty rate before states can be re-opened. While such

statements are self-evident, tools like the one we have developed in this paper can be used to inform policy decisions.

5 Conclusion

The main message is that the growth/decay of cases in a pandemic is governed by a log-normal distribution. This distribution changes in time according to a generalized diffusion equation. The log-normal distribution arises from the fact that an epidemic is rather like a chain reaction in a fission bomb. There are signs of an increase in hospitalizations in some states away from either coast in the U.S.

A recent preprint [10] suggests that the European strain of COVID 19 may be more transmissible than the Wuhan strain. In light of this we examine the progress of the disease in California (presumably caused by the Wuhan strain), and New York (presumably caused by the Italian strain). Based on a comparison of Figs. 1a and 2a, it may be tempting to say that the higher Ro in NY implies a higher transmissibility in that state, compared to CA. But we know that social distancing was imposed in NY later than in CA. So the difference could have arisen for this reason. Furthermore, we cannot distinguish the Ro between the two coasts within the estimated RMS error. In any event, as the curves were flattened, Ro diminishes below one. The implication is that even if there is a difference in the transmission of the two strains, social distancing has been able to control the progression of COVID 19.

Acknowledgments: I would like to acknowledge discussions with John Scott, and I thank LTC Daniel Bahaghidat for encouragement. I appreciate the references provided to me by LTC Nicholas Clark. Last but not least, I want to acknowledge Maneesha Chitanvis who helped me understand subtle implications of epidemiological models.

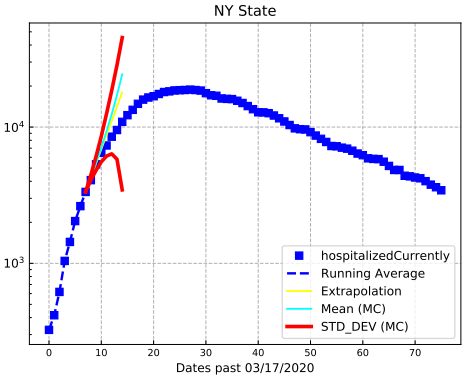
The work described in this paper was done under the auspices of the DOE while the author was stationed at the United States Military Academy at West Point in the Department of Physics and Nuclear Engineering.

Conflict of interest: The author states no conflict of interest.

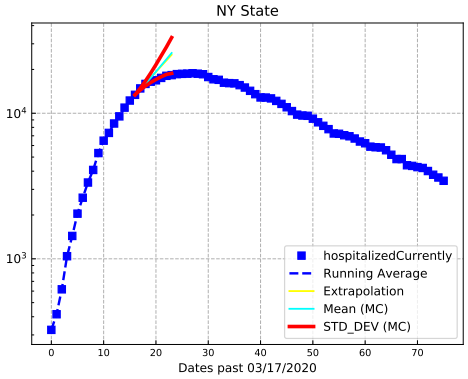
A Back testing to validate our model

We show the results of back testing the time-series analysis for NY, extrapolating seven days. Observe that the 85% confidence bands adequately describe the evolution over one week.

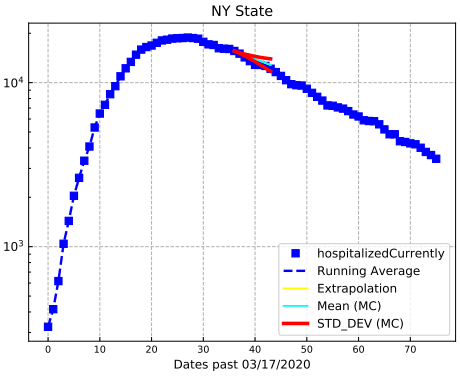
It is important to note the usefulness of the confidence bands, as exemplified in Figs. 3b-3d. These bands adequately captured the evolution of COVID-19 in NY over the respective periods, without overestimating the uncertainty.



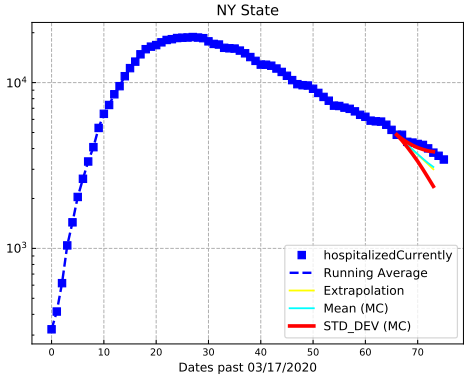
(a) 8 days past the start of data collection.



(b) 28 days past the start of data collection.



(c) 48 days past the start of data collection.



(d) 58 days past the start of data collection.

Figure 3

References

[1] CC Branas and et al. Flattening the curve before it flattens us: hospital critical care capacity limits and mortality from novel coronavirus (SARS-Cov2) cases in US counties - 3 and 6 week projections from April 2, 2020. 2020. URL https://behcolumbia.files.wordpress.com/2020/04/flattening-the-curve-before-it-flattens-us-20200405b.pdf.

[2] J Hullman. Leading with the Unknowns in COVID-19 Models. *Scientific American*, 2020. URL https://blogs.scientificamerican.com/observations/leading-with-the-unknowns-in-covid-19-models/.

[3] IHME team and Christopher JL Murray. Forecasting the impact of the first wave of the COVID-19 pandemic on hospital demand and deaths for the USA and European Economic Aea countries. *medRxiv*, 2020. 10.1101/2020.04.21.20074732.

[4] H.J. Unwin, S. Mishra, V.C. Bradley, and et al. State-level tracking of COVID-19 in the United States. 2020. https://doi.org/10.25561/79231. URL https://www.imperial.ac.uk/media/imperial-college/medicine/mrc-gida/2020-05-21-COVID19-Report-23.pdf.

[5] R Sosa and W Sosa. A non-parametric mathematical model to investigate the dynamic of a pandemic. *medRxiv*, 2020. https://doi.org/10.1101/2020.04.30.20086199. URL https://www.medrxiv.org/content/10.1101/2020.04.30.20086199v1.full.pdf.

[6] WO Kermack and AG McKendrick. A contribution to the mathematical theory of epidemics. *Roy. Soc. Proc, A*, pages 700–720, 1927.

[7] JT Wu, K Leung, and GM Leung. Nowcast and Forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modeling study. *Lancet*, 395:689–697, 2020. URL https://doi.org/10.1016/S0140-6736(20)30260-9.

[8] Yao Yu Yeo, Yao-Rui Yeo, and Wan-Jin Yeo. A computational model for estimating the progression of COVID-19 cases in the US West and East Coasts. *medRxiv*, 2020. 10.1101/2020.03.24.20043026. URL https://www.medrxiv.org/content/early/2020/03/31/2020.03.24.20043026.

[9] R Ke, S Sanche, E Romero-Stevenson, and N Hengartner. Fast spread of Covid-19 in Europe and the U.S. and its implications: Even modest public health goals require comprehensive intervention. *medRxiv*, 2020. URL https://doi.org/10.1101/2020.04.04.20050427.

[10] B Korber, WM Fischer, S Gnanakaran, H Yoon, J Theiler, W Abfalterer, B Foley, EE Giorgi, T Bhattacharya, MD Parker, DG Partridge, CM Evans, TM Freeman, TI de Silva, CC LaBranche, and DC Montefiori. Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV2. *bioRxiv*, 2020. 10.1101/2020.04.29.069054. URL https://www.biorxiv.org/content/early/2020/05/05/2020.04.29.069054.