

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

Shirish M. Chitanvis\*

*Los Alamos National Laboratory,  
Los Alamos, N.M. 87545.*

(Dated: May 18, 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 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 generalized diffusion equation which incorporates the effect of social distancing to forecast 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 publication and our forecasts indicate possible precursors of increased hospitalizations.

Additionally we studied the reproducibility  $R_0$  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.

---

\* Also at Department of Physics and Nuclear Engineering, U.S. Military Academy at West Point, N.Y.; [shirish.chitanvis@westpoint.edu](mailto:shirish.chitanvis@westpoint.edu)

## I. INTRODUCTION

One goal while tracking the COVID-19 epidemic is to make forecasts of hospitalizations[1]. This calls for caution[2] as the uncertainty in the forecast has to be estimated for a virgin virus. The uncertainty can be reduced by using what has been learned from the history of hospitalizations. 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.

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[3]. Indeed, it would be desirable 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 Fokker-Planck/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 than do standard distributions.

## II. METHODS

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

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

where  $\alpha/\ln(2)$  is the inverse doubling time.  $\alpha$  can be reduced via social distancing and is related to the reproducibility  $R_0$ :

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

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:

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

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. 1a, which displays a “square root x” behavior near the beginning of the data):

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

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  $\xi_{rms}$  traveled by a particle is proportional to  $t^{1/2}$ . In our case  $\xi_{rms} \equiv \ln \left( \frac{I(t)}{I_0} \right)$ . And we know that the probability  $P(x, t)$  of finding a diffusing particle a distance  $x$  after a time  $t$  is described by:

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

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  $\xi \equiv \ln(I/I_0)$ , so that the general master equation we seek follows by inspection:

$$\frac{\partial P}{\partial \tau} - \frac{\partial^2 P}{\partial \xi^2} = \mathcal{S}(\xi, \tau)$$

$$\xi \equiv \ln \left( \frac{I}{I_0} \right) \quad (6)$$

where  $\tau \equiv \tau(t)$ ,  $P(\xi, \tau)$  is the probability that  $\xi$  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.

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

$$P(\xi, \tau) = \frac{1}{\sqrt{2\pi(\tau - \tau_0)}} \exp \left[ -\frac{(\xi - \xi_0)^2}{4(\tau - \tau_0)} \right] \quad (7)$$

And since  $\xi$  is logarithmic in  $I$ , it follows that the number of hospitalizations will be determined by a log-normal distribution. This is especially meaningful since the pandemic shows signs of producing an immense number of infections. 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  $\xi_0$ . Here the diffusion constant  $D = 1$  [*dimensionless*].

## A. Data extraction

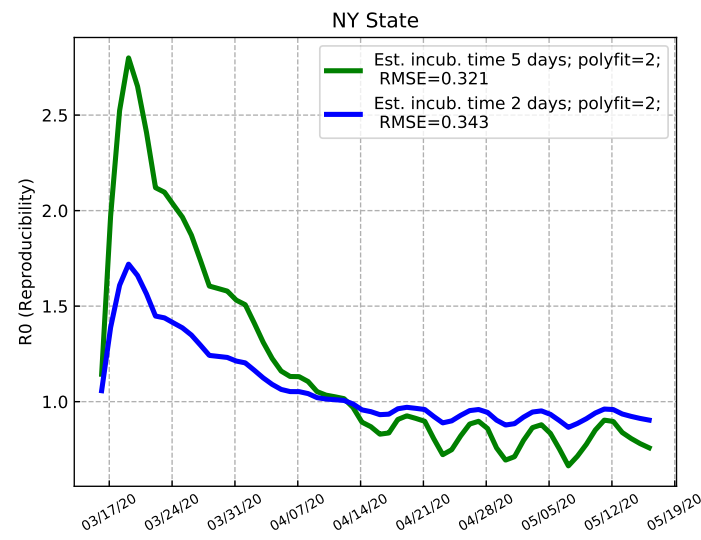
We used COVID19 state-wide hospitalization data from the Covid tracking project: <https://covidtracking.com/api/v1/states/daily.csv> to plot and obtain fits to  $\tau(t)^{1/2}$ .

## B. Data Analysis

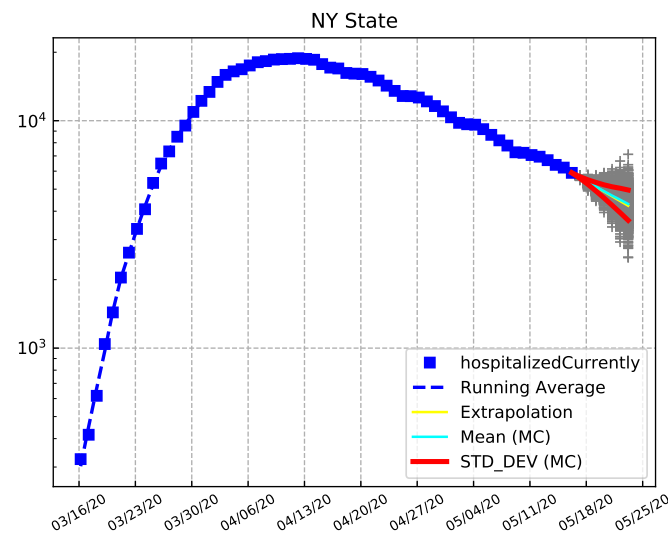
A simple IR (Infections (hospitalizations), Recovered cases) model[4], a subset of the SEIR model[5], was used to obtain  $R_0$  from this data. By dealing directly with a population of hospitalized patients, we are not required to track susceptible and exposed persons. We find from data that the recovery 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[6, 7] 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.

### III. RESULTS

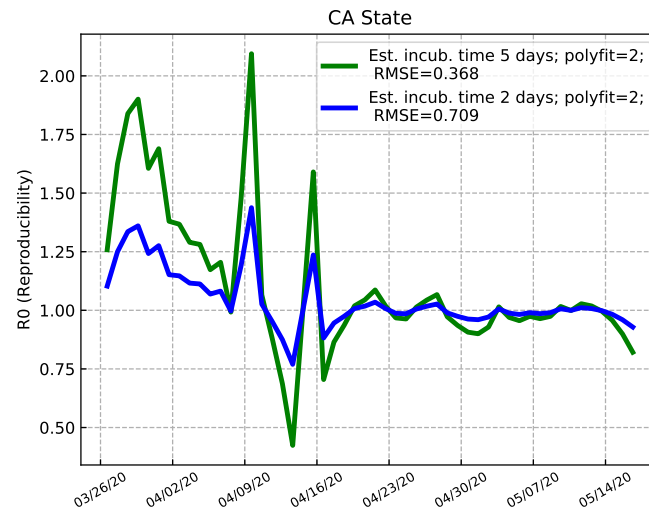


(a) Reproducibility via an IR model. This measure is a leading indicator of an apex in (b).

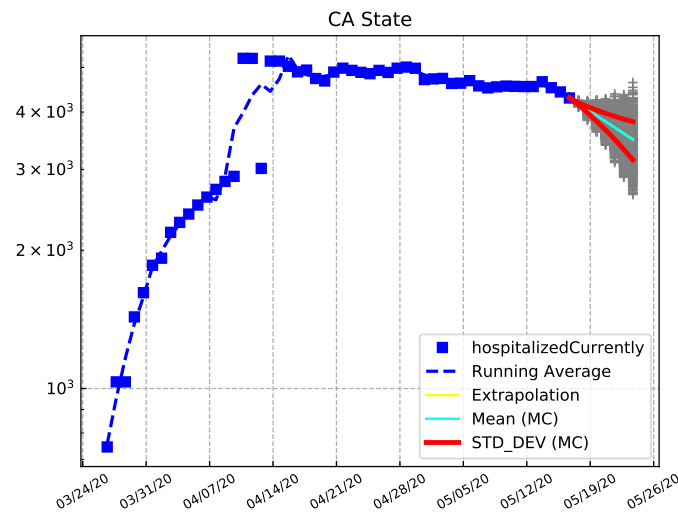


(b) Current hospitalizations in the state.

FIG. 1

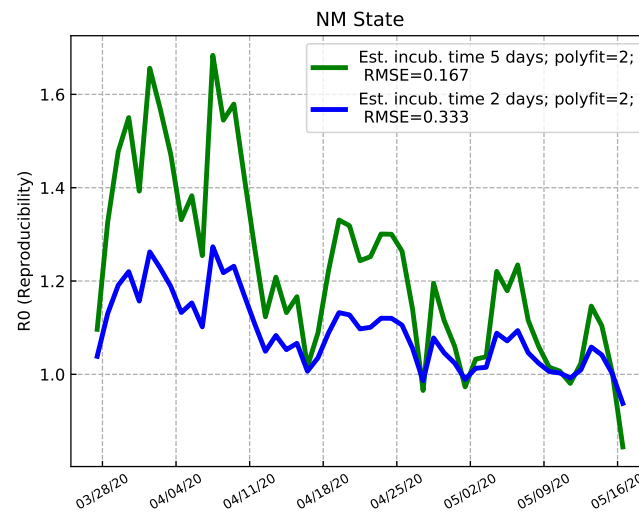


(a) Reproducibility via an IR model. Data fluctuations around the middle of April (see Fig. 2b) causes the global RMS error to be high. This measure is a leading indicator of an apex in (b).

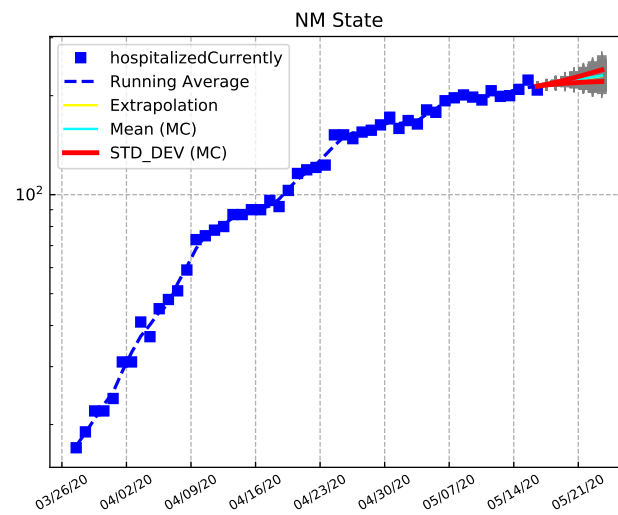


(b) Current hospitalizations in the state. Notice the slow drop off, as the  $R_0$  stays close to unity.

FIG. 2



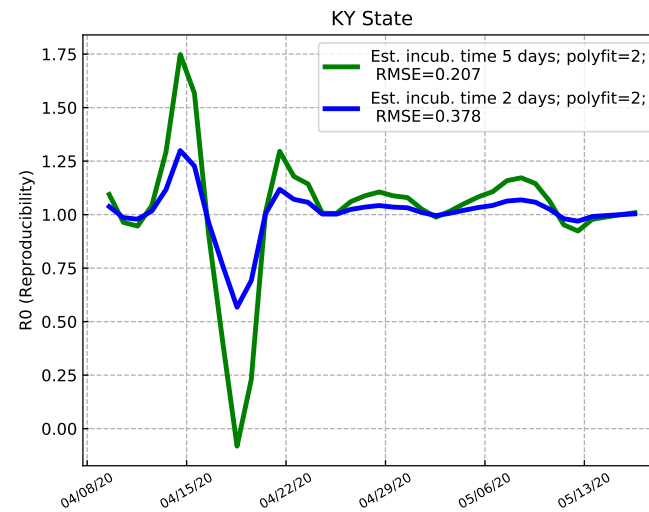
(a) Reproducibility via an IR model. This measure is a leading indicator of an apex in (b). Observe the fluctuations in  $R_0$  as social distancing changes over time.



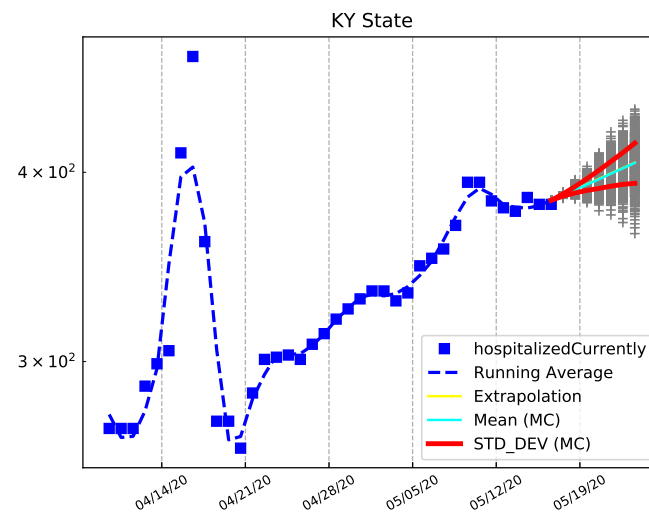
(b) Current hospitalizations in the state. A full apex has not yet formed.

FIG. 3



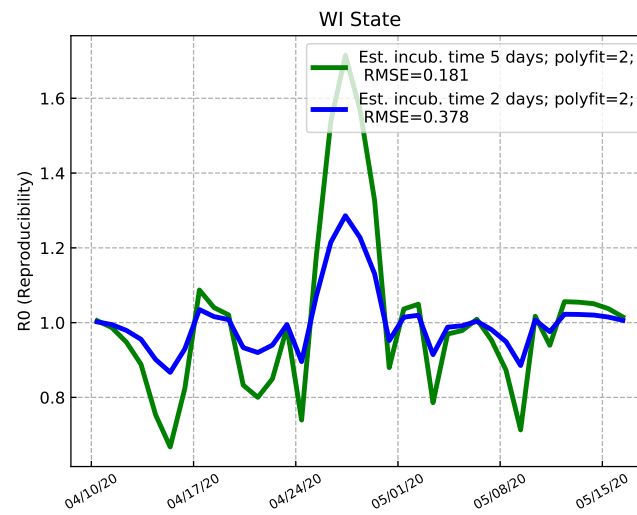


(a) Reproducibility via an IR model. This measure is a leading indicator of an apex in (b).

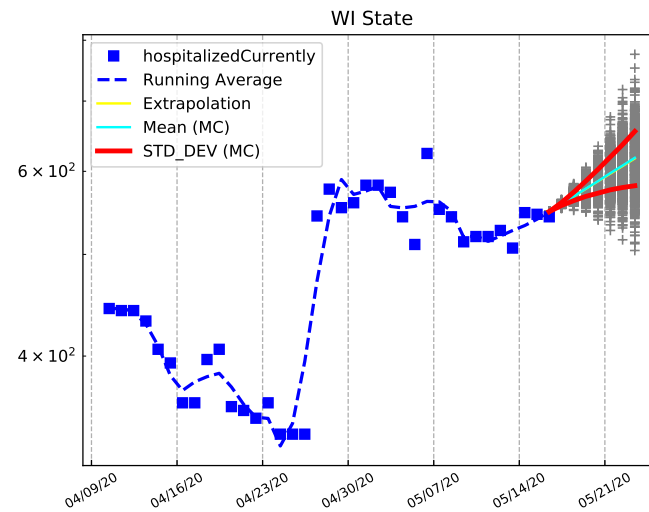


(b) Current hospitalizations in the state.

FIG. 4

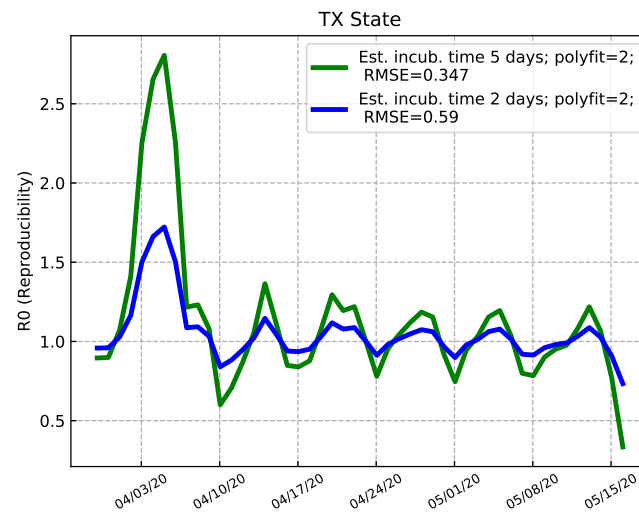


(a) Reproducibility via an IR model. This measure is a leading indicator of an apex in (b).

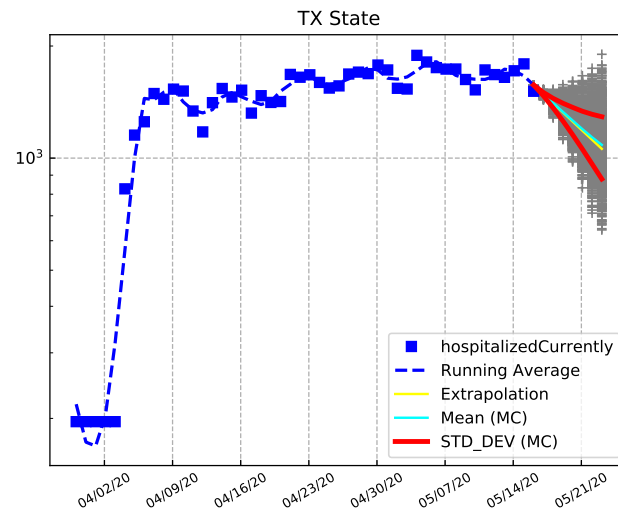


(b) Current hospitalizations in the state are increasing.

FIG. 5



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



(b) Current hospitalizations in the state are increasing, as is  $R_0$ .

FIG. 6

## IV. DISCUSSION

The state of the COVID-19 epidemic is currently fluid. At the time of publication states around the U.S. were beginning to reopen. And you can see possible precursors of resurgence of hospitalizations in some of them. California continues to reside on a plateau with  $Ro \approx 1$ . The reported hospitalizations in Kentucky appear to be relatively low. Observe the intra-week systematic “waves” in hospitalizations for almost every state.

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.

A recent preprint[8] 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.

## V. 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.

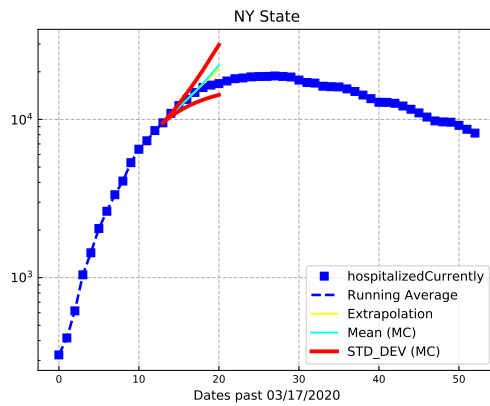
## VI. ACKNOWLEDGMENTS

I would like to acknowledge discussions with John Scott, and I thank LTC Bahaghidat for encouragement. Last, but not least, I appreciate the references provided to me by LTC Clark.

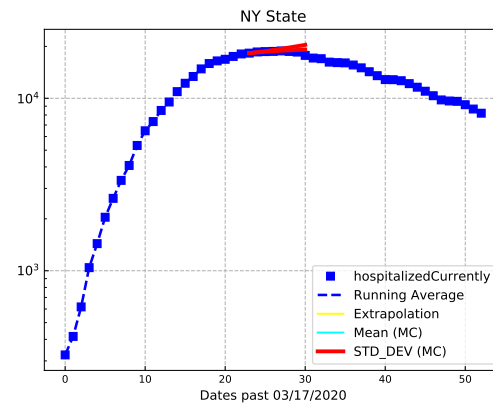
The work described in this paper was done under the auspices of the DOE while the author was stationed at West Point in the Department of Physics and Nuclear Engineering.

## Appendix A: Back testing to validate our model

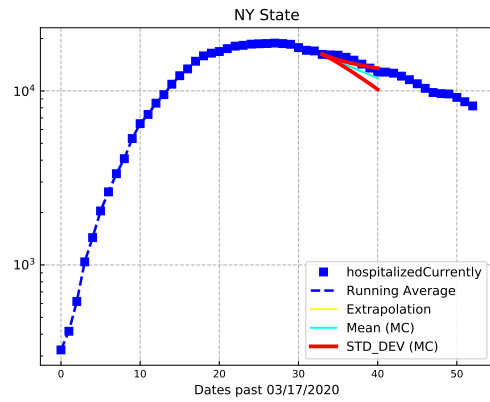
We show the results of back testing the time-series analysis for NY, extrapolating seven days. Observe that the 85% confidence bands adequately describe the evolution over one week.



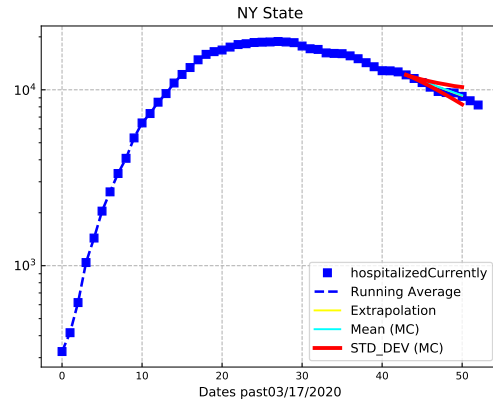
(a) 14 days past the start of data collection.



(b) 24 days past the start of data collection.



(c) 34 days past the start of data collection.



(d) 44 days past the start of data collection.

FIG. 7

- 
- [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] R Sosa and W Sosa. A non-parametric mathematical model to investigate the dynamic of a pandemic. *medRxiv*, 2020. doi:<https://doi.org/10.1101/2020.04.30.20086199>. URL <https://www.medrxiv.org/content/10.1101/2020.04.30.20086199v1.full.pdf>.
  - [4] WO Kermack and AG McKendrick. A contribution to the mathematical theory of epidemics. *Roy. Soc. Proc. A*, pages 700–720, 1927.
  - [5] 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](https://doi.org/10.1016/S0140-6736(20)30260-9).
  - [6] 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. doi:10.1101/2020.03.24.20043026. URL <https://www.medrxiv.org/content/early/2020/03/31/2020.03.24.20043026>.
  - [7] 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>.
  - [8] 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-cov-2. *bioRxiv*, 2020. doi:10.1101/2020.04.29.069054. URL <https://www.biorxiv.org/content/early/2020/05/05/2020.04.29.069054>.