2020 Fourth International Conference on Multimedia Computing, Networking and Applications (MCNA)

# Improved Estimation of Daily COVID-19 Rate from Incomplete Data

Ian McCulloh*, Kevin Kiernan†, Trevor Kent§
*Applied Intelligence*
*Accenture*
Washington DC
*https://orcid.org/0000-0003-2916-3914
†kevin.kiernan@accenturefederal.com
§trevor.g.kent@accenturefederal.com

***Abstract*—Novel diseases such as COVID-19 present challenges for identifying and assessing the impact of public health interventions due to incomplete and inaccurate data. Many infected persons may be asymptomatic, pre-symptomatic, or may choose to not seek medical treatment. Insufficient testing and reporting standards coupled with reporting delays may also affect the accuracy of case count, recovery rate, fatalities and other key metrics used to model the disease. High error in these metrics are propagated to all aspects of public health response including estimates of daily transmission rates. We propose a method that integrates Monte Carlo simulation based on clinical studies, linear noise approximation (LNA), and Hidden Markov Models (HMMs) to estimate daily reproductive number. Results are validated against known state population behavior, such as social distancing and stay-at-home orders. The proposed approach provides improved model initial conditions resulting in reduced error and superior modeling of COVID-19 disease dynamics, notably including the effective reproduction rate $R_t$.**

***Keywords—COVID-19, COVID-19, Monte Carlo, Linear Noise Approximation, LNA, Bayesian Data Augmentation, Partially Observed Markov Process, Coronavirus***

## I. Introduction

National and local governments must balance numerous cross-cutting equities in choosing interventions to address a pandemic disease. Social distancing and stay-at-home orders are highly effective in limiting the spread of infectious diseases like COVID-19, but have harmful impact on the economy. Other interventions, such as increased hand-washing and mask-wearing, are more compatible with commercial activity but may be less effective at combating the pandemic. In epidemiological terms, an intervention's utility to policymakers is a cost-benefit tradeoff between reduction in $R_t$ --the daily basic reproductive number of the virus-- and the social and economic costs of the intervention itself. The most effective interventions are the ones that maximally suppress $R$ with the least social and economic cost. While this is an appealingly simple framework for policymakers, it is difficult to implement in practice due to the fact that $R$ is not a directly measurable quantity; it must be derived from infected case counts, which have proven unreliable in the early stages of the outbreak.

In this paper, we propose a method to more accurately estimate $R_t$ and validate the method against known data from New York and Florida. The paper is organized as follows. The background section will review key epidemic models, methods and shortcomings. The method section specifies our assumptions, Markov model specification, application of linear noise approximation (LNA), and Monte Carlo methods. The results section presents the method applied to New York and Florida for expository purpose. Finally, conclusions for modeling and assessing the impact of COVID-19 interventions are discussed.

## II. Background

### A. Mechanistic Models

Mechanistic models use known dynamics of infectious diseases, coupled with parameters estimated from an outbreak to predict outcomes based on potential interventions [1,5]. In contrast to empirical models, they mitigate the risk of overfitting and are better for evaluating the potential impact of interventions such as mandatory mask-wearing or stay at home orders. An early Markov Chain model was effective at modeling the spread of COVID-19 in Spain and Italy, accounting for the high number of asymptomatic patients, those shedding virus, yet showing no symptoms [1]. Attempts to apply this model to U.S. populations, however, proved ineffective [5]. A key reason was due to inaccurate reporting of case counts due to lack of testing differences between U.S. and European medical systems ability to track health data, leading to incorrect model initial conditions [5]. Effective "what-if" analyses depend upon accurate parameter estimation and correct initial conditions (e.g. number susceptible, exposed, infected, recovered). Errors in estimating correct initial conditions are propagated through the models and prevent effective outcome prediction.

### B. Overcoming Data Quality Issues

McCulloh et al overcame these model issues by using the most reliable data reported, death rates, and sampling from clinical estimates of fatality rates to create a Monte Carlo estimate of true case counts [5]. They validated these findings against serology studies conducted in several states including New York. The accuracy of this method hinges on correct counts of COVID-19 related deaths. If deaths are inaccurately reported, the error is propagated through the model.

An article from the Washington Post highlights potential under-reporting of COVID-19 related deaths as shown in Fig1 [2]. The figure shows a sharp increase in U.S. deaths above seasonal norms. The dark-shaded region indicates those deaths that are reported as COVID-19, while the lighter shaded region is not reported as COVID-19. Based on these data, we may infer that not all true COVID-19 deaths may have been accurately attributed to the disease. While it is undeniably true that some fraction of the excess deaths are due to pandemic-related but non-COVID-19 causes (e.g. suicides, accidents, or failure to seek medical treatment due to fear of infection), we assert that the dominant share of the excess deaths observed in our sample were COVID-19-related. Our reasons for this inference are threefold:

978-1-7281-8373-2/20/$31.00 ©2020 IEEE

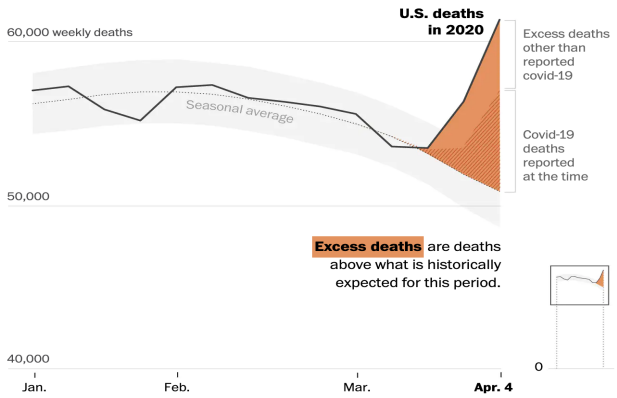


*Figure 1:Excess COVID-19 deaths as reported by the Washington Post*

First, the early stages of the outbreak were characterized by a lack of testing resources and confusion over how to classify deaths.

Second, COVID-19's increased lethality in patients with comorbidities increases the probability that a death may be misclassified as resulting from the comorbidities themselves.

Third, even taking into account the reasons COVID-19 deaths are susceptible to under-counting, COVID-19 became the leading cause of death in the United States by April 2020 [12]. This effect was especially pronounced in the areas which were experiencing the worst outbreaks.

There are additional limitations with prior approaches to model COVID-19. The time that infections first arrived, $t_0$, in a population is unknown and difficult to estimate. While contact tracing can be used to track the origin of the disease, the low fatality rates and high asymptomatic patients make this difficult and potentially inaccurate. The fatality rate, $\omega$, is also problematic due to poor data quality linking positive COVID-19 to cause of death, inaccurate counting of those testing positive for the disease, and the aforementioned observation of deaths, or death detection rate, $\rho$.

### *C. Maximum A Posteriori Estimation*

A maximum a posteriori (MAP) estimate is a Bayesian approach to approximate an unknown quantity. MAP uses an optimization objective function augmented by prior knowledge such as a mechanistic model and improved initial conditions to estimate the unknown quantity. We utilize the **stemr** package in R for MAP estimation [8]. The **stemr** package implements a Bayesian data augmentation algorithm for fitting stochastic epidemic models (SEMs) with arbitrary dynamics to partially observed epidemic count data via linear noise approximation (LNA). The LNA approach improves approximation of the one–step conditional transition densities of the Markov process by restarting the approximation at the beginning of each inter–observation interval as shown by [4,8].

### *D. Linear Noise Approximation*

The challenge with traditional fitting approaches is that the dimensionality of the state space makes inferring the transition densities of the entire system an intractable problem to solve. This stems from the fact that when the data is underreported, one must sum over all paths from which the observed data could have arisen [8]. The Linear Noise Approximation (LNA) algorithm significantly reduces the computational requirements for inferring the transition densities of cumulative infection, exposure, and recovery incidence.

The **stemr** package implements a Bayesian data augmentation algorithm for fitting stochastic epidemic models (SEMs) with arbitrary dynamics to partially observed epidemic count data via LNA [8]. Cumulative incidence is assumed to follow a Markov Jump Process (MJP) whose dynamics are modeled with a stochastic differential equation via the diffusion approximation. The LNA method works by Taylor expanding the stochastic residual of the SDE centered around the deterministic ODE limit and removing higher order terms. This permits the convergence on a solution without having to explore the entire state space. This has an explicit solution as a Gaussian random variable and can approximate the stochastic aspects of a Markov process under certain conditions [8,9]. Let $S(t)$, $E(t)$, $I(t)$, $R(t)$, and $D(t)$ represent the current counts of the population that are considered to be susceptible, exposed, infected, recovered or dead. Define the cumulative $S \to E, E \to I, I \to R$, and $I \to D$ transitions up to time $t$ by:

$$\mathbf{N}(t) = (N_{SE}(t), N_{EI}(t), N_{IR}(t), N_{ID}(t)) \quad (1)$$

and let:

$$\Delta\mathbf{N}(t) = \mathbf{N}(t) - \mathbf{N}(t-1) = (\Delta N_{SE}(t), \Delta N_{EI}(t), \Delta N_{IR}(t), \Delta N_{ID}(t)) \quad (2)$$

denote the change in cumulative incidence transitions over the time period $(t-1, t)$.

For a SEIRD model the state vector can be written as:

$$\mathbf{X}(t) = \begin{pmatrix} S(t) \\ E(t) \\ I(t) \\ R(t) \\ D(t) \end{pmatrix} = \begin{pmatrix} S_0 - N_{SE}(t) \\ E_0 + N_{SE}(t) - N_{EI}(t) \\ I_0 + N_{EI}(t) - N_{IR}(t) - N_{ID}(t) \\ R_0 + N_{IR}(t) \\ D_0 + N_{ID}(t) \end{pmatrix} \quad (3)$$

With underlying state space:

$$S_X = \{(s,e,i,r,d): s,e,i,r,d \in \{0,\dots,P\}, \; s+e+i+r+d = P\} \quad (4)$$

where $P$ represents the population size.

The waiting times between state transitions are assumed to be exponentially distributed which means $\mathbf{X}(t)$ evolves according to a Markov Jump Process (MJP). The integer valued MJPs are approximated with the real-valued diffusion processes associated with the Chemical Langevin Equation (CLE) [8]. This is a well-studied stochastic differential equation that describes Brownian motion of particles in a fluid due to collisions with other molecules in the fluid.

Given the vector of transition rates at time $t$:

$$\boldsymbol{\lambda}(\mathbf{X}(t)) = \begin{pmatrix} \beta(t)S(t) \\ \sigma E(t) \\ (1-\omega)\gamma I(t) \\ \omega\delta I(t) \end{pmatrix} \quad (5)$$

and $\mathbf{\Lambda}(\mathbf{X}(t)) = \text{diag}\,(\boldsymbol{\lambda}(t))$, the CLE is given by:

$$d\mathbf{N}(t) = \boldsymbol{\lambda}(\mathbf{X}(t))dt + \mathbf{\Lambda}(\mathbf{X}(t))^{\frac{1}{2}}d\mathbf{W_t} \quad (6)$$

where $\mathbf{W}_t$ is distributed a bivariate Brownian motion with independent components.

$\mathbf{X}(t)$ as defined in above can then be decomposed into a vector of initial conditions $\mathbf{x_0} = \mathbf{X}(t_0)$ and $\mathbf{N}(t)$ along with the matrix $\mathbf{A}$ that specifies the direction of flow for the system subjected to a new infection, exposure, recovery, or death.

$$\mathbf{X}(t) = \mathbf{x_0} + \mathbf{A}^T\mathbf{N}(t) \quad (7)$$

Where:

$$\mathbf{A} = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 0 & 1 \end{pmatrix} \quad (8)$$

Combining the above yields:

$$d\mathbf{N}(t) = \boldsymbol{\lambda}(\mathbf{x_0} + \mathbf{A}^T\mathbf{N}(t))dt + \Lambda(\mathbf{x_0} + \mathbf{A}^T\mathbf{N}(t))^{\frac{1}{2}}d\mathbf{W}_t \quad (9)$$

Taking the log transform of the system above yields a new system that better accounts for the multiplicative effects compartment volumes have on transition rates [8]. The LNA is then derived by decomposing the result into the deterministic ODE limit and stochastic residual via a Taylor series expansion. Once the LNA is defined, slice sampling can be used to generate a distribution of LNA paths to uncover the MAP estimate of a latent variable of interest such as exposed incidence. The ability to uncover the exposed proportion of the population is invaluable given that non-pharmaceutical interventions like mask wearing and social distancing directly affect the rate of exposure to the virus.

## III. Method

We implement a simple Markov model with compartments susceptible, exposed, infected, recovered, dead as shown in Fig 2. We choose explicitly not to model the asymptomatic population as a separate compartment in the model, noting recent clinical findings that the viral load of an infected subject does not differ meaningfully between the asymptomatic and symptomatic populations [13]. A separate Asymptomatic-Infected compartment would add an unnecessary layer of complexity; instead, we attribute these cases to the overall miscount and assume all infectious individuals are subjected to the same transition densities. Empirical data for the model is obtained from the Johns Hopkins University Coronavirus Resource Center as a publicly available source of globally reported COVID-19 case count and related deaths [3]. We focus on New York and Florida data for expository purpose, but results have been replicated for multiple locations within the U.S. The **stemr** package constructs the data series containing all subsequent values of disease transmission rate, $\beta(t)$, by Bayesian inference. This $\beta(t)$ data series is one of the most important results of this analysis, as it permits improved estimation of the reproductive rate $R_t$.

The method accepts an observed incidence and solves for the MAP estimate of true incidence $Y_l$ over a given time period. In our case we treat deaths as the observed variable given that it is far more likely for a death to be observed and recorded. However, this still reflects only a fraction of the true incidence

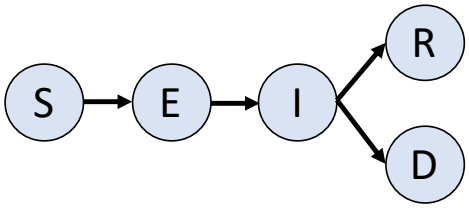


*Figure 2: S-E-I-R-D Markov Model*

due to factors such as the comorbidity of other life threatening illnesses associated with COVID-19. Thus true incidence is assumed to follow a negative binomial distribution with mean death detection rate $\rho$ and overdispersion parameter $\phi$ conditional on the observed incidence $\Delta N_{ID}(t)$.

$$Y_l|\Delta N_{ID}(t) \sim NB(\mu = \rho\Delta N_{ID}(t), \sigma^2 = \mu + \frac{\mu^2}{\phi}) \quad (10)$$

A sensitivity analysis was conducted for the overdispersion parameter and no significant difference in results were obtained for sufficiently large values. Using death incidence as a source of truth, we uncover the latent epidemic process by applying a constrained random walk (i.e. subject to the Markov property) to sample from an unknown distribution. We use Markov chain Monte Carlo (MCMC) to estimate the transition densities of the system. Once the transition density from susceptible to exposed has been solved for, $\beta(t = t_0)$ can be obtained directly from the transition rate equation defined above. Finally, $\beta(t = t_0)$ is a function of the reproductive number $R_t$ from the previous observed time step to the current time as shown in Eq (17) below. MCMC estimates yield MAP fit parameters along with a time series for all incidence compartments of the SEIRD Markov model.

The transition matrix for our Markov model is specified as shown in Eq. (11) and difference equations as specified in Eqs. (12)-(16). The parameters of the model were either fixed based on a review of the literature or estimated via the **stemr** package and are shown in Table 1. The daily reproductive number, $R_t$, is derived from Eq (17). The value of $R_t$ is an effective measure of the relative transmission risk and we posit that this value can be used as a variable to evaluate the effectiveness of non-pharmaceutical interventions such as mask-wearing mandates or stay-at-home orders.

$$T(t) = \begin{bmatrix} 1-\beta(t) & \beta(t) & 0 & 0 & 0 \\ 0 & 1-\sigma & \sigma & 0 & 0 \\ 0 & 0 & 1-\gamma(1-\omega)-\omega\delta & \gamma(1-\omega) & \omega\delta \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (11)$$

$$S(t+1) = (1-\beta(t))S(t) \quad (12)$$

$$E(t+1) = \beta(t)S(t) + (1-\sigma)E(t) \quad (13)$$

$$I(t+1) = \sigma E(t) + (1-\gamma(1-\omega)-\omega\delta)I(t) \quad (14)$$

$$R(t+1) = (1-\omega)\gamma I(t) \quad (15)$$

$$D(t+1) = \omega\delta I(t) \quad (16)$$

$$\beta(t) = \frac{R_t\delta}{P} \quad (17)$$

TABLE I. MARKOV PARAMETER SPECIFICATION

| Prm | Description | Value | Fixed | Source |
|---|---|---|---|---|
| $s$ | Serial interval | 5.1dy | Fixed | [7,11] |
| $g$ | Mean recovery time | 21.0dy | Fixed | [7] |
| $d$ | Mean time to death | 17.8dy | Fixed | [7] |
| $f$ | Overdispersion term | 50 | Fixed | |
| $t_0$ | Introduction date | -25dy | Fixed | |
| $R_t$ | Reproductive number at time $t$ | -- | Est | -- |
| $r$ | Death detection prob. | 66% | Fixed | [2] |
| $w$ | Infection fatality rate | 1.3% | Fixed | [7] |

## IV. RESULTS

MCMC estimates yield MAP fit parameters, time series for all incidence compartments of the SEIRD model. Fig 3-5 show the estimated transition counts for New York. The reported case counts are overlaid on Fig 4, which illustrates the magnitude of under reporting. The shaded area represents the time frame that New York was in a state of lockdown, where the government-imposed restrictions on social gathering and certain businesses. Fig. 5 shows observed death counts as points overlaid on the model estimates. Based on the model fit, it appears deaths are also under reported in late April through May 2020 and improves over time as expected. The epidemic peak of this model occurs in early April, which is consistent with our knowledge of New York's COVID-19 epidemic. Fig. 6 displays $R_t$ over time. Though we primarily focus on presenting results from New York for expository purposes, the Bayesian Data Augmentation approach is robust and flexible enough to model the COVID-19 outbreaks in other states without significant modification. As a demonstration of this capability, we have elected to include similar plots for the state of Florida in Fig 7-9.

Results of our model can be validated against serology reports that sampled the general population to infer the magnitude of the pandemic. Our estimate of the total infection rate at the end of May 2020 is calculated by summing Eqs (15) and (16), which accounts for 10.8% of the population in New York. This estimate is much closer to the rate of 12.3% found in a seropositivity study conducted around the same time than the rate of 1.8% reported on public tracking sites [6]. Given issues with seropositivity sampling, such as bias towards recovered individuals participating in studies, it is possible that the estimate of 10.8% is a more accurate statewide estimate than the 12.3% found in the study. Similar results exist for Florida and other states.

The results from Florida are particularly interesting, since the epidemic peak is multi-modal occurring in mid-April and late July 2020. As shown in Fig 7, the lack of adequate testing

### A. New York

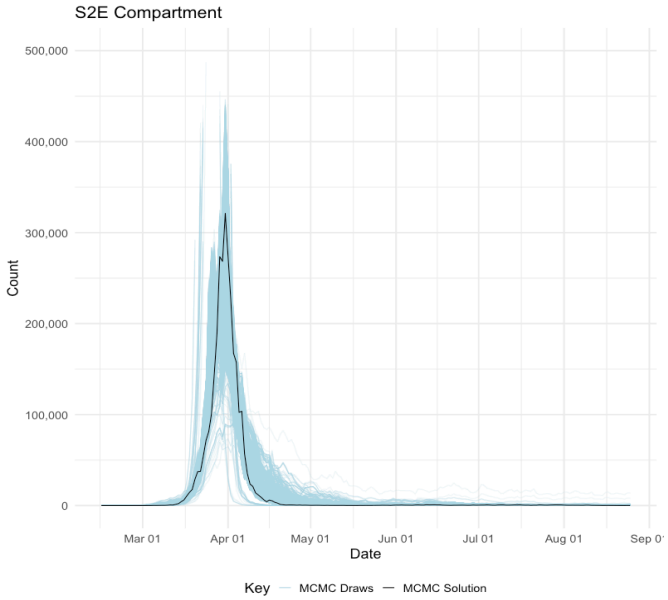


*Figure 3: Daily Susceptible to Exposed in NY State*

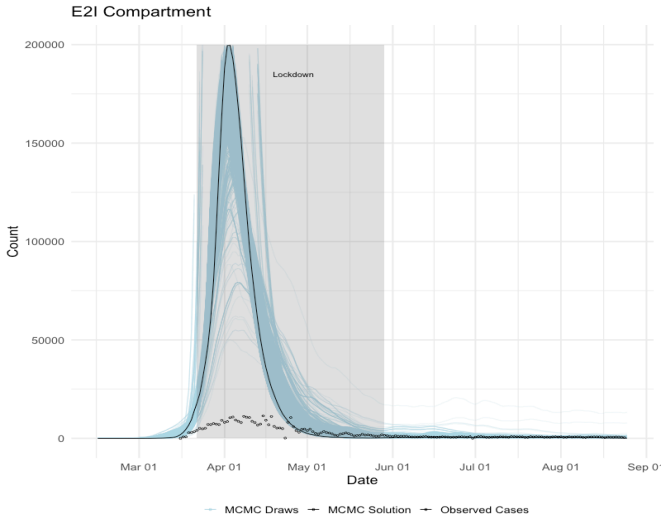


*Figure 4: Daily Exposed to Infected in NY State*

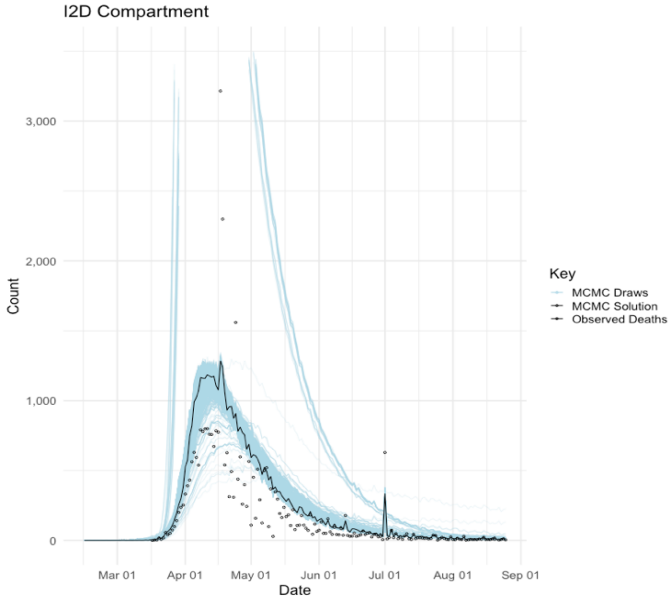


*Figure 5: Daily Infected to Dead in NY State*

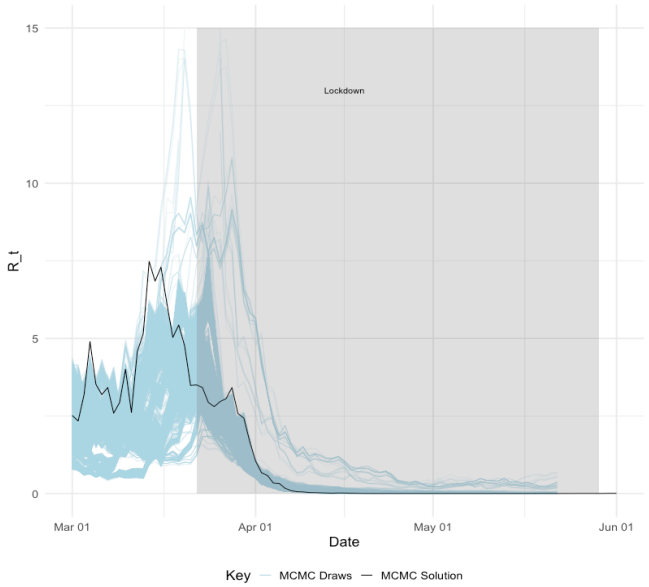


*Figure 6: Effective Reproductive Rate ($R_t$) in NY State*

## B. Florida

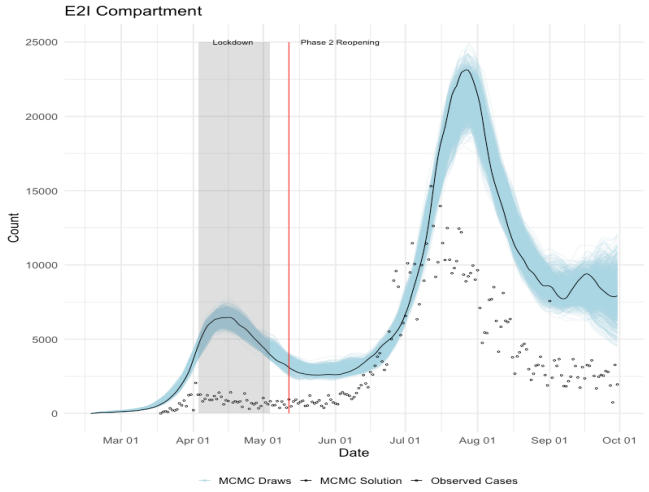


*Figure 7: Daily Exposed to Infected in Florida*

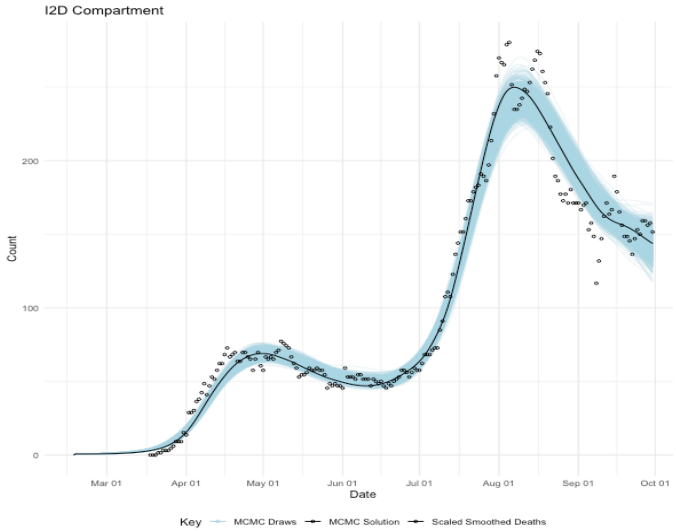


*Figure 8: Daily Infected to Dead in Florida*

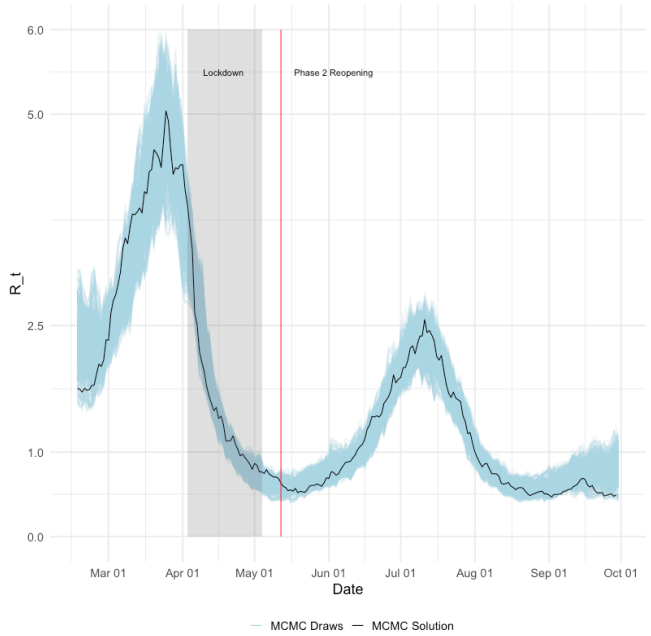


*Figure 9: Effective Reproductive Rate ($R_t$) in Florida*

resources early in the pandemic is clear, showing significant case-count under-reporting. Testing improves through June and early July, but under-reporting again becomes problematic in late July through the writing of this paper. We speculate that under-reporting early in the pandemic is largely due to a lack of testing resources, while under-reporting later in the pandemic is due to reduced COVID-19 fear and a lack of willingness for people to be tested.

The refined estimate of $R_t$ as shown in Figs 6 and 9 provides another intriguing insight. The initial epidemic peak at the state level occurs days before any government imposed lockdown. This suggests that populations are self-selecting to adopt non-pharmaceutical interventions such as social distancing and mask wearing well in advance of any government directive to do so. The government officials, likely reacting to the same information available to the public are delayed in implementing their response and there is little evidence to suggest that they are actually effective in decreasing the risk of transmission with these data. In contrast, lifting government restrictions may create a false sense of security and lead to increased transmission risk as seen in Florida in Fig. 9. Perhaps government investment in public health communication and conveying the true risk of transmission is more impactful than other policy actions. We recommend the use of the proposed estimate of $R_t$ for assessing risks to the population.

We recognize that there are many social factors contributing to COVID-19 dynamics that have not been captured in this study. One important consideration with government lockdowns is the sense of freedom people have to adopt non-pharmaceutical precautions. Without a lockdown or government policy, people may not have the choice to work from home, social distance, or avoid high risk situations for fear of losing their job. An assessment of these factors is well beyond the scope of this paper. We simply use this discussion to highlight the potential benefit of a refined

measure of $R_t$ as one of many considerations for assessing the impact of government policies on the spread of COVID-19.

## V. Conclusion

This work extends a Bayesian data augmentation method leveraging LNA for Markov transition probability estimation to the novel COVID-19 pandemic. Traditional Markov processes are difficult to fit, due to inaccurate/unknown initial conditions such as the number of exposed people, date of initial disease onset, and transmission probabilities. Empirical data, when combined with known clinical parameters, produce infeasible solutions or fail to converge. The proposed method provides an approach to estimate true case counts and unknown parameters, resulting in functioning models that appear more valid than reported case counts or alternative models. The resulting estimates of daily transmission risk as measured by the daily reproduction number $R_t$ provide an improved measure for the impact of public health interventions and policy decisions such as stay-at-home orders or mandatory mask wearing.

A key limitation of this approach is the lack of existing data for disease co-morbidity. Pre-existing conditions such as respiratory disease, hypertension, diabetes and old age are known to increase the risk of death in clinical studies, yet these data are absent at scale for inclusion in population models. With improved data quality, more complex models could be constructed using the same approach to reduce induced error from inaccurate reporting.

Using this improved estimate of $R_t$ future research may include county-level estimation and regression against census data for population rates of tobacco use, population density, and other contributing factors. With reduced error in $R_t$ estimation, improved county-level risk models may help authorities make better policy decisions such as where to relieve restrictions and allow economic activities to resume.

Finally, we argue that the proposed estimate of $R_t$ provides a better measure of COVID-19 transmission risk than reported case counts with known data quality issues. The proposed model is therefore preferred over reported case-count data for informing government policy makers and commercial industry when making decisions regarding the safety of citizens and employees respectively.

## Acknowledgment

This research was funded by Accenture Applied Intelligence in an effort to better support clients responding to COVID-19and better estimate future impacts.

## VI. References

[1] A. Arenas, W. Cota, J.Gomez-Gardenes, S. Gómez, C. Granell, … and B. Steinegger. A mathematical model for the spatiotemporal epidemic spreading of COVID19. *MedRxiv*. 2020.

[2] E. Brown, A. Tran, and R. Thebault. Excess U.S. Deaths Hit Estimated 37,100 in Pandemic's Early Days, Far More than Previously Known. *The Washington Post*, WP Company, 2 May 2020, Retrieved September 14 2020: www.washingtonpost.com/investigations/2020/05/02/excess-deaths-during-covid-19/.

[3] E. Dong, H. Du, and L.Gardner. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis*; published online Feb 19.

[4] P. Fearnhead, V. Giagos, and C. Sherlock. Inference for reaction networks using the linear noise approximation. *Biometrics*, 70:457–466, 2014.

[5] I. McCulloh, K. Kiernan, T. Kent. Inferring True COVID19 Infection Rates from Deaths. *Frontiers in Big Data, Public Health, Medicine.* 2020

[6] New York Press Office. *Amid Ongoing COVID-19 Pandemic, Governor Cuomo Announces Results of Completed Antibody Testing Study of 15,000 People Showing 12.3 Percent of Population Has COVID-19 Antibodies*. Retrieved May 12, 2020: https://www.governor.ny.gov/news/amid-ongoing-covid-19-pandemic-governor-cuomo-announces-results-completed-antibody-testing

[7] R. Verity, L.C. Okell, I. Dorigatti, P. Winskill, C. Whittaker, … and A. Dighe. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet Infectious Diseases*. 2020.

[8] J. Fintzi, J. Wakefield, and V.N. Minin. A linear noise approximation for stochastic epidemic models fit to partially observed incidence counts. *arXiv.* 2020.

[9] E. W. J. Wallace, D. T. Gillespie, K. R. Sanft, and L. R. Petzold. Linear noise approximation is valid over limited times for any chemical system that is sufficiently large. *IET systems biology*, 6:102–115, 2012.

[10] D. J. Wilkinson. Stochastic Modelling for Systems Biology. *CRC Press*, Boca Raton, 2011

[11] J. Zhang, M. Litvinova, W. Wang, Y. Wang, X. Deng, X., ... and Q. Wu. Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. *The Lancet Infectious Diseases*. 2020

[12] D. Keating, C. Esteban. Covid-19 is rapidly becoming America's leading cause of death. The Washington Post. 2020. https://www.washingtonpost.com/outlook/2020/04/16/coronavirus-leading-cause-death/

[13] Lavezzo, E., Franchin, E., Chrisanti, A., et al. (2020). Suppression of a COVID-19Outbreak in the Italian Municipality of Vo, Nature 584:425–429(2020)