The New England Journal of Statistics in Data Science Volume 1, 110–121 (2023)
DOI: https://doi.org/10.51387/21-NEJSDS2

A Compartment Model of Human Mobility and Early Covid-19 Dynamics in NYC

Ian FRANKENBURG and Sudipto BANERJEE

Abstract

In this paper, we build a mechanistic system to understand the relation between a reduction in human mobility and Covid-19 spread dynamics within New York City. To this end, we propose a multivariate compartmental system that jointly models smartphone mobility data and case counts during the first 90 days of the epidemic. Parameter calibration is achieved through the formulation of a general statistical-mechanistic Bayesian hierarchical model. The open-source probabilistic programming language Stan is used for the requisite computation. Through sensitivity analysis and out-of-sample forecasting, we find our simple and interpretable model provides quantifiable evidence for how reductions in human mobility altered early case dynamics in New York City.

KEYWORDS AND PHRASES: Bayesian Analysis, Statistical-Mechanistic Modeling, Covid-19.

1. INTRODUCTION

The global Covid-19 pandemic has underscored the importance of mathematical and statistical models in understanding disease dynamics, assessing policy efficacy, and examining counterfactual scenarios to formulate thorough cost-benefit analyses. Lockdown measures can have drastic impact on individual well-being as well as society and the economy at large [2], [4]. Therefore a retrospective study of Covid-19 lockdown and mitigation measures can help policymakers and public health officials understand to what end such efforts were effective. The formulation of a mechanistic compartmental model is a pathway towards such goals. In this article, we review compartmental model methodology, construct our new Bayesian hierarchical model, and discuss numerical methods relevant for implementation and fitting to real-world data. The new compartmental model is a simple modification of the classical susceptible-infectious-removed (SIR) model and enables a mechanistic correspondence between smartphone mobility data and infection dynamics. This can provide evidence of how reduced mobility due to early lockdowns or mitigation measures within New York City influenced Covid-19 spread dynamics. Case count data is obtained from the official website of the City of New York, available at https://www1.nyc.gov/site/doh/covid/covid-19-data.page. Population transit mobility data is obtained from https://covid19.apple.com/mobility and consists of anonymized Apple iPhone transit usage reported as a percent relative to baseline. Starting from the day after Governor Andrew Cuomo declared a state of emergency in New York State on March 7th, 2020, both the raw case count and transit mobility time series are presented below. Our end goal is then to establish a relationship between the two series shown in Figure 1.

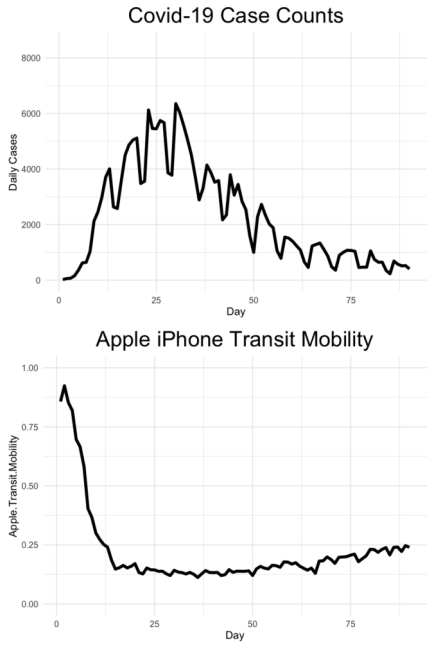


Figure 1: Raw Daily Case Counts and Mobility Time Series.

2. BACKGROUND

Mathematical modeling in epidemiology has a long history, famously dating back to the eighteenth century with the work of Bernoulli [6] or the mid-nineteenth century through John Snow's modeling of the cholera outbreak in London [24]. However, at the turn of the early twentieth century, mathematical epidemiology turned to the modern theory of dynamical systems analysis to understand out-

break evolution. In this section, we review the popular SIR model and subsequent mathematical analysis used to glean both qualitative and quantitative understanding of the dynamical system. This simple framework provides the necessary foundation for more complicated compartmental models with more population states. For examples of other compartment models designed to study early Covid-19 outbreak dynamics, see [11] or [30].

2.1 SIR Compartmental Model

In modeling population-level data with a compartmental system, the population is typically subdivided into separate homogeneous groups. Here we focus on reviewing the simple SIR model developed in 1927 by A. G. McKendrick and W. O. Kermack to model a plague outbreak in Bombay [15]. In such a model, the population is divided into susceptible (S), infectious (I), and removed (R) groups. Individuals then progress through the various states at certain rates over time. The mathematical description of this changing system is the coupled set of ordinary differential equations

$$\begin{cases} \dfrac{dS(t)}{dt} &= -\beta S(t)I(t)/N \\ \dfrac{dI(t)}{dt} &= \beta S(t)I(t)/N - \gamma I(t) \\ \dfrac{dR(t)}{dt} &= \gamma I(t). \end{cases} \tag{2.1}$$

The progression of the disease throughout the population depends upon the contact rate between susceptible and infectious individuals, the probability of transmission upon contact, and the prevalence of disease. To mathematically capture these factors and express the rate of new infections, let λ be defined as a per capita contact rate among individuals. In this way, $\lambda S(t)$ will give the average number of susceptible contacts over time. Now let p be the probability that a contact results in a new infection. Finally, the prevalence of the disease at time t is by definition $I(t)/N$. Combining these terms gives $p\lambda S(t)I(t)/N$ as the incidence rate. In defining the *effective contact rate* β as the product of the per capita contact rate λ and transmission probability p, the necessary form in equation (2.1) is recovered.

The remaining parameter γ is interpreted by considering that $1/\gamma$ is the average sojourn time of an individuals within compartment I. It should also be noted the population is fixed throughout time since $N = S + I + R$. This can alternatively be seen since adding the terms in (2.1) gives zero. With meaning associated to the compartmental parameters, we next turn to an overview of the mathematical analysis involved in analyzing the basic SIR dynamical system.

2.2 Linear Stability Analysis and the Basic Reproductive Number

The system of differential equations in (2.1) are nonlinear, arising from the term $S(t)I(t)$. Linear stability analysis is the workhorse to understand the behavior of nonlinear dynamical systems and has a fundamental connection to the *basic reproductive number* $\mathcal{R}_0$, popularized recently through media coverage of the Covid-19 pandemic. $\mathcal{R}_0$ is defined roughly to be the expected number of subsequent infections resulting from a single infected individual. In this section, we briefly review linear stability analysis and make the connection to $\mathcal{R}_0$. In the next section, we highlight the computation of $\mathcal{R}_0$ for a general class of compartmental models.

A steady state or equilibrium of a dynamical system is a point $\mathbf{x}^*$ where the system of differential equations evaluated at $\mathbf{x}^*$ is zero for all t. In this way, compartmental contents within the system are not changing over time. In the SIR model of (2.1), an important steady state is the so-called disease-free equilibrium of $\{(S^*, 0, 0) : S^* \geq 0\}$. A natural subsequent question is the behavior of the system around small perturbations of the equilibrium. In computing the Jacobian about such a point, we linearize and are afforded tractable analysis. A heuristic justification arises from considering a Taylor expansion of the system about the disease-free equilibrium and ignoring high-order terms since the perturbation is assumed small. After dropping the explicit dependence on time t from equation (2.1) to avoid clutter, the Jacobian of the system is computed as

$$\mathbf{J} = \begin{pmatrix} \dfrac{\partial \dot{S}}{\partial S} & \dfrac{\partial \dot{S}}{\partial I} & \dfrac{\partial \dot{S}}{\partial R} \\ \dfrac{\partial \dot{I}}{\partial S} & \dfrac{\partial \dot{I}}{\partial I} & \dfrac{\partial \dot{I}}{\partial R} \\ \dfrac{\partial \dot{R}}{\partial S} & \dfrac{\partial \dot{R}}{\partial I} & \dfrac{\partial \dot{R}}{\partial R} \end{pmatrix} = \begin{pmatrix} -\beta I/N & -\beta S/N & 0 \\ \beta I/N & \beta S/N - \gamma & 0 \\ 0 & \gamma & 0 \end{pmatrix}. \tag{2.2}$$

Through elementary matrix operations, $\mathbf{J}$ can be transformed to a triangular matrix. The eigenvalues are then found from inspection. Evaluating the Jacobian $\mathbf{J}$ at $(S^*, 0, 0)$ to linearize about the steady state results in eigenvalues of $\lambda_1 = 0$ and $\lambda_2 = \beta S^*/N - \gamma$. The sign of these eigenvalues then determine the stability of the equilibrium point. The eigenvalue of λ_1 is ignored, as it corresponds to a line of equilibrium values S^*. In this way, the second eigenvalue of λ_2 is of main interest. If $N/S^* < \beta/\gamma$, then $\lambda_2 > 0$ and the steady state is unstable; otherwise it is stable. In epidemic terms, an outbreak occurs if the disease-free equilibrium is unstable. This ratio β/γ acts as a bifurcation parameter in determining if an outbreak will occur and is thus afforded the fancy title of basic reproductive number. Letting the equilibrium point S^* be the population size so that $S^* = N$, a simple relation emerges: if $\mathcal{R}_0 > 1$ the disease continues to spread but dies out otherwise. The effective reproductive number $\mathcal{R}_t$ then extends $\mathcal{R}_0$ by accounting for a changing susceptible population over time and is defined as $\mathcal{R}_t := \mathcal{R}_0 S(t)/N$. A general method to compute $\mathcal{R}_0$ for more elaborate compartmental models will be discussed in the next subsection.

2.3 Spectral Radius or Next Generation Matrix Method

For general compartment models that extend the simplistic SIR framework, computing the basic reproductive number can be difficult. [5] and [12] describe a general method to compute $\mathcal{R}_0$ called the *Next Generation Matrix* or *Spectral Radius Method*, which we briefly review and compute for the SIR model.

Let $d\mathbf{X}(t)/dt$ represent a general coupled system of differential equations describing a compartmental model with n components and m infectious states. Define a vector-valued $\mathbf{F}(\mathbf{X}(t))$ to be a function where each component specifies flow rate into one of the respective m infected compartments. Similarly, define a function $\mathbf{V}(\mathbf{X}(t))$ where each component specifies the flow rate out of a respective infectious compartment.

Next, $\mathbf{F}$ and $\mathbf{V}$ are linearized about the disease-free equilibrium point by computing the Jacobian. [5] prove the Jacobian with respect to each infectious state will take the form

$$\mathbf{J}(\mathbf{F}) = \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \text{ and } \mathbf{J}(\mathbf{V}) = \begin{pmatrix} \mathbf{B} & \mathbf{0} \\ * & * \end{pmatrix}, \tag{2.3}$$

where $\mathbf{A}$ and $\mathbf{B}$ are $m \times m$ matrices. The next generation matrix is then defined as $\mathbf{AB}^{-1}$. The basic reproductive value of $\mathcal{R}_0$ is subsequently the spectral radius or largest eigenvalue of the next generation matrix. In the case of the SIR model, $\mathbf{F}(\mathbf{X}(t)) := -\beta SI/N$, while $\mathbf{V}(\mathbf{X}(t)) := \gamma I$. It follows that the necessary Jacobians evaluated at the disease-free equilibrium of $(N, 0, 0)$ results in $\mathbf{AB}^{-1} = \beta/\gamma$ and agree with the previous section.

3. METHODS

In this section, we detail the construction of our compartmental model designed to formulate an understanding of how reduction in human mobility might have altered early infection dynamics within New York City. The model is parsimonious, in that it consists of only four compartments and shares many well-established mathematical properties of the SIR model discussed in the background section. The dynamical system proposed comprises a closed population divided into susceptible, lockdown (L), infectious, and removed states. Since the time horizon under investigation is short, we choose to ignore demographic factors such as birth, death, and migration. Below in Figure 2 is a visualization of the population progression through the different states.

The system of differential equations describing the changing system are

$$\begin{cases} \dfrac{dS(t)}{dt} &= -\beta S(t)I(t)/N - aS(t) + bL(t) \\ \dfrac{dL(t)}{dt} &= aS(t) - bL(t) \\ \dfrac{dI(t)}{dt} &= \beta S(t)I(t)/N - \gamma I(t) \\ \dfrac{dR(t)}{dt} &= \gamma I(t), \end{cases} \tag{3.1}$$

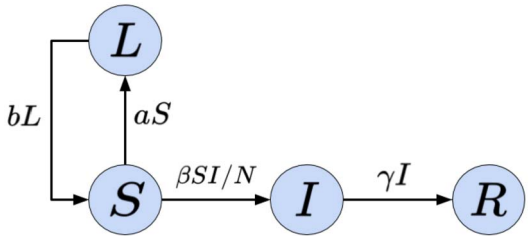
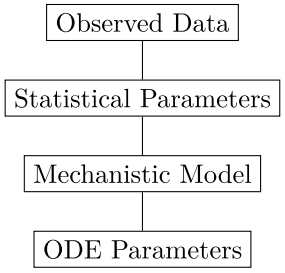


Figure 2: Proposed SLIR compartmental model.

with initial conditions of $S(0) = N - i_0$, $L(0) = 0$, $I(0) = i_0$, and $R(0) = 0$. An important qualitative feature of our model is that susceptible individuals are temporarily moved into the lockdown compartment to reflect social distancing, mitigation measures, and reduced mobility. Over time, individuals are reintroduced into the susceptible population out of the lockdown state. Through an application of the next generation matrix method described in section 2, $\mathcal{R}_0$ can be seen to be equivalent to the standard SIR model, i.e. $\mathcal{R}_0 = \beta/\gamma$.

3.1 A Bayesian Hierarchical Model

We present our methodology in a general hierarchical framework to facilitate Bayesian inference of compartmental system parameters in equation (3.1). In this hierarchical formulation, we can be explicit about the role of the mechanistic system, requisite numerical integration, and underlying process parameters. This section will construct the statistical model piece-by-piece. The hierarchy of connected components in the model can be visualized bottom-up as follows,



We first establish notation to represent the mechanistic system of differential equations in the middle of the hierarchy. Let the equations in (3.1) be denoted by $\mathbf{F}$, where

$$\frac{d}{dt}\mathbf{X}(t) = \mathbf{F}(\mathbf{X}(t), t), \text{ where } \mathbf{X}(t) = \begin{pmatrix} S(t) \\ L(t) \\ I(t) \\ R(t) \end{pmatrix} \tag{3.2}$$

and

$$\mathbf{F}(\mathbf{X}(t), t) = \begin{pmatrix} -\gamma \mathcal{R}_0 S(t)I(t)/N - aS(t) + bL(t) \\ aS(t) - bL(t) \\ \gamma \mathcal{R}_0 S(t)I(t)/N - \gamma I(t) \\ \gamma I(t) \end{pmatrix} \tag{3.3}$$

and we have suppressed the dependence of $\{\mathcal{R}_0, \gamma, a, b\}$ in $\mathbf{F}(\mathbf{X}(t))$. The system is reparameterized in terms of $\mathcal{R}_0$ rather than β to be more epidemiologically interpretable and results from a simple transformation of $\beta = \gamma\mathcal{R}_0$. To recover the system states $\mathbf{X}(t)$, the system of differential equations must be solved. Given fixed values of $\mathcal{R}_0$, γ, a, and b, the solution to the system of differential equations is a vector-valued function

$$\mathbf{X}(t) = \int \mathbf{F}(\mathbf{X}(t), t; \mathcal{R}_0, \gamma, a, b)dt. \tag{3.4}$$

This solution will be necessary to connect with the observed data.

The top of the hierarchy is described by formulating a measurement process for the two outcome variables. Let the first outcome of interest be labeled $Y_L(t)$ and represent the percent of the population removed from the susceptible compartment by adhering to mitigation protocol. The subscript L is used for a reminder that this data is used to gain information on the lockdown compartment. Likewise, the second outcome is labeled $Y_I(t)$ and denotes the observed case counts over time. To model observation error in $Y_L(t)$, we choose a Beta distribution dependent upon a parameter ϕ_1 to control dispersion, i.e.,

$$Y_L(t)|L(t), \phi_1 \sim \text{Beta}\big(\phi_1 L(t)/N, \phi_1(1 - L(t)/N)\big). \tag{3.5}$$

Notice $L(t)$ is necessarily scaled by the population size N to respect the support of the beta distribution. As we seek to inform the L compartment through cell phone mobility data, the beta distribution is a natural choice because the data will be anonymized and reported as a percentage of nominal movement. This parameterization of the beta distribution has expectation and variance

$$\begin{aligned}\mathbb{E}[Y_L(t)|L(t), \phi_1] &= L(t)/N \\ \text{Var}(Y_L(t)|L(t), \phi_1) &= \frac{L(t)/N\big(1 - L(t)/N\big)}{\phi_1 + 1}.\end{aligned}$$

To model observation noise in $Y_I(t)$, we use a negative binomial to account for overdispersion,

$$Y_I(t)|I(t), \phi_1 \sim \text{Negative Binomial}(I(t), \phi_2). \tag{3.6}$$

Stan provides an alternative parameterization of the negative binomial called `neg_binomial_2` with first two moments of

$$\begin{aligned}\mathbb{E}[Y_I(t)|I(t), \phi_2] &= I(t) \\ \text{Var}(Y_I(t)|I(t), \phi_2) &= I(t) + \frac{I(t)^2}{\phi_2}.\end{aligned}$$

In this way, ϕ_2 is viewed as a dispersion parameter. The full hierarchical description of the model can be completed by introducing prior distributions on the system parameters governing the differential equations. Writing the model in full,

$$\begin{aligned} Y_L(t)|L(t), \phi_1 &\sim \text{Beta}\big(\phi_1 L(t)/N, \phi_1(1 - L(t)/N)\big) \\ Y_I(t)|I(t), \phi_2 &\sim \text{Negative Binomial}(I(t), \phi_2) \\ \frac{d\mathbf{X}(t)}{dt} &= \mathbf{F}(\mathbf{X}(t), t; \mathcal{R}_0, \gamma, a, b) \\ \mathcal{R}_0|\gamma &\sim \text{log-normal}(0, 1) \\ \gamma &\sim \text{Uniform}(0, 1) \\ \phi_1 &\sim \text{Inverse Gamma}(0.1, 0.1) \\ \phi_2 &\sim \text{Inverse Gamma}(0.1, 0.1) \\ a &\sim \text{Beta}(1, 5) \\ b &\sim \text{Uniform}(0, 1). \end{aligned} \tag{3.7}$$

The prior distributions on system parameters are weakly-informative. However, the prior distribution on a might at first appear suspect. Through prior predictive checks, we find that placing a uniform prior on a results in the SLIR model a priori favoring no epidemic breakout, as susceptibilities are removed from the population too quickly. Since the classical SIR model is a special case of our SLIR model as $a \to 0$, we place mass closer to 0 through the Beta(1,5) distribution to ensure the model generates reasonable predictions before seeing the data. The hierarchical model of the previous section crucially depends upon the numerical solution to a coupled set of differential equations of (3.1). In the appendix section, we detail the internal workings of Stan's numerical optimization routines. Finally, efficient Bayesian analysis of parameters within the set of nonlinear differential equations relies upon the efficiencies gained through Hamiltonian Monte Carlo (HMC). A detailed review of this methodology is also included in the appendix section.

4. ANALYSIS AND RESULTS

In this section, we present two simulation studies and conclude with the New York City analysis. First, the proposed SLIR compartmental model is used to simulate data from two lockdown scenarios that affect human mobility differently. We then fit our Bayesian model to assess whether the true parameter values are adequately recovered. After, we analyze the real-world mobility and case count data that initially inspired the model formulation.

4.1 Simulated Data

To illustrate the nonlinear dynamics of which our model can capture, the first simulation reflects the idealized scenario of strict adherence to lockdown and mitigation measures, when population movement is quickly reduced in the early stages of the outbreak and remains reduced for the next 90 days. In this case, individuals move from the S compartment to the L compartment quickly and are slowly reintroduced into the susceptible population so that population movement decreases to about 60% relative to baseline within the first 20 days.

Table 1. Simulation Study.

Data	Parameter	Truth	Median	95%-credible interval	G-R $\hat{R}$
Simulation 1	$\mathcal{R}_0$	5	4.93	4.720–5.130	1.00
	γ	0.1	0.090	0.086–0.150	1.00
	a	0.05	0.045	0.040–0.051	1.00
	b	0.1	0.095	0.040–0.051	1.00
Simulation 2	$\mathcal{R}_0$	5	4.79	4.590–5.010	1.00
	γ	0.1	0.099	0.096–0.104	1.00
	a	0.05	0.045	0.040–0.051	1.00
	b	0.1	0.095	0.086–0.106	1.00

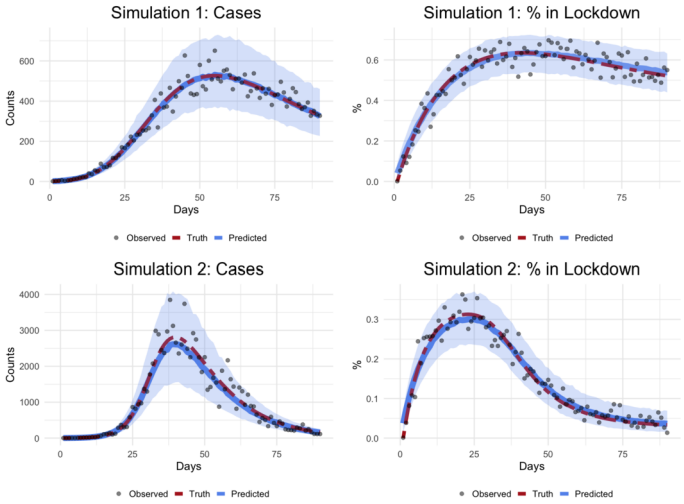


Figure 3: Model Fit to Simulated Data.

In the second scenario, we consider weak adherence to mitigation measures and illustrate the substantial change in dynamics by only altering the speed of flow back into the susceptible population from the L compartment. In this case, the peak percentage of the population in the lockdown compartment is 30% but quickly diminishes. In both simulations, the population size is fixed to $N = 10{,}000$, $i_0 = 1$, and the true data-generating process is shown below as a dashed red line. The median of the posterior predictive distribution and 95% credible intervals are shown in blue in Figure 3.

We fit our model to both scenarios using Stan's NUTS algorithm with 4 chains and 5,000 iterations each, the first half of which are discarded as warm-up. The convergence of the parameter chains are judged by inspecting the trace plots along with the Gelman-Rubin $\hat{R}$ values, which compares the variation between chains to the variation within [9]. Ideally, the $\hat{R}$ value is close to one. These simulation results are displayed in Table 1.

In both cases, the Bayesian hierarchical model is able to infer the structural parameters of the SLIR model. Notice the change in case counts of both scenarios resulting from different susceptible population sizes.

4.2 New York City Analysis

The entire lead-up thus far was requisite background material for compartmental model inference and application to real-world data. As mentioned in the introduction, our main motivation for this article was to understand how a reduction in mobility affected early Covid-19 dynamics specifically within NYC. For convenience, we restate the data sources. Case counts are reported by the official website of the City of New York, available at https://www1.nyc.gov/site/doh/covid/covid-19-data.page and mobility data is hosted at https://covid19.apple.com/mobility. Since the Apple mobility data reflects a percent decrease in movement, it must first be transformed by subtraction from unity to adhere with the SLIR compartmental model structure. In other words, to prepare the mobility data for use in the hierarchical model, it must first be subtracted from one so that it no longer represents a percent decrease in transit mobility but rather a percent increase in individuals adhering to mitigation measures. Finally, we mention again that although the first case of Covid-19 in New York City was recorded on February 29th, we align our movement and case data to begin on March 8th, 2020, the day after Governor Andrew Cuomo declared a state of emergency in New York State. Finally, we take the initial number of cases to be the cases recorded on March 8th, 2020, and the population is fixed at 8,336,817 as determined by the US Census [29].

To achieve parameter calibration, we fit in Stan our full hierarchical model using four chains run for 10,000 iterations each. We discard the first 5,000 as warm-up. Sufficient posterior exploration is assessed by examining parameter chain plots below and assessing $\hat{R}$ values, shown in Table 2.

The fitted time series are presented below on the right, along with the raw data used to train the model. On the left, prior predictive distributions are included to illustrate the degree in which Bayesian learning occurs after observing the data. We also include the fit of an SIR model to illustrate its inability to capture the dynamics.

To assess the structural fit of our hypothesized SLIR mechanism, we next interpret parameter values to ensure they are logical and consistent with outside literature. The

Table 2. New York City Analysis.

Data	Parameter	Median	95%-credible interval	G-R $\hat{R}$
New York City	$\mathcal{R}_0$	5.130	4.841–5.448	1.00
	γ	0.212	0.191–0.234	1.00
	a	0.115	0.106–0.124	1.00
	b	0.022	0.019–0.024	1.00

$\mathcal{R}_0$ estimate is cross-referenced with those of other popular online models. Using only death statistics as reported by Johns Hopkins University, [10] embeds a SEIR model in a machine learning framework for many regions across the United States and 70 countries. In this work, $\mathcal{R}_0$ is estimated for NYC to be between 5.0 and 5.8, in close agreement with our model. As an additional point of reference, [14] provide an alternative methodology that fits a time series state-space model to death counts and explicitly accounts for reporting delays. $\mathcal{R}_0$ is estimated to be 6.3 with a 95% confidence interval of 4.5–9. Our model thus has the added benefit of mechanistic interpretability as well as a tighter credible interval.

The posterior of γ has a median value of 0.21 and a 95% credible interval of (0.191, 0.234). From this, we arrive at an estimate of approximately 5 days for the average infectious removal time. This estimate could be reasonable assuming asymptomatic or presymptomatic transmission is possible and that individuals isolate at the onset of symptoms; see [8] for evidence. Outside work estimates symptom onset time to also be around 5 days. For example, [17] use 181 confirmed Covid-19 cases to estimate a median symptom onset time of 5.1 days with a 95% confidence interval (4.5, 5.8) days. A systematic review by [18] estimates a mean symptom onset time of 5.2 days with a 95% confidence interval of (4.1, 7.0) days. These estimates provide credence in establishing a correspondence between population movement reduction and infection dynamics with the mechanistic SLIR model.

4.3 Sensitivity Analysis and Out-of-Sample Forecasting

We conclude the New York City analysis by assessing the sensitivity of the model to changing mobility levels. Additionally, we assess out-of-sample predictive capacity of the SLIR model. To perform the sensitivity analysis, we first generate a range of hypothetical mobility scenarios, from full mobility reduction through extremely stringent lockdown measures to a more mild decline. This is displayed in Figure 4, with the true, real-world observed mobility levels highlighted as blue. It is important to note the explosive case growth with mobility reduction levels under 60% due to the nonlinear dynamics present in SIR-type models. This is evidence that a mobility reduction significantly altered infections within the city, assuming SIR-type dynamics.

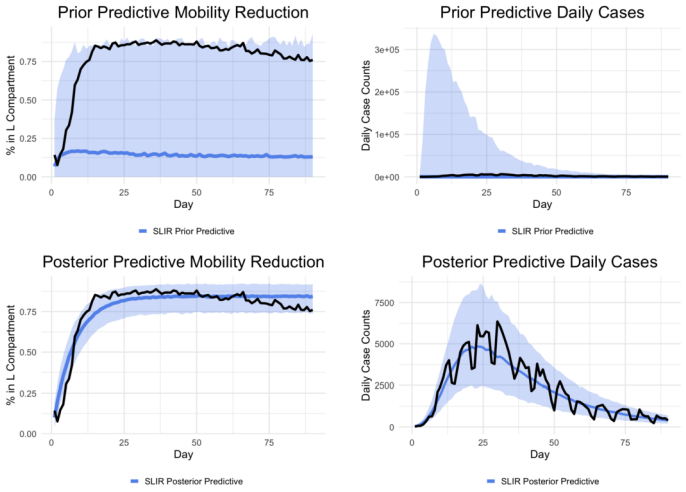


Figure 4: SLIR Prior and Posterior Predictive Checks.

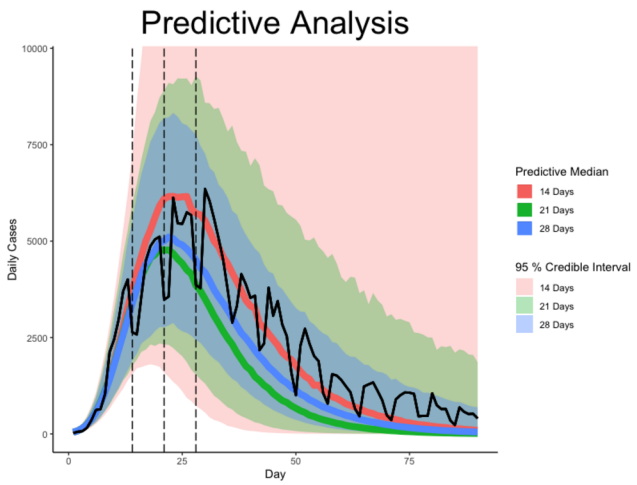


Figure 5: SLIR Out-of-sample Forecasts.

To conclude this section, we analyse the out-of-sample forecasting ability of the SLIR model when trained on only a subset of the ninety day period. We consider for illustration a two, three, and four week training intervals. These are indicated by vertical dashed lines in Figure 5.

The predictive median is illustrated with a solid line. With only two weeks observed, the predictive median is rea-

sonably representative of the future trajectory but with very large uncertainty. The upper 95% predictive curve for the two week window reaches roughly 400,000 infections, but is clear that the predictive interval quickly contracts as more data is observed (cf. Figure 4). Both the three week and four week training periods have contracted credible interval forecasts that contain the observed data. Finally, we mention that we were unable to fit the standard SIR model to the New York City data. Using the quasi-Newton optimization functions available in Stan, we found the SIR model fit was highly unstable across a range of starting values indicating extreme multi-modality in the likelihood surface.

5. DISCUSSION AND FUTURE WORK

5.1 New York City Analysis

In this work, we have formulated a basic extension of the classical SIR model to jointly fit cell phone transit mobility data and case counts. By jointly modeling two outcome variables, we establish a mechanistic correspondence between reduced mobility and infection dynamics. The applied analysis and findings, however, are limited to NYC during the first 90 days. The disease progression throughout the region was well-approximated by deterministic dynamics and modeling with a simple compartment system. In other geographic regions, the dynamics might not be as suitably well-behaved or understood. It is difficult to capture the myriad of factors contributing to disease spread throughout a population, and often a stochastic model may be more appropriate. Additionally, the time horizon considered in our applied analysis is relatively short. As a disease progresses throughout the population and becomes more widespread, the strategy of informing the susceptible population numbers through cell phone mobility data becomes more difficult. It is also worth consideration about whether in general smartphone mobility data can be considered a representative sample of individuals' adherence to mitigation protocols. It is therefore difficult to build a quantitative understanding of the dynamics between lockdown measures and disease progression over long periods of time. A direction for future research is to establish modeling strategies for such scenarios. This will introduce additional challenges of accounting for population immunity, vaccinations, and natural seasonality effects.

5.2 Stochastic Model Extensions

We briefly discuss for both discrete and continuous time how the deterministic SLIR model can be relaxed through a state-space (or Hidden Markov Model) framework, where stochasticity is introduced into the underlying driving epidemic process. We first discuss a discrete-time extension by modifying equation (3.7) so that the numerical solution to $\frac{d\mathbf{X}(t)}{dt}$ is embedded within a likelihood function at the discrete observation time points. In this way, equation (3.7) is modified to include $p(\mathbf{X}(t)|\boldsymbol{\theta}, \mathcal{R}_0, \gamma, a, b)$, where $\boldsymbol{\theta}$ is introduce to accommodate new parameters. In other words, stochasticity is introduced to the driving dynamics through the choice of suitable probability distribution that is a function of the numerical solution of the system of differential equations. This approach has been successfully applied to forecast seasonal influenza [21], [7], as well as to assess Covid-19 interventions in China [31] and Japan [16].

An alternative approach is to incorporate continuous-time stochasticity directly by expressing $d\mathbf{X}(t)$ as a stochastic differential equation (SDEs) [20]. In this case, one modification of equation (3.7) could be

$$d\mathbf{X}(t) = \boldsymbol{\mu}(\mathbf{X}(t), t)dt + \mathbf{L}(\mathbf{X}(t), t)d\mathbf{W}(t), \tag{5.1}$$

where $\mathbf{W}(t) = (W_1(t), W_2(t), W_3(t), W_4(t))^\top$ is a vector of standard Wiener processes or Brownian motions, $\boldsymbol{\mu}$ is some vector-valued function, and $\mathbf{L}$ is a compatible matrix. Depending on the structure of $\boldsymbol{\mu}$, there may or may not exist an explicit solution. A SDE that affords a closed form solution is the Ornstein-Uhlenbeck (OU) process. See [3] for Stan code in the case where the infectious compartment of the SIR model is endowed an OU process to allow for random fluctuations. See also [32] for an analysis of the susceptible-infectious-susceptible model where time-varying parameters are introduced through an OU process. In both examples, the analytic nature of the solution to the SDE enables efficient implementation in Stan or alternative software. In cases where no closed-form solution to (5.1) exists, inference in such a system is closely related to Bayesian filtering and smoothing, amenable to such methods as the Kalman filter and extensions [27]. The New York City data was captured well by deterministic dynamics, but often elsewhere disease progression is not suitably captured within such a framework. Bayesian inference for SDEs modeling infectious disease dynamics is thus an attractive area of future research.

APPENDIX

A.1 Numerical Solvers

The hierarchical model of the previous section crucially depends upon the numerical solution to a coupled set of differential equations. We briefly review the popular numerical integration schemes and connect them with our hierarchical model of the compartmental system in (3.1). In practice, solutions to the SIR model are numerically approximated and entail specific computational challenges. Runge-Kutta (RK) is a classical numerical method [26, 23] in which we employ to numerically solve this set of nonlinear differential equations. RK generalizes the well-known Euler method for iteratively solving systems of differential equations. In the following, we formulate the RK method in vector notation for the proposed SLIR compartmental model.

Consider $\mathbf{X}(t)$ in (3.2). From the fundamental theorem of Calculus, we know that

$$\mathbf{X}(t + \Delta t) = \mathbf{X}(t) + \int_t^{t+\Delta t} \mathbf{F}(\mathbf{X}(u), u)du\ . \tag{A.1}$$

Given the initial value $\mathbf{X}(t_0)$ at time t_0, numerically solving for $\mathbf{X}(t)$ amounts to constructing a sequence $\{\mathbf{X}_n : n = 0, 1, \ldots, T\}$, where $\mathbf{X}_n := \mathbf{X}(t_n)$ and $t_n = t_{n-1} + \Delta t$ are equispaced time points for $n = 0, 1, \ldots, T$, by approximating the integral of $\mathbf{F}(\mathbf{X}(t), t)$ in (A.1). Using this approximation, a sequence is generated starting from some t_0 as,

$$\mathbf{X}_{n+1} = \mathbf{X}_n + \int_{t_n}^{t_n+\Delta t} \mathbf{F}(\mathbf{X}(u), u)du \ , \tag{A.2}$$

where the initial condition $\mathbf{X}_0 := \mathbf{X}(t_0)$ is given. There are several customary choices for such approximations, but for most practical purposes one need not look beyond the following:

$$\int_t^{t+\Delta t} \mathbf{F}(\mathbf{X}(t), t)dt = \begin{cases} (\Delta t)\mathbf{F}(\mathbf{X}(t), t) \\ \frac{(\Delta t)}{2}\left(\mathbf{F}(\mathbf{X}(t), t) + \mathbf{F}(\mathbf{X}(t + \Delta t), t + \Delta t)\right) \\ (\Delta t)\mathbf{F}(\mathbf{X}(t + (\Delta t)/2), t + (\Delta t)/2) \ . \end{cases} \tag{A.3}$$

Euler's approximation (first equation in (A.3)) is the simplest of the approximations which yields Euler's Method. In this case, the sequence is constructed for $n = 0, 1, \ldots, T-1$ as

$$\mathbf{X}_{n+1} = \mathbf{X}_n + (\Delta t)\mathbf{F}(\mathbf{X}_n, t_n) \ . \tag{A.4}$$

Starting with $\mathbf{X}_0$, each element of the sequence in (A.4) is computed since $\mathbf{F}(\mathbf{X}_n, t_n)$ is available. The spacing between the time points, Δt, is specified by the user and controls the resolution of the numerical solution.

Euler's approximation is easy to execute and based upon a first order Taylor expansion. It is also the least accurate. The Trapezoidal and Modified Euler approximations in (A.3) are based upon numerical integration using trapezoidal areas and the midpoint approximation. Both these methods help improve Euler's method with the Modified Euler outperforming the trapezoidal rule in terms of accuracy. However, the Trapezoidal and the Modified Euler methods involve $\mathbf{X}_{n+1}$ and $\mathbf{X}_{n+1/2} := \mathbf{X}(t_{n+1/2})$ with $t_{n+1/2} := t_n + (\Delta)t/2$, respectively, which are unknown at iteration n. In fact, $\mathbf{X}_{n+1}$ is the very quantity we wish to compute at iteration n. Therefore, we substitute these unknown quantities with their first order (Euler) approximations in (A.4) that are available at iteration n. For each $n = 0, 1, \ldots, T-1$, we compute

$$\begin{aligned} \mathbf{a}_n &= \mathbf{F}\left(\mathbf{X}_n, t_n\right) \\ \mathbf{b}_n &= \mathbf{F}\left(\mathbf{X}_n + (\Delta t)\mathbf{a}_n, t_{n+1}\right) \\ \mathbf{c}_n &= \mathbf{F}\left(\mathbf{X}_n + (\Delta t/2)\mathbf{a}_n, t_{n+1/2}\right) \end{aligned} \tag{A.5}$$

and the transition from $\mathbf{X}_n$ to $\mathbf{X}_{n+1}$ as

$$\mathbf{X}_{n+1} = \begin{cases} \mathbf{X}_n + (\Delta t)\frac{(\mathbf{a}_n+\mathbf{b}_n)}{2} & \text{(Trapezoidal)} \ ; \\ \mathbf{X}_n + (\Delta t)\mathbf{c}_n & \text{(Modified Euler)} \end{cases} \tag{A.6}$$

Both methods in (A.6) deliver noticeable improvements over Euler's method in (A.4). Numerical error from (A.6) are much smaller in magnitude and grow less quickly. This can be explained by observing that while (A.4) depends only upon the data available at t_n (only one data point), the two methods in (A.6) use current data at t_n along with estimates of the slope at a point that lies in the future. While these estimates are computed using only the currently available data, they still produce substantially improved estimates. Also see [28] for comparisons among different numerical integration schemes.

Higher order Taylor expansions produce other iterative schemes. Thus, second order methods emerge from

$$\begin{aligned} \mathbf{X}(t + \Delta t) = \mathbf{X}(t) &+ (\Delta t)\mathbf{F}(\mathbf{X}(t), t) \\ &+ \frac{(\Delta t)^2}{2}\frac{d}{dt}\mathbf{F}(\mathbf{X}(t), t) + O((\Delta t)^2) \ . \end{aligned} \tag{A.7}$$

A second order iterative scheme corresponding to (A.7) updates

$$\begin{aligned} \mathbf{X}_{n+1} = \mathbf{X}_n &+ (\Delta t)\mathbf{F}(\mathbf{X}_n, t_n) \\ &+ \frac{(\Delta t)^2}{2}\left\{ \left.\frac{\partial \mathbf{F}}{\partial t}\right|_{t=t_n} + \left.\left[\frac{\partial \mathbf{F}}{\partial \mathbf{X}}\right]\right|_{\mathbf{X}=\mathbf{X}_n} \left.\frac{d}{dt}\mathbf{X}(t)\right|_{t=t_n} \right\} \ , \end{aligned} \tag{A.8}$$

where we have used the multivariable chain-rule of derivatives to evaluate the derivative of $\mathbf{F}(\mathbf{X}(t), t)$ with respect to t, $\left[\frac{\partial \mathbf{F}}{\partial \mathbf{X}}\right]$ is the matrix with (i, j)-th element being the partial derivative of the i-th element of $\mathbf{F}(\mathbf{X}(t), t)$ with respect to the j-th variable in $\mathbf{X}$, and $\left.\frac{d}{dt}\mathbf{X}(t)\right|_{t=t_n} = \mathbf{F}(\mathbf{X}_n, t_n)$. Unfortunately, computing (A.8) requires the derivatives of $\mathbf{F}(\mathbf{X}(t), t)$ and may, in general, be numerically cumbersome.

The Runge-Kutta methods are among the most conspicuous of numerical methods for solving systems of ordinary differential equations. The underlying idea is to achieve the same accuracy as Taylor series updates without requiring higher order derivatives of $\mathbf{F}(\mathbf{X}(t))$. We can motivate this approach from the earlier methods. In (A.6), the Trapezoidal and Modified Euler methods define the updates using $\mathbf{a}_n$, $\mathbf{b}_n$ and $\mathbf{c}_n$ that are completely specified. In particular, observe that the Trapezoidal method updates using a weighted average of $\mathbf{a}_n$ and $\mathbf{b}_n$. Instead of prescribing $\mathbf{a}_n$ and $\mathbf{b}_n$, the Runge-Kutta approach prefers to find weighted averages to ensure that the approximation matches that from a Taylor series expansion such as in (A.7) or (A.8). Therefore, a second-order Runge-Kutta method (RK2) writes

$$\begin{aligned} \mathbf{X}_{n+1} = \mathbf{X}_n &+ (\Delta t)\left\{\omega_1 \mathbf{a}_n \right. \\ &+ \left. \omega_2 \mathbf{F}(\mathbf{X}_n + (\Delta t)\beta \mathbf{a}_n, t_n + (\Delta t)\alpha)\right\} \end{aligned} \tag{A.9}$$

and seeks to find ω_1, ω_2, α and β so that the approximation matches (A.8). Substituting the first-order expansion,

$$\begin{aligned}
&\mathbf{F}(\mathbf{X}_n + (\Delta t)\beta \mathbf{a}_n, t_n + \alpha(\Delta t)) = \mathbf{F}(\mathbf{X}_n, t_n) \\
&\quad + (\Delta t)\beta \left[\frac{\partial \mathbf{F}}{\partial \mathbf{X}}\right]\Big|_{\mathbf{X}=\mathbf{X}_n} \mathbf{a}_n + (\Delta t)\alpha \frac{\partial \mathbf{F}}{\partial t}\Big|_{t=t_n} + O((\Delta t)^2) \\
&= \mathbf{a}_n + (\Delta t)\beta \left[\frac{\partial \mathbf{F}}{\partial \mathbf{X}}\right]\Big|_{\mathbf{X}=\mathbf{X}_n} \mathbf{a}_n \\
&\quad + (\Delta t)\alpha \left\{ \frac{\partial \mathbf{F}}{\partial t}\Big|_{t=t_n} + \left[\frac{\partial \mathbf{F}}{\partial \mathbf{X}}\right] \frac{d}{dt}\mathbf{X}(t)\Big|_{t=t_n} \right\} + O((\Delta t)^2) \,,
\end{aligned} \tag{A.10}$$

into the right hand side of (A.9) and comparing with the expansion (A.8) we find that the two expansions are equivalent if

$$\omega_1 + \omega_2 = 1 \,; \quad \omega_2\beta = \omega_2\alpha = 1/2 \,. \tag{A.11}$$

RK2 specifies $\omega_1 = \omega_2 = 1/2$ and $\alpha = \beta = 1$, which, when substituted into (A.9), yields the Trapezoidal approximation.

More generally, the explicit RK methods of order s specify updating schemes

$$\mathbf{X}_{n+1} = \mathbf{X}_n + (\Delta t) \sum_{i=1}^{s} \omega_i \mathbf{k}_i \,, \tag{A.12}$$

where

$$\begin{aligned}
\mathbf{k}_1 &= \mathbf{F}(\mathbf{X}_n, t_n + (\Delta t)\alpha_1) \\
\mathbf{k}_i &= \mathbf{F}\left(\mathbf{X}_n + (\Delta t)\sum_{j=1}^{i-1} \beta_{ij}\mathbf{k}_j, t_n + (\Delta t)\alpha_i\right) \; i = 2, \ldots, s.
\end{aligned}$$

The coefficients are found from an s-th order Taylor expansion. A popular choice sets $\alpha_1 = 0$ and solves

$$\sum_{i=1}^{s} \omega_i = 1 \;\text{ and }\; \sum_{j=1}^{i-1} \beta_{ij} = \alpha_i \;\text{ for }\; i = 2, 3, \ldots, s \,. \tag{A.13}$$

In particular, the RK4 method specifies the following values after taking $s = 4$: (A.12):

$$\begin{gathered}
\begin{bmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \\ \omega_4 \end{bmatrix} = \frac{1}{6}\begin{bmatrix} 1 \\ 2 \\ 2 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 0 \\ 1 \\ 1 \\ 2 \end{bmatrix}, \\
\begin{bmatrix} \beta_{21} & & \\ \beta_{31} & \beta_{32} & \\ \beta_{41} & \beta_{42} & \beta_{43} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & & \\ 0 & \frac{1}{2} & \\ 0 & 0 & 1 \end{bmatrix} .
\end{gathered} \tag{A.14}$$

The appropriate step size Δt is hard to determine. An advantage of Stan's implementation is an adaptive step size is used by comparing the solution obtained using the four term approximation of above as well as a five term approximation. If these approximations agree, the algorithm proceeds, otherwise a new step size is calculated. More details can be found in the user manual [25], but the result is a fast, efficient procedure of high accuracy.

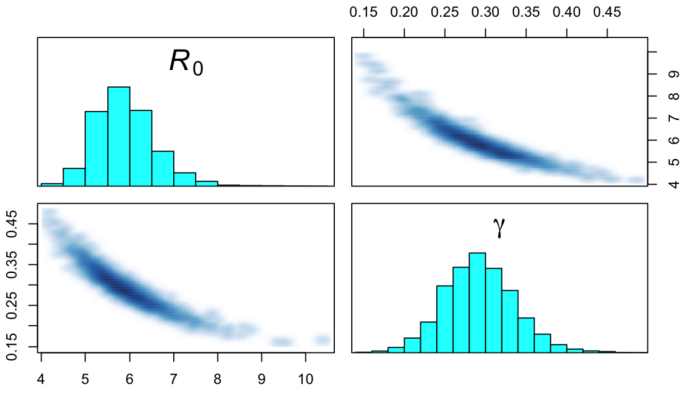


Figure 6: $\mathcal{R}_0$ and γ joint posterior.

A.2 Hamiltonian Monte Carlo and the No-U-Turn Sampler

Our choice of implementation in Stan, as opposed to more traditional BUGS or JAGS [22], is pragmatic. First, the latter languages are declarative and built upon graphical models. In contrast, Stan is a fully imperative programming language. Additionally, built-in differential equation routines are included such as the Runge-Kutta numerical solver described in the previous section. This makes the software implementation of our model more natural and readable. More importantly, however, the parameters in a nonlinear compartmental model are often highly correlated, as demonstrated in below in Figure 6.

Hamiltonian Monte Carlo (HMC) is more equipped to sample from complex posterior distributions with high autocorrelations than standard Metropolis schemes. The most popular presentation of Hamiltonian Monte Carlo is by way of analogy with statistical mechanics. Let $\boldsymbol{\theta}$ be an arbitrary d-dimensional parameter vector. To sample efficiently from the posterior of $\boldsymbol{\theta}$ after conditioning on data, an idealized physical system is introduced to leverage the geometry of the underlying manifold on which $\boldsymbol{\theta}$ lives. We will not embark upon a comprehensive development of HMC here, referring the reader to excellent introductory expository articles by [19] and [1]. Instead, we provide a heuristic account of the HMC algorithm and why it works.

We begin by recalling the more conspicuous Metropolis random walk and the concept of detailed balance. Let $p(\boldsymbol{\theta} \,|\, \mathbf{Y})$ be the posterior distribution from which we wish to sample. As it may be difficult to directly sample from $p(\boldsymbol{\theta} \,|\, \mathbf{Y})$, the Metropolis random-walk algorithm constructs a Markov chain with $p(\boldsymbol{\theta} \,|\, \mathbf{Y})$ as its stationary distribution.

Given an initial value $\boldsymbol{\theta}^{(0)}$, at iteration t we draw a "proposed" value $\boldsymbol{\theta}^*$ from a symmetric distribution $q(\cdot \,|\, \boldsymbol{\theta}^{(t-1)})$. A simple yet effective choice for many applications is $q(\cdot \,|\, \boldsymbol{\theta}^{(t-1)}) = N(\cdot \,|\, \boldsymbol{\theta}^{(t-1)}, \mathbf{V})$, where $\mathbf{V}$ is a fixed variance covariance matrix that helps tune the algorithm, although in general the proposal can be generated from any symmetric distribution. After generating $\boldsymbol{\theta}^*$ we simulate a coin with probability of heads $\min\left(1, \dfrac{p(\boldsymbol{\theta}^* \,|\, \mathbf{Y})}{p(\boldsymbol{\theta}^{(t-1)} \,|\, \mathbf{Y})}\right)$ and set $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^*$ if it is a head. Otherwise we set $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^{(t-1)}$. That $p(\boldsymbol{\theta} \,|\, \mathbf{Y})$ is indeed the desired stationary distribution can be seen as follows. Let us assume that the current state $\boldsymbol{\theta}^{(t-1)} = \mathbf{a}$ is a draw from $p(\boldsymbol{\theta} \,|\, \mathbf{Y})$ and consider the possibility of moving to $\boldsymbol{\theta}^{(t)} = \mathbf{b}$. This conditional probability is given by the transition probability that a value of $\mathbf{b}$ is proposed from $q(\cdot \,|\, \mathbf{a})$ and that this value is accepted. Hence,

$$\begin{aligned} T(\mathbf{a} \to \mathbf{b}) &= P(\boldsymbol{\theta}^{(t)} = \mathbf{b} \,|\, \boldsymbol{\theta}^{(t-1)} = \mathbf{a}) \\ &= q(\mathbf{b} \,|\, \mathbf{a}) \min\left(1, \frac{p(\mathbf{b} \,|\, \mathbf{Y})}{p(\mathbf{a} \,|\, \mathbf{Y})}\right) \\ &= \min\left(q(\mathbf{b} \,|\, \mathbf{a}), q(\mathbf{b} \,|\, \mathbf{a}) \frac{p(\mathbf{b} \,|\, \mathbf{Y})}{p(\mathbf{a} \,|\, \mathbf{Y})}\right) . \end{aligned} \tag{A.15}$$

The form of the transition probability in (A.15) implies time-reversibility (or detailed balance) in the following sense:

$$\begin{aligned} P(\boldsymbol{\theta}^{(t-1)} = \mathbf{a}, \boldsymbol{\theta}^{(t)} = \mathbf{b}) &= P(\boldsymbol{\theta}^{(t-1)} = \mathbf{a}) T(\mathbf{a} \to \mathbf{b}) \\ &= p(\mathbf{a} \,|\, \mathbf{Y}) \min\left(q(\mathbf{b} \,|\, \mathbf{a}), q(\mathbf{b} \,|\, \mathbf{a}) \frac{p(\mathbf{b} \,|\, \mathbf{Y})}{p(\mathbf{a} \,|\, \mathbf{Y})}\right) \\ &= \min\left(p(\mathbf{a} \,|\, \mathbf{Y}) q(\mathbf{b} \,|\, \mathbf{a}), q(\mathbf{b} \,|\, \mathbf{a}) p(\mathbf{b} \,|\, \mathbf{Y})\right) \\ &= \min\left(p(\mathbf{a} \,|\, \mathbf{Y}) q(\mathbf{a} \,|\, \mathbf{b}), q(\mathbf{a} \,|\, \mathbf{b}) p(\mathbf{b} \,|\, \mathbf{Y})\right) \\ &= p(\mathbf{b} \,|\, \mathbf{Y}) q(\mathbf{a} \,|\, \mathbf{b}) \min\left(\frac{p(\mathbf{a} \,|\, \mathbf{Y})}{p(\mathbf{b} \,|\, \mathbf{Y})}, 1\right) \\ &= P(\boldsymbol{\theta}^{(t-1)} = \mathbf{b}) T(\mathbf{b} \to \mathbf{a}) \\ &= P(\boldsymbol{\theta}^{(t-1)} = \mathbf{b}, \boldsymbol{\theta}^{(t)} = \mathbf{a}) , \end{aligned} \tag{A.16}$$

where we have used the symmetry $q(\mathbf{a} \,|\, \mathbf{b}) = q(\mathbf{b} \,|\, \mathbf{a})$ in the fourth equality in (A.16). It follows that the draw of $\boldsymbol{\theta}^{(t)}$ is also from $p(\boldsymbol{\theta} \,|\, \mathbf{Y})$ because

$$\begin{aligned} P(\boldsymbol{\theta}^{(t)} = \mathbf{b}) &= \int P(\boldsymbol{\theta}^{(t-1)} = \mathbf{a}, \boldsymbol{\theta}^{(t)} = \mathbf{b}) d\mathbf{a} \\ &= \int P(\boldsymbol{\theta}^{(t-1)} = \mathbf{b}, \boldsymbol{\theta}^{(t)} = \mathbf{a}) d\mathbf{a} \\ &= P(\boldsymbol{\theta}^{(t-1)} = \mathbf{b}) \\ &= p(\mathbf{b} \,|\, \mathbf{Y}) . \end{aligned} \tag{A.17}$$

The underlying idea behind HMC is that instead of generating the proposed value from a random distribution, we use a deterministic *symplectic integrator* to propose $\boldsymbol{\theta}^*$. This symplectic integrator is designed based upon Hamiltonian dynamics. Suppose that we wish to sample from $p(\boldsymbol{\theta} \,|\, \mathbf{Y})$, where $\boldsymbol{\theta} \in \mathbb{R}^d$. We introduce an auxiliary variable $\mathbf{r} \in \mathbb{R}^d$ so that we can efficiently sample from the joint density $p(\boldsymbol{\theta}, \mathbf{r} \,|\, \mathbf{Y})$. If $(\boldsymbol{\theta}^{(t)}, \mathbf{r}^{(t)}) \sim p(\boldsymbol{\theta}, \mathbf{r} \,|\, \mathbf{Y})$, then

$$\begin{aligned} P(\boldsymbol{\theta}^{(t)} = \mathbf{b}) &= \int P(\boldsymbol{\theta}^{(t)} = \mathbf{b}, \mathbf{r}^{(t)} = \mathbf{u}) d\mathbf{u} \\ &= \int p(\mathbf{b}, \mathbf{u} \,|\, \mathbf{Y}) d\mathbf{u} \\ &= p(\mathbf{b} \,|\, \mathbf{Y}) . \end{aligned} \tag{A.18}$$

Hence, sampling from the joint density $p(\boldsymbol{\theta}, \mathbf{r} \,|\, \mathbf{Y})$ results in samples from $p(\boldsymbol{\theta} \,|\, \mathbf{Y})$.

The auxiliary variable, $\mathbf{r}$, is also called the "momentum" in Hamilton dynamics. For our purposes, it suffices to specify that $p(\boldsymbol{\theta}, \mathbf{r} \,|\, \mathbf{Y}) = p(\boldsymbol{\theta} \,|\, \mathbf{Y}) \times p(\mathbf{r})$. Therefore, $p(\mathbf{r} \,|\, \boldsymbol{\theta}, \mathbf{Y}) = p(\mathbf{r})$ which means that $\mathbf{r}$ is independent of the data $\mathbf{Y}$ and the model parameters $\boldsymbol{\theta}$. More specifically, we assume that $p(\mathbf{r}) = N(\mathbf{r} \,|\, \mathbf{0}, \mathbf{I}_d) \propto \exp\left(-\frac{1}{2}\mathbf{r}^\top \mathbf{r}\right)$. Therefore,

$$\log p(\boldsymbol{\theta}, \mathbf{r} \,|\, \mathbf{Y}) = \text{constant} + \log p(\boldsymbol{\theta} \,|\, \mathbf{Y}) - \frac{1}{2}\mathbf{r}^\top \mathbf{r} . \tag{A.19}$$

The above density can be looked upon as a physical system subject to Hamiltonian dynamics, where $\boldsymbol{\theta}$ is a particle's position in $\mathbb{R}^d$ and $\mathbf{r}$ is the particle's momentum.

In order to sample from (A.19), a simple HMC algorithm proceeds closely on the lines of the Metropolis random walk described earlier, but replaces the random generation of a proposed value for $\boldsymbol{\theta}$ by a symplectic integrator constructed from Hamiltonian dynamics. With the current state $(\boldsymbol{\theta}^{(t-1)}, \mathbf{r}^{(t-1)})$, we begin iteration t by drawing the momentum variable $\mathbf{r}^* \sim N(\mathbf{0}, \mathbf{I}_d)$. Setting $\mathbf{r}^{(t)} = \mathbf{r}^*$ we perform L steps of a symplectic integrator (also known as "leapfrog"), where each step comprises the following:

$$\begin{aligned} \mathbf{r}^{(t+\epsilon/2)} &= \mathbf{r}^{(t)} + \frac{\epsilon}{2} \nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}); \\ \boldsymbol{\theta}^{(t-1+\epsilon)} &= \boldsymbol{\theta}^{(t-1)} + \epsilon \mathbf{r}^{(t+\epsilon/2)}; \\ \mathbf{r}^{(t+\epsilon)} &= \mathbf{r}^{(t+\epsilon/2)} + \frac{\epsilon}{2} \nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}) , \end{aligned} \tag{A.20}$$

where $\mathcal{L}(\boldsymbol{\theta}) = \log p(\boldsymbol{\theta} \,|\, \mathbf{Y})$. Let $\tilde{\boldsymbol{\theta}}$ and $\tilde{\mathbf{r}}$ be the output of (A.20) at the end of L steps. The values of $\tilde{\boldsymbol{\theta}}$ and $\tilde{\mathbf{r}}$ are considered the "proposed" values at iteration t and accepted as $(\boldsymbol{\theta}^{(t)}, \mathbf{r}^{(t)}) = (\tilde{\boldsymbol{\theta}}, -\tilde{\mathbf{r}})$ with acceptance probability $\min\left(1, \dfrac{p(\tilde{\boldsymbol{\theta}} \,|\, \mathbf{Y}) p(\tilde{\mathbf{r}})}{p(\boldsymbol{\theta}^{(t-1)} \,|\, \mathbf{Y}) p(\mathbf{r}^*)}\right)$. This last Metropolis step together with the negation of the momentum variable in the final update ensures time-reversibility as in (A.16) and, as seen in (A.17), maintains $p(\boldsymbol{\theta} \,|\, \mathbf{Y})$ as the stationary distribution.

We provide some further intuition on the time-reversibility of the simple HMC algorithm. The key to this result is that the leapfrog iteration in (A.20) preserves volumes. To be slightly more precise, let $\mathcal{D}$ be a small region

in the $(\boldsymbol{\theta}, \mathbf{r})$ space and suppose the L leapfrog steps maps $\mathcal{D}$ to a region $\tilde{\mathcal{D}}$. Then $\mathcal{D}$ and $\tilde{\mathcal{D}}$ both have the same volume. We write the transition probability from $\mathcal{D}$ to $\tilde{\mathcal{D}}$ as

$$\begin{aligned} T(\mathcal{D} \to \tilde{\mathcal{D}}) &= (\delta V) \min\left(1, \frac{\exp(-H(\tilde{\mathcal{D}}))}{\exp(-H(\mathcal{D}))}\right) \\ &= (\delta V) \min\left(1, \exp\left(-H(\tilde{\mathcal{D}}) + H(\mathcal{D})\right)\right) , \end{aligned} \tag{A.21}$$

where δV represents the volume of $\mathcal{D}$ and $\tilde{\mathcal{D}}$, and $\int_{\mathcal{D}} p(\boldsymbol{\theta}, \mathbf{r}) d\boldsymbol{\theta} d\mathbf{r} = \exp\left(-H(\mathcal{D})\right)$. If $(\boldsymbol{\theta}^{(t-1)}, \mathbf{r}^*)$ is drawn from the joint density in (A.19), then $P((\boldsymbol{\theta}^{(t-1)}, \mathbf{r}^*) \in \mathcal{D}) = \int_{\mathcal{D}} p(\boldsymbol{\theta}, \mathbf{r}) d\boldsymbol{\theta} d\mathbf{r} = \exp\left(-H(\mathcal{D})\right)$. Therefore,

$$\begin{aligned} &P((\boldsymbol{\theta}^{(t-1)}, \mathbf{r}^*) \in \mathcal{D}, (\boldsymbol{\theta}^{(t)}, \mathbf{r}^{(t)}) \in \tilde{\mathcal{D}}) \\ &\quad = P((\boldsymbol{\theta}^{(t-1)}, \mathbf{r}^*) \in \mathcal{D}) T(\mathcal{D} \to \tilde{\mathcal{D}}) \\ &\quad = \exp\left(-H(\mathcal{D})\right) (\delta V) \min\left(1, \exp\left(-H(\tilde{\mathcal{D}}) + H(\mathcal{D})\right)\right) \\ &\quad = (\delta V) \min\left(\exp\left(-H(\mathcal{D})\right), \exp\left(-H(\tilde{\mathcal{D}})\right)\right) \\ &\quad = P((\boldsymbol{\theta}^{(t-1)}, \mathbf{r}^*) \in \tilde{\mathcal{D}}, (\boldsymbol{\theta}^{(t)}, \mathbf{r}^{(t)}) \in \mathcal{D}) , \end{aligned} \tag{A.22}$$

where the last equality follows from the symmetry in the expression above it.

This procedure, while maintaining the stationary distribution through time-reversibility, introduces a host of complexities. Perhaps most importantly, tuning the many parameters needed in this process is inherently difficult. This motivated the development of an automatic procedure known as the No-U-Turn-Sampler (NUTS) [13]. This algorithm achieves significant efficiency over the simple HMC algorithm described above by either explicitly avoiding a U-turn to previously explored region or terminating after a pre-defined number of exploration steps. In this way, the algorithm is guaranteed to only explore new areas of the space. This efficient exploration results in typically faster convergence and higher effective sample sizes per iteration as compared to classical MCMC.

Accepted 20 June 2021

REFERENCES

[1] Betancourt, M. (2018). *A Conceptual Introduction to Hamiltonian Monte Carlo*. 1701.02434.

[2] Bonaccorsi, G., Pierri, F., Cinelli, M., Flori, A., Galeazzi, A., Porcelli, F., Schmidt, A. L., Valensise, C. M., Scala, A., Quattrociocchi, W. and Pammolli, F. (2020). Economic and social consequences of human mobility restrictions under COVID-19. *Proceedings of the National Academy of Sciences* **117**(27) 15530–15535. https://doi.org/10.1073/pnas.2007658117. https://www.pnas.org/content/117/27/15530.full.pdf.

[3] Chatzilena, A., van Leeuwen, E., Ratmann, O., Baguelin, M. and Demiris, N. (2019). Contemporary statistical inference for infectious disease models using Stan. *Epidemics* **29** 100367. https://doi.org/10.1016/j.epidem.2019.100367.

[4] Cutler, D. M. and Summers, L. H. (2020). The COVID-19 Pandemic and the $16 Trillion Virus. *JAMA* **324**(15) 1495–1496. https://doi.org/10.1001/jama.2020.19759.

[5] Diekmann, O., Heesterbeek, J. A. P. and Roberts, M. G. (2009). The construction of next-generation matrices for compartmental epidemic models. *Journal of the Royal Society, Interface / the Royal Society* **7** 873–885. https://doi.org/10.1098/rsif.2009.0386.

[6] Dietz, K. and Heesterbeek, J. A. P. (2002). Daniel Bernoulli's epidemiological model revisited. *Mathematical biosciences* **180** 1–21. https://doi.org/10.1016/S0025-5564(02)00122-0. MR1950745

[7] Dukic, V., Lopes, H. and Polson, N. (2012). Tracking Epidemics with Google Flu Trends Data and a State-Space SEIR Model. *Journal of The American Statistical Association* **107**. https://doi.org/10.1080/01621459.2012.713876. MR3036404

[8] Gandhi, M., Yokoe, D. and Havlir, D. (2020). Asymptomatic Transmission, the Achilles' Heel of Current Strategies to Control Covid-19. *New England Journal of Medicine* **382**. https://doi.org/10.1056/NEJMe2009758.

[9] Gelman, A. and Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science* **7**(4) 457–472. https://doi.org/10.1214/ss/1177011136.

[10] Gu, Y. (2020). *COVID-19 Projections Using Machine Learning*. https://covid19-projections.com.

[11] Hao, X., Cheng, S., Wu, D., Wu, T., Lin, X. and Wang, C. (2020). Reconstruction of the full transmission dynamics of COVID-19 in Wuhan. *Nature* **584** 1–7. https://doi.org/10.1038/s41586-020-2554-8.

[12] Heffernan, J., Smith, R. J. and Wahl, L. M. (2005). Perspectives on the Basic Reproductive Ratio. *Journal of the Royal Society, Interface / the Royal Society* **2** 281–293. https://doi.org/10.1098/rsif.2005.0042.

[13] Hoffman, M. and Gelman, A. (2011). The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research* **15**. MR3214779

[14] Ives, A. R. and Bozzuto, C. (2020). State-by-State estimates of R0 at the start of COVID-19 outbreaks in the USA. medRxiv. https://doi.org/10.1101/2020.05.17.20104653.

[15] Kermack, W. O. and McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences* **115**(772) 700–721.

[16] Kobayashi, G., Sugasawa, S., Tamae, H. and Ozu, T. (2020). Predicting intervention effect for COVID-19 in Japan: state space modeling approach. *BioScience Trends* **14**. https://doi.org/10.5582/bst.2020.03133.

[17] Lauer, S., Grantz, K., Bi, Q., Jones, F., Zheng, Q., Meredith, H., Azman, A., Reich, N. and Lessler, J. (2020). The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Annals of internal medicine* **172**. https://doi.org/10.7326/M20-0504.

[18] Madjid, M., Safavi-Naeini, P., Solomon, S. and Vardeny, O. (2020). Potential Effects of Coronaviruses on the Cardiovascular System: A Review. *JAMA Cardiology* **5**. https://doi.org/10.1001/jamacardio.2020.1286.

[19] Neal, R. (2012). MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*. https://doi.org/10.1201/b10905-6. MR2858447

[20] Øksendal, B. (2003) *Stochastic Differential Equations: An Introduction with Applications. Hochschultext / Universitext*. Springer. https://books.google.com/books?id=kXw9hB4EEpUC.

[21] Osthus, D., Hickmann, K., Caragea, P., Higdon, D. and Del Valle, S. (2017). Forecasting seasonal influenza with a state-space SIR model. *The Annals of Applied Statistics* **11** 202–224. https://doi.org/10.1214/16-AOAS1000. MR3634321

[22] Plummer, M. (2003). *JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling*.

[23] Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flan-

nery, B. P. (2007) *Numerical Recipes 3rd Edition: The Art of Scientific Computing*, 3 ed. Cambridge University Press, USA. MR2371990

[24] Shiode, N., Shiode, S., Rod-Thatcher, E., Rana, S. and Vinten-Johansen, P. (2015). The mortality rates and the space-time patterns of John Snow's cholera epidemic map. *International Journal of Health Geographics* **14** 21. https://doi.org/10.1186/s12942-015-0011-y.

[25] Stan Development Team (2020) Stan Modeling Language Users Guide and Reference Manual, Version 2.21.0. http://mc-stan.org/.

[26] Struthers, A. and Potter, M. (2019) *Differential Equations: For Scientists and Engineers*. https://doi.org/10.1007/978-3-030-20506-5. MR3967751

[27] Särkkä, S. and Solin, A. (2019) *Applied Stochastic Differential Equations. Institute of Mathematical Statistics Textbooks*. Cambridge University Press. https://doi.org/10.1017/9781108186735. MR3931353

[28] Treiber, M. and Kanagaraj, V. (2015). Comparing numerical integration schemes for time-continuous car-following models. *Physica A: Statistical Mechanics and its Applications* **419** 183–195. https://doi.org/10.1016/j.physa.2014.09.061.

[29] U. S. Census Bureau (2019). *2010 Census*.

[30] Wang, C., Liu, L., Hao, X., Guo, H., Wang, Q., Huang, J., He, N., Yu, H., Lin, X., Pan, A., Wei, S. and Wu, T. (2020). Evolving Epidemiology and Impact of Non-pharmaceutical Interventions on the Outbreak of Coronavirus Disease 2019 in Wuhan, China. https://doi.org/10.1101/2020.03.03.20030593.

[31] Wang, L., Zhou, Y., He, J., Zhu, B., Wang, F., Tang, L., Eisenberg, M. and Song, P. X. K. (2020). An epidemiological forecast model and software assessing interventions on COVID-19 epidemic in China. medRxiv. https://doi.org/10.1101/2020.02.29.20029421.

[32] Wang, W., Ding, Z. and Gui, Z. (2018). A stochastic differential equation SIS epidemic model incorporating Ornstein-Uhlenbeck process. *Physica A: Statistical Mechanics and its Applications* **509**. https://doi.org/10.1016/j.physa.2018.06.099. MR3834095

Ian Frankenburg. Department of Biostatistics, Fielding School of Public Health, University of California Los Angeles, USA. E-mail address: ian.frankenburg@ucla.edu

Sudipto Banerjee. Department of Biostatistics, Fielding School of Public Health, University of California Los Angeles, USA. E-mail address: sudipto@ucla.edu