Faculty Publications, Department of Statistics   Statistics, Department of

7-20-2022

Spatio-temporal models of infectious disease with high rates of asymptomatic transmission

Aminur Rahman

Angela Peace

Ramesh Kesawan

Souparno Ghosh

Spatio-temporal models of infectious disease with high rates of asymptomatic transmission

Aminur Rahman *†, Angela Peace‡, Ramesh Kesawan§, Souparno Ghosh§

Abstract

The surprisingly mercurial Covid-19 pandemic has highlighted the need to not only accelerate research on infectious disease, but to also study them using novel techniques and perspectives. A major contributor to the difficulty of containing the current pandemic is due to the highly asymptomatic nature of the disease. In this investigation, we develop a modeling framework to study the spatio-temporal evolution of diseases with high rates of asymptomatic transmission, and we apply this framework to a hypothetical country with mathematically tractable geography; namely, square counties uniformly organized into a rectangle. We first derive a model for the temporal dynamics of susceptible, infected, and recovered populations, which is applied at the county level. Next we use likelihood-based parameter estimation to derive temporally varying disease transmission parameters on the state-wide level. While these two methods give us some spatial structure and show the effects of behavioral and policy changes, they miss the evolution of hot zones that have caused significant difficulties in resource allocation during the current pandemic. It is evident that the distribution of cases will not be stagnantly based on the population density, as with many other diseases, but will continuously evolve. We model this as a diffusive process where the diffusivity is spatially varying based on the population distribution, and temporally varying based on the current number of simulated asymptomatic cases. With this final addition coupled to the SIR model with temporally varying transmission parameters, we capture the evolution of "hot zones" in our hypothetical setup.

*Corresponding, arahman2@uw.edu, http://faculty.uw.edu/arahman2

†Department of Applied Mathematics, University of Washington

‡Department of Mathematics and Statistics, Texas Tech University, Lubbock, TX

§Department of Statistics, University of Nebraska - Lincoln

1 Introduction

Similar phenomena often appear across many different branches of Science. Infectious disease spreads within a population through probabilistic interactions, and it has been shown that encounters between organisms may be modeled as finite-speed Brownian-like motion [1]. Spatially discrete interactions between patches of populations have also been modeled [2] using a Susceptible, Infected, Recovered (SIR) model within a patch. However, in addition to local Brownian-like motion and interactions between population groups, humans travel long distances, which has its own transport dynamics as shown by Brockmann *et al.* [3]. Similar to Statistical Mechanics [4], these interactions appear random at the scale of a few agents, but average out at the scale of an entire population.

Throughout the Covid-19 pandemic there was an emphasis on modeling the averaged spread of infections at either the federal or state levels. However, it is evident that in the beginning of the pandemic resources needed to be distributed to the areas that were going to be hit the hardest as it went from one hot zone to the next. This would have required fast fine-grained spatio-temporal predictions. However, for current spatio-temporal models [2], the computational expense would impede fast predictions. Using statistical techniques may mitigate much of the computational expense while still providing fine-grained predictions. Using Brownian motion, for example, we have temporal dynamics that provide an averaged picture of the transport, and spatial distributions, which can be represented using a probability density function. Taking the product provides spatio-temporal dynamics within the system without having to keep track of each molecule. Similarly, for infectious disease we have temporal dynamics that can be informed by an SIR-type model, and spatial distributions, which naturally has a probability density function associated with it.

A major issue with SARS CoV2 is the high rates of asymptomatic transmission. This is due to its unusually long latency period and the likelihood of an asymptomatic case. If a disease is highly symptomatic, we expect the carrier would know that they have the disease thereby voluntarily limiting contact with others, and they may also be unwell enough to involuntarily reduce contact with others. On the other hand, with highly asymptomatic diseases, even if someone intends to limit contact if they were to ever get infected, they may not even know that they were infected. This creates an environment conducive to spatio-temporal transport rather than isolated community spread.

Any complex phenomenon is difficult to model, and the complications present in the spread of SARS CoV2 make the Covid-19 pandemic particularly challenging to model [5, 6]. This manuscript presents new modeling techniques inspired by the study of Statistical Mechanics, and provides preliminary results as motivation for the use of these techniques on infectious disease, especially those with high rates of asymptomatic transmission. The remainder of the paper is organized as follows: starting in Sec. 2 we derive the SIR-like model for the county-level temporal dynamics. Next in Sec. 3 we conduct a likelihood-based parameter estimation of the transmission parameters from the model in Sec. 2. This reveals how the evolu-

tion landscape can change with behavioral or policy decisions such as masking and lockdowns. Finally, section 4 captures the evolution of "hot zones". We model the redistribution of the simulated asymptomatic cases from the SIR model in Sec. 2 as a diffusive process that is dependent on the population distribution and current number of asymptomatic cases in each county. In Sec. 5 we conclude with a discussion on the modeling framework and promising avenues for future work.

2 County-wise epidemic temporal dynamics

Let us first derive an SIR-type model of viral transmission of a highly asymptomatic disease. Consider dividing the population into subpopulations of susceptible individuals S, latently infected individuals who aren't yet infectious L, infectious individuals who are asympotimatic I_A and symptomatic I_S, and recovered individuals with temporary immunity R. The base model has the following structure:

$$\frac{dS}{dt} = \underbrace{-\beta_A S I_A}_{\substack{\text{asymptomatic} \\ \text{transmission}}} \quad \underbrace{-\beta_S S I_S}_{\substack{\text{symptomatic} \\ \text{transmission}}} \quad + \underbrace{\eta R}_{\substack{\text{immunity} \\ \text{loss}}} \tag{1a}$$

$$\frac{dL}{dt} = \underbrace{\beta_A S I_A}_{\substack{\text{asymptomatic} \\ \text{transmission}}} + \underbrace{\beta_S S_S^I}_{\substack{\text{symptomatic} \\ \text{transmission}}} - \underbrace{\varepsilon L}_{\substack{\text{latent get} \\ \text{infectious}}} \tag{1b}$$

$$\frac{dI_A}{dt} = \underbrace{\rho \varepsilon L}_{\substack{\text{latent get} \\ \text{asym. infectious}}} - \underbrace{\gamma_A I_A}_{\text{recovery}} - \underbrace{\mu_A I_A}_{\substack{\text{disease induced} \\ \text{mortality}}} \tag{1c}$$

$$\frac{dI_S}{dt} = \underbrace{(1-\rho)\varepsilon L}_{\substack{\text{latent get} \\ \text{sym. infectious}}} - \underbrace{\gamma_S I_S}_{\text{recovery}} - \underbrace{\mu_S I_S}_{\substack{\text{disease induced} \\ \text{mortality}}} \tag{1d}$$

$$\frac{dR}{dt} = \underbrace{\gamma_A I_A + \gamma_S I_S}_{\text{recovery}} - \underbrace{\eta R}_{\text{immunity loss}} \tag{1e}$$

where β_A and β_S are the asymptomatic and symptomatic transmission rates respectively, ε^{-1} is the incubation period, ρ is the percentage of cases that are asymptomatic, γ_A and γ_S are the asymptomatic and symptomatic recovery rates, μ_A and μ_S are the disease induced mortality rates for asymptomatic and symptomatic cases, and η is the waning rate of the temporary immunity gained from recovery.

While the disease discussed in this investigation is hypothetical, we will, whenever possible, use features of the Covid-19 pandemic to inform parameters and modeling strategies. The results however, are intended to be proof-of-concept demonstrations of this hypothetical disease and not predictors of Covid-19 dynamics.

We often think of infectious disease parameters as stagnant – a property of the disease itself. However, the environment in which it spreads can have a significant effect

on these parameters as we have noticed with the current pandemic. Consequently, specifying time-varying parameters make the model more flexible and perhaps more realistic. However, we do not have sufficiently resolved observational information that would enable us to infer the evolution of the said parameters. Hence, we need to fix some of these parameters and infer about the dynamics of the rest. In the current study, we assume disease recovery rates (γ_s and γ_a) and percentage of asymptomatic cases (ρ) are constants. As such, we specify $\gamma_s = 1/25.2$ and $\gamma_a = 1/22.6$ [7], and $\rho = 0.25$ [7, 8]. The asymptomatic mortality parameter μ_A is set to zero, and the symptomatic mortality paramter is set to $\mu_S = 0.005\gamma_S$ [9]. Further, we let the rate leaving the latent class be $\varepsilon = 1/5.8$ [10]. Finally, the time-varying transmission parameters β_S and $\beta_A = \rho\beta_S/(1-\rho)$ are estimated using the functional regression technique described in Section 3.

Consider a fictional country, let us call it Ω, where all the counties are uniform squares organized into a rectangular grid of 42×74 forming 48 contiguous states running sequentially down the columns. We shall use data from usafacts.org [11] for the population and initial conditions for each county. The county-wise and state-wide population distribution of Ω is shown in Fig. 1(a) and (b), respectively. In the heat maps the data for the states are taken in alphabetical order from left to right with each county represented down the column. So the Ω state that uses data from California will take up multiple columns, whereas the equivalent of Rhode Island will only take up three rows of a column. As a more concrete example, we can see California in deep red towards the left of the heat map, Fig. 1 (b), with its 58 counties and Texas towards the right with, in the words of Beto O'Rourke, "all 254 counties".
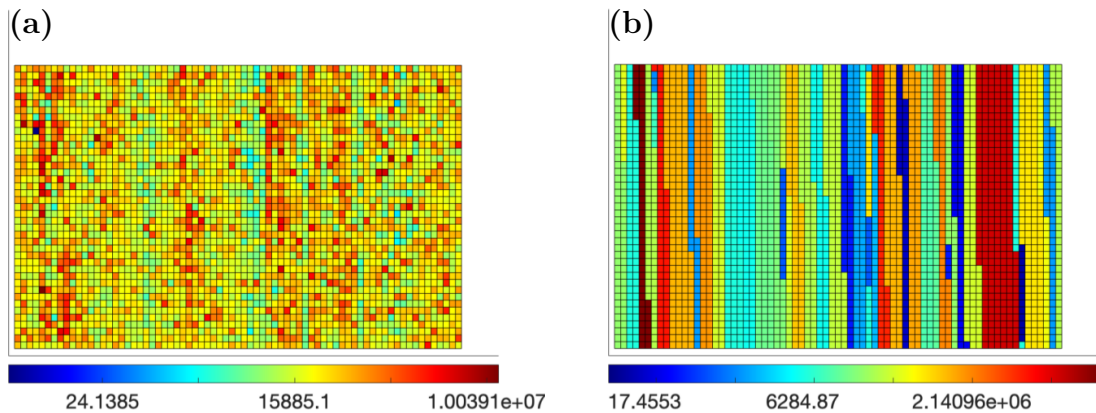

Figure 1: Population distribution of Ω. (a) County-wise populations. (b) State-wide populations. The colormap, with color bars on the bottom, for the heat maps are on a log-scale for improved visibility.

Using a representative value of $\beta_S = 0.02$ we may simulate the SIR model for self-contained counties; i.e., it is assumed that there there is no interaction between the counties. Heat maps for the simulation of the number of symptomatically infected individuals, I_S, are produced by solving (1) through the standard MATLAB Runge-

Kutta [12, 13] solver `ode45`. The heat maps are shown in Fig. 2(**a - d**) with the initial condition in (**a**) calculated at the 100 day mark from the `usafacts.org` data set [11]. Fig. 2(**b - d**) show the solutions to the SIR model (1) at day 110, 130, and 300. The colormap for the heat maps are on a log-scale to highlight the large differences in disease proliferation between counties, which is also what has been observed for the current pandemic. The daily totals for the entirety of Ω is shown in Fig. 2(**e**) where we observe an initial logistic-like growth eventually decaying to a steady-state.

In addition to the predicted evolution of a disease, epidemiologists often estimate a basic reproduction number, $\mathcal{R}_0$. With our SIR modeling setup, we can compute the $\mathcal{R}_0$ using the "next generation matrix approach" of Diekmann and co-workers [14]:

$$\mathcal{R}_0 = \left(\frac{\beta_a \rho}{\gamma_a + \mu_a} + \frac{\beta_s(1-\rho)}{\gamma_s + \mu_s} \right) N \tag{2}$$

where N is the normalized population of the region of interest. Once we estimate the temporal evolution of the transmission parameters β_S and β_A in Sec. 3, we can estimate the $\mathcal{R}_0$, and investigate how it too evolves over time.

3   State-wide likelihood-based parameter estimation

With an SIR-type population model and accurately estimated parameters, we would be able to describe the averaged effects of the virus on a community. Standard SIR-type models, like System (1) can be spatially expanded using spatially discrete patches each with coupled systems of ordinary differential equations (ODEs) [15]. However, if we use a full patch modeling approach the dimension of the system can be extremely large as the number of patches is increased, and the populations are further divided into subpopulation compartments within each patch. Additionally, the size of parameter space will also increase as the model dimension increase leading to computational instability, or even intractability.

Thus, to reduce computational complexity some disease specific epidemiological characteristics could be used as auxiliary information. For instance, it is observed that SARS-CoV2 symptomatically infects individuals with some probability that is dependent on age, number of encounters with infected individuals, respiratory health, population density, and proximity to an infection "hot zone". We hypothesize that this individual-level probability manifests itself as a spatial continuum through a spatial stochastic process. Let $\Psi(x, y)$ be the joint probability density function (pdf) for a finite collection of random variables operating in space. Applying the pdf locally (near a "hot zone") to an averaged symptomatically infected population yields a spatial distribution of the most vulnerable individuals; that is, $u(x, y; t) = \Psi(x, y)I_s(t)$. Observe that, for a given t, the formulation of $u(x, y; t)$ is reminiscent of intensity function driving a non-homogeneous spatial point process in
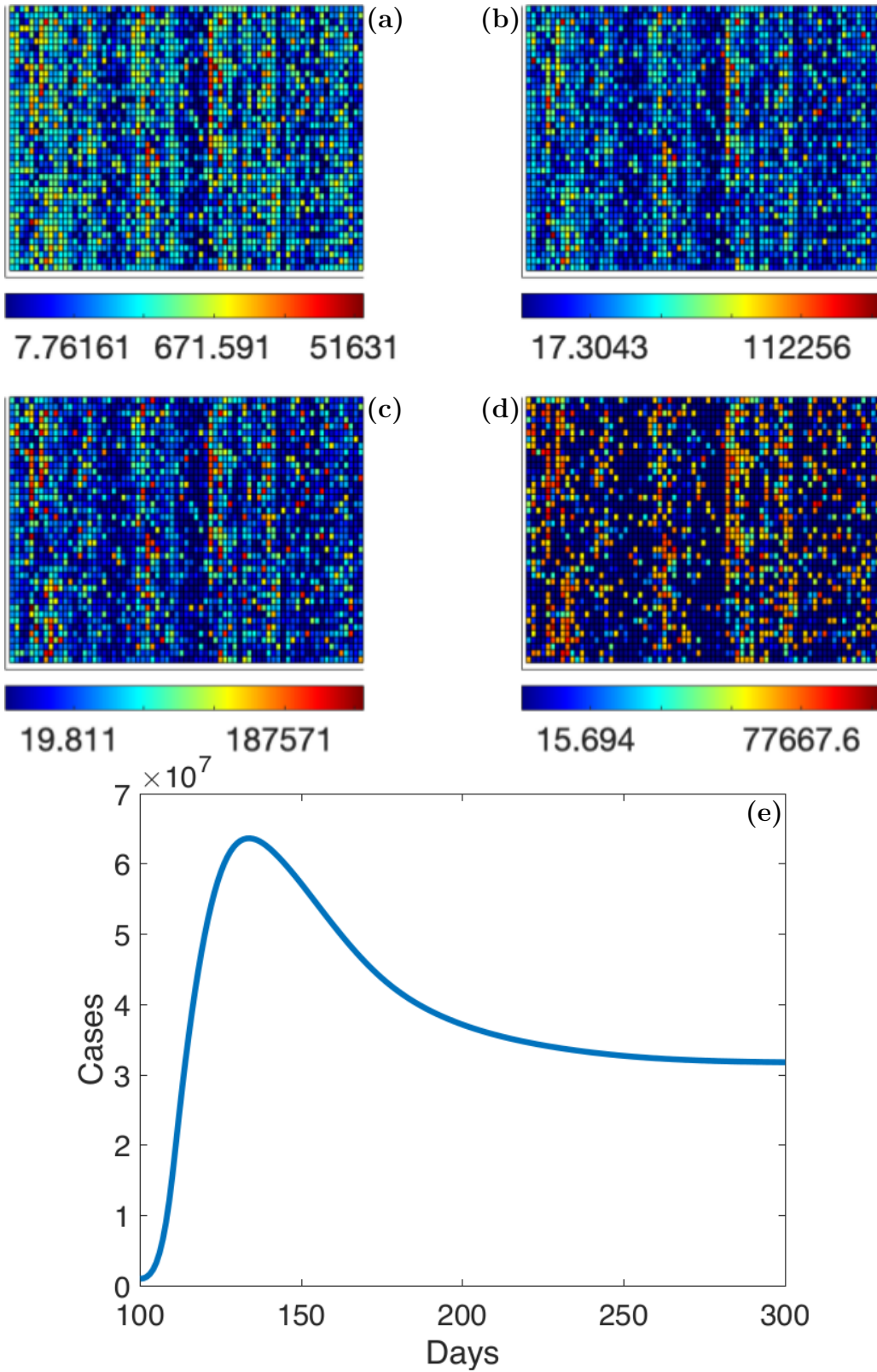
5

Figure 2: Simulation of the SIR model (1) with initial condition **(a)** taken from the 100 day mark. **(b)** - **(d)** simulated heat maps at the 110, 130, and 300 day marks. The colormap, with color bars on the bottom, for the heat maps are on a log-scale. **(e)** Daily case totals for Ω showing logistic-like growth for the first few weeks, and then a decline to a steady-state.

the sense that $\int_{(x,y)\in\mathcal{A}} u(x,y,t)dxdy = I_s(t)$[16]. In other words, at time t, the spatial process $\Psi(.,.)$ redistributes $I_s(t)$ over a bounded spatial domain $\mathcal{A}$. Of course, these "hot zones" expand and/or move over time, and for long time dynamics of a pandemic we may need to introduce non-separability of space and time in $\Psi(.)$, that is, $u(x,y;t) = \Psi_t(x,y)I_s(t)$ where $\Psi_t(.)$ can be constructed using non-separable space-time covariance matrices [17, 18]. However, even with the short time localized distribution, we may presumably be able to quickly implement policies that mitigates an epidemic or prevents a pandemic and send resources to necessary locations to save as many lives as possible.

Customarily, the transmission rate parameters β_A, β_S in (1a) are assumed to be constant. However, in reality, they can change under external forcing (for example: policy intervention). Hence, we propose to statistically estimate a temporal evolution of β_A and β_S, explicitly taking into account a policy intervention event - the implementation of mask mandates in several US states. In estimating β, we only focus on (1a) because the time series of susceptible population can be observed and therefore an empirical model can be formulated[1]. However, fitting a statistical model with the parameterization in (1a) is problematic because β_A and β_S are not identifiable individually. Hence, we reparameterize the statistical model in the following way:

Let $S_j(t)$ be the susceptible population of state j at day t. Since the observed time series is at daily resolution, we define $Y_j(t) := S_j(t) - S_j(t-1), t = 1, 2, .., T$ as a discrete approximation of the LHS of (1a). We now model the time series $Y_j(t)$ using the functional predictors $S_j^*(t-1)$, that quantifies the latent population level evolution of the susceptible individuals and a binary vector $Z_{j,t}$ that takes the value 1 (0) if mask mandate was (was not) in effect at time t in the jth state. Although Z is time indexed, it difficult to envision it as a function. Hence given a time point, we simply view $Z_{.,t}$ as fixed predictor available to us at that time point.

Observe that, for the jth state, we view $Y_j(t)$ as a random response function dependent on two predictor functions and therefore can be statistically modeled using the function-on-function regression techniques. Formally, we write the regression model as follows,

$$Y_j(t) = f_0(t) + S_j^*(t-1)\beta(t) + Z_{j,t}f_1(t) + b_j(t) + \varepsilon_j(t), \tag{3}$$

where $f_0(t)$ is the population level functional intercept term, $\beta(t)$ is the slope function modeling the *concurrent* association between $S^*(t-1)$ and $Y(t)$ and therefore can be viewed as the composition of β_A and β_S in (1a), $f_1(t)$ quantifies the association between mask mandate and $Y(t)$ at the population, $b_j(t)$ is the state-specific random fluctuations from the $f_0(t)$, and $\varepsilon_j(t)$ are pure noise.

To ensure smoothness $b_j(t)$ is modeled as a zero mean Gaussian process with an AR(1) covariance function, $\varepsilon_j(t)$ are assumed to be independent $N(0,\sigma_\varepsilon^2)$ with $b_j(t)$ and $\varepsilon_i(t)$ are mutually independent across j. Furthermore, since model $S_j^*(t-1)$

---

[1]The trajectory of L in (1b) is guided by unobserved latent processes and therefore observational models cannot be specified.

in the RHS of (3) is a latent process, we propose the following process level specifi-caiton of S^* process $S_{j,obs}(t) = \mu_S(t) + b_{j,S}(t) + \varepsilon_j^s(t)$, with $S_j^*(.) = \mu_S(.) + b_{j,S}\mu_S(.)$ and $\varepsilon_j^s(.)$ is again a pure noise process. $\mu_S(.)$ the population-level mean of the S^* process and $b_{j,S}(.)$ is the state-specific deviation from the population mean of the S^* process.The regression coefficient functions are modeled using cubic B-splines with 30 interior knots. Thus, denoting $\chi(t) = \{\chi_1(t), \cdots, \chi_{30}(t)\}'$ as a sequence of B-spline basis functions evaluated at t, then $\beta(t) = \chi(t)'\boldsymbol{\theta}$ with $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_{30})'$ be-ing the vector of coefficients. Similarly, state-specific intercept curves $b_i(t)$, are also approximated using cubic B-spline basis functions as $b_j(t) = \sum_{k=1}^{30} u_{jk}\chi_k(t)$, where $\mathbf{u}_i = (u_{i1}, \cdots, u_{i30})'$ are the coefficient vectors.

In order to estimate a policy-specific β curve, we can either introduce an interac-tion term $(Z \times S^*)$ in the RHS of (3) or we can fit two separate models for Y series separately, i.e

$$Y_j^{(1)}(t) = f_0^{(1)}(t) + S_j^{*(1)}(t-1)\beta^{(1)}(t) + b_j^{(1)}(t) + \varepsilon_j^{(1)}(t) \tag{4}$$

$$Y_j^{(2)}(t) = f_0^{(2)}(t) + S_j^{*(2)}(t-1)\beta^{(2)}(t) + b_j^{(2)}(t) + \varepsilon_j^{(2)}(t) \tag{5}$$

where $Y_j^{(1)}(.)$ is the series associated with $Z = 0$ and $Y_j^{(1)}(.)$ is the series associated with $Z = 1$. Evidently, coefficient curves $\beta^{(1)}(t)$ and $\beta^{(2)}(t)$ will capture the differ-ential impact of policy intervention. Since (3) can handle unequally sampled time series [19], we have used, separately, models of Y series to extract the policy-specific coefficient curves. The fitting procedure was carried out using the R package *fcr*. All of the estimated β values can be found in the supplemental materials located in the file `Complete_result_beta.csv`.

Recall, the β curves obtained using functional regression does not directly yield β_A and β_S that we need to compute $\mathcal{R}_0$ in (2). Hence, a post-hoc estimate of the curves associated with these two parameters are generated using the formula $\beta(t) = w \times \beta_A(t) + (1-w) \times \beta_S(t)$ where w is dependent on the proportion of asymptomatic cases in the population with a further constraint that population mean of β_A curve associated with the regime $Z = 1$ will be dominated by the population mean of β_A curve associated with the regime $Z = 0$ [20].

Now that we have obtained the time-evolving transmission parameters β_A and β_S for each state, we have another way to analyze the spatio-temporal dynamics of the disease. In Sec. 2 the redistribution function Ψ was unity as the dynamics within the separate patches (counties) were the same. With the varying β's we have some non-trivial redistribution $u(x, y, t) = \Psi(x, y, t)I_s(t)$ that is implicitly dependent on β_A and β_S. First let us run our SIR model (1) with the new *beta*'s on a state-wide level. The state-wide heat maps for the number of symptomatically infected individuals, I_S, are shown in Fig. 3 with the initial condition in **(a)** calculated at the 100 day mark from the usafacts.org data set [11]. Fig. 2**(b - d)** show the solutions to the SIR model (1) with the varying β values at day 110, 130, and 300. The colormap for the heat maps are on a log-scale to highlight the large differences in disease proliferation between states, which is also what has been observed for the current pandemic. In the heat maps the data for the states are taken in alphabetical order
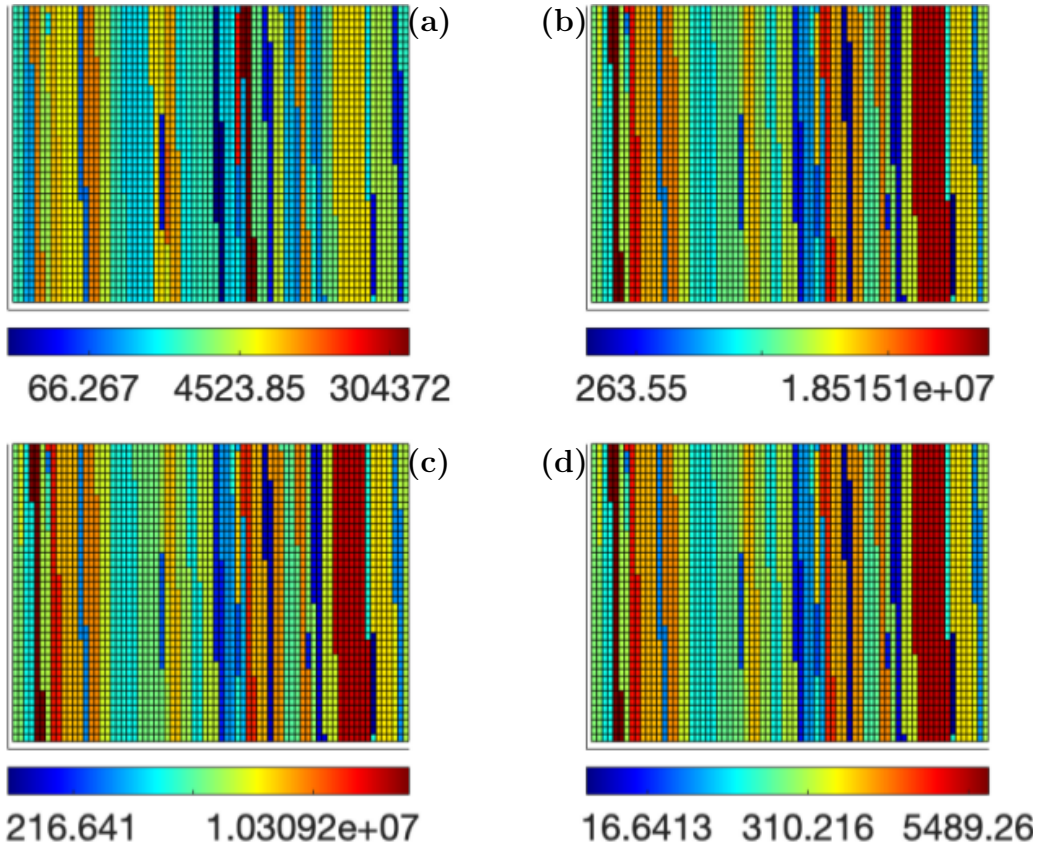
Figure 3: Simulation of the SIR model (1) with state-wise temporally evolving β_A and β_S with initial condition **(a)** taken from the 100 day mark. **(b)** - **(d)** simulated heat maps at the 110, 130, and 300 day marks. The colormap for the heat maps are on a log-scale.

from left to right with each county represented down the column. So the Ω state that uses data from California will take up multiple columns whereas the equivalent of Rhode Island will only take up three rows of a column.

While this is compelling, there is a lot of variance within individual states. Although the β values are state-wide we can downscale to the county level. This downscaling can be done in many different ways by modifying the redistribution function Ψ. For simplicity let us redistribute the state-wide results by taking the product with the population distribution. This yields the county-wise number of symptomatically infected individuals, I_S, which is shown in Fig. 4. As before, the initial condition in **(a)** is calculated at the 100 day mark from the `usafacts.org` data set [11]. Fig. 2**(b - d)** show the solutions to the SIR model (1) with the varying β values at day 110, 130, and 300. The colormap, with color bars on the bottom, for the heat maps are on a log-scale to highlight the large differences in disease proliferation between states, which is also what has been observed for the current pandemic. Even with this very
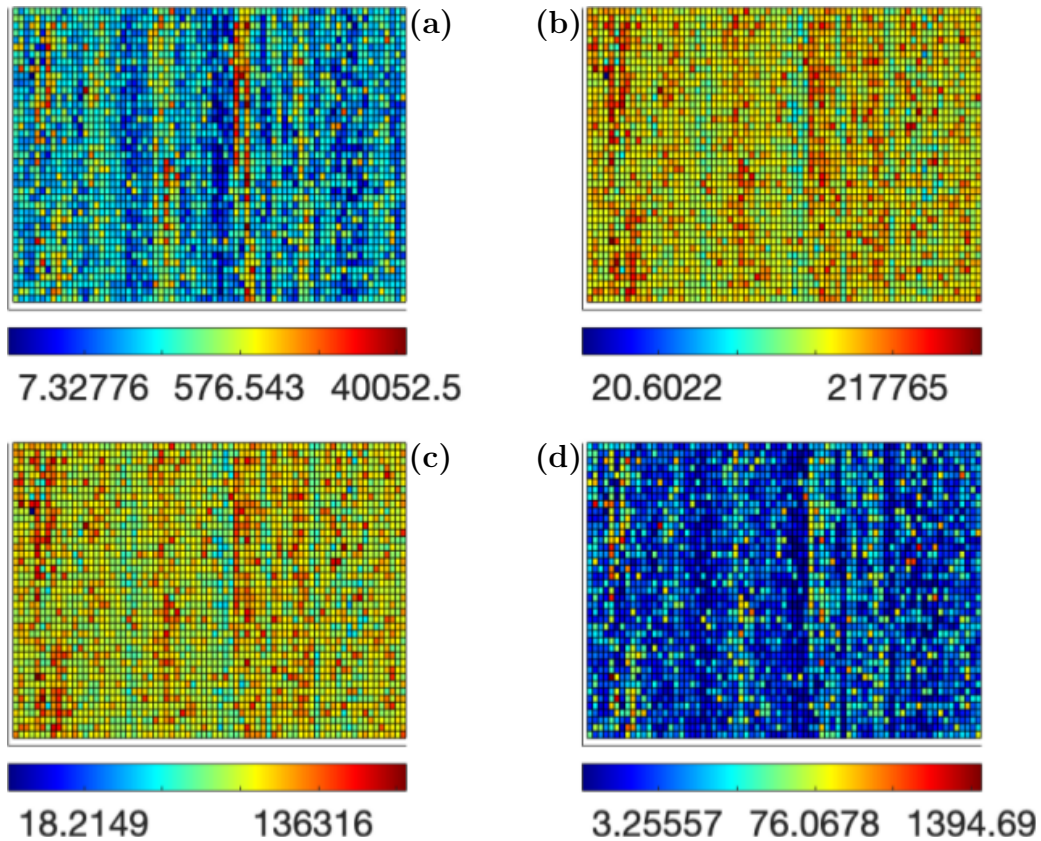

Figure 4: Simulation of the SIR model (1) with state-wise temporally evolving β_A and β_S with a county-wise downscaling through a simple product with the population distribution. The initial condition **(a)** is taken from the 100 day mark. **(b)** - **(d)** simulated heat maps at the 110, 130, and 300 day marks. The colormap, with color bars on the bottom, for the heat maps are on a log-scale.

simple down-scaling we see a lot more structure in the heat maps. This intensity

structure is precisely what could help policy makers dictate location-specific policy and send resources where they are predicted to be needed the most.

As mentioned in Sec. 2, epidemiologists often estimate a basic reproduction number, $\mathcal{R}_0$, to indicate the overall course of a disease, and it is often used to inform policy decisions. Now, that we have the β values, we have the tools to calculate the spatio-temporally evolving $\mathcal{R}_0$. We present a heat map of the $\mathcal{R}_0$ in Fig. 5. The 48 states of Ω are organized in alphabetical order from 1 to 48 (increasing ordinate direction). The abscissa is marked from day 93 to day 560 corresponding to tick marks 0 to 467. Finally, the colormap, with the color bar on the right, is on a linear scale from 0 to 3. Notice that the basic reproduction number is similar to that of the current pandemic even though the calculation was done for a hypothetical country. We also observe that as the pandemic carries on, the $\mathcal{R}_0$ increases, but then there is a sudden decrease. This would be due to policy changes in the various states of Ω, which to compare with our real-world situation, is most likely due to increased masking and locking down.
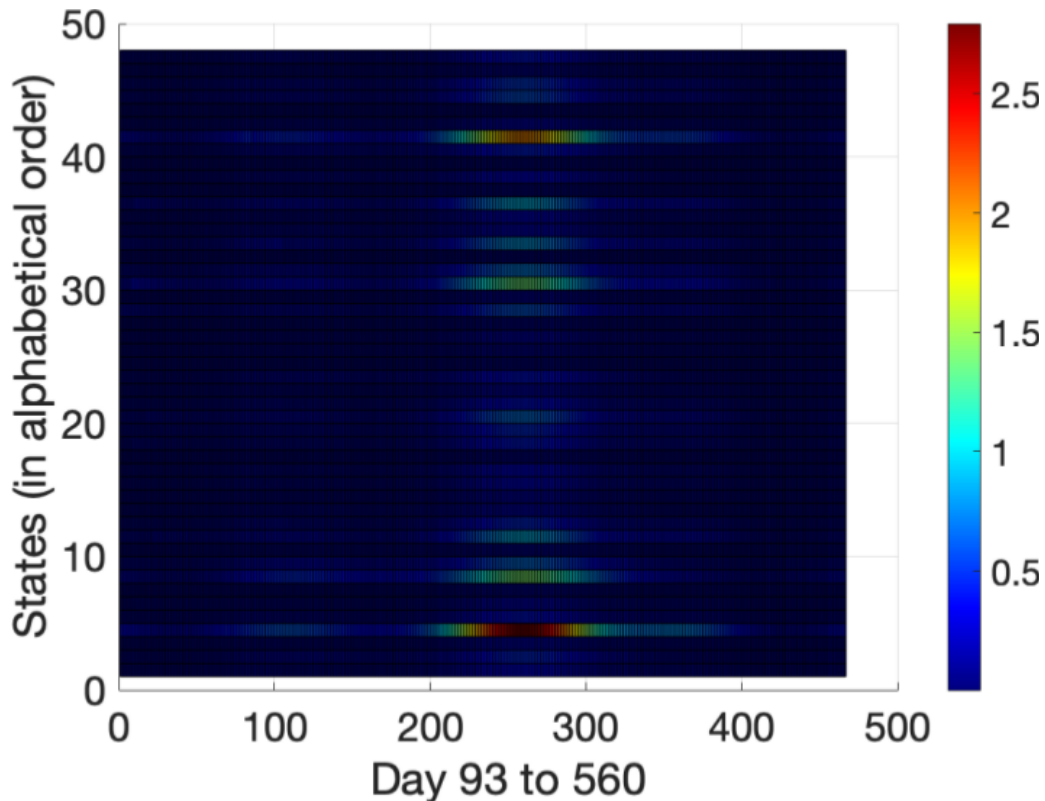

Figure 5: Heat map of the basic reproduction number, $\mathcal{R}_0$. The 48 states of Ω are organized in alphabetical order from 1 to 48 (increasing ordinate direction). The abscissa is marked from day 93 to day 560 corresponding to tick marks 0 to 467. Finally, the colormap, with the color bar on the right, is on a linear scale from 0 to 3, which while calculated for a proof-of-concept toy model, is on par with that of the $\mathcal{R}_0$ of Covid-19.

4   Diffusion-driven redistribution between counties

In Sec. 2 we simulated spatio-temporal behavior in a purely dynamical manner by considering the individual counties as uniform and non-interacting, then Sec. 3 considered non-uniformity in the form of temporally varying state-wide transmission parameters, β_A and β_S. One major issue of the current pandemic was the speed at which it spread across the world, effectively catching us by surprise. We hypothesize that the speed of the spread was due to the significantly high rate of asymptomatic transmission compared to other similar diseases, which indicates the importance of considering interaction between locations. Indeed we observed Covid-19 quickly spreading from one "hot zone" to another, however we only small indications of such hot zones in the simulations of the previous two sections.

Suppose that highly asymptomatic diseases do not significantly reduce travel between locations since individuals may not even know they are carriers. This is conducive to close interactions with other individuals thereby transmitting the disease even without the presence of symptoms. Although there are many transport-dynamics models such as that of Brockmann *et al.* [3], we notice that long distance travel reduced significantly and even halted once Covid-19 accelerated. However, the hot zones kept evolving, which indicates many short distance interactions spatially propagating through the population, similar to Brownian-like motion. Therefore, let us model the redistribution of asymptomatic individuals as a diffusive process, which will then implicitly (through the SIR model (1)) redistribute symptomatic individuals.

If we consider $I_A(t)$ in (1) as the "stationary" proliferation of asymptomatic infections and $\Psi(x, y, t = 0)$ as the standard population-wise distribution, then $u(x, y, t) = \Psi(x, y, t)I_A(t)$ is the redistributed proliferation. Now we suppose the redistribution is diffusion-driven; i.e., $u(x, y, t)$ satisfies the diffusion equation

$$\frac{\partial u}{\partial t} = \nabla \cdot (D(x, y)\nabla u)\,; \qquad u(x, y, 0) = \Psi(x, y, 0)I_A(0), \tag{6}$$

where $D(x, y)$ is the diffusivity, which is proportional to the population, and the proliferation is driven by the SIR model (1). We set the diffusivity to be proportional to the population since the disease spreads faster in populations of higher densities; i.e., we expect the disease from a particular hot zone to spread to nearby places, but also eventually to far away places with a preference towards locations of high density. Further, since we observed that Covid-19 started to spread from a particular hot zone once enough individuals were infected, and therefore we use a piecewise diffusivity

$$D(x, y) = \begin{cases} P(x, y)/\max(P) & \text{if } u(x, y) > u_*, \\ 0 & \text{otherwise;} \end{cases}$$

where $P(x, y)$ is the count-wise population and u_* is the threshold for diffusivity.

While we model the redistribution as one physical process, it need not be a physical process at all, but rather a mathematical construct which implicitly houses

the physical information of the system, in a possibly convoluted manner, such as with Schrodinger's equation.

Since we need to solve a partial differential equation (PDE) coupled to a system of ordinary differential equations (ODEs) with spatio-temporally varying parameters calculated through a statistical model, while also potentially hamstrung by geographic and data analytic considerations, the numerical methods could become quite complex. In the interest of simplicity, we solve the SIR model (1) with `ode45`, as done in Sec. 2, over a single day, and with single day temporal discretization we solve the PDE (6) using the Forward in Time Centered in Space method (FTCS). We acknowledge that there are convergence and accuracy considerations, and that the numerics could be made far more sophisticated, and we leave that to a future study. To guarantee convergence we take care to satisfy the Courant–Friedrichs–Lewy (CFL) condition [21] by scaling the county-wise population by the population of the largest county to produce the diffusivity $D(x, y)$. Even with the heavily simplified numerics, some of the simulations took several hours to run on an Early 2015 Macbook Pro.

We illustrate the numerical scheme in Fig. 6. We plot subsequent progression of the scheme from **(a, b)** the 100 day mark to **(e, f)** the 102 day mark. The left hand column **(a, c, e)** represent the stationary proliferation with the standard distribution based on population, and the right hand column **(b, d, f)** are the post-diffusion distributions. From **(a)** to **(b)** we observe the majority of the diffusion coming from counties with a large number of cases. Once it diffuses, the disease continues to proliferate (**(b)** to **(c)**). Then it diffuses more, then proliferates, and the process continues. Of course, this is due to the discretization, and analytically the proliferation and redistribution would happen at the same time in a coupled manner.

Similar to those of the previous sections, we present the simulation the symptomatic cases by solving (6) coupled with (1) using the β values of Sec. 3 in Fig. 7. The initial condition **(a)** is taken from the 100 day mark, and then **(b)** - **(d)** are simulated at the 110, 120, 130 day marks. The colormap, with color bars on the bottom, for the heat maps are on a log-scale to highlight the large differences in disease proliferation between counties. It was necessary to shorten the prediction window because of computational expenditure, which a more sophisticated numerical scheme, as acknowledged in the previous paragraph, would rectify. Despite these limitations, we see the hot zones more clearly in the heat maps.

Although the asymptomatic cases were presented in Fig. 6, it is perhaps more compelling to get the full picture of the evolution of asymptomatic cases, and hence present the evolution in Fig. 8. Just like in Fig. 7 we solve the PDE redistribution model (6) coupled with the SIR model (1) with state-wise temporally evolving β_A and β_S. Once again, the initial condition is taken from the 100 day mark from the `usafacts.org` data set [11]. The 110, 120, and 130 day marks are simulated in **(b)** - **(d)**. The colormap, with color bars on the bottom, for the heat maps are on a log-scale to highlight the large differences in disease proliferation between counties, which will make the hot zones clearer. Here we see a quite compelling illustration of
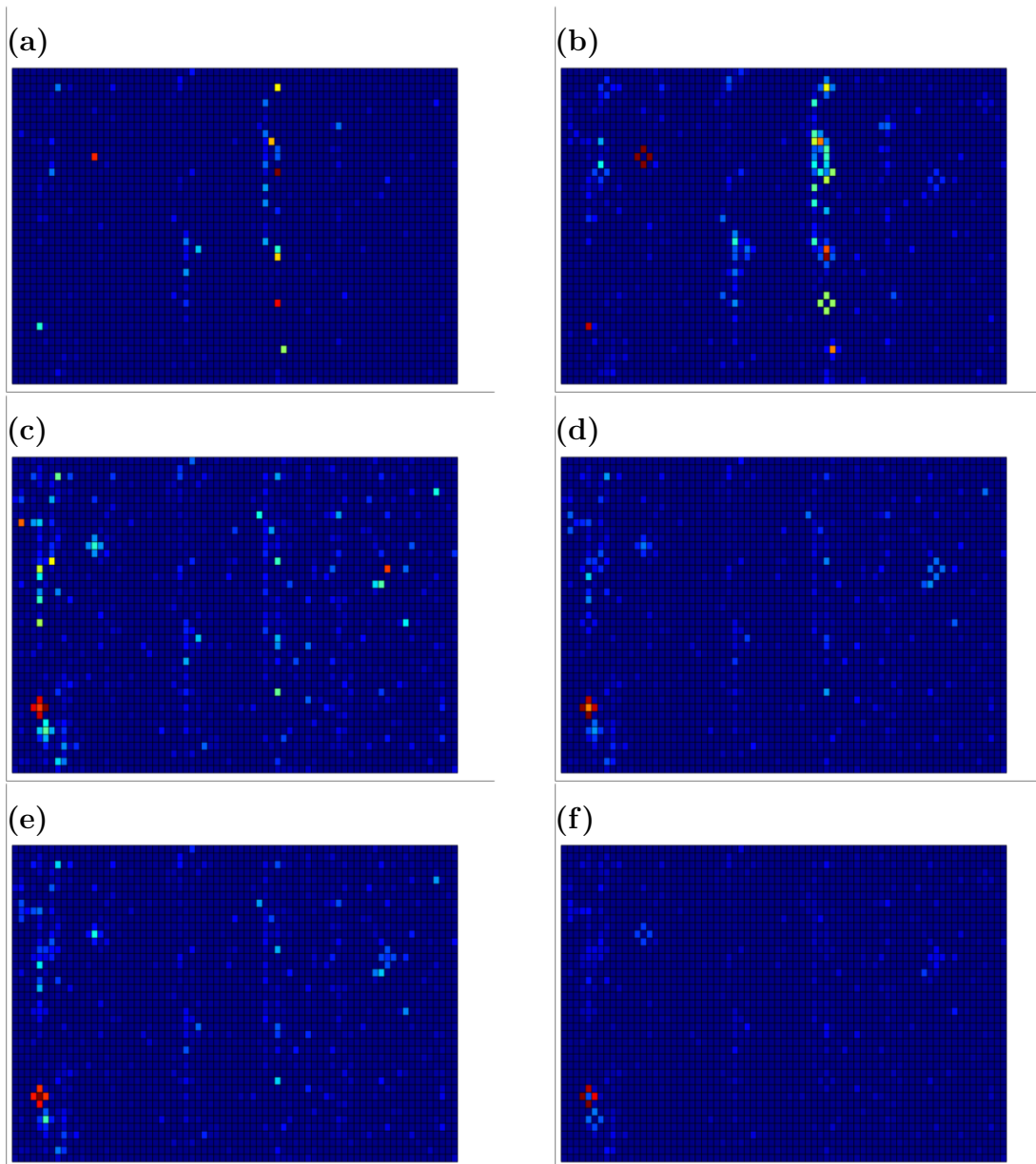
Figure 6: Illustration of the numerical scheme to solve the PDE redistribution model (6) coupled with the SIR model (1) using FTCS and Runge-Kutta. **(a) - (f)** Subsequent progression of the scheme through three timesteps: **(a, b)** 100 day mark, **(c, d)** 101 day mark, **(e, f)** 102 day mark. The heat maps in **(a, c, e)** are of the stationary proliferation with the standard distribution based on population, and **(b, d, f)** are post-diffusion distributions. Since there was no way to accurately track asymptomatic cases, all of the heat maps are simulated. The colormap is linear for the purposes of the demonstration.
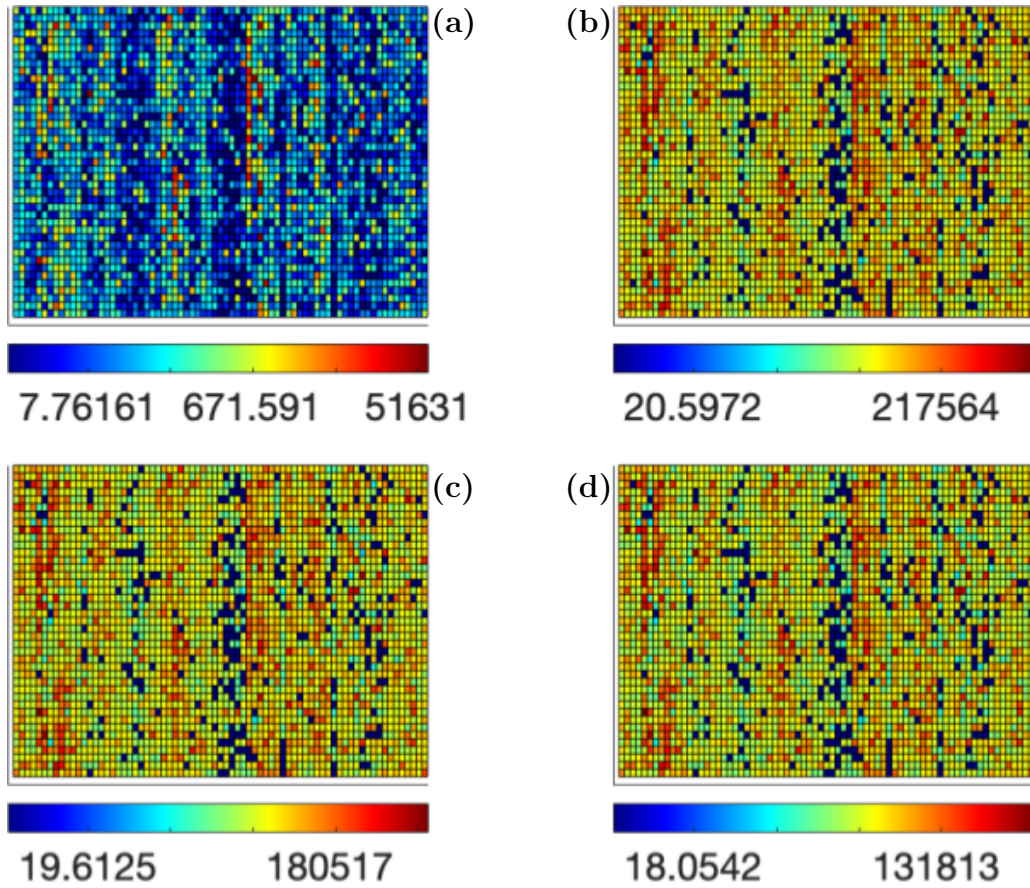
Figure 7: Simulation of the symptomatic cases by solving the PDE redistribution model (6) coupled with the SIR model (1) with state-wise temporally evolving β_A and β_S. The initial condition **(a)** is taken from the 100 day mark. **(b)** - **(d)** simulated heat maps at the 110, 120, and 130 day marks. The colormap, with color bars on the bottom, for the heat maps are on a log-scale.
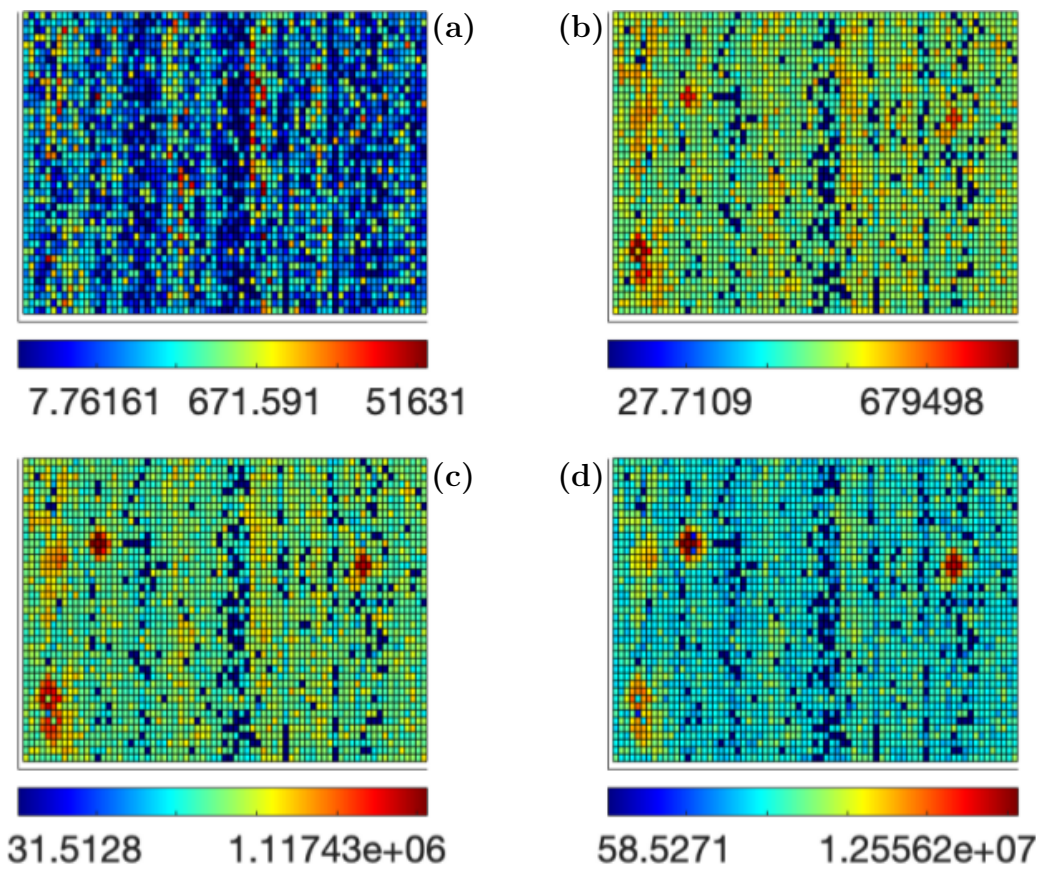
Figure 8: Simulation of asymptomatic cases by solving the PDE redistribution model (6) coupled with the SIR model (1) with state-wise temporally evolving β_A and β_S. **(a)** - **(d)** simulated heat maps at the 100, 110, 120, and 130 day marks. The colormap, with color bars on the bottom, for the heat maps are on a log-scale.

the hot zones, and even their evolution. Hypothetically, if there is a disease where the symptomatic cases lag behind the asymptomatic cases, and we do not have an accurate way of tracking asympotmatic cases, a model such as the one presented here could indicate where resources may be necessary before a hot zone arises and the healthcare system becomes overloaded.

It should be noted that in Fig. 8 the case numbers may seem unrealistic, but the reader is reminded that this is simply a proof-of-concept toy model to demonstrate that such a modelling framework can predict and hopefully help mitigate a disaster when applied to a real-world situation.

5 Discussion and Future Work

Although infectious disease has been studied for centuries, and at an especially accelerated rate in modern times, the current Covid-19 pandemic has revealed that there is still a lot we do not know about infectious disease. In particular, the highly asymptomatic nature of the disease caught the world by surprise. Our globalized society creates an environment conducive to the swift spread of diseases over long distances. The ability for the disease to cause an unbearable load on local medical infrastructure, and then move unpredictably to another hot zone, created problems of resource allocation. Even when the resources were available to release the pressure on our infrastructure, they did not always get to where it was needed. It is evident that more spatio-temporal models for infectious disease will be necessary in the future.

In this investigation, we progress through three types of spatio-temporal models for highly asymptomatic diseases. We first derive an SIR-like model (1) with fixed parameter values in Sec. 2. The model is applied at the county-wise level where all the counties are assumed to be uniform and non-interacting, which reveals some spatial structure, but misses more subtle structure, such as hot zones. Then in Sec. 3 we derive the state-wide spatio-temporal evolution of the transmission parameters, β_A and β_S, through likelihood based estimation. The variance in the transmission shows the effect of behavioral and policy changes, such as masking and lockdowns. This is particularly revealing in the evolution of the basic reproduction number, $\mathcal{R}_0$, (2) in Fig. 5. Finally, in Sec. 4 we derive a complete (including interactions between regions) spatio-temporal model by coupling a PDE governing redistribution (6) to the SIR model (1) with the state-wide spatio-temporally varying transmission parameters. This is where we see hot zones appearing and evolving, which strengthens the modeling framework as a proof-of-concept.

We foresee many more future projects arising from the modeling framework presented here. For example, notice that even at the state level population density is the strongest indicator of disease proliferation, regardless of clearly beneficial policy implementation such as masking. So does masking not work? There is an abundance of articles that show masking is effective (e.g., [22, 23, 24]). One way to see this is by comparing predicted results for no mitigation policies versus the location by location

differences in the actual case counts due to the implemented mitigation techniques in those areas. While, this was not included in the manuscript, it shows the importance of spatio-temporal models for disease transmission, and as more data is produced and more sophisticated models are derived, the possibility for such comparisons becomes very promising.

Moreover, the numerical methods in Sec. 4 can be improved significantly. We propose a future study with rigorous numerical analysis on solving the coupled PDE-SIR model. Most importantly, the SIR ODE system and the PDE should be discretized together. Runge-Kutta can still be used to integrate the ODE and Crank-Nicolson [25] can be used to integrate the PDE, which should be done in unison over one timestep. Crank-Nicolson was avoided (in favor of FTCS) due to the 2-dimensional nature of the problem. While the matrix equation can certainly be written down, the pentadiagonal system would be impossible to solve in any practical sense even for a highly optimized algorithm such as that of Strassen [26]. We would need to employ an alternating-direction implicit method, similar to that of Peaceman and Rachford Jr. [27], in order to avoid being hamstrung by the computational expense. Even then, the computational complexity would be quite high, and we foresee the need to parallelize several of the processes to run on a supercomputing cluster.

A major test for the modeling framework will be its performance when applied to real-world geographical constraints. At any other time the data for such a problem would not be available. The unique circumstances of the current pandemic provides us with multiple years of epidemiological data. By carefully constructing numerical methods for the highly non-uniform grid formed by the counties of the United States, we would have direct comparisons between the simulations and the data. Section 4 also only considers the diffusion of asymptomatic cases, but there are other factors that could affect the deviation of the distribution of cases from that of the standard population-wise distribution. In future studies, we need not be beholden to a physical process, but rather use the redistribution Ψ as a mathematical construct similar to the way Schrodinger's equation is used in Quantum Mechanics.

Finally, it is not unlikely that the ebb and flow of such a complex system is chaotic, especially as the pandemic has carried on for so long and the virus has mutated several times. As with the weather, due to errors in measurement (no matter how small) we can only hope to reproduce the qualitative long term behavior rather than precise long term predictions. However, we may be able to reliably forecast a week or a month ahead, which is effective for weather-related preparations, and it may be effective for infectious disease as well. Further, the manuscript does not consider mutations or vaccination, but as we are currently observing, this pandemic is uniquely complex due to the rate at which SARS CoV2 mutates, which occasionally evades vaccination. We foresee the inclusion of mutations and vaccinations in future studies as an important step in understanding the spatio-temporal evolution of a highly contagious disease.

Acknowledgments

References

[1] Eliezer Gurarie and Otso Ovaskainen. Towards a general formalization of encounter rates in ecology. *Theor. Ecol.*, 6:189–202, 2013.

[2] Alun L Lloyd and Vincent AA Jansen. Spatiotemporal dynamics of epidemics: synchrony in metapopulation models. *Mathematical biosciences*, 188(1-2):1–16, 2004.

[3] D. Brockmann, L. Hufnagel, and T. Geisel. The scaling laws of human travel. *Nature*, 439:462–465, 2006.

[4] A. I. Khinchin. *Statistical Mechanics*. Dover Publications, Inc., 1949.

[5] David Adam. Modelers struggle to predict the future of the covid-19 pandemic. *The Scientist*, 2020.

[6] A. Bertozzi, E. Franco, G. Mohler, M. B. Short, and D. Sledge. The challenges of modeling and forecasting the spread of covid-19. *Proc. Nat. Acad. Sci.*, 117(29):16732–16738, 2020.

[7] J. Y. Noh, J. G. Yoon, H. Seong, W. S. Choi, J. W. Sohn, W. J. Kim, and J. Y. Song. Asymptomatic infection and atypical manifestations of covid-19: Comparison of viral shedding duration. *J. Infect.*, 81(5):816–846, 2020.

[8] M. Alene, L. Yismaw, M. A. Assemie, D. B. Katema, B. Mengist, B. Kassie, and T. Y. Birhan. Magnitude of asymptomatic covid-19 cases throughout the course of infection: A systematic review and meta-analysis. *PLOS ONE*, page 0249090, 2021.

[9] T. Usherwood, Z. LaJoie, and V. Srivastava. A model and predictions for covid-19 considering population behavior and vaccination. *Sci. Rep.*, 11:12051, 2021.

[10] C. McAloon, A. Collins, K. Hunt, A. Barber, A. W. Byrne, F. Butler, M. Casey, J. Griffin, E. Lane, D. McEvoy, P. Wall, Green M., L. O'Grady, and S. J. More. Incubation period of covid-19: a rapid systematic review and meta-analysis of observational research. *Epidemiology*, 10(8):e039652, 2020.

[11] Us covid-19 cases and deaths by state, 2020.

[12] C. D. T. Runge. Über die numerische auflösung von differentialgleichungen. *Mathematische Annalen*, 46(2):167–178, 1895.

[13] M. Kutta. Beitrag zur näherungsweisen integration totaler differentialgleichungen. *Zeitschrift für Mathematik und Physik*, 46:435–453, 1901.

[14] Odo Diekmann, JAP Heesterbeek, and Michael G Roberts. The construction of next-generation matrices for compartmental epidemic models. *Journal of the Royal Society Interface*, 7(47):873–885, 2010.

[15] P Van den Driessche. Spatial structure: Patch models. In *Mathematical epidemiology*, pages 179–189. Springer, 2008.

[16] Janine Illian, Antti Penttinen, Helga Stoyan, and Dietrich Stoyan. *Statistical analysis and modelling of spatial point patterns*, volume 70. John Wiley & Sons, 2008.

[17] Francesca Bruno, Peter Guttorp, Paul D Sampson, and Daniela Cocchi. A simple non-separable, non-stationary spatiotemporal model for ozone. *Environmental and ecological statistics*, 16(4):515–529, 2009.

[18] Alexandre Rodrigues and Peter J Diggle. A class of convolution-based models for spatio-temporal processes with non-separable covariance structure. *Scandinavian Journal of Statistics*, 37(4):553–567, 2010.

[19] A. Leroux, L. Xiao, C. Crainiceanu, and W. Checkley. Dynamic prediction in functional concurrent regression with an application to child growth. *Statistics in Medicine*, 37(8):1376–1388, 2017.

[20] A. Catching, S. Capponi, M. T. Yeh, S. Bianco, and R. Andino. Examining the interplay between face mask usage, asymptomatic transmission, and social distancing on the spread of covid-19. *Sci. Rep.*, 11:15998, 2021.

[21] R. Courant, K. Friedrichs, and H. Lewy. Über die partiellen differenzengleichungen der mathematischen physik. *Mathematische Annalen*, 100:32–74, 1928.

[22] L. Peeples. Face masks: what the data say. *Nature*, 586:186–189, 2020.

[23] Y. Cheng, N. Ma, C. Witt, S. Rapp, P. S. Wild, M. O. Andreae, U. Pöschl, and H. Su. Face masks effectively limit the probability of sars-cov-2 transmission. *Science*, 372(6549):1439–1443, 2021.

[24] J. T. Brooks and J. C. Butler. Effectiveness of mask wearing to control community spread of sars-cov-2. *JAMA Insights*, 325(10):998–999, 2021.

[25] J. Crank and P. Nicolson. A practical method for numerical evaluation of solutions of partial differential equations of the heat conduction type. *Proc. Camb. Phil. Soc.*, 43(1):50–67, 1947.

[26] V. Strassen. Gaussian elimination is not optimal. *Numer. Math*, 13(4):354–356, 1969.

[27] D. W. Peaceman and H. H. Rachford Jr. The numerical solution of parabolic and elliptic differential equations. *SIAM J. Appl. Math.*, 3(1):28–41, 1955.