Reaction-diffusion spatial modeling of COVID-19 in Chicago

Trent Gerew*
Department of Applied Mathematics, Illinois Institute of Technology, Chicago, Illinois

January 19, 2022

Abstract

We investigate whether a reaction-diffusion model considering only meanly daily movement is sufficient to describe the spread of the COVID-19 virus in Chicago. The model is calibrated using publicly available health data published by the city. We first study the system of partial differential equations, then derive the basic reproduction number $\mathcal{R}_0$. Then we consider the numerical simulations conducted from March 18 to June 24, 2020. These simulations show that this model may not be sufficient to describe COVID-19 in Chicago. Finally, we discuss shortcomings of the model, and offer some potential solutions.

1 Introduction

Emerging from Wuhan, China in late 2019, the COVID-19 virus rapidly spread throughout the globe [17]. Since the outbreak, governments have been trying to contain the pandemic. Now, with safe and effective vaccines being distributed, it is hoped the pandemic can be overcome [10]. Before the introduction of the vaccines, so-called "non-therapeutic" interventions [16] were frequently deployed. Prominent among these are social distancing, self quarantining, lockdowns, gathering limitations, and the use of personal protective equipment. Now with the Omicron variant on the rise [20], these interventions are once again becoming prominent with some European countries reintroducing lockdowns and travel restrictions [15].

Many mathematical models have been proposed to study the spread of the virus. Most are compartment models, where a population is divided into groups according to the state of individuals. These are generally known as SIR (susceptible-infected-removed) type models. The majority of such models ignore any spatial components. [3] employs a SAIR (asymptomatic) model and includes mobility terms based on the position of cell phones to model the propagation of the disease through Spain. Similarly, [8] incorporates daily movements into an SEIR (exposed) model. In a different approach, [11] considers a statistical model to handle diffusion of COVID-19 in Italy.

The goal of our model is to strike a reasonable balance between representation of the pandemic and the populations involved, and feasibility of data collection and computation.

To do this, we follow [13] and [14] in building a virus model where the spatial propagation is modeled by a diffusion and the reactions are derived from an SAIR model. Specifically, we examine the spread of COVID-19 in the city of Chicago. Does a reaction-diffusion model, which considers only the average daily movements of the population, correctly describe the spread of the virus in Chicago? If it does, we can use the model to assess possible vaccination strategies or reopening scenarios.

A successful model can also be used to examine how social and economic disparities play a role in the transmission of COVID-19. One can merge race, income, or health-care accessibility data from the Chicago area with the results from the model to ask and hopefully answer questions like

*tgerew@hawk.iit.edu

the following: Does community structure play a role? How effective are vaccines? (Or what is thwarting them?) Are care facilities properly allocated?

The paper is organized as follows. Section 2 outlines the methodology, including data collection, modeling assumptions, and parameter estimates. Section 3 contains a brief analysis of the model and the results. In Section 4, we offer a discussion of the shortcomings of the model, and how they can be improved.

2 Materials and methods

2.1 Confirmed and death data

In this study, we used the publicly available data sets of COVID-19 metrics provided by the City of Chicago Data Portal. [7] includes daily counts of city-wide confirmed infected cases, hospitalizations, and deaths. Weekly cases, tests and deaths by ZIP code are logged in [6].

2.2 Mathematical model

We focus our study on four components of the epidemic flow (Figure 1). That is, the populations of Susceptible individuals (S), Asymptomatic infected individuals (A), symptomatic Infected individuals (I), and Removed individuals (R).

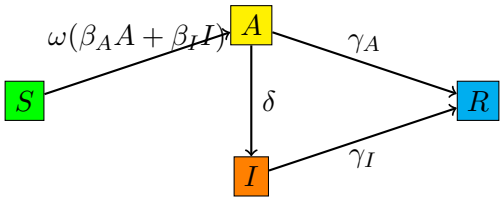


Figure 1: Compartmental representation of the SAIR model.

Our model is known as the SAIR model [3], which substitutes the E compartment of the SEIR model by the A compartment. This model is relevant when there are many undetected asymptomatic infectious individuals, which is known to be the case for COVID-19.

A few notes motivating this very simple model are necessary.

- We do not consider the "exposed" E group in this work. Because the members of A have separate infection and recovery rates, and because there is a possibility they are detected and move to I, we consider E to be merged with A.
- We do not distinguish between quarantined, incarcerated, hospitalized, or nursing home populations. Nursing homes are now known to be epicenters of the virus, however the true case rates are notoriously difficult to track [12]. Additionally, these four population groups can be assumed to be stationary in nature, and in the case of quarantined and hospitalized, isolated. Therefore we ignore their individual contributions.
- In general, the way in which data has been collected and is provided by authorities has varied over time, making its usage rather difficult[1] [18], [2], [4]. There is also the question of case counts being influenced by access to testing, especially early in the pandemic. Including additional compartments would over-parameterize the model making it more difficult to verify, and would not provide useful information.

[1] For an in-depth analysis of the data collection problem, see [19].

- Lastly, adding additional compartments and parameters reduces the computational feasibility of the optimization problem.

To build the mathematical model, we followed the standard strategy developed in the literature concerning SIR models [9]. We assume that individuals in S can be infected by both members of A and I. We suppose that the individuals in the A and I compartments may have different contact rates β_A and β_I, and different recovery rates γ_A and γ_I. Furthermore, we consider a rate δ at which individuals in A may develop symptoms or are otherwise detected and so will move to the I compartment. Lastly, we assume that only members of S and A are mobile.

The average number of contacts ω is described in Equation 2. The diffusion coefficient D is assumed to be the same for both A and I, and is defined by Equation 4.

The dynamics is governed by a system of two partial differential equations (PDE) and two ordinary differential equations (ODE) as follows, for $\mathbf{x} = (x, y) \in \Omega \subset \mathbf{R}^2, t > 0$,

$$\begin{aligned} S_t - D(t)\Delta S &= -\omega(t)\left(\beta_A A + \beta_I I\right) S, \\ A_t - D(t)\Delta A &= \omega(t)\left(\beta_A A + \beta_I I\right) S - (\gamma_A - \delta)A, \\ I_t &= -\gamma_I I + \delta A, \\ R_t &= \gamma_A A + \gamma_I. \end{aligned} \tag{1}$$

Since travel into the city of Chicago was heavily restricted for the early stages of the pandemic, the homogeneous Neumann boundary condition is imposed [14]. The population compartments are considered fractions, such that $S + A + I + R = 1$.

2.3 Parameter estimation

To account for the lockdown, the average number of of contacts is updated as in [13]

$$\omega(t) = \omega_0 \left[\eta + (1 - \eta)\frac{1 - \tanh\left[2(t - t_q)\right]}{2}\right]. \tag{2}$$

Here, $t_q = (t_{\text{eol}} + t_{\text{bol}})/5$, where bol denotes the beginning of the lockdown and eol denotes the end of the lockdown. The parameter $0 \leq \eta \leq 1$ is a varying coefficient translating respect for social distancing and other preventative measures. Note that ω_0 is the average number of contacts before any intervention, and is a constant.

The parameters ω_0, β_A, and β_I are not independently identifiable, so the optimization problem reduces to five parameters $\theta = (\omega_0\beta_A, \omega_0\beta_I, \delta, \gamma_A, \gamma_I)$. Given the observations $I_{\text{obs}}(t_i)$ over n days, we have the minimization problem

$$\begin{aligned} \min_{\theta} \quad & \sum_{i=1}^{n} \left[I_{\text{obs}}(t_i) - I(t_i)\right]^2 \\ \text{s.t.} \quad & \mathbf{0} \leq \theta \leq \mathbf{1} \end{aligned} \tag{3}$$

$I(t_i)$ denotes the output of the mathematical model at time t_i computed with parameters θ.

The optimization problem is solved using the MATLAB nonlinear optimization function `fmincon`. The initial parameter guesses to seed `fmincon` were randomly sampled from a uniform distribution over 1000 iterations. The median of each resulting parameter was selected. Figure 3 shows the estimated parameters.

The diffusion coefficient is also assumed to be altered by the lockdown, and is updated as a simple step function

$$D(t) = D_0\left[1 - (1 - \eta)H(t - t_{\text{bol}})\right] \tag{4}$$

where $H(t)$ is the Heaviside step function. The average one-way commute in Chicago is about 5 miles [1]. Thus we set the diffusion coefficient D_0 to the fixed value of $5/0.72$ to convert to the spacing used for the discretization.

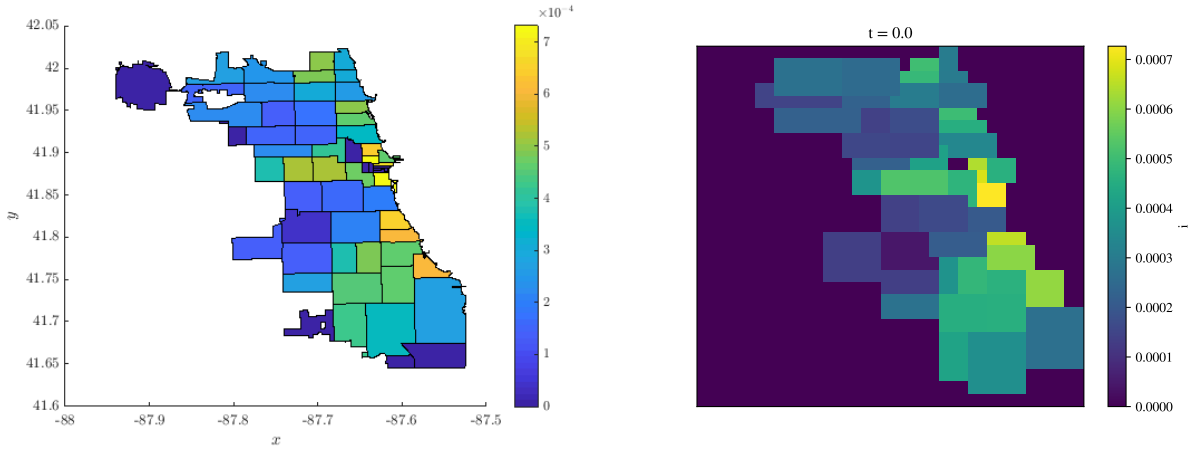
2.4 Numerical discretization

From the map of the city, the computational domain Ω is defined as the minimum square enclosing the city. The city is contained by $\Omega' = \{(X, Y) \mid X \in [-87.9397, -87.5245],\ Y \in [41.6447, 42.023]\}$ where X represents degrees latitude and Y represents degrees longitude. Note that degrees latitude and longitude are not equal when converted to miles. Therefore, we define the computational domain such that the grid spacing is approximately 0.72 miles in both axes. This gives the domain

$$\Omega = \left\{\mathbf{x} \in \mathbf{R}^2 \mid x \in [0, 40],\ y \in [0, 29]\right\}.$$

The initial spatial distribution of the infected population is determined by sampling the ZIP code data of confirmed cases [6] at the grid points, as shown in Figure 2.

The model is solved via finite differences, using the `scikit-fdiff` Python module [5]. It is important to note that due to the limitations of the solver, the simulation had to be broken into time blocks where the values of $\omega(t)\beta_A$, $\omega(t)\beta_I$, and $D(t)$ are evaluated for the initial time in each block and held constant.



(a) ZIP data.

(b) Seeded model.

Figure 2: Initial seeding of the model with data from March 18, 2020.

3 Results

3.1 Existence of solutions and basic reproduction numbers

We show that we are justified in searching for suitable parameters to solve the model. Let $\mathbf{x} = (S, A, I, R)^\intercal$ and $\mathbf{x}_0 = (S_0, A_0, I_0, R_0)^\intercal$. First we show that the solution of the initial value problem (1) exists for the case with no diffusion.

Theorem 1. *Let $\mathbf{0} \leq \mathbf{x}_0 \leq \mathbf{1}$ be the initial datum. Then there exists a unique in time solution $\mathbf{x}$ of the initial value problem (1) without diffusion over $C \subset U \subset \mathbf{R}^4 \times \mathbf{R}^1$ where C is a compact set that contains $(\mathbf{x}_0, t_0)$. Moreover, the solution is $\mathbf{C}^1$.*

Proof. Since $\mathbf{x}_t = f(\mathbf{x}, t)$ is $\mathbf{C}^1$ on the open set $U \subset \mathbf{R}^4 \times \mathbf{R}^1$, it follows that there exists a solution of (1) without diffusion through the point $\mathbf{x}_0$ at $t = t_0$ for $|t - t_0|$ sufficiently small. Moreover, $\mathbf{x}(t, t_0, \mathbf{x_0})$ is a $\mathbf{C}^1$ function [22]. Since C is a compact set containing $(\mathbf{x}_0, t_0)$, then the solution $\mathbf{x}(t, t_0, \mathbf{x}_0)$ can be uniquely extended backward and forward in t up to the boundary of C [22]. □

We do not provide a proof for the existence of solutions to the initial boundary value problem (1) with diffusion terms, but we assume it exists.

Conjecture 1. *Let* $\mathbf{0} \leq \mathbf{x}_0 \leq \mathbf{1}$ *the initial datum. Then there exists a unique global in time solution* $\mathbf{x}$ *of the initial boundary value problem (1).*

A proof for this conjecture will be similar to the one given in [14].

An important parameter in understanding the initial growth of the virus is the basic reproduction number.

Definition 1 (Basic Reproduction Number)**.** *The basic reproduction number* $\mathcal{R}_0$ *can be computed using the next generation of the model without diffusion. Since the infected individuals are in* A *and* I*, new infections* ($\mathcal{F}$) *and transitions between compartments* ($\mathcal{V}$) *can be rewritten as*

$$\mathcal{F} = \begin{pmatrix} \omega(\beta_A A + \beta_I I)S \\ 0 \end{pmatrix}, \qquad \mathcal{V} = \begin{pmatrix} (\gamma_A + \delta)A \\ \gamma_I - \delta A \end{pmatrix}.$$

Then F *and* V *are the Jacobians of* $\mathcal{F}$ *and* $\mathcal{V}$ *respectively evaluated at the disease free equilibrium. Thus,* $\mathcal{R}_0 = \rho(\mathsf{F}\mathsf{V}^{-1})$ *of the next generation matrix*

$$\mathsf{F}\mathsf{V}^{-1} = \begin{pmatrix} \frac{S_0\omega_0\beta_A}{\gamma_A+\delta} + \frac{S_0\omega_0\beta_I\delta}{\gamma_I(\gamma_A+\delta)} & \frac{S_0\omega_0\beta_I}{\gamma_I} \\ 0 & 0 \end{pmatrix}.$$

So,

$$\mathcal{R}_0 = \frac{S_0\omega_0}{\gamma_A + \delta}\left(\beta_A + \beta_I\frac{\delta}{\gamma_I}\right).$$

This number is dimensionless and has an epidemiological meaning. The first term represents the transmission rate by asymptomatic individuals, and the second term represents the transmission rate by symptomatic individuals.

3.2 Model resolution

Calibration of the model is done from March 18, 2020 to June 24, 2020. This range corresponds approximately to the first "wave" of cases seen in Chicago. The lockdown time points match exactly to the imposed lockdown of the city. That is, t_{bol} = March 21, 2020 and t_{eol} = May 29, 2020. In Figure 3, the table shows the estimated parameters. Figure 4 compares the data and the fitted symptomatic infected populations.

$\omega_0\beta_A$	transmission rate due to A	0.1969	days^{-1}
$\omega_0\beta_I$	transmission rate due to I	0.3061	days^{-1}
η	lockdown scale factor	0.5514	1
γ_A	removal rate of A	0.3632	days^{-1}
γ_I	removal rate of I	0.1385	days^{-1}
δ	symptom onset rate	0.0939	days^{-1}

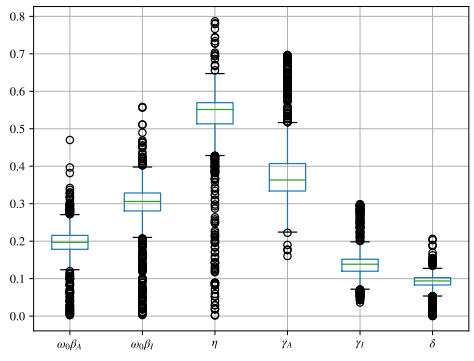


Figure 3: Parameters calibrated according to data from the Chicago Data Portal. On the right is a boxplot of the parameter distribution from 1000 optimization iterations.

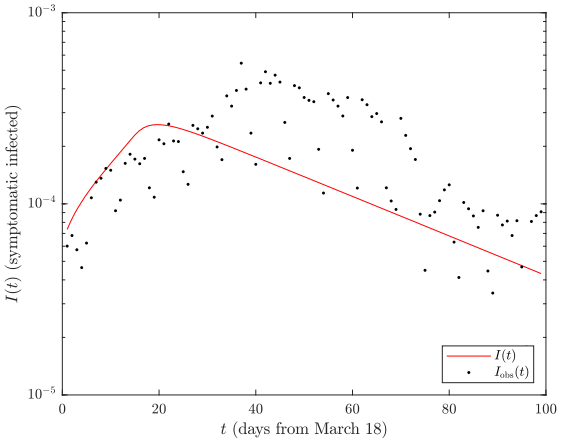


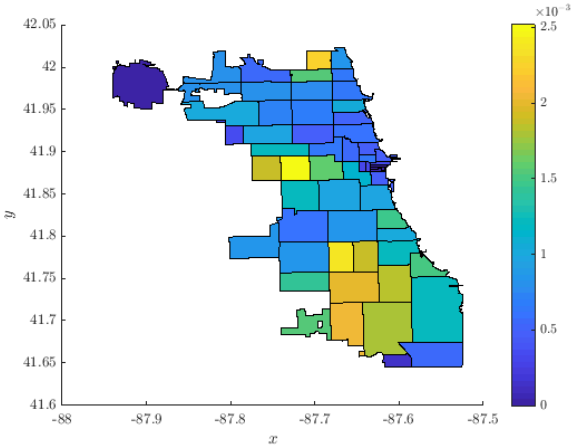
Figure 4: Fitting symptomatic infected by the median value. Note the logarithmic scale on the y-axis.

3.3 Spatial spread of COVID-19

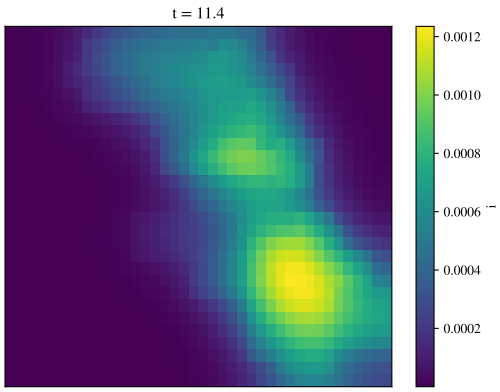
Simulations are performed from March 18, 2020 to June 24, 2020. The time step is $\Delta t = 0.2$ [days], chosen to satisfy the convergence condition [21].

Figure 5 presents three of the days from the simulation time: the effective lockdown t_q day, the last day of simulation, and arbitrary day from the period in between. The observed data is shown on the left, while the model results are on the right. Comparing the model results to the data, we see that the model's diffusion mostly misses the diseases west-ward movement between days 15 and 35. Additionally, the model shows significantly fewer cases than are seen in the data (note the difference in the scale of the colorbars).

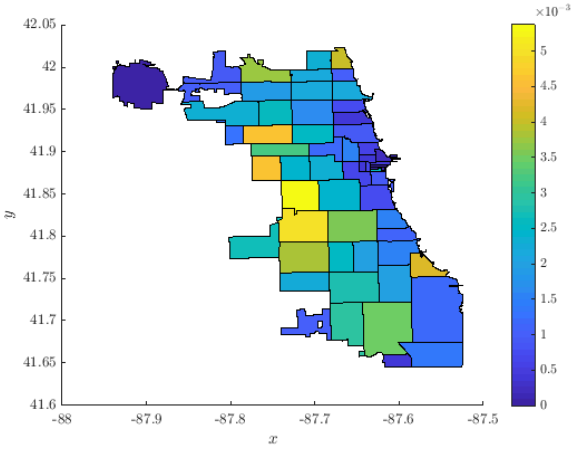
In short, this model does not produce a reasonable reproduction of the spatial spread of COVID-19.



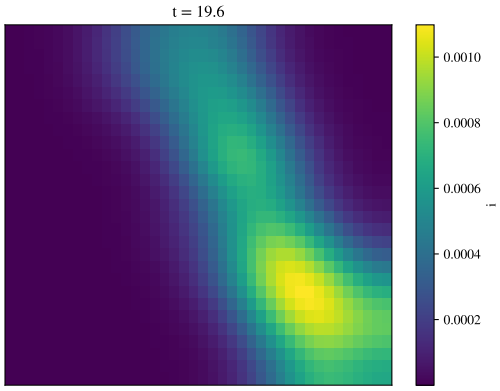
(a) Observed Day 15 (t_q): April 2, 2020



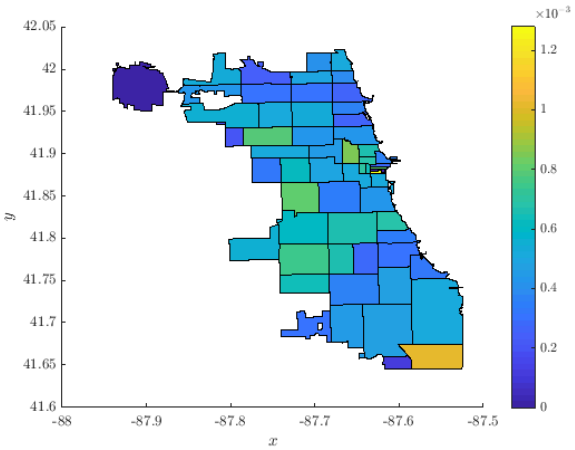
(b) Model Day 15 (t_q): April 2, 2020



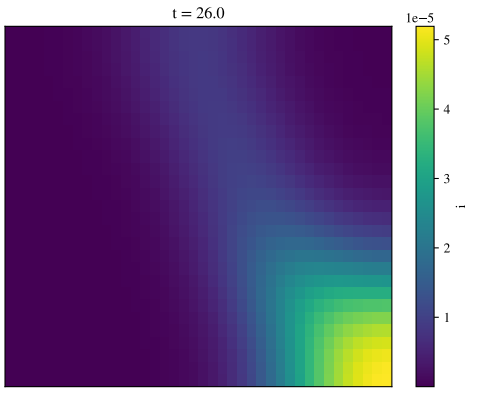
(c) Observed Day 35: April 22, 2020



(d) Model Day 35: April 22, 2020



(e) Observed Day 99: June 24, 2020



(f) Model Day 99: June 24, 2020

Figure 5: Comparison of observed infected cases and the model results.

4 Discussion

As mentioned in Section 3.3, the proposed model does not reproduce the spatial propagation of the virus. In this section, we address possible causes of these inaccuracies and discuss proposals for

improving the results.

The selected model populations may not, contrary to the assumption, be sufficient to describe the virus dynamics. Figure 4 suggests this conclusion, since the growth of the model population only roughly describes the observed data. Note that after Day 20, the model and data begin to diverge substantially. This seems to correspond to the discrepancy in the number of cases seen in the spatial results in Figure 5. Likely, the Exposed compartment describing the latent period is necessary, as in [14]. It is possible that the assumption that Deceased and Recovered populations can be modeled by the same population compartment is errant. On the other hand, in either case these populations have no effect on the act of transmission, so the model dynamics would likely be similar for both cases. A Deceased population may be useful for other reasons, as discussed below.

The boundaries imposed on the computational domain (Equation 2.4) are likely not sufficient to force the diffusion process to replicate the true virus propagation. Figure 6 shows the S and A populations at the same time points. There is a notable amount of diffusion of the S population out of the boundaries of the city, including into Lake Michigan. Since the S population diffuses into areas not reachable by the I population, the reaction rates as described in Equation 1 are small. This may also explain the differences in case numbers seen in Figure 5. It seems a tightening of the bounds of the computational domain is in order. Ideally, we would define the computational domain to be exactly the boundaries of the city. In practice, however, this is infeasible due to the irregularities of the shape. A better alternative is to define a polygon that approximates the shape of the city to be the computational domain. In any case, a new solver will be needed as the `scikit-fdiff` module only allows rectangular domains.

The optimization method can be improved in two ways. First, by adding a Deceased population we can modify the objective function (Equation 3) to compare both cases and deaths, as in [13]. COVID-19 death data is readily available, though requires caution to work with due to reporting inconsistencies. Second, we can perform the optimization directly on the spatial model, as in [14], rather than relying on parameters from a temporal-only model.

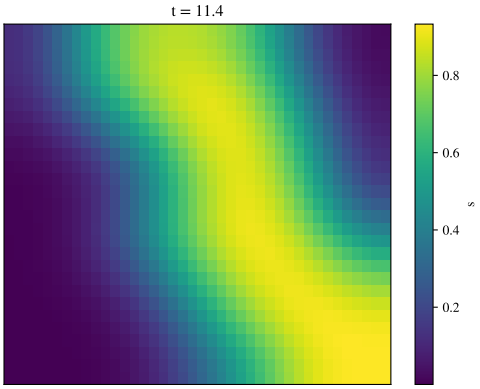
Appendix

Nomenclature (Units are indicated in brackets.)

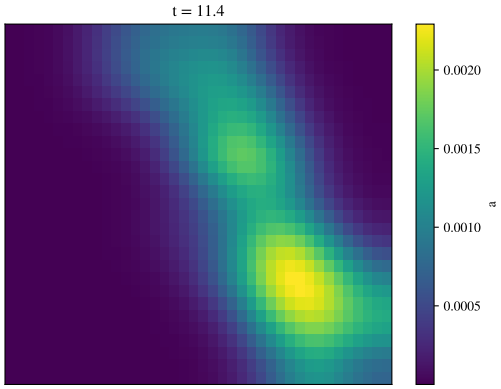
Latin symbols		
A	Asymptomatic infected individuals	[1]
D	Diffusivity	
$\mathcal{F}$	Rate of appearance of new infections	[days^{-1}]
F	Jacobian of $\mathcal{F}$ evaluated at the disease-free equilibrium	N/A
I	Symptomatic infected individuals	[1]
R	Removed individuals	[1]
$\mathcal{R}_0$	Basic reproduction number	[1]
S	Susceptible individuals	[1]
$\mathcal{V}$	Rate of transfer of individuals	[days^{-1}]
V	Jacobian of $\mathcal{V}$ evaluated at the disease-free equilibrium	N/A
Greek symbols		
β_j	Contact rate of compartment j	[days^{-1}]
γ_j	Recovery rate of compartment j	[days^{-1}]
Δ	Laplacian operator	N/A
δ	Rate of asymptomatic individuals that may develop symptoms	[days^{-1}]
η	Lockdown scale factor	[1]
θ	Parameter vector	N/A
ρ	Spectral radius	N/A

Ω	Computational domain	N/A
ω	Contacts	[1]

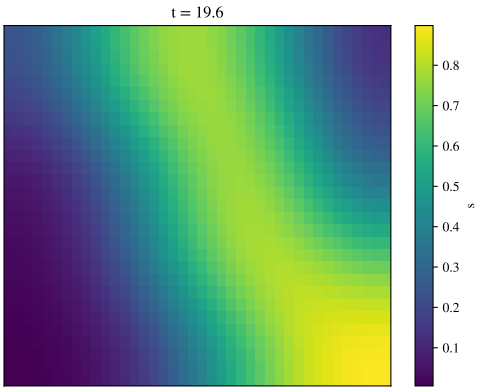
Model Populations



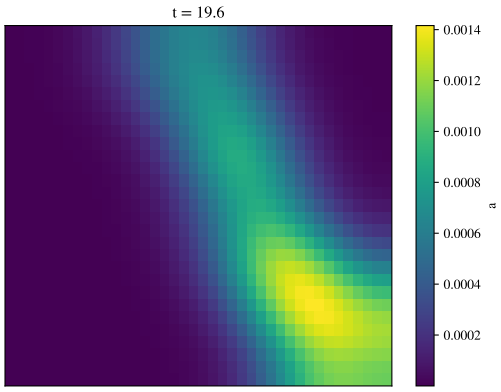
(a) Susceptible Day 15 (t_q): April 2, 2020



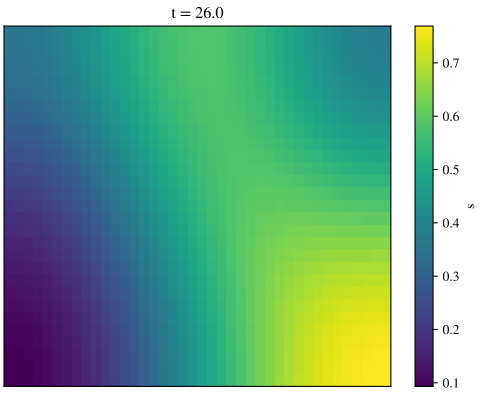
(b) Asymptomatic Day 15 (t_q): April 2, 2020



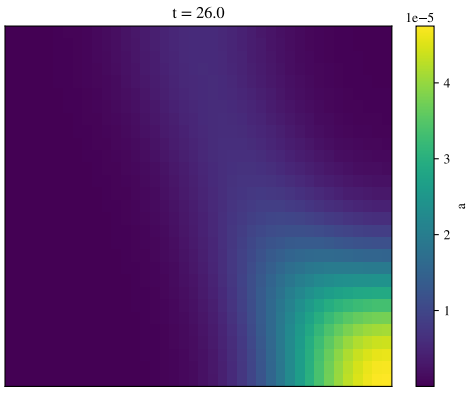
(c) Susceptible Day 35: April 22, 2020



(d) Asymptomatic Day 35: April 22, 2020



(e) Susceptible Day 99: June 24, 2020



(f) Asymptomatic Day 99: June 24, 2020

Figure 6: Model results for susceptible and asymptomatic populations.

References

[1] Travel trends: Understanding how our region moves. *Chicago Metropolitan Agency for Planning*, September 2016.

[2] J. Achenbach and Y. Abutaleb. Messy, incomplete u.s. data hobbles pandemic response. *Washington Post*, September 2021.

[3] F. Aràndiga, A. Baeza, I. Cordero-Carrión, R. Donat, M. C. Martí, P. Mulet, and D. F. Yáñez. A spatial-temporal model for the evolution of the covid-19 pandemic in spain including mobility. *Mathematics*, 8(10), 2020.

[4] R. Badker, K. Miller, and C. Pardee. Challenges in reported covid-19 data: Best practices and recommendations for future epidemics. *BMJ Global Heath*, 2021.

[5] N. Cellier and C. Ruyer-Quil. scikit-finite-diff, a new tool for pde solving, June 2019.

[6] City of Chicago. Covid-19 cases, tests, and deaths by zip code, 2021. Data retrieved from Chicago Data Portal, `https://data.cityofchicago.org/Health-Human-Services/COVID-19-Cases-Tests-and-Deaths-by-ZIP-Code/yhhz-zm2v`.

[7] City of Chicago. Daily chicago covid-19 cases, deaths, and hospitalizations, 2021. Data retrieved from Chicago Data Portal, `https://data.cityofchicago.org/Health-Human-Services/Daily-Chicago-COVID-19-Cases-Deaths-and-Hospitaliz/kxzd-kd6a`.

[8] L. Danon, E. Brooks-Pollock, M. Bailey, and M. Keeling. A spatial model of covid-19 transmission in england and wales: early spread and peak timing. *medRxiv*, 2020.

[9] L. Edelstein-Keshet. *Mathematical Models in Biology*. SIAM, 2005.

[10] C. for Disease Control and Prevention. Effectiveness. November 2021. `https://www.cdc.gov/coronavirus/2019-ncov/vaccines/effectiveness/index.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fvaccines%2Feffectiveness.html`.

[11] D. Giuliani, M. M. Dickson, G. Espa, and F. Santi. Modelling and predicting the spatio-temporal spread of coronavirus disease 2019 (covid-19) in italy, 2020.

[12] C. A. Grimm. Cms's covid-19 data included required imformation from the vast majority of nursing homes, but cms could take actions to improve completeness and accuracy of the data. *Department of Health and Human Services*, September 2021.

[13] P. Kevrekidis, J. Cuevas-Maraver, Y. Drossinos, Z. Rapti, and G. Kevrekidis. Spatial modeling of covid-19: Greece and andalusia as case examples. *Physical Review E*, April 2021.

[14] Y. Mammeri. A reaction-diffusion system to better comprehend the unlockdown: Application of seir-type model with diffusion to the spatial spread of covid-19 in france. *Computational and Mathematical Biophysics*, 8(1):102–113, 2020.

[15] D. Meyer. Lockdowns spread across europe in the face of omicron's 'lightning' advance. *Fortune*, December 2021. `https://fortune.com/2021/12/20/omicron-lockdown-netherlands-covid-europe-france-germany-italy-ireland/`.

[16] W. H. Organization. *Emerging Infectious Diseases*, 12(88), 2006.

[17] W. H. Organization. Naming the coronavirus disease (covid-19) and the virus that causes it. 2020.

[18] K. Schechtman and S. Simon. America's entire understanding of the pandemic was shaped by messy data. *The Atlantic*, May 2021.

[19] S. Simon. Covid-19 data in historical perspective: Lessons from chicago's 1995 heat wave. *Socially Responsible, Computation, and Design*, 2(1), 2021.

[20] M. Stobbe. Omicron sweeps across nation, now 73 *Associated Press*, December 2021. `https://apnews.com/article/omicron-majority-us-cases-833001ef99862bd6ac17935f65c896cf`.

[21] W. A. Strauss. *Partial Differential Equations: An Introduction*. Wiley, 2008.

[22] S. Wiggins. *Introduction to Applied Nonlinear Dynamical Systems and Chaos*. Springer, 2003.