Reaction-diffusion spatial modeling of COVID-19 in Chicago

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

December 19, 2021

Abstract

content...

1 Introduction

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. [8] includes daily counts of city-wide confirmed infected cases, hospitalizations, and deaths. Weekly cases, tests and deaths by ZIP code are logged in [7].

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). Our model is known as the SAIR model [3], which

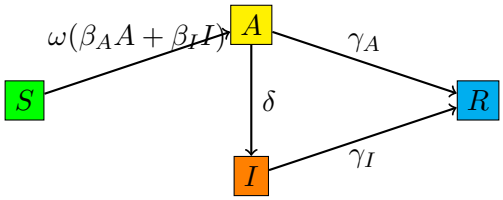


Figure 1: Compartmental representation of the SAIR model.

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.

*tgerew@hawk.iit.edu

- We do not distinguish between quarantined, 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 [10]. Additionally, these three 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] [13], [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.

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

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

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

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

Since travel into the city of Chicago was heavily restricted for the early stages of the pandemic, the homogeneous Neumann boundary condition is imposed [12]. 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 [11]

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

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, for n days, the observations $I_{\text{obs}}(t_i)$,

[1] For an in-depth analysis of the data collection problem, see [14].

the objective function is the least square function

$$J(\theta) = \sum_{i=1}^{n} [I_{\text{obs}}(t_i) - I(t_i)]^2 \tag{3}$$

with constraints $0 \leq \theta \leq 1$. $I(t_i)$ 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

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

where $H(t)$ is the Heaviside step function. The average one-way commute in Chicago is about 5 miles [1]. Thus, we set D_0 to the fixed value of $(5/0.72)^2/4\Delta t$ where Δt is the time step [6].

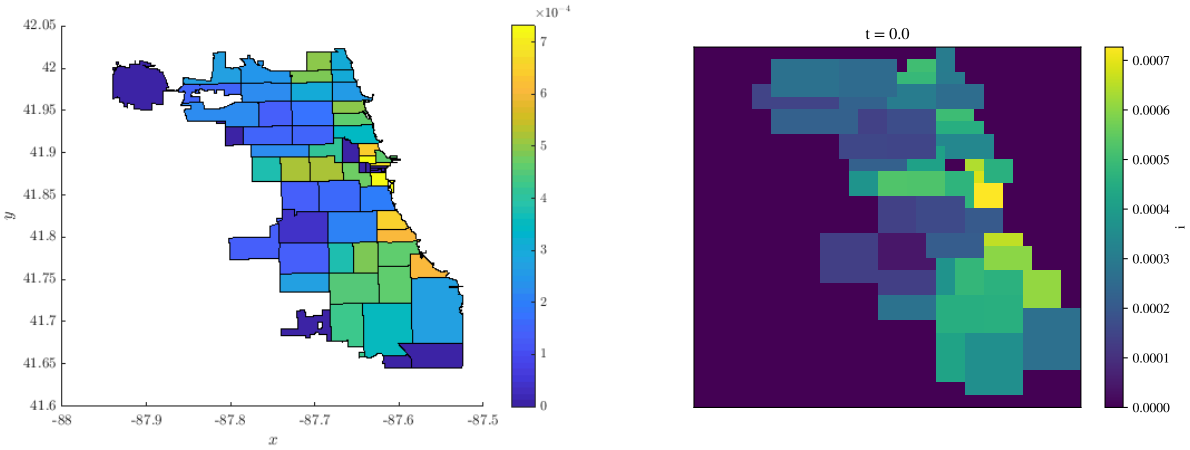
2.4 Numerical discretization

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

$$\Omega = \left\{\mathbf{x} \in \mathbf{R}^2 \mid x \in [0, 40],\ y \in [0, 29]\right\}. \tag{5}$$

The initial spatial distribution of the infected population is determined by sampling the ZIP code data of confirmed cases [7] at the grid points, as shown in Figure 2.

The model is solved via finite differences, using the `scikit-fdiff` Python module [5].



(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 $\mathbf{x} = (S, A, I, R)$ and $\mathbf{x}_0 = (S_0, A_0, I_0, R_0)$. First we show that the solution of the initial value problem (1) exists for the case with no diffusion.

Theorem 1. *Let $0 \leq \mathbf{x}_0 \leq N$ 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 [15]. 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 [15]. □

We do 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 $0 \leq \mathbf{x}_0 \leq N$ 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 [12].

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*

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

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,

$$\mathcal{R}_0 = \frac{S_0\omega_0}{\gamma_A + \delta}\left(\beta_A + \beta_I \frac{\delta}{\gamma_I}\right). \tag{6}$$

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
δ	symptom onset rate	0.0939	days^{-1}
γ_A	removal rate of A	0.3632	days^{-1}
γ_I	removal rate of I	0.1385	days^{-1}

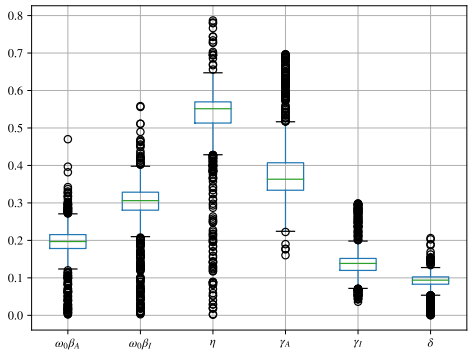


Figure 3: Parameters calibrated according to data from the Chicago Data Portal. On the right is a boxplot of the parameter distribution from 1000 optimization iterations.

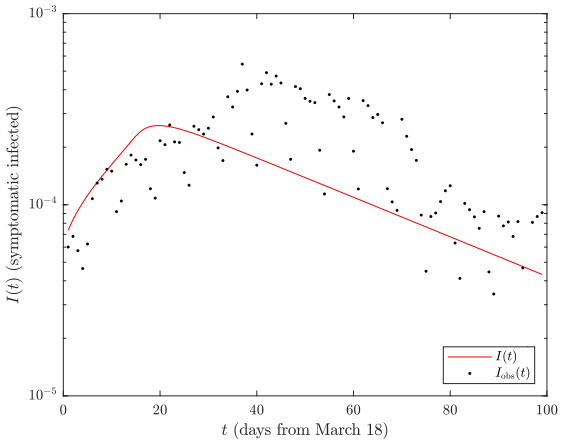


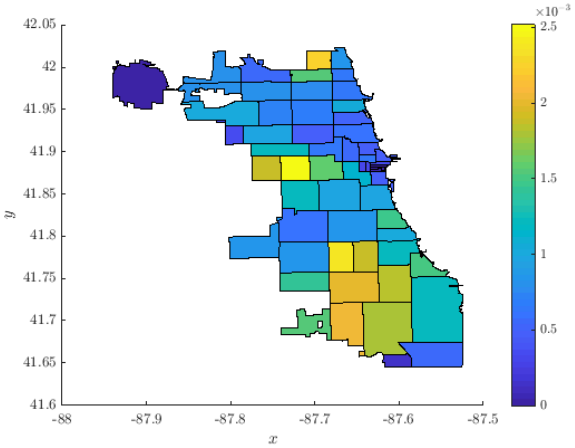
Figure 4: Fitting symptomatic infected by the median value. Note the logarithmic scale on the y-axis.

3.3 Spatial spread of COVID-19

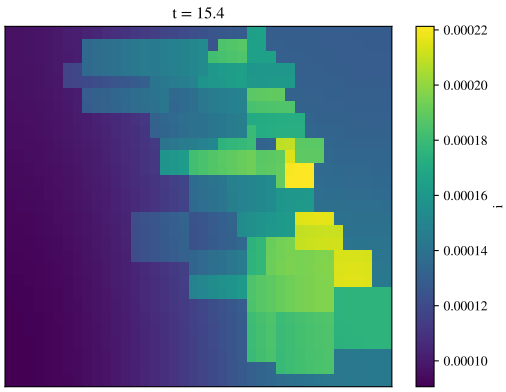
Simulations are performed from March 18, 2020 to June 24, 2020. The time step is $\Delta t = 0.2$ [days].

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. The data shows that the main progression of the disease is from the lake, to the southwest, then to center-west, then slowly decreases through most of the city. In contrast, the model infected population is an order of magnitude different than the true population after only 15 days, and the difference increases as the simulation continues. Furthermore, the hoped-for diffusion behavior is not present in the model results.

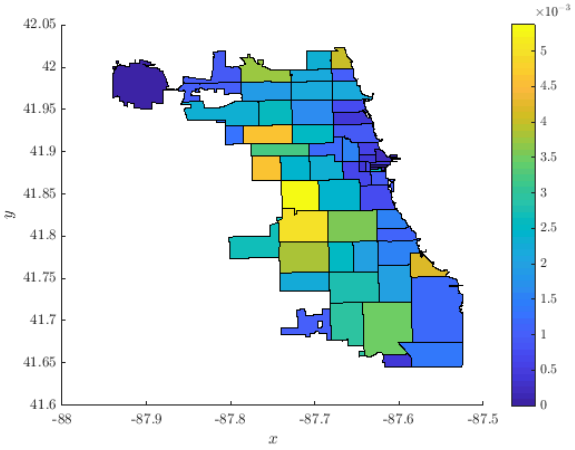
In short, the proposed model does *not* reproduce the spatial propagation of the virus.



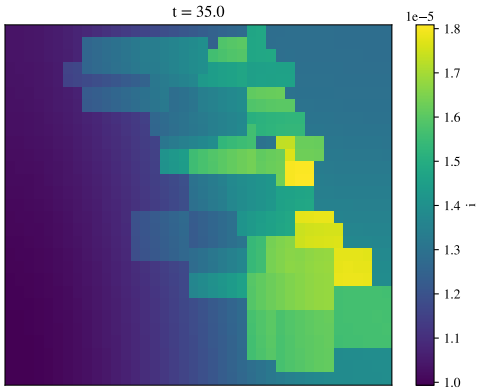
(a) Observed Day 15 (t_q): April 2, 2020



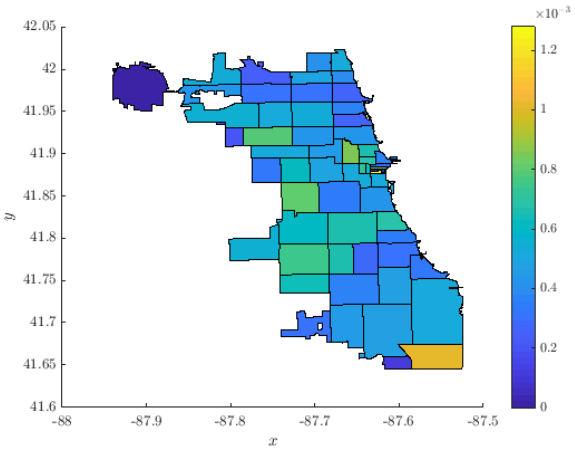
(b) Model Day 15 (t_q): April 2, 2020



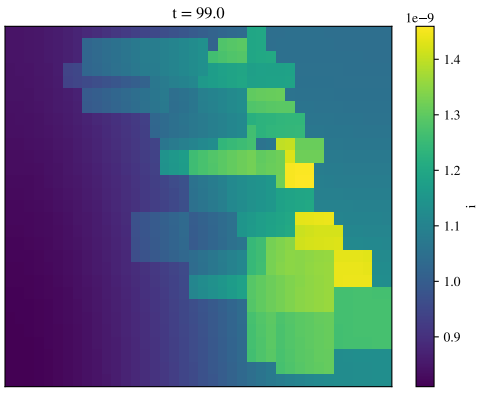
(c) Observed Day 35: April 22, 2020



(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 §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.

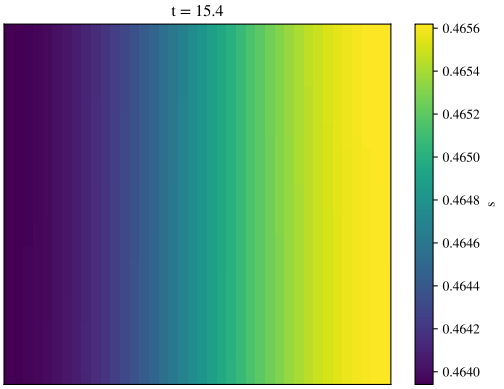
The selected model populations may not, contrary to the assumption, be sufficient to describe the virus dynamics. Figure 4 suggests this conclusion, since the growth of the model population only roughly describes the observed data. Likely, the Exposed compartment describing the latent period is necessary, as in [12]. 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.

Appendix

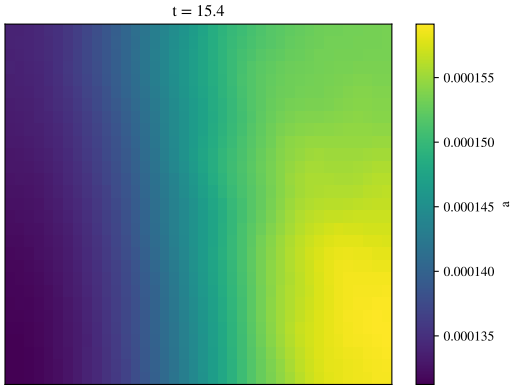
Nomenclature (Units are indicated in brackets.)

Latin symbols		
A	Asymptomatic infected individuals	[1]
D	Diffusivity	
I	Symptomatic infected individuals	[1]
R	Removed individuals	[1]
S	Susceptible individuals	[1]
Greek symbols		
β_j	Contact rate of compartment j	[days^{-1}]
γ_j	Recovery rate of compartment j	[days^{-1}]
δ	Rate of asymptomatic individuals that may develop symptoms	[days^{-1}]
ω	Contacts	[1]

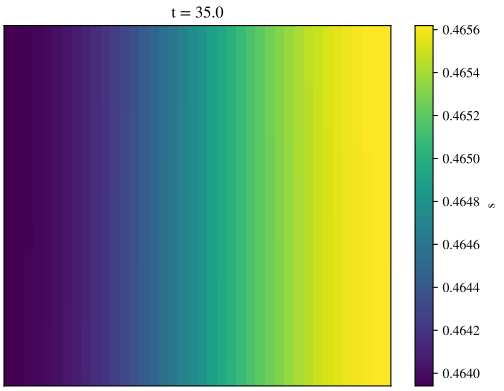
Model Populations



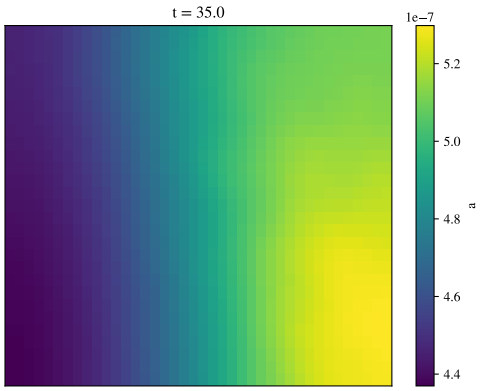
(a) Susceptible Day 15 (t_q): April 2, 2020



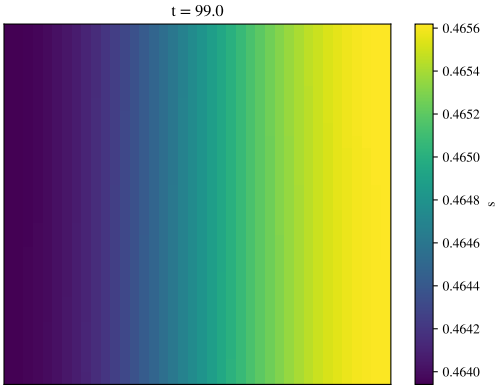
(b) Asymptomatic Day 15 (t_q): April 2, 2020



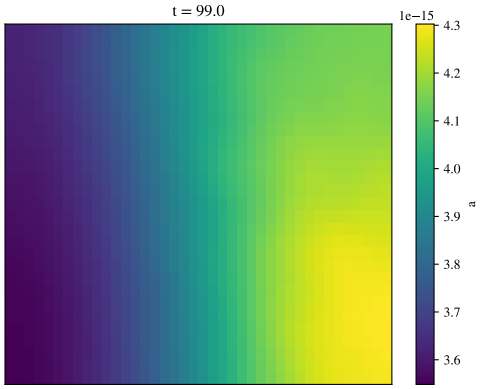
(c) Susceptible Day 35: April 22, 2020



(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] S. C. Chapra and R. P. Canale. *Numerical Methods for Engineers*. McGraw-Hill, 2006.

[7] 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`.

[8] 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`.

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

[10] 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.

[11] 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.

[12] 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.

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

[14] S. Simon. tba. *Socially Responsible, Computation, and Design*, 2(1), 2021.

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