

# The impact of seasonality on the time and size of peak SARS-CoV-2 infections\*

Physics Department, University of California, Santa Barbara

(Dated: May 11, 2020)

In order to guide future policy-making decisions regarding the relaxation of quarantine measures established in response to the COVID-19 pandemic, newer models which account for long-term factors in virulent transmission are necessary. Seasonal pressure is investigated as a factor in the arrival of a second wave of infections in a stochastic implementation of the classic SIR model, using a seasonal forcing amplitude of  $0.4\langle R_0 \rangle$ . The maximum number of infected individuals was found to relate sinusoidally to the time delay between peak infection and the maximum value of  $R_0$ , fitted with  $\chi^2_\nu = 1.657$ , indicating an accurate fit. The greatest peak infection of 3,220/10,000 individuals was predicted to correspond to a peak  $R_0$  time of  $200 \pm 20$  days after the first case. The time delay between the peak maximum infected and the maximum  $R_0$  value was found to be directly proportional to time of peak  $R_0$  measured in days after the first case, with an  $R^2 = 0.994$ . The slope  $\alpha$  was found to be  $0.82 \pm 0.02$ , which is an estimate of the increase in delay between the day of peak infections and the day of  $R_{0,\max}$  for a given increase in the time of peak infection since the start of an epidemic. Such parameters can be easily obtained from real-world data to make prediction of the future dynamics of an epidemic.

## I. BACKGROUND

The novel SARS-CoV-2 has disrupted nearly every major industry over the course of a few months (an impressive feat that most startup companies claim to achieve yet rarely accomplish). One third of the global population is currently under some form of quarantine measures [1], ranging from issuing voluntary guidelines to severe lockdown, yet many nations are finally starting to roll back these restrictions. In most cases, these nations have evidence indicating the peak of total infections has passed. However, a number of projections [2-4] have suggested that a second wave of infections is plausible, if not likely. In order to mitigate the potential for a second wave and the necessity of a second quarantine, modeling that factor in long term properties of SARS-CoV-2 is necessary to guide future policy.

### A. The SIR Model

The starting point for nearly all models of an epidemic is the SIR model, with SIR standing for Susceptible, Infected, and Removed (or Recovered). A closed system of  $N$  people

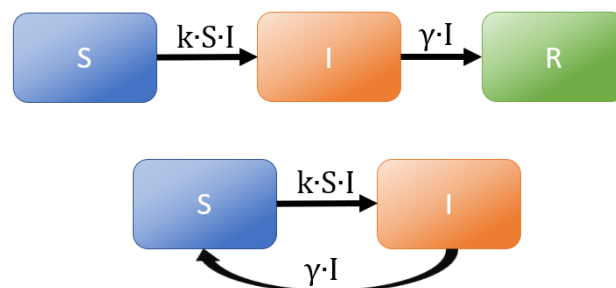
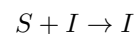


FIG. 1: The SIR model divides the population into the susceptibles (S), infected (I), and removed (R). The SIS model is a variant of SIR that assumes no long-lasting immunity is acquired after infection.

are categorized under one of the three groups, and a set of differential equations are used to relate the three populations' dynamics in time.

These dynamical relationships are derived by treating the populations as "reactants" and "products" as in chemical reactions, and applying the law of mass action. For this model, to produce one infected individual, you need one susceptible individual and one infected. One recovered individual requires only one infected individual. This can be summarized as:



\* Many thanks to Kelly Pawlak for her guidance and time invested towards helping us finish our graduation requirements.

† apeng@ucsb.edu

I'm super impressed you learned this on your own. Indeed, all nanophysics are integer problems and integer problems can be solved as reaction-diffusion problems!

The law of mass action assumes that all reactants are homogeneously dispersed, so each individual from a population has an equal likelihood of interacting with an individual from another population. From this, we estimate the rate of change of a species as directly proportional to the amount of its reactants. Thus, using the above relationships, the SIR model differential equations can be obtained as follows:

$$\frac{dS}{dt} = -k \cdot S(t) \cdot I(t) \quad (1a)$$

$$\frac{dI}{dt} = k \cdot S(t) \cdot I(t) - \gamma \cdot I(t) \quad (1b)$$

$$\frac{dR}{dt} = \gamma I(t) \quad (1c)$$

The constants of proportionality  $k$  and  $\gamma$  can be understood physically as the rates of infection and recovery respectively, both per infected individual per unit time. Their ratio  $\frac{k}{\gamma}$  is called the basic reproduction ratio  $R_0$ , and represents the average number of new cases an infected individual will cause throughout the course of the infection. This parameter is key to understanding the dynamics of an epidemic: the inflection point of  $\frac{dI}{dt}$  occurs at  $R_0 = 1$ . For  $R_0 < 1$ , the rate of infection is decreasing, stifling the spread of the disease, while an  $R_0 > 1$  corresponds to an increasing rate of infection, allowing the epidemic to successfully take hold of the total population. Major influenza pandemics such as the 1918 Spanish Flu and the 2009-10 Swine Flu pandemics have been estimated to have a median value of  $R_0$  of just below 2.0 ( $R_0 = 1.8$  and  $1.46$  respectively). The  $R_0$  for SARS-CoV-2 has been determined to be between 2.0 and 2.5 on average [4], but ranges widely between regions. In major COVID-19 hotspots such as Italy and mainland China,  $R_0$  has been estimated between 2.43 and 3.10 prior to the enforcement of quarantine measures (Fig. 2), highlighting the unprecedented infectiousness of the novel coronavirus.

Notice from equation set 1 that  $\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0$ , which follows from the fact that the total population of people is a constant  $N$ . This model neglects what is called "vital statistics," which are birth and death rates in the population. These are assumed to be much slower than the dynamics of the virus, and are thus able to be ignored. Furthermore, the flow of individuals is in one direction (Fig. 1); removed individuals are assumed to either confer permanent immunity to the disease, or are deceased. In either case, they have zero probability of infecting someone else or getting

| City                | Population | $R_0$     | $r^2$ |
|---------------------|------------|-----------|-------|
| Bergamo             | 122,161    | 2.52      | 0.98  |
| Lodi                | 45,872     | 3.09      | 0.91  |
| Cremona             | 72,680     | 2.93      | 0.94  |
| Brescia             | 198,536    | 2.61      | 0.85  |
| Piacenza            | 103,942    | 2.76      | 0.90  |
| Milano              | 1,389,834  | 2.70      | 0.91  |
| Pavia               | 73,195     | 2.43      | 0.91  |
| Parma               | 197,499    | 2.46      | 0.99  |
| Italy               | 60,483,973 | 3.10      | 0.99  |
| Majumder et al. [4] | –          | 2.00–3.10 | –     |
| Li et al. [5]       | –          | 1.40–3.90 | –     |
| Wu et al. [7]       | –          | 2.47–2.86 | –     |
| Zhao et al. [8]     | –          | 2.24–3.58 | –     |
| Chen et al. [9]     | –          | 3.58      | –     |

FIG. 2: Table adapted from D'Arienzo et al. [5] listing the estimates of  $R_0$  from various regions of Italy. The  $r^2$  for the linear fit to the  $R_0$  values' exponential growth curve is also provided. While there is clear variation between each region, the  $R_0$  values are consistently above 2.4 by the authors' analysis.

reinfected. While the SIR model serves as a reliable starting point, incorporating more details can greatly enhance its accuracy and predictive power. For instance, some diseases confer no long term immunity, thus instead having infected individuals flow into the recovered population, they are put back into the susceptible group, from which they can become re-infected; this variation is named the SIS model. Eqns. 1a, 1b and 1c can be solved deterministically, but lose the inherently probabilistic nature of the course of an infection. A stochastic SIR model will be able to provide uncertainties in its future projections, which is valuable in providing a range of estimates, and are continually updated in the presence of new data.

## B. Markov Chains

One way to implement a stochastic SIR model is using a Markov chain, the strategy used in the following report. A Markov chain generates a series of population counts for a finite number of time steps. The primary assumption in a Markov chain is that the next time step only depends on the state of the previous time step. Mathematically, the probability of the state reaching a state  $\psi$  at time  $t$  can be represented as:

$$\text{Probability}(\psi, t) \equiv p_\psi(t)$$

$$p_\psi(t) = \sum_{\{\phi_i\}} p_{\phi_i}(t - \delta t) \cdot p_{\psi|\phi_i}(t - \delta t)$$

Where  $\{\phi_i\}$  is the set of all possible previous states which

have a nonzero probability of leading to the next state  $\psi$ . The deterministic set Eqn. 1 can be adapted to the Markov model as the so called Master Equations:

$$\begin{aligned} P_{S,I,R}(t) = & p_{S,I,R|S,I,R} \cdot P_{S,I,R}(t - \delta t) + \\ & p_{S,I,R|S+1,I-1,R} \cdot P_{S+1,I-1,R}(t - \delta t) \\ & + p_{S,I,R|S,I+1,R-1} \cdot P_{S,I+1,R-1}(t - \delta t) \end{aligned} \quad (2)$$

with

$$p_{S,I,R|S+1,I-1,R} = p_{\text{infect}} = \frac{k}{N} \delta t (S+1)(I-1) \quad (3a)$$

$$p_{S,I,R|S,I+1,R-1} = p_{\text{recover}} = \gamma \delta t (I+1) \quad (3b)$$

$$p_{S,I,R|S,I,R} = p_{\text{null}} = 1 - \frac{k}{N} \delta t S I - \gamma \delta t I \quad (3c)$$

$p_{\text{infect}}$  is the probability of infecting a new susceptible given the previous state  $\phi_i = \{S(t - \delta t), I(t - \delta t), R(t - \delta t)\} \equiv \{S+1, I-1, R\}$ ,  $p_{\text{recover}}$  is the probability of generating a recovered individual given  $\phi_i = \{S, I+1, R-1\}$ , and  $p_{\text{null}} = 1 - p_{\text{infect}} - p_{\text{recover}}$  is the probability nothing happens given  $\phi_i = \{S, I, R\}$ . Note that the Markov chain also assumes only one person can move between populations per time step, restricting the set  $\{\phi_i\}$  to these three states.

This approach can be easily implemented in code, and is essentially an implementation of a Monte-Carlo simulation.

### C. Seasonality

Seasonal forcing is a complicated phenomena that, as a whole, results in a cyclical infection rate in a host of diseases, including the four most common human coronaviruses [2]. While there is no consensus on the exact cause of seasonal forcing, a number of potential contributors include the density of people indoors during cold seasons, school system scheduling, and the virus' tolerance for different climates [3]. It is still undetermined whether SARS-CoV-2 is also a seasonal virus, however, a number of recent papers [2, 4] have explored such scenarios. In Neher et al., the  $k$  value (or  $\beta$  in their paper) was multiplied by the following time dependence:

$$k(t) = k_0(1 + \varepsilon \cos(2\pi(t - \theta))) \quad (4)$$

with  $k_0$  as the average infection rate,  $\varepsilon$  as the maximum amplitude of seasonal forcing, and  $\theta$  as the point in peak transmissibility (in units of years). A similar approach will be used in our model, with the extension of using a stochastic model.

It has been suggested [6] that there is a potential correlation between seasonal humidity and temperature variations and the onset of influenza season in the United States. Using the stochastic SIR model, we will attempt to illuminate the relationship between these factors.

## II. METHODS

A stochastic SIR model was implemented in Python based on the concept of Markov chains. The probabilities of infection, recovery, and null were calculated in each time step as given by equation set 3. A random fraction `rand_num` was then chosen from a uniform distribution ranging between 0 and 1. If `rand_num` <  $p_{\text{infect}}$ , then the infected population increased by one while the susceptible population decreased by one. If  $p_{\text{infect}} < \text{rand\_num} < p_{\text{infect}} + p_{\text{recover}}$ , then the recovered population increased by one, and the infected population was reduced by one. Otherwise, all populations were kept constant. This procedure was run in a for-loop for `t_max` times, but only stored values that fell on day increments to reduce memory requirements and runtime. `rand_num`,  $p_{\text{infect}}$ , and  $p_{\text{recover}}$  were adapted to an array of `num_trials` values to allow for running `num_trials` trials simultaneously. This vectorized implementation also has drastically reduced runtime compared to a nested for-loop. Unfortunately, as each time step depends on the previous one, the usage of one for-loop was unavoidable.

The value of  $\gamma$  is equivalent to the inverse of the average recovery rate of an infected individual. In Neher et al., a  $\gamma$  value of 0.2 (recovery time 5 days) was used, however, it has been observed [7] that traces of the virus had been detected via reverse transcriptase-polymerase chain reaction tests (RT-PCR) up to fourteen days after being discharged from intensive care due to a lack of symptoms, indicating a slight potential for viral transmission. Thus, a slightly more conservative  $\gamma$  would be to assume an average recovery time of 10 days, or  $\gamma = 0.1$ .

Using the mid-range of  $R_0$  estimates taken from Moore et al.,  $\langle R_0 \rangle = 2.25$  was chosen, with a seasonal amplitude

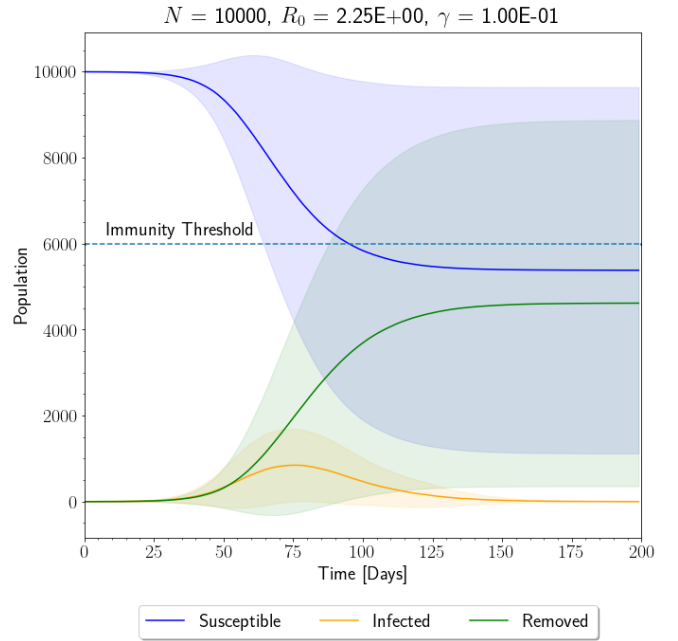
fraction  $\varepsilon = 0.4$ , within the conservative range used in Neher et al. A modest population of 10,000 was used, as computation times were drastically increased as the number of trials were increased. This is likely due to the use of a small time step ( $\delta t = 1$  minute or  $\frac{1}{1440}$  days). However, this time step was chosen as a reasonable rate at which a single person may be changing between S, I, and R among  $N = 10,000$  people.

To verify the chosen parameter ranges were satisfactory, a basic SIR model was implemented and run over fifty trials at  $N = 10,000$  and  $\gamma = 0.1 \text{ days}^{-1}$  (Fig. 3). The probability of infection and recovery were monitored over time to ensure that they or their sum did not exceed 1, which proved to be an issue for a ratio of  $\frac{N}{\gamma} \geq 10^6$  or  $\frac{N}{\gamma} \leq 10^2$  for  $N \geq 10^3$  (Further analysis of this issue is found in Section IV). Drastic improvement of the error bars was demonstrated by only considering the trials where an epidemic had actually taken place, as seen in Fig. 3b. From this observation, all estimated parameters are based on the subset of successful epidemic trials, rather than over all runs, which would consistently underestimate the scope of the epidemic. Additionally, since the COVID-19 pandemic has already successfully taken hold of the population, data from such "null" trials would not provide an accurate representation of areas affected by SARS-CoV-2.

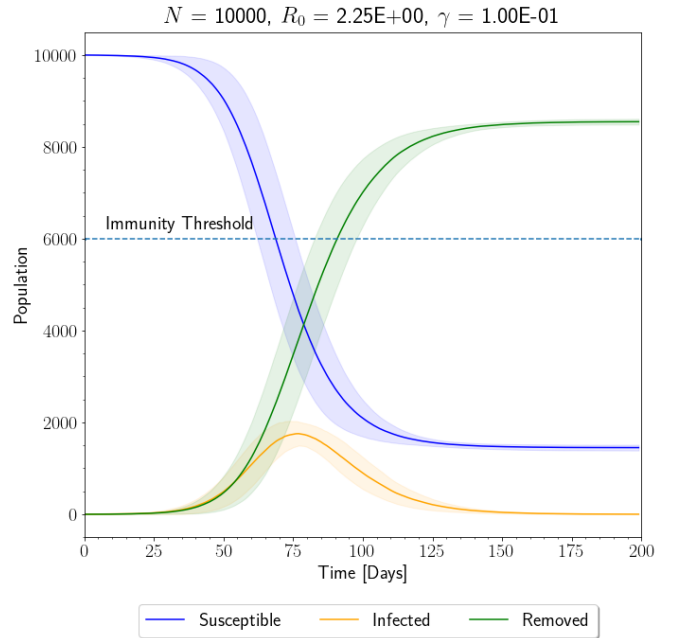
In order to study the relationship between the time of the peak seasonal  $R_0$  and the onset of an epidemic ( $\theta$  in Eqn. 4),  $n_{\text{thetas}} = 12$  different points in the year were chosen to represent  $R_{0,\text{max}}$ , starting from day 15 in the calendar year (Jan. 15th), at evenly spaced intervals. Each value of  $\theta$  (converted to units of days, and shifted from the calendar date to the start of the simulation) was averaged over  $n_{\text{each}} = 30$  trials. For each trial, the mean peak infection population, time of peak infection, and the proportion of immunity at the end of the simulation was recorded. The distance between the time of peak infection and  $\theta$  was calculated as  $\min|t_{I_{\text{max}}} - t_{R_{0,\text{max}}}| \equiv \Delta t$ ; thus, the closest  $R_0$  peak to the the time of maximum infection would be used, whether it was before or after.

### III. DATA ANALYSIS

The thirty trials for each value of  $\theta$  were averaged over in both  $\Delta t$  and  $I_{\text{peak}}$ , the results of which are summarized in



(a) SIR populations plotted over 200 days and averaged over fifty trials. Because the stochastic model includes trials where the virus could not take hold, the  $1 \sigma$  error bars are large, and do encompass the immunity threshold of 60% of the population.



(b) However, removing the cases where no epidemic takes hold clearly reveals that immunity has taken place. The chosen threshold of  $I_{\text{max}}$  for a trial to have experienced an epidemic was 1% of the total population.

FIG. 3: Preliminary SIR trial comparison between fitting all trials and only the successful epidemics, using the "immunity threshold" (the proportion of recovered required to stop the epidemic as stated in Moore et al. [4]) as a benchmark for how helpful the data is.

Table. II and graphically in Fig. 4. This treatment assumes that on average, for a given value of  $\theta$ ,  $\Delta t$ , the time between the peak infection and peak  $R_0$  should be the same, and deviations are purely statistical. Given that we are working with a stochastic model, this is a fairly reasonable estimate.

The data was then fit to a cosine function of the form  $f(\Delta t, a, b, c, d) = a \cos(b \Delta t - c) + d$  using least-squares regression analysis. The uncertainty from each  $\theta$  is given by the equation:

$$\sigma_{I_{\max}} = \sqrt{\frac{\sum_i^n (I_{\max,i} - \bar{I}_{\max})^2}{n - 2}} \quad (5)$$

where  $n = \text{n\_each} = 30$ . Two degrees of freedom are removed from the denominator because they are used to estimate the mean and standard deviation of the sample.

Replacing  $I_{\max}$  with  $\Delta t$  gives the uncertainty in the time between peak infection and max  $R_0$ . From there, the weighted reduced chi square value can be calculated as follows:

$$\chi_\nu^2 = \frac{1}{\nu} \sum_i^m \frac{(I_{\max,i} - f(\theta_i, \bar{a}, \bar{b}, \bar{c}, \bar{d}))^2}{\sigma_i} \quad (6)$$

Where  $\bar{a}, \bar{b}, \bar{c}, \bar{d}$  are the best estimate of the parameters of the function  $f(\theta, a, b, c, d)$ , and  $m = \text{n\_thetas} = 12$ . Using the uncertainty in  $I_{\max}$  as the set  $\{\sigma_i\}$ , the reduced chi-squared value was calculated to be  $\chi_\nu^2 = 1.657$  (Table. III), indicating a relatively accurate fit. This is visually supported by Fig. 5a. This indicates that there is an optimal time for an infection to take hold after the first case is established in a community, assuming  $R_0$  follows a periodic distribution in time.

From Fig 5b it is clear that a linear fit is suitable to describe the relationship between  $\theta$  and  $\Delta t$ . This is supported by a high  $R^2$  value, which represents the proportion of the variance in  $\Delta t$  that can be attributed to the variance in  $\theta$ .

## IV. DISCUSSION

### A. Simulation Results

The approach used in the paper is admittedly "fitting the model to the data," rather than the other way around. While this paper provides the most accurate outcome given that the  $R_0$  value of SARS-CoV-2 is seasonal, further real-

world data must be collected in order to verify this assumption.

However, it is clear from the data that given the basic reproduction ratio  $R_0$  varies in time sinusoidally, some useful relationships can be derived. From Fig. 5a the highest number of cases for this set of parameters (in particular,  $\varepsilon$  and  $\langle R_0 \rangle$ ), is found at  $\theta = \frac{\varepsilon}{b} = (200 \pm 20)$  days between the peak  $R_0$  and the first case. This result would prove helpful in predicting when to implement quarantine measures and other intervention strategies. Secondly, there is a clear linear relationship between the time between peak infection and the maximum  $R_0$  and when that peak  $R_0$  value is reached relative to the time of first case. The ratio  $\alpha = 0.86 \pm 0.02$  is roughly the relative rate the time delay between the max infection peak and peak  $R_0$  increases as the time of peak  $R_0$  since the first infection increases. This parameter would be useful in predicting the dynamics of the virus under seasonal conditions using real world data.

Future studies would benefit in exploring the relationship between the parameters  $\varepsilon, \langle R_0 \rangle$  and the time of maximum infection, delay between peak  $R_0$ , etc. with the intention of fitting real world measurements of the latter parameters to estimate  $\varepsilon$  and  $\langle R_0 \rangle$ .

### B. Modeling Limitations

While the stochastic approach proved useful in estimating the uncertainty of the calculated parameters, it experienced some limitations in the limit of high total population. During the collection of data, it was observed that the expected exponential curves describing the populations S, I, and R started to flatten into linear, almost triangular trajectories. Further inspection indicated that the probabilities of infection and recovery were larger than 1 (Fig. 6). This is likely due to some sort of normalization error, or reflects some of the limitations of the stochastic approach. Some additional normalization factor would have to account for the fact that  $I(t)$  grows very large, while  $\gamma$  and  $\delta t$  stay constant. One potential solution would be to decrease the step size quadratically with the number of total time steps elapsed. In theory, as  $I(t)$  increases, the rate of transmission should also increase, as there is a greater likelihood any one person will infect another individual at any given moment in time.

Look into continuous time new cov algs that use arrival times.

5.0/5.0

more room  
 7 is you only print axis titles once!

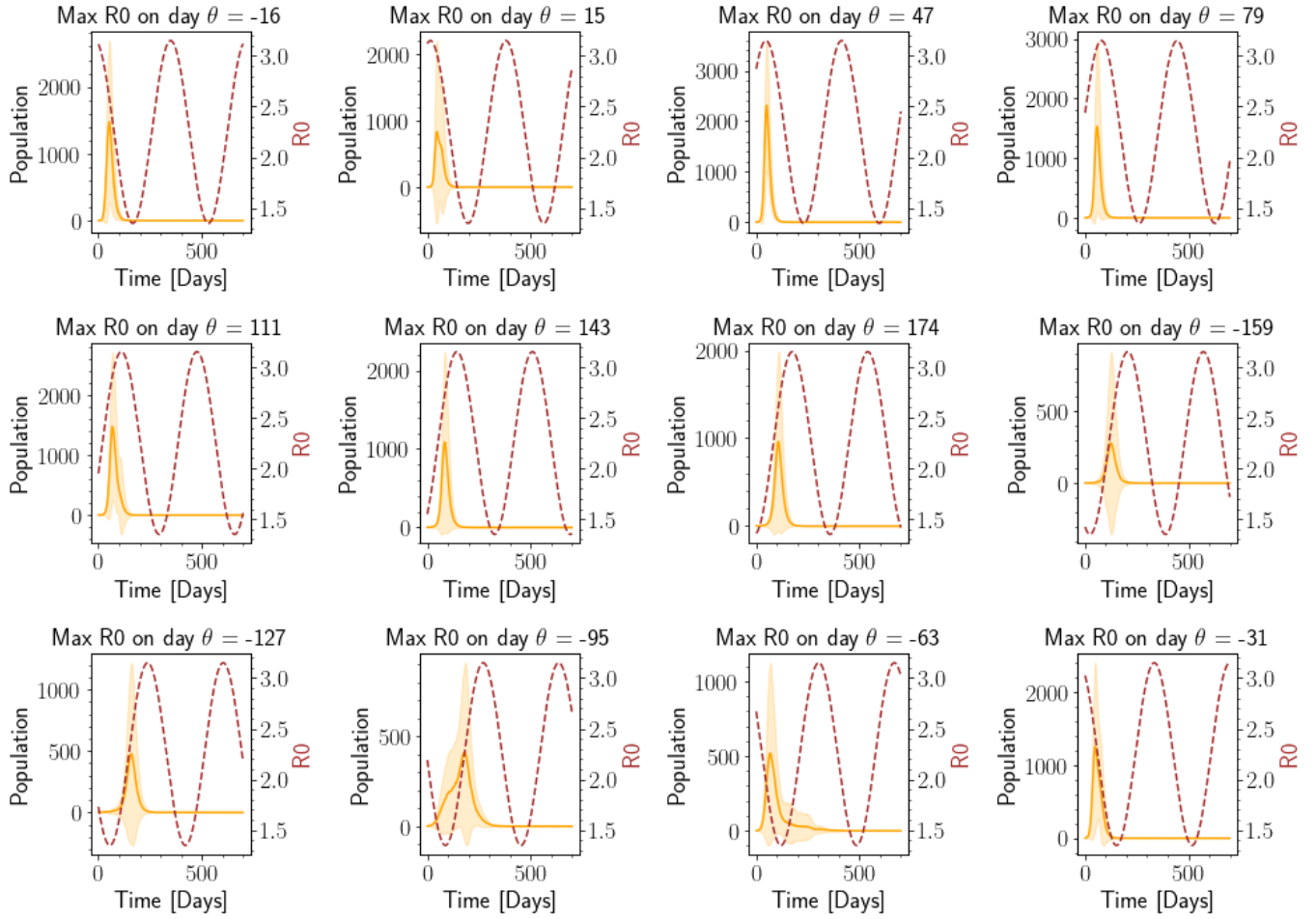
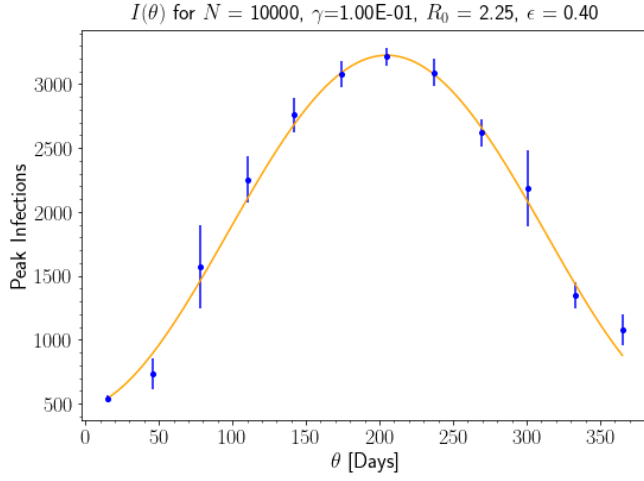


FIG. 4: All trials were run over a 700 day period with a time step of 1 minute, and total population  $N = 10,000$ . Shaded areas indicate the  $1\sigma$  uncertainty in  $I(t)$  over the thirty trials.

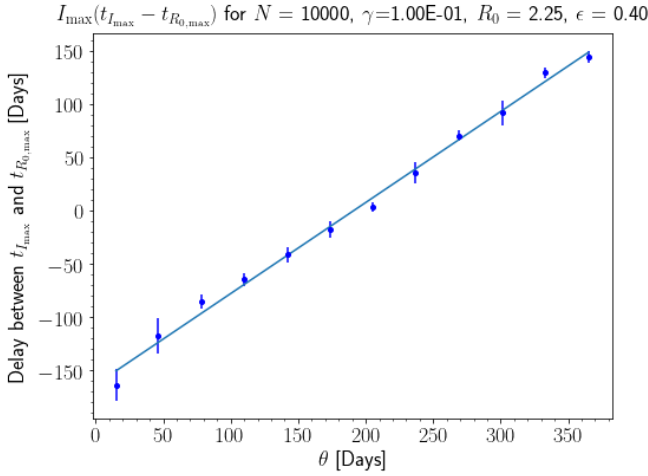
| Time $\theta$ of $R_{0,\max}$ after first case (Days) | $\Delta t$ (Days) | $\sigma_{\Delta t}$ (Days) | $\bar{I}_{\max}$ (People) | $\sigma_{I_{\max}}$ (People) |
|---|-------------------|----------------------------|---------------------------|------------------------------|
| -159  | 40                | 10                         | 3,100                     | 100                          |
| -127  | 70                | 5                          | 2,600                     | 100                          |
| -95   | 90                | 10                         | 2,200                     | 300                          |
| -62   | 130               | 5                          | 1,300                     | 100                          |
| -31   | 145               | 6                          | 1,100                     | 100                          |
| -16   | -160              | 20                         | 540                       | 30                           |
| 15  | -120              | 20                         | 700                       | 100                          |
| 47  | -85               | 7                          | 1,600                     | 300                          |
| 79  | -64               | 6                          | 2,300                     | 200                          |
| 111   | -41               | 7                          | 2,800                     | 100                          |
| 143   | -18               | 8                          | 3,100                     | 100                          |
| 174   | 4                 | 5                          | 3,220                     | 70                           |

TABLE I: Summary of average data collected over 30 trials per value of  $\theta$  (the time of the closest  $R_0$  peak measured from the start of the simulation, with  $I(t=0) = 1$ ), and a total of 12 different times of peak  $R_0$ . Negative  $\theta$  values indicate the closest peak  $R_0$  value was prior to the start of the simulation, while negative  $\Delta t$ 's correspond to times when the peak infection happened before the closest peak  $R_0$  value.





(a) Fit using the time of peak  $R_0$  with respect to the start of epidemic ( $\theta$ ) as the independent variable for maximum infection. The  $\chi^2_\nu = 1.657 \approx 1$  suggests a good fit.



(b) The time delay between  $t_{I_{\max}}$  and  $t_{R_0, \max}$  was calculated using the closest peak  $R_0$  value to the time of peak infection. Negative values indicate that the peak infection occurred before the closest  $R_0$  value.

FIG. 5

### C. Future Directions

Further extensions to the model could be used to improve the fit to real-world data. For instance, the addition of an exposed class would more accurately reflect the incubation period of the virus. There is a small time delay in viral infections before an individual becomes infectious because the virus needs time to replicate inside the body. The SEIR model places an "exposed" class in between the flow of susceptible to infected populations. The growth of the exposed population is not dependent on its own size, only the size of susceptibles and infected. The SIR model implemented

$$\text{Fit: } I_{\max} = a \cos(b\theta - c) + d$$

| Parameter            | Estimate            | Units   |
|----------------------|---------------------|---------|
| $a$                  | $1,390 \pm 50$      | people  |
| $b$                  | $0.0145 \pm 0.0007$ | rad/day |
| $c$                  | $3.0 \pm 0.2$       | rad     |
| $d$                  | $1830 \pm 60$       | people  |
| $\chi^2_\nu = 1.657$ |                     |         |

TABLE II: Fit data for  $I_{\max}$  as a function of  $\theta$ , assuming a sinusoidal dependence.

$$\text{Fit: } \Delta t = \alpha \theta + \beta$$

| Parameter     | Estimate        | Units |
|---------------|-----------------|-------|
| $\alpha$      | $0.86 \pm 0.02$ | 1     |
| $\beta$       | -160            | days  |
| $R^2 = 0.994$ |                 |       |

TABLE III: Parameter fit data for  $t_{I_{\max}} - t_{R_0, \max} \equiv \Delta t$  was calculated using the least squares method assuming a linear relationship between the peak  $R_0$  value and the time rate of infection.

in this paper also assumes indefinite immunity to the virus. As we do not fully understand the long-lasting immunity conferred to recovered individuals, modeling the cases of modest and short-term immunity would provide best and worst-case scenarios from which to shape intervention policy.

Another important parameter to consider for future policymakers is the degree of "herd immunity" required to stop a pandemic, which has been estimated by the following [8]:

$$r_{crit} = 1 - \frac{1}{R_t}; R_t = s(t) \cdot R_0 \quad (7)$$

with  $r(t) = \frac{1}{N}R(t)$ ,  $s(t) = \frac{1}{N}S(t)$ , and  $R_t$  as the effective reproduction number (a quantity more reflective of real-time data). From this, Kwok et al. concluded the virus will not stop until 60% to 70% of the population has received immunity on average. Thus, an important question to ask is with an "immune" population, if we relax the restriction of a closed system, what number of imported cases would be able to start a second wave of infections? If certain intervention strategies are implemented, how will they effect the required fraction of recovered individuals to establish herd immunity?

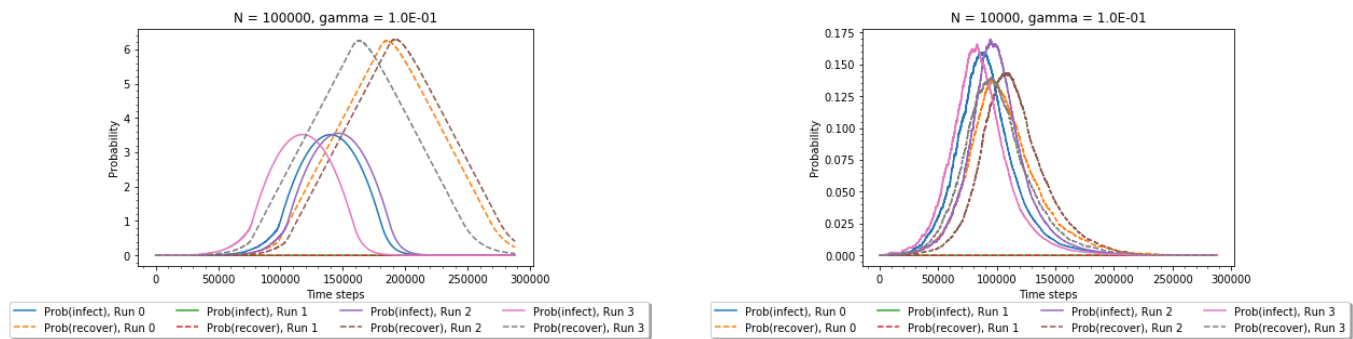


FIG. 6: The exponential growth of infections were stunted by the calculations of probabilities greater than one for infection and recovery. While  $p_{\text{infect}}$  seem to be slightly blunted by the factor of  $\frac{1}{N}$ , since  $\gamma$  was a constant,  $p_{\text{recover}}$  was allowed to grow as fast as  $I(t)$  would allow it. To the right of the run at  $N = 100,000$  is a run with all the same parameters except  $N = 10,000$  for a comparison.

Lastly, a challenging yet insightful extension to this model would be to remove the assumption that all individuals are equally likely to come into contact with one another. This can be implemented by using data structures linking different populations, or one could separate the susceptible and infected class into weighted categories based on how likely their behavior is to get infected and infect others. This can be achieved by including more subgroups within the susceptible and infected populations. Additionally, each person is not likely to have the same infection rate  $k$  or recovery rate  $\gamma$ . Thus, for each time step, values of  $k$  and  $\gamma$  could be chosen from a probabilistic distribution; the distribution to use

would be determined by comparing real world infection and recovery rates among the population of verified COVID-19 cases.

In the limited time available for this study, useful relationships between the time of peak infections and the maximum number infections with respect to the peak seasonally-varying basic reproduction  $R_0$  were uncovered. There are plenty of open-ended questions left and extensions to the SIR model that can be pursued to more accurately predict the spread of SARS-CoV-2 and construct effective policy to reduce its negative impacts on the healthcare system and society at large.

- [1] M. J. K. McFall-Johnsen, Lauren Frias, [A third of the global population is on coronavirus lockdown — here's our constantly updated list of countries locking down and opening up](#)
- [2] R. A. Neher, R. Dyrda, V. Druelle, E. B. Hodcroft, and J. Albert, Potential impact of seasonal forcing on a SARS-CoV-2 pandemic, , 8.
- [3] S. M. Kissler, C. Tedijanto, M. Lipsitch, and Y. Grad, [Social distancing strategies for curbing the COVID-19 epidemic](#),
- [4] K. A. Moore, M. Lipsitch, J. M. Barry, and M. T. Osterholm, Part 1: The future of the covid-19 pandemic: Lessons learned from the pandemic influenza, [1](#)
- [5] M. D'Arienzo and A. Coniglio, Assessment of the SARS-CoV-2 basic reproduction number,  $R_0$ , based on the early phase of COVID-19 outbreak in Italy [10.1016/j.bsheal.2020.03.004](#)
- [6] D. Fisman, Seasonality of viral infections: Mechanisms and unknowns, [18](#), 946
- [7] L. Lan, D. Xu, G. Ye, C. Xia, S. Wang, Y. Li, and H. Xu, Positive RT-PCR Test Results in Patients Recovered From COVID-19, [323](#), 1502
- [8] K. O. Kwok, F. Lai, W. I. Wei, S. Y. S. Wong, and J. W. Tang, Herd immunity – estimating the level required to halt the COVID-19 epidemics in affected countries [10.1016/j.jinf.2020.03.027](#) [32209383](#)

|                     |           |
|---------------------|-----------|
| Title and Abstract: | 3.0 / 3.0 |
| Introduction:       | 3.0 / 3.0 |
| Methods:            | 4.0 / 4.0 |
| Results/Analysis:   | 5.0 / 5.0 |
| Discussion:         | 5.0 / 5.0 |

Total: 20 / 20

First 100% ever given!  
congrats.