SANDIA REPORT
SAND2020-8172

Printed August 2020

Sandia National Laboratories

A Novel use of Direct Simulation Monte-Carlo to Model Dynamics of COVID-19 Pandemic Spread

Jose L. Pacheco, Zakari S. Eckert, Russell W. Hooper, Melissa Finley, Ronald P. Manginell

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185
Livermore, California 94550

Issued by Sandia National Laboratories, operated for the United States Department of Energy by National Technology & Engineering Solutions of Sandia, LLC.

Available to DOE and DOE contractors from

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@osti.gov
Online ordering: http://www.osti.gov/scitech

Available to the public from

U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Road
Alexandria, VA 22312

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.gov
Online order: https://classic.ntis.gov/help/order-methods

ABSTRACT

In this report, we evaluate a novel method for modeling the spread of COVID-19 pandemic. In this new approach we leverage methods and algorithms developed for fully-kinetic plasma physics simulations using Particle-In-Cell (PIC) Direct Simulation Monte-Carlo (DSMC) models. This approach then leverages Sandia-unique simulation capabilities, and High-Performance Computer (HPC) resources and expertise in particle-particle interactions using stochastic processes. Our hypothesis is that this approach would provide a more efficient platform with assumptions based on physical data that would then enable the user to assess the impact of mitigation strategies and forecast different phases of infection. This work addresses key scientific questions related to the assumptions this new approach must make to model the interactions of people using algorithms typically used for modeling particle interactions in physics codes (kinetic plasma, gas dynamics). The model developed uses rational/physical inputs while also providing critical insight; the results could serve as inputs to, or alternatives for, existing models. The model work presented was developed over a four-week time frame, thus far showing promising results and many ways in which this model/approach could be improved. This work is aimed at providing a proof-of-concept for this new pandemic modeling approach, which could have an immediate impact on the COVID-19 pandemic modeling, while laying a basis to model future pandemic scenarios in a manner that is timely and efficient. Additionally, this new approach provides new visualization tools to help epidemiologists comprehend and articulate the spread of this and other pandemics as well as a more general tool to determine key parameters needed in order to better predict pandemic modeling in the future. In the report we describe our model for pandemic modeling, apply this model to COVID-19 data for New York City (NYC), assess model sensitivities to different inputs and parameters and , finally, propagate the model forward under different conditions to assess the effects of mitigation and associated timing. Finally, our approach will help understand the role of asymptomatic cases, and could be extended to elucidate the role of recovered individuals in the second round of the infection, which is currently being ignored.

ACKNOWLEDGMENT

We found many Sandians willing to help with this endeavor, from asking questions an providing suggestions to providing data, and bringing emergent/relevant information to the attention of the team. In particular, we want to thank, Ed Cole, Pat Finley, Kathy Simonson, Michael Gallis, William Miller, Keith Matzen and Walt Witkowski.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

The purpose of this report is to determine if the Direct Simulation Monte Carlo (DSMC) method can be used to model dynamics and characteristics of the COVID-19 pandemic. The many unknowns about the COVID-19 pandemic, and the lack and/or lag of test data, are a challenge for making predictions with models and make policy decisions uncertain and difficult. Existing pandemic/epidemic tools were used to model COVID-19[4], as well as to inform government officials on public health policy [see W. H. briefing from 3/30/2020]. The aforementioned model is constantly being updated as new information becomes available, is informed by infection characteristics based on previous epidemics or pandemics and assumes a general trend, or curve, for the pandemic evolution. This, and most other modeling approaches, face significant challenges that arise from test data that carries a number of uncertainties, gaps, and biases. Calibrating a model with data for which there is little to no control over data acquisition practices can result in inherent biases. Given the novelty and all the other challenges associated with the COVID-19 pandemic, widespread testing remains elusive and 'good' data collection practices have suffered, likely due to other priorities/demands placed on the health care system.

The trends and characteristics of the COVID-19 pandemic appear to be qualitatively well captured in a nation-wide study completed in 2006 for a hypothetical pandemic but rooted in the H5N1 characteristics[14]. This study was carried out post-H5N1 and evaluates the effects that a pandemic could have on the US population as well as the benefits of different mitigation strategies and immunization. In that study, the sensitivity of pandemic propagation was assessed assuming a range of values for the so-called R_o value (average number of people infected by one person). Many other modeling approaches exist that could help provide answers, or reduce uncertainties about, COVID-19. These modeling approaches range from purely statistical to mechanistic (such as SEIR (Susceptible Exposed Infected Recovered) modeling [3]), Agent-Based Models [15], Individual-Level-Models [13], etc.). Modeling approaches similar to that used by [4] rely on historical and/or early pandemic data to make projections/predictions, the accuracy of which is clearly affected by lagged or skewed test data. Although the study presented in [14], identifies the R_o value as a sensitive parameter, Individual-Level Model (ILM) predictions for this value can be significantly different. It is, therefore, reasonable to expect that local variations of this value based on density and behaviors of a population may significantly affect the evolution of a pandemic. Higher-fidelity Agent Based Models use literal inputs, use minimal assumptions, but can require massive computational resources such that simulation throughput becomes a challenge. Of course, each of these modeling tools provides a different level of fidelity with high fidelity models requiring larger efforts in terms of data acquisition and modeling resources but providing 'numbers' (approaching quantitative), and lower fidelity making a significant number of assumptions but providing 'trends' relatively quickly.

The modeling approach presented in this report uses the DSMC algorithm to model individual interactions for a large population of particles (representing individual people). DSMC is typically employed for kinetic plasma physics and rarefied gas dynamics simulations, where the number of particles (in this case representing gas particles) can easily number in the billions. As such, these algorithms are computationally efficient, massively parallel, and run on state-of-the-art HPC platforms, at scales that easily resolve large populations of people for modeling of COVID-19.

Our new approach is to use existing infrastructure for Direct Simulation Monte-Carlo (DSMC) simulations to model the spread of pandemics with higher fidelity (as compared to[4] & [3]). In DSMC, input for interactions between particles (representing people) are physical data for the interaction in question, (a particle diameter in gas dynamics, or an interaction frequency for people). The simulation propagates a population of people in time, and captures events such as the passing of infection and progression of an individual's disease in a statistically accurate way with respect to the input data. In addition, other factors in the simulation are considered such as density gradients, disturbances and dynamically evolving populations such that the initial assumptions are minimal and key parameters, such as local R_0 value, are simulation results. Making the transition from gas particles to people is primarily a matter of semantics. Then capturing sufficient demographic traits and using physical parameters and disease progression characteristics are key to model disease spread with DSMC.

The report is arranged as follows: In section 2, existing modeling approaches are briefly discussed with some emphasis in SEIR (readily available from web). In Section 3, COVID-19 characteristics that are key for this work are outlined. Section 4 describes the model and how COVID-19 characteristics are included as well as showing model validation against NYC data. This section is completed by assessing model sensitivities to assumptions in terms of initial conditions and details of interactions among different populations in the pandemic. In section 5 we present model predictions obtained by looking at hypothetical 'what if' scenarios. Lastly, in Section 6, we conclude by briefly discussing future work and drawing some conclusions about the modeling tool developed and the potential significance of results obtained.

2. MODELING METHODS

2.1. SEIR Pandemic Modeling

Consider the use of an SEIR model and the so-called basic reproduction number, R_o, for modeling COVID-19 pandemic. Fig. 2-1 shows the basic reproduction number estimate for New York State as a function of time [7]. The first estimate was reported (2020/03/11) as ≈ 5, which then decreases to < 1 (on 2020/4/15).

The R_o value early in the pandemic, when people were not as informed, started much higher, then reduced as awareness developed and mitigation strategies were implemented. Taking the R_o values estimated for New York, we can apply these to a population of $8.4M$ to represent NYC and use available SEIR [3] models for predictions of how the pandemic will progress. Figure 2-2 shows one such a calculation using values that appear reasonable from Fig 2-1. However, when compared to NYC published data [5], significant differences between model and data are clear. Figure 2-3 shows NYC daily data for confirmed cases, number of hospitalizations, and deaths. Daily hospitalizations and deaths are the two quantities where it is reasonable to make comparisons since it is part of the process to report these numbers when people reach a hospital. The number of confirmed cases depends on whom and why these tests are performed. Note that the number of exposed or infectious people from the epidemic model have no clear relationship to the number cases as reported in NYC data. Hospitalized patients emerge from the

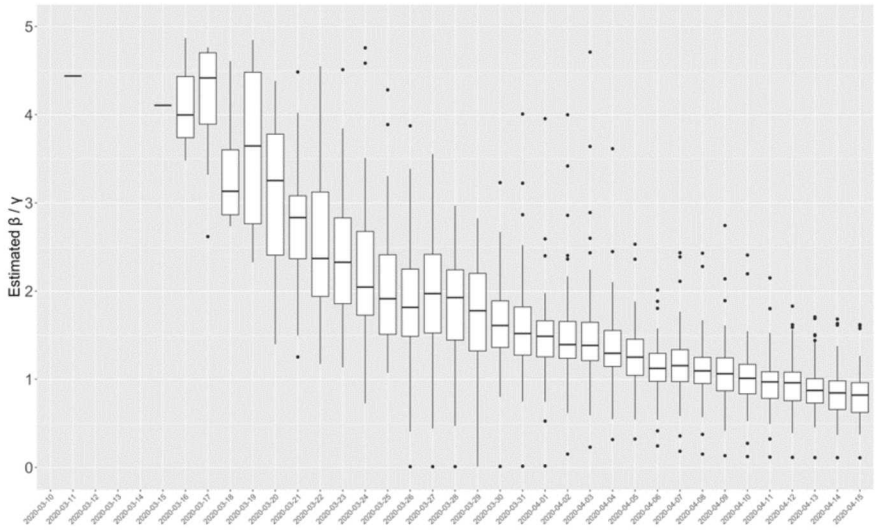

Figure 2-1 Estimated Ro for New York as a function of time. Note the large error bars and relatively large Ro value in the early stages of the pandemic. During later stages of the pandemic, error bars are tighter, likely due to the larger available data sets but the value is changing due to enacted mitigation strategies.

'Hospitalization Rate' parameter. While evidence strongly suggests that the hospitalization rate is age dependent, the 20% value chosen is close to an average value (i.e. $50-60\%$ for elderly down to 5% for individuals below 20 years old). Importantly, hospitalization rate vs age demographic is a key piece of information that is, arguably, less subjected to issues with data compilation (described before). However, including this information in SEIR would require individual model for each age group and the associated progression rates, achieving a self-consistent solution between the different age groups (models) would be a challenge. Other values used to produce that figure appear very reasonable. However, the number of peak hospitalizations between epidemic model shown and NYC data differ by more than two orders of magnitude.

Noting that R_o is decreased on the intervention day, including a time dependent R_o value (see Fig. 2-1) is a direct extension of SEIR models that could give more accurate results. This would require timely and accurate data since the model appears sensitive to the timing and actual value used. The paper by Germann et. al. [14] identifies these parameters as key to understand what is required in order to have control over a pandemic.

Additionally, as defined, R_o value will depend on the population spatial distribution. Further, because of the stochastic nature of personal interactions, and the fact that infectious period is finite, the number of people that can be infected by one person will follow Poisson statistics [13]. At an average value of 2, and assuming all individuals are susceptible, there is a chance that an infected person could infect >5 individuals. Under similar circumstances, this number can increase as the population density increases. The disease progression with an initial infected population of 1 vs 10 is significantly different. As such, starting a deterministic calculation (such as SEIR) with inherently stochastic initial conditions may require some sort of initial model

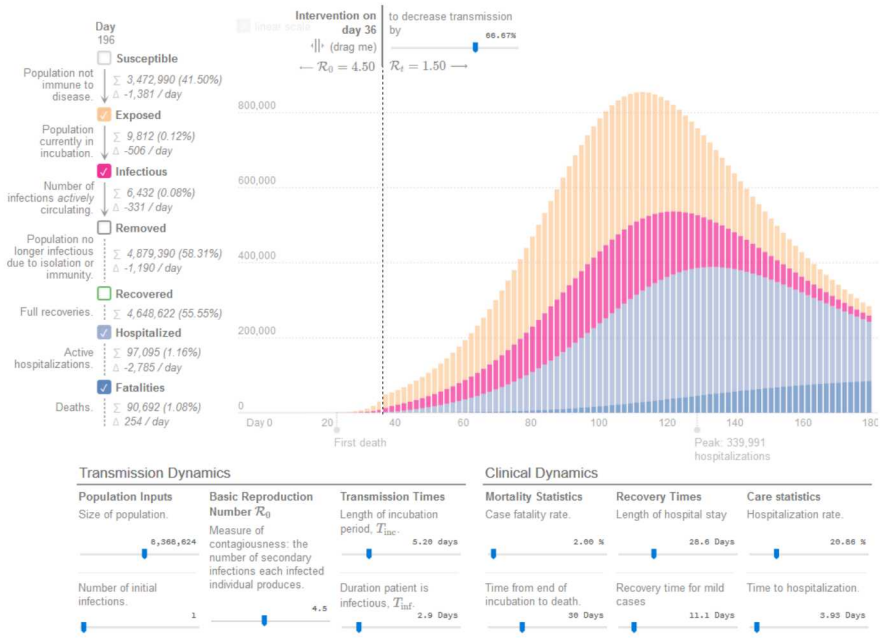

Figure 2-2 Epidemic calculation using web-available SEIR modeling tool [3]

calibration.

3. PANDEMIC PROGRESSION CHARACTERISTICS

Because of the way in which particle interactions are handled in DSMC, it seems possible to develop a model that sidesteps some of the issues outlined while providing higher fidelity results (via individual interactions) using modest computational resources (when compared to ABM). Further, there appears to be a gap between SEIR and ABM where the former can be done in a relatively light 'app' and the former can require a large HPC to run a single simulation. It is expected that the work in this report could fill that gap with a relatively low cost (computationally) tool that can provide information to, or, in some cases, be an alternative for, SEIR and ABMs.

3.1. Pandemic Characteristics

Capturing the disease characteristics in terms of statistically accurate disease progression probabilities and rates could yield a model that is more robust to, or independent of, variations in data available (when compared to 'curve fitting' approaches to modeling the pandemic). It is worth noting that with the method employed herein, probabilities and rates could be extended to capture emerging trends based on co-morbidities, race, etc., at the expense of a more complex input.

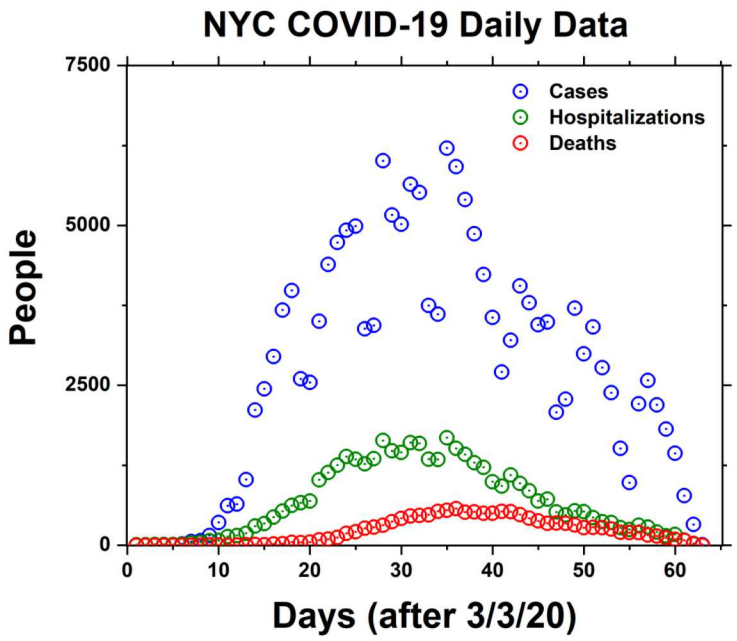

Figure 2-3 Daily reported data for NYC, available from [5]

3.1.1. Asymptomatic Population

Anecdotal evidence suggests that once it was realized that a respiratory disease was spreading in the Diamond Princess cruise ship, everyone was tested. This piece of evidence is key since it represents the likely progression characteristics and state of a pandemic for a naive population (where naive in this context indicates lack of knowledge of the presence of the disease and therefore absence of precautions or mitigation). The following describes a snapshot of the pandemic once everyone was tested:

'Among 3,711 Diamond Princess passengers and crew, 712 (19.2%) had positive test results for SARS-CoV-2 ... Of these, 331 (46.5%) were asymptomatic at the time of testing. Among 381 symptomatic patients, 37 (9.7%) required intensive care, and nine (1.3%) died (8). Infections also occurred among three Japanese responders, including one nurse, one quarantine officer, and one administrative officer (9). As of March 13, among 428 U.S. passengers and crew, 107 (25.0%) had positive test results for COVID-19; 11 U.S. passengers remain hospitalized in Japan (median age = 75 years), including seven in serious condition (median age = 76 years)'... extracted from [6]

- 19.2% of the total population tested positive
- 46.5% of people who tested positive for COVID-19 were asymptomatic
- These data pertain to a group with median age in the 70's

With regards to NYC data and a a different age group, we quote key findings from a New York hospital:

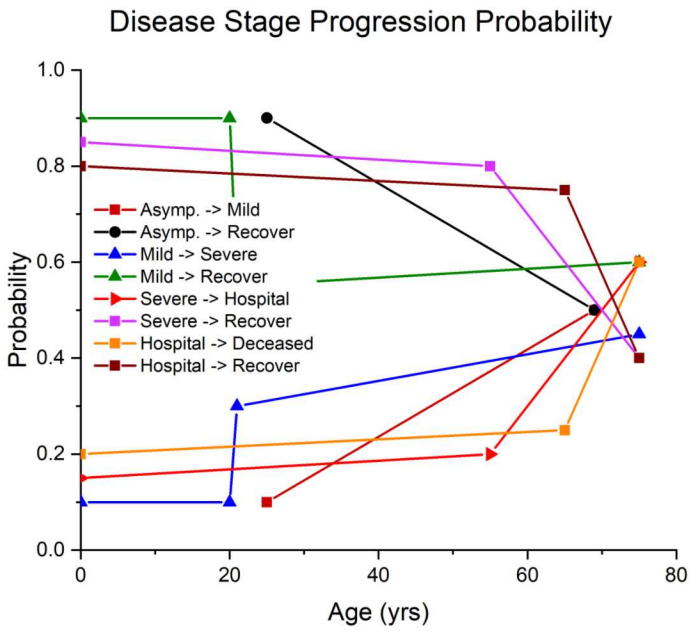

Figure 3-1 Disease progression probability for different age groups.

'This week, we published our findings after two weeks of universal screening at New York-Presbyterian/Columbia University Irving Medical Center. In our area, which includes upper Manhattan and the Bronx, about 15 percent of patients who came to us for delivery tested positive for the coronavirus, but around 88 percent of these women had no symptoms of infection. That means 13.5 percent of all our patients during this time were infected with the coronavirus but were not exhibiting symptoms' ... extracted from [9]

- 15% of the total population tested positive
- 88% of people who tested postive for COVID-19 were asymptomatic
- These data pertain to a pregnant population demographic, which we assume are in an age group between 20 and 30 years old.

With this information and details in the following section, disease progression probabilities and rates can be obtained. Having an estimate for the asymptomatic population is key for self-consistent evolution of the disease since test data can be biased (testing only those experiencing symptoms, or due to the relative availability of testing, as examples). An estimate of asymptomatic population, disease progression rates, and hospitalization data for recovered and deceased places a large number of constraints on the model such that initial conditions must adhere (statistically) to all of these constraints simultaneously.

3.1.2. Disease progression

Based on the data summarized above, disease progression probabilities from initial stages (pre-contagious and asymptomatic) were developed as follows: on average 50% of infected people in their 70's are asymptomatic, and that, on average, 88% of infected people in their 20-30's are asymptomatic. Simulations then assume a linear extrapolation for the other age groups. Other disease progression probabilities are extracted from the statements above and made to match hospital records for hospitalized, recovered, and diseased vs demographic group. These probabilities are constrained since, for example, hospitalized patients result in recovery or death, such that knowing one of these probabilities is sufficient.

The time dependence of disease progression rates are assumed from CDC published statements:

'The respiratory symptoms of COVID-19 typically appear an average of 5-6 days after exposure, but may appear in as few as 2 days or as long as 14 days after exposure, according to the U.S. Centers for Disease Control and Prevention (CDC).'

'Using available preliminary data, the median time from onset to clinical recovery for mild cases is approximately 2 weeks and is 3-6 weeks for patients with severe or critical disease.'

Again, to use DSMC for modeling the pandemic spread, statistically accurate progression probabilities and rates appear sufficient, model depends less on details of testing, and could provide an estimate for asymptomatic population (if hospitalizations and deaths can be matched).

3.1.3. Demographics

Population demographic ratios can be estimated from the census data. An example is shown below for demographic distribution in New York City. Table 3-1 shows the NYC demographic details used in the model.

Table 3-1 New York City Age Demographics

Age	Percent of Population
0-17	25%
18-44	41%
45-64	22%
65-74	6%
75 or older	5%

In a similar manner, the model could accommodate other demographic details relevant to disease progression and fatalities such as race and co-morbidities.

4. DSMC FOR COVID-19 MODELING

4.1. Model

Aleph is a particle-in-cell, direct simulation Monte-Carlo (PIC-DSMC) [12] [11] code that was developed for large scale simulations of fully kinetic, low temperature plasma physics [10]. In these simulations, PIC self-consistently accounts for charged particle contributions to electrostatic field. In turn, the newly established field moves the particles. DSMC is implemented in a 'substride' to PIC, where DSMC typically modifies particle properties as well as create new, or destroy reacted particles. Physically, interactions result in modified velocity vector (scatter), additional charged particles (i.e. ionization of neutral atoms), excited particle states, and a range of other 'particle types'. Particle creation, destruction, scatter, and interactions in general are handled with the DSMC algorithm. For the simulations presented in this report, only the parts of the code necessary for DSMC were used (particle movement and interactions).

In DSMC, the number of potential interactions between two different particle types is:

$$n_{int} = n_A n_B \langle v\sigma \rangle_{AB} V dt \tag{1}$$

Where n_A, n_B are the number of particles densities of type A, type B, respectively, that reside within volume V. v is the relative velocity between these two particles and *sigma* is the 'cross section', which is a measure of how likely the interaction is to happen given all other parameters (all of which are known). dt is the simulation time step. It is worth noting here that Eq. 1 captures all the parameters that affect the number of interactions. The factors n_A, n_B, and V are controlled by the population density of the region being simulated. While v and σ_{AB} have different physical meanings for gas dynamic simulations, here these can be best interpreted together as an effective rate of interaction between people. This is analagous to the R_o value divided by population density, which helps to isolate the different effects that can affect R_o in SEIR models.

The simulation domain is divided into many smaller individual volume elements (simulation 'mesh') and, along with the time step chosen, the number of potential interactions between two particles is limited to a small fraction (0.1) to reduce the likelihood of a particle interacting twice in a single time-step. Eq. 1 is evaluated within each element in the domain, and DSMC provides an efficient algorithm for choosing the particles which participate in the n_{int} interactions. Once two particles are chosen to have interacted, their characteristics (demographics and disease stage) are used to calculate the possible results of the interaction (e.g. no change or passing of the infection). In a single interaction, the result it chosen randomly, however as more interactions occur, the statistics converge consistent with the local population density and input interaction rate.

Leveraging the DSMC algorithm, we can envision using particles to model people, so long as we can capture sufficient human traits. The DSMC method is capable of including an essentially unlimited number of different traits. However, with each added trait combination, the number of interaction paths increases dramatically (as the number of trait combination squared). As has been stated, we include as many people traits that our model can accommodate, which are based

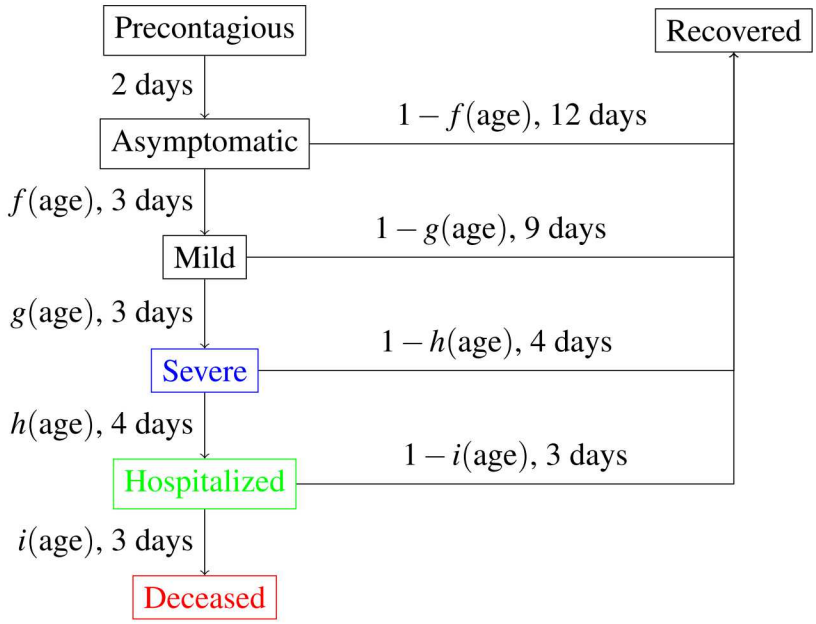

Figure 4-1 Model of disease progression rates and probabilities with the functions of age shown in Fig. 3-1. Colors of later states represent colors used in later results plots.

on data available (e.g. progression rates and probabilities). It is likely that additional granularity in demographics and co-morbidities would improve the model results that follow and is a reasonable starting point for future work.

4.1.1. Interactions

The disease was spread through interactions between healthy and contagious populations in the following manner:

$$Healthy + \textbf{contagious} \rightarrow Precontagious + \textbf{contagious}, \tag{2}$$

where **contagious** is one of Asymptomatic, Mild, or Severe. In addition to becoming infected through the interaction above, people of all demographics and disease stages interact with each other and 'scatter', which effectively randomizes their motion. Therefore, the interaction probability includes 'scattering' and 'infection', meaning that a subset of all interactions can result in infection of an individual.

4.1.2. Probabilities & Rates

Progression for each newly infected individual follows according to the disease progression rates and probabilities as shown in Fig. 4-1.

4.1.3. Changing Conditions due to Mitigation

Implementation of different mitigation strategies will inevitably result in modifications to the model inputs. Our model can accommodate some aspects of mitigation. For example, a 'lock down' can be mapped to a reduced people/particle temperature. 'Face coverings' can be mapped to a smaller effective diameter, resulting in a lower interaction probability. As the population becomes more aware, it seems likely that both of these conditions are changing simultaneously. However, it will be shown later that the model is highly sensitive to both of these parameters since effective diameter and temperature both increase the number of interactions. As mitigation is relaxed, or compliance with mitigation is challenged, these two parameters, temperature and effective diameter, can be first estimated from different temporal segments in the disease progression and then re-used to propagate the pandemic forward.

4.2. Model Validation

4.2.1. New York City data

There are roughly $8.4M$ people in New York City. For all simulation results in this report, each person is treated individually, such that the simulation uses $8.4M$ particles. The particles are created in the different age groups in the percentages shown in Table 3-1. An instance of initial disease state assumes 6k asymptomatic individuals, 800 individuals with mild symptoms, and 240 individuals with severe symptoms, distributed proportionally among the different age groups. There are 7 hospitalized individuals distributed among the three older age groups. This last initial condition detail closely represents what was recorded in New York City on 3/3/20.

We assume a uniform population density over $784km^2$ which results in of 10715 people per km^2, matching the area, population, and density of NYC. The simulation time step used is 1hr. The time period simulated is defined by the user but covers a span of 60 days for most simulations shown in this report. The interaction rate was chosen such that the population each experience a few interactions per day. Individuals in the population can have higher or lower interaction rates, that are effectively sampled from a global distribution.

Using the New York City details above, disease progression fractions as a function of age are shown in Fig. 3-1, and progression rates shown in Fig. 4-1, simulations were conducted to begin 'tuning' some model inputs for which data was unavailable (such as initial populations in different disease stages).

The effective probability for interaction was also varied to better match the data using the following rationale. NYC officials closed schools and a large number of public services buildings on March 14th, and, on March 22nd, made the anouncement that 'shelter in place' order would go in effect on March 22nd. In the results that follow, the simulation progresses through 3 phases, phase 1 represents 'no mitigation' or disease spread in a population that is not concerned with slowing disease spread. In phase 2, the probability of interactions in the simulation would need to be decreased to capture higer general awareness that the disease is spreading and mandatory closures. In phase 3, probability of interactions is further decreased to capture the most restrictive

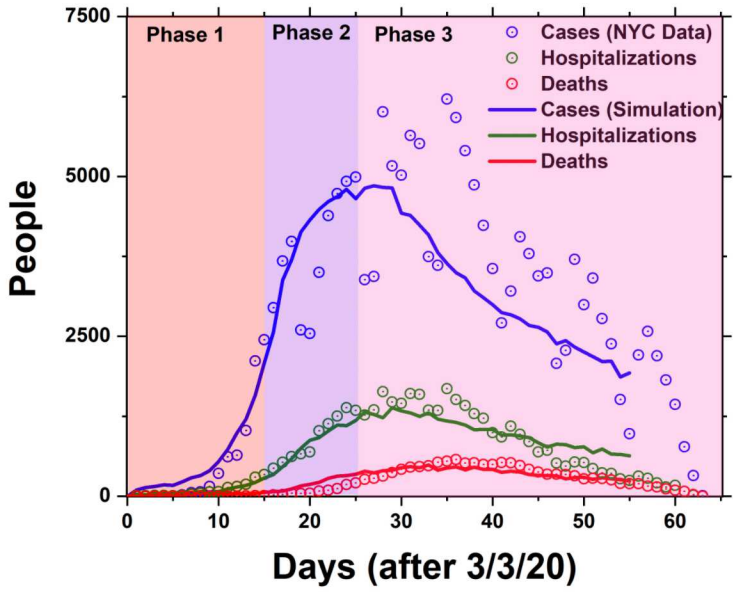

Figure 4-2 Daily NYC data for positive cases (blue), hospitalizations (green), and deaths (red). Data points were published by NYC public health officials, lines are simulation results. The three different background colors represent the phases in the simulation, see text for details.

'shelter in place' restrictions as well as the more wide-spread use of face coverings. At the point in the simulation coincident with the start of mandatory closures implemented in New York City, the simulation is stopped, all particles and their properties are recorded, and the simulation continues (from the stoppage point before) with the same simulation details, except that the effective interaction frequency is decreased. This first phase in the simulation lasts for 15 days (interaction diameter = 0.4), the second phase lasts for 10 days (interaction diameter = 0.2) and the third phase continues until the end of the simulation (interaction diameter = 0.15).

Figure 4-2 shows results where the initial population and effective diameter were varied to the point where model results show good agreement with the daily data. In addition to matching the daily numbers, this approach is capable of tracking demographic-based progression rates and probabilities. Figure. 4-3 illustrates good comparisons between published data and model predictions for age demographics.

Except for the interaction rate, other inputs to the model emerge from general characteristics of the pandemic. Though initial disease population conditions are also assumed, these are partly based on disease progression probabilities. The progression rates emerge from CDC publications [2]. Model sensitivity to these parameters is assessed in the following section. However, it is worth noting that the results thus far show a quite favorable agreement.

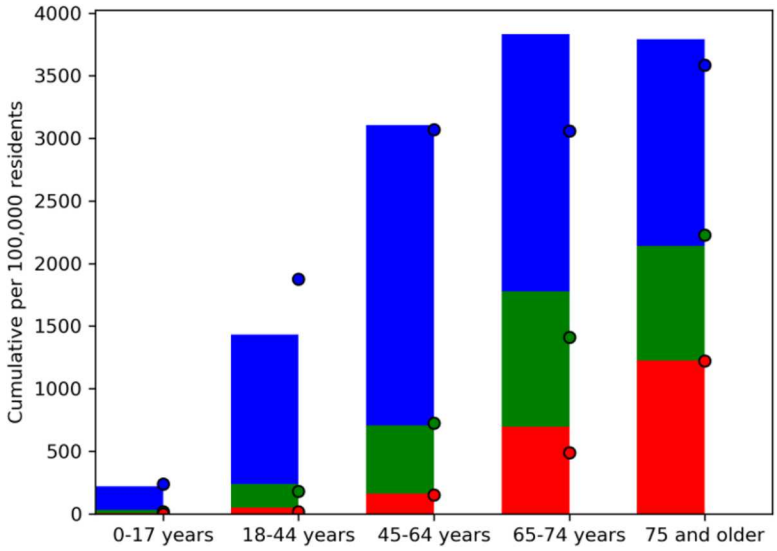

Figure 4-3 Demographic disease data for New York City. Data plotted as cumulative number per 100k residents vs age group. Bars represent simulation results, dots represent data published by NYC health officials. The blue represents number of positive cases, green is number of hospitalized patients and red is number of deceased.

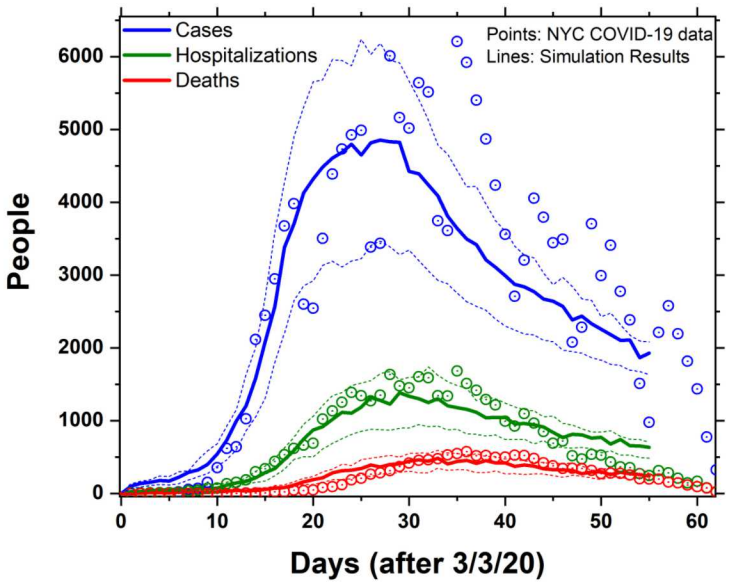

Figure 4-4 Model sensitivity to initial number of people in infected categories. Solid lines are simulation results for 7k people in the different infected categories, evenly spread over all age groups, where the majority (6k) of these are asymptomatic. The dashed lines are simulation results when this initial population is increased or decreases by 1k (from 7k).

4.3. Sensitivity

4.3.1. Initial Conditions for Disease

We now assess the model sensitivity to initial conditions including initial number of infected population and demographics, interaction probabilities, and disease progression rates. Figure 4-4 shows simulation results as lines and NYC COVID-19 published data as points. Solid lines represent the 'tuned' case (described in previous section) and dashed lines are +/- 1500 (from 7k in 'tuned' case) people infected initially. An estimate based on NYC data indicates that there are ≈1k infected individuals for every hospitalized patient, suggesting that the initial conditions in the 'tuned' case are reasonable.

Notice that the these variations in initial conditions bracket reported NYC data for hospitalizations and deaths, in both magnitude an timing. 7 hospitalized patients and 7k people distributed in the other infected categories (1k infected for every hospitalized patient) were used as initial conditions, the simulation results show that this is a reasonable estimate but could be as high [low] as 1.2k [0.8k] infected individuals per hospitalized patient. Further, the model is sensitive to 1.5k initially infected individuals in a population of 8.4 M!

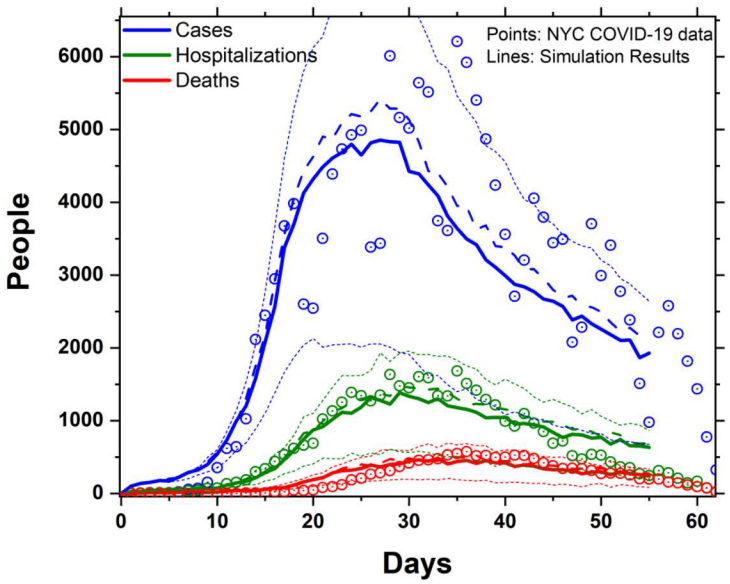

Figure 4-5 Sensitivity interaction diameter variations. 'D5' is the best result in the interaction diameter parameter sweep (shown in thick dashed lines) but a marginal improvement over results shown in Fig. 4-2. See text for details.

4.3.2. Interaction Area & Mobility

Intervention strategies in the form of stay-home orders, social distancing, and face coverings will reduce the overall number of interactions, thereby reducing infections. There have been several experiments where droplets generated after a cough or sneeze event are tracked and the distances traveled recorded. It was shown that some of these droplets may travel as much as 12 ft [16]. Assuming droplets follow $\approx \frac{1}{r^n}(1 < n < 2)$, ($n = 2$ for point emitter into 4π) spatial distribution, where r is the distance from emitting event (person sneezes), 'social distancing' or physical distancing of 6ft or greater significantly reduces the likelihood that a bystander interacts with droplets from an uncontrolled sneeze or cough event. Talking, yelling, singing, or otherwise normal breathing can also generate droplets, all of which become more highly concentrated in closer proximity to an infected individual. In our model, this region in the vicinity of an infected individual within which a bystander has a high probability of becoming infected is called the 'interaction diameter.' Clearly, wearing face coverings will decrease the distance droplets can travel (in the impractical limit of a thick solid mask, no particles escape, interaction diameter is the extent of the face covering itself). Figure 4-5 shows modeling results when the interaction diameter is varied. The dotted lines represent a 5% variation (high/low, respectively) in interaction diameter vs. that used in the tuned case. The best fit is plotted as a dashed lines and was obtained with the simulation 'D5' which increases the interaction diameter by 1.3% compared to the 'tuned' case.

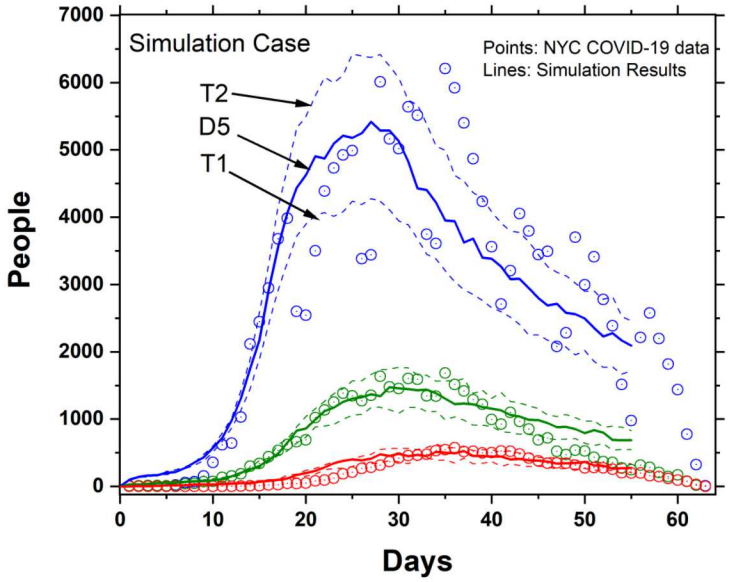

Figure 4-6 Sensitivity to variations in reference temperature. T_0 was chosen for particle speeds of $3m/s$, T1 is 10% lower, T2 is 10% higher. D5 is the best result shown in the interaction area parameter sweep.

Stay-home orders will directly impact the mobility of an individual. Effectively, this strategy will reduce the temperature used in our model ($T\ v^2$, and interactions proportional to the relative velocity between particles, see 1). The number of interactions will increase as the individual mobility or 'temperature' increases, and vice versa. However, reducing an individual's mobility does not change the interaction area discussed before. Reducing the mobility of an infected individual in the presence of many other healthy individuals does not make sense unless face coverings and social distancing are strictly enforced. For the tuned case presented in Sec. 4.2.1, the temperature was chosen such that individuals would have an average speed of 3m/s. Fig. 4-6 shows modeling results for two additional cases, where the temperature is varied by $\pm 10\%$.

The results above show that the model is sensitive to both the interaction diameter and temperature. The current values used were obtained empirically. In typical PIC-DSMC simulations, the gas or plasma temperature is measured/known, the interaction diameter determined experimentally or theoretically. It appears possible to arrive at COVID-19 interaction area and the mobility of people experimentally but that is beyond the scope of this project.

4.3.3. Progression Rates

All the details about pandemic test data acquisition make model comparisons difficult and/or suspect, this is particularly true with the timing of disease progression. As an example, in NYC, there were 7 hospitalized individuals on 3/3/20 and no positive tests outside of the hospital, turn around time for results varies and often takes a few days, and daily positive cases drop with a 7

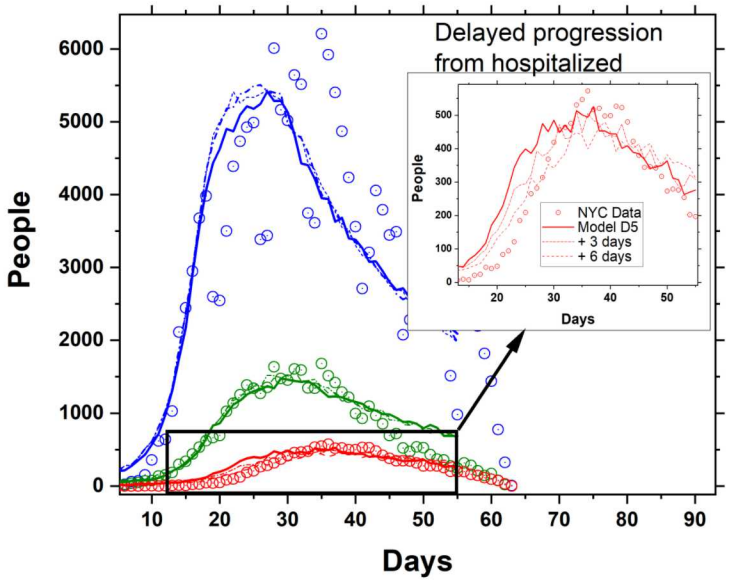

Figure 4-7 Model sensitivity to progression rates as assessed by increasing delay from hospitalizations to recoveries and deaths. The progression rates are increased by 3 and 6 days in simulation case D5. Varying the progression rate (delay) by 3 days is visibly a best match to the data. Data then 'leads' simulations results (worse match) when delay increased to 6 days.

day period with two days (Saturday and Sunday) showing significantly lower number of positive cases than expected from the data trend. In what follows, we assess timing sensitivities for disease to progress from hospitalized to deceased. Here we use the best case from the diameter study as shown in Fig. 4-5 as a starting point and vary the disease progression timing for hospitalized patients to progress to either of the last two stages. Figure 4-7 shows the base case and two additional cases for deceased status to be reached, first with a 3 day delay, second with a 6 day delay.

The results show that the model is also sensitive to progression rate details. A difference of 3 days shows that further delaying the progression from hospitalized to deceased appears to better represent timing details associated with hospitalizations,

5. PREDICTIONS

5.1. No mitigation

The initial conditions and some of the assumptions in the model are based on some anecdotal and/or documented evidence, particularly as it pertains to the progression rates and probabilities for disease, but these come from statements by CDC and other health officials. However, as stated in the previous section, the interaction diameter and the reference temperature were determined empirically. Ideally, one or both of these values would be determined experimentally. With out any further evidence for what these two values may be in reality, we push forward to address 'what if' scenarios with regards to different paths the pandemic spread can take given implemented mitigation strategies, different levels of compliance, re-opening, etc.

First, what if no mitigation strategies were implemented? How *bad* could it get? For the following simulation, the reference temperature and interaction diameter are held constant. The values that matched the daily data for the initial phase of the simulation, before any official orders for mitigation were announced, are used for the same simulation time period to observe model predictions of disease progression without mitigation. Figure 5-1 shows these results.

The staggering numbers in each of these categories gives a sense how critical mitigation is to keep the health care system from being over-run.

5.2. Medium and Low compliance at 'phase 3'

No or low compliance with mitigation strategies clearly affects the pandemic dynamics. As previously discussed, we assumed initial conditions of roughly 6k infected individuals was required in order to match the rates and magnitudes of both hospitalizations and deaths. If behavior had returned toward normal, either due to mitigation strategies being removed or compliance with them dropping, when the curve had 'flattened' then effects of mitigation would quickly be undone. Essentially re-starting disease spread with the same exponential growth but with a much larger baseline. For the simulation results shown in Fig. 5-2, the model assumes the initial conditions are the same as for the best fit simulation for phase 1 and phase 2 but, for phase 3, the interaction diameter is increased to 0.4 or 0.25 in an attempt to demonstrate what different levels of compliance with mitigation could do to the daily cases in the pandemic. Diameter of 0.4 corresponds to little or no mitigation. Figure 5-2 shows the results from such a simulation.

Implementing mitigation initially (first two phases) and relaxing mitigation completely thereafter effectively undermines all mitigation efforts done during the first two phases (without formal evaluation or investigation, it appears this was shown in a Washington Post report where NM and AZ data were compared [1]). The peak cases, hospitalizations, and deaths if no mitigation is enforced/practiced when the curve is flat ('re-opening too soon') are very close in value to the respective peaks in the case with no mitigation at all.

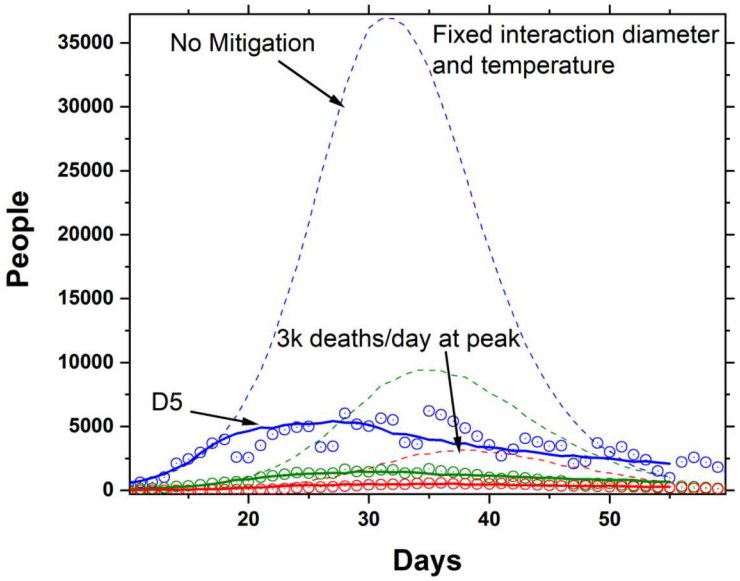

Figure 5-1 Model predictions for NYC with no mitigation strategy implemented. Both the temperature and initial interaction diameter persist for the duration of simulation. The data points are NYC daily data. The solid lines is the 'best fit' of model to data (D5 in interaction diameter PS). The dashed lines are simulation results if no mitigation was implemented. The number of daily deaths increases drastically to roughly 3100 at peak. Also notable is that the number of hospitalizations increases to 10k per day for about 30 days.

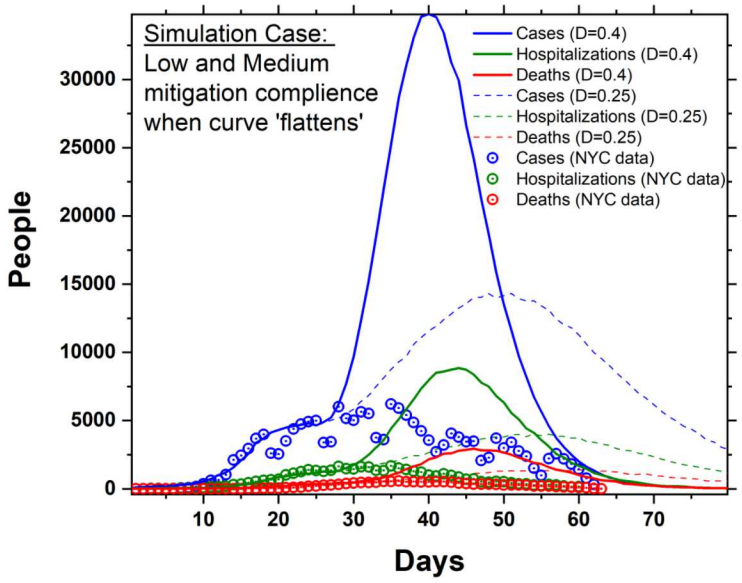

Figure 5-2 Model predictions for different compliance with mitigation at the point when number of positive tests begins to 'flatten'. In the plot, NYC data shows the number of people who tested positive reached a plateau near day 25 (at end of phase 2, see Fig. 2-3), at which point, in the model, conditions are set for low or medium complience with mitigation which means that the interaction diameter, D = 0.4 or 0.25, respectively.

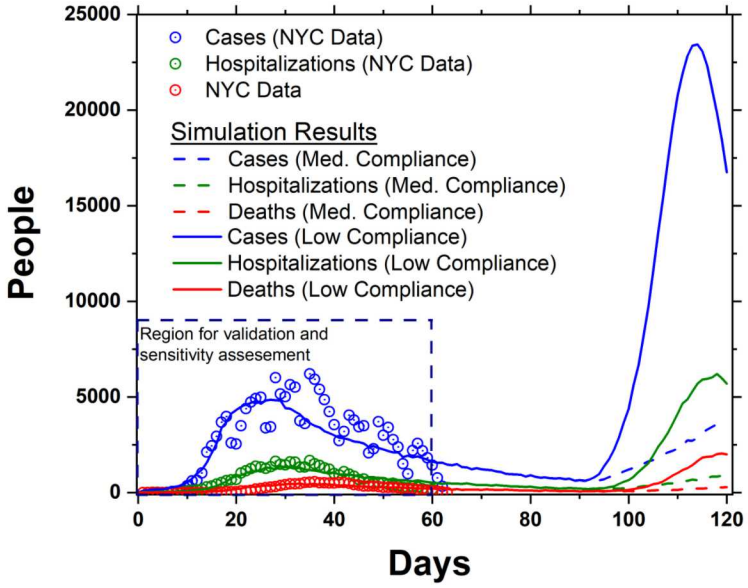

Figure 5-3 Model predictions for different compliance with mitigation at a hypothetical 're-opening' date, arbitrarily chosen to be in early July. NYC data is shown inside dashed lined box. Model is propagated in time for an additional 30 days (month of June) from day 60 to day 90. At that point, similarly to Fig. 5-2, the model is further progressed forward under low or medium compliance with mitigation (D = 0.4 or 0.25, respectively). Solid [dashed] lines represent low [medium] compliance with mitigation.

5.3. Medium and Low compliance at 'reopening'

Obviously, from the previous results, the mitigation must be continued into the decline of daily cases numbers, but the question remains when to re-open? In the model, we can add one more phase such that re-opening is now phase 4. Phase 3 is prolonged to ensure that both the magnitude of cases is lower and the slope of the curve is decreasing. End of phase 3 and start of phase 4 occurs roughly 90 days into the simulation, such that this phase transition would occur in early July. Once phase 3 is complete, the simulaiton is re-started with no or low compliance with mitigation (diameter of 0.4 or 0.25, respectively). The simulation results are shown in Fig. 5-3.

The model suggests a very unfortunate scenario if re-opening means going back to behaviors similar to before the pandemic. Going back to 'normal' behaviors can undo three months of work in a matter of weeks. Results obtained by propagating the model forward with some reasonable level of compliance certainly suggests a much more favorable outcome in terms of daily cases. Even with moderate compliance with mitigation, the model suggests a situation similar to that observed in NYC late March could repeat itself in August.

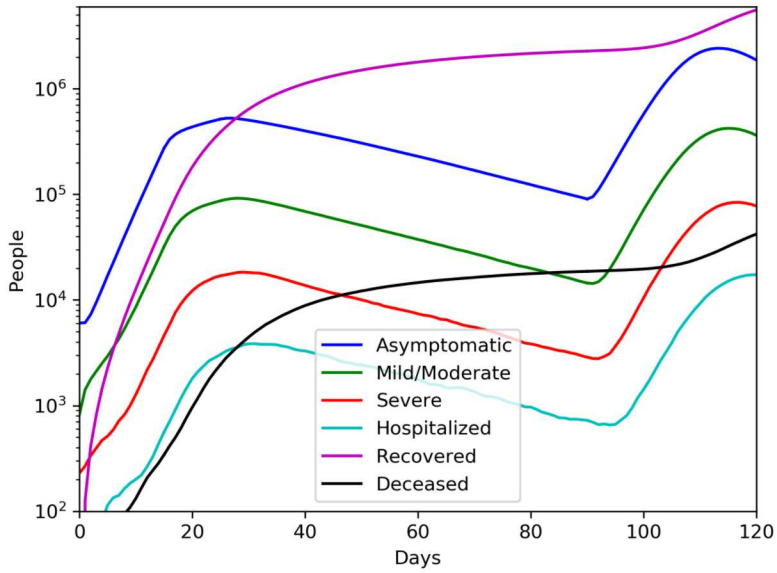

Figure 5-4 Model predictions for asymptomatic, mild and severe population dynamics. Notice the plot is semi-log. This figure is the 'global' population in the pandemic case modeled in Fig. 5-3. The peak for asymptomatic population leads the hospitalized population by $\approx$ 10 days. Once hospitalizations begin to increase, model suggests that there is a (two orders of magnitude) larger asymptomatic population amidst.

5.4. Estimate for Asymptomatic Population

How can this be? How can the infection begin to spread so rapidly after 3 moths of mitigation? Doesn't the low number of positive cases, hospitalizations, and deaths suggest the disease spread is more controlled? From the model inputs and results, it appears these three populations (cases, hospitalizations and deaths) is just a fraction of the story. The rest is governed by the dynamics of mild, asymptomatic and healthy populations. In Fig. 5-4 we plot the total number of people in each of the disease stages, as outlined in Fig. 4-1, versus time for the case of low compliance with mitigation at 'reopening' (early July arbitrarily chosen as re-opening for demonstration). These plots show total values (not daily) which can increase or decrease depending on mitigation and progression rates.

The missing part of the story is that the number of asymptomatic, mild/moderate and severe combined is much larger than the hospitalized and deceased cases. Re-opening without a good strategy to deal with these populations means that there are a large number of people who resume normal operations, unknowingly infecting people they come in contact with. For safe re-opening, it may be just as (or more) important to know how the asymptomatic and mild/moderate population is trending.

An additional estimate of what the asymptomatic population is doing in the 'background' was somewhat brought to light by serological testing results. Recall that the asymptomatic population estimate arises from the Diamond Princess cruise ship and the hospital in New York (and linear interpolations for the age groups in between). NYC reported estimated 21% of the population had been exposed to the virus in early April. In our model, it seems reasonable that the recovered and asymptomatic population would account for the population that would test positive during serological testing (remember that asymptomatic population reaches the recovered stage according to the rates and probabilities pertinent). Simulation results for mid-April show that these two populations amount to $\approx 2M$ individuals or roughly 23% of the population. Removing the asymptomatic population from the calculation bring the estimate to 18%.

6. CONCLUDING REMARKS

6.1. Future Work

There are quite a few and obvious improvements that could be made to the model and work flow presented. The initial scope of this project was to assess if DSMC infrastructure could be used for modeling the pandemic. Though it appears that is possible, the interaction model is very basic, the parameters for the interaction itself were determined empirically, and validation data could certainly be improved. Next we describe relatively simple additions to the model that could improve the model and the people traits that are assigned to the particles in the simulation.

Further developing the model could include emerging details about race, co-morbidities, and other factor that appear to play a role on how the disease affects different people according to ethnicity or under-lying health conditions. This can be accomplished by augmenting the demographics to include these and, potentially, other relevant factors so long as these can be assessed in a statistically accurate manner.

Mobility vs. age group is another relatively simple change to the model that we could implement. This basically requires no additional code development, and could be handled by the model as is, however, again, finding high quality data to inform the inputs would be vital. Potentially, this improvement would enable estimates of what it means for the pandemic progression in general when a select age group is allowed to ‘re-open’ under different mitigation conditions.

Mobility vs infectious stage seems to be another modification that would improve model accuracy. Consider an individual that progresses through the disease stages eventually reaching the hospital. As this individual becomes more and more ill, the less this person would continue to interact with other individuals, eventually reaching the hospital and interacting, potentially, only with people such as Doctor and Nurses whom presumably are wearing PPE and other hospitalized patients that are already infected. At present the model assumes complete quarantine for hospitalized patients. Adding in a low but non-zero mobility would clearly impact the likelihood that these people continue to infect other individuals. Though this appears a reasonable modification to the model, and simple to implement, it would likely affect the results very minimally since the number of severe and hospitalized patients is much lower than the estimated asymptomatic and mild population.

The number of interactions that result in additional infections increases with population density. Capturing population density gradients in the current model and work flow is not possible and implementation of this capability does require a significant effort. Restricting people movement such that these density gradients develop could be achieved by using shape manifolds (random motion restricted to these shapes [17], shapes represent city blocks, cities, states, countries), cell phone data could also be used for this purpose, as well as using the PIC algorithm and fields (Aleph is a PIC-DSMC code) to confine particles to pre-determined regions.

Once more details are known about immunity and virus mutation, these are also two phenomena that could be modeled with relatively simple additions to the model itself. Considering Fig. 4-1, the addition would be to allow recovered individuals to ’loop-back’ and become susceptible (at

the appropriate rate) if immunity holds for a relatively short time period; modeling the effects on pandemic dynamics due to virus mutation would follow in a similar manner.

6.2. Discussion and Conclusions

Existing methods and models for epidemic/pandemic modeling cover a wide range of required computational effort as well as fidelity. Using DSMC for modeling disease progression dynamics as shown in this report can potentially bridge an existing gap between SEIR (continuum) and ABM/ILM (individual) modeling in a way that enables capturing a range of people characteristics while being computationally efficient.

The model presented uses physical inputs in the way of areas, population densities and demographics, disease progression rates and probabilities. Estimates for interaction diameters and temperatures (i.e. people mobilities) were determined empirically but could very well be deduced experimentally. Of course, one would have to devise experiments to determine these values in a manner that does not place test subjects at risk. The work presented here convincingly shows that the DSMC approach can be used to model the dynamical characteristics of COVID-19 and extended to model other epidemics/pandemics where the rates and probabilities for disease spread are known. This new modeling tool developed still has many ways in which it can be improved but the relatively short time period for its development ($\approx$4 weeks) did not allow the authors to explore many additions/improvements.

The model progresses in three phases, one with a high probability of infection (large interaction diameter) corresponding to disease spread with no mitigation, next with a medium diameter corresponding to schools and other public services closing as well as people becoming aware of disease spread, and the third phase uses an even smaller interaction diameter that may be associated with stay home order in addition to the mitigation strategies of the second phase. In return, we can then evaluate disease progression after a certain time period by falling back to either of these phases. The effect of medium and low compliance with mitigation in NYC in early June (day 90) shows that maintaining a decreasing slope is relatively sensitive to compliance with mitigation. Figure 5-3, shows that over a period of 2-3 months with strict mitigation measures, the curve begins to slope downwards and tending toward small numbers of infected individuals. However, this figure also shows that low mitigation compliance during reopening, at the point where there is still a significant number of infected individuals, will result in large increases for all stages of the pandemic within 1-2 week period, essentially nullifying all the efforts during 2-3 month period of mitigation (lockdown).

The results and sensitivities of the model suggest that it may be possible to achieve a highly consistent model with a full sensitivity analysis on rates, probabilities, interaction details coupled to timely and accurate data.

Figure 5-4 shows that peak of positive cases leads peak of deaths by $\approx$ 10 days and leads hospitalizations by $\approx$ 1 week. The probability that a certain individual will be asymptomatic or not is assumed in the model to follow from test results obtained from the Diamond Princess cruise ship and a NYC hospital, as discussed earlier. A problem with ‘re-opening too soon’ without sufficient testing is that the asymptomatic population can still be large. The number of

asymptomatic individuals is significantly larger with respect to the rest of the infected population. While the hospitalized and positive cases can be trending downward and low numbers maybe achieved, the number of asymptomatic individuals also trends downward but the magnitude of the asymptomatic population is predicted to be significantly higher. It appears then that safe reopening would have to consider 1) low number of hospitalizations and deaths as well as a 2) low number of positive cases (including asymptomatic, mild and severe). Understanding the last observation (item 2 above) is, of course, only possible with randomized, wide-spread testing.

Serological test results in NYC in early April showed that the number of people exposed to the virus was 21.2% among NYC population, with lower positive seroligical test cases in nearby regions ($4\% - 17\%$ for rest of NY State and for Long Island, respectively) [8]. This figure would likely correlate to the ratio of recovered plus asymptomatic population to the total population of NYC. Fig. 5-4 suggests that in early April, the results presented here are in close agreement with the serological results. We believe this is a figure of merit of this report since model assumptions are minimal, asymptomatic population plays such a key role in the pandemic, and has not, to our knowledge, been estimated by other models.

The model presented here appears to be robust, sensitive to inputs and is computationally efficient. Since the model depends only on statistically accurate rates and probabilities, statistical fluctuations of test data are smoothed out by the inherent statistical Monte-Carlo sampling of these values. Consider, for example, using 'fits' to the modeling results presented in earlier sections as estimates for the R_0 value used in SEIR models, or using DSMC-like modeling for local effects, coupled with ABM, for global effects. The variability in the R_0 used in SEIR models and the likelihood that wrong assumptions enter ABM simulations could be reduced. In combination with these other modeling tools (SEIR or ABM), it is possible to envision significantly more powerful tools for disease spread modeling. By itself, however, it appears this approach would be a great tool for mitigation strategy design. Importantly, using existing infrastructure for DSMC (in terms of software and hardware) makes this approach computationally efficient and is massively parallel. Multiple small cases could run simultaneously (i.e. simulations shown in this report use 60 processors for 10's of minutes) by taking advantage of infrastructure developed for parameter and/or sensitivity studies. Larger cases (i.e. U.S. population with each person modeled individually) could be accomplished with the respective increase in the number of processors since the PIC-DSMC algorithm in Aleph scales well with number of processors. Even to model the entire U.S. population, we would require relatively modest computational resources. For NYC, simulating 8M people over 60 days uses 60 processors for 10's of minutes (computational time). We expect the entire U.S. population would require less than 512 processors for the same amount of computational time. Aleph has been demonstrated to scale well over 100k processors for runs that use hours/days of computer time. This suggests that, if this modeling approach captures sufficient people traits during a pandemic, the entire world population could be modeled with relatively modest computational resources.

To conclude, we have demonstrated that DSMC can be used for modeling dynamics of a pandemic, the model is robust with respect to inputs and can track people and pandemic characteristics (demographics and stages). Decades of software and hardware and DSMC algorithm development are leveraged here to provide an efficient tool. The model captures NYC data well using three different phases. The details from these phases can be used to assess 'what

if' scenarios when considering re-opening plans. A potential figure of merit (if it can be futher validated) of this modeling approach, and this model in particular, is that we can estimate the asymptomatic population as well as the role it plays in the pandemic dynamics. The model results suggest that, for re-opening, hospitalizations, positive cases and deaths are a good indicators but understanding how the rest of infected population is trending in the background is key; model suggests the asymptomatic population has the same trend as hospitalizations but leads by roughly 10 days and is two orders of magnitude larger. Without widespread and random testing, people in the asymptomatic phase, and unbeknown to them or to public health officials, are potentially spreading the disease. As a consequence, with hospitalizations and deaths as the only evidence for disease spread, it is difficult to understand how the rest of disease phases are trending. The model suggests that the larger number of asymptomatic cases can still generate spikes even after relatively few hospitalizations or deaths are recorded.

REFERENCES

[1] As arizona struggles, neighboring new mexico found a more cautios path to sustained growth. https://www.washingtonpost.com/business/2020/07/21/arizona-struggles-neighboring-new-mexico-found-more-cautious-path-sustained-growth/. Accessed: 2020/08/05.

[2] Covid-19 disease progression rates as reported by cdc. https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html#:~:text=Symptoms%20may%20appear%202%2D,exposure%20to%20the%20virus. Accessed: 2020/07/30.

[3] Epidemic calculator. https://gabgoh.github.io/COVID/. Accessed: 2020-06-22.

[4] Ihme model, uw medicine, university of washington. https://covid19.healthdata.org/united-states-of-americ a. Accessed: 2020-07-01.

[5] Nyc daily data for cases, hospitalizations and deaths. https://www1.nyc.gov/site/doh/covid/covid-19-data.page. Accessed: 2020/07/10.

[6] Public health responses to covid-19 outbreaks on cruise ships âĂŤ worldwide. https://www.cdc.gov/mmwr/volumes/69/wr/mm6912e3.htm. Accessed: 2020-07-01.

[7] r_0 estimated for ny as a function of time. Internal SLN communication. Center 1400 Town Hall. Obtained: 2020/06/01.

[8] Serological testing results for new york and california. https://www.360dx.com/infectious-disease/new-york-california-serology-studies-give-early-estimates-covid-19-prevalence. Accessed: 2020-06-01.

[9] We tested all our patients for coronavirus âĂŤ and found lots of asymptomatic cases. https://www.washingtonpost.com/outlook/2020/04/20/we-tested-all-our-patients-covid-19-found-lots-asymptomatic-cases/. Accessed: 2020-07-01.

[10] Matthew T. Bettencourt, Jeremiah J. Boerner, Paul S. Crozier, Andrew S. Fierro, Anne M. Grillet, Russell W. Hooper, Matthew M. Hopkins, Thomas P. Hughes, Harold E. Meyer, Christopher H. Moore, Stan G. Moore, Lawrence C. Musson, and Jose L. Pacheco. Aleph manual. Technical report SAND2017-10343, Sandia National Laboratories, Albuquerque, New Mexico 87185 and Livermore, California 94550, May 2017.

[11] GA Bird. Molecular gas dynamics and the direct simulation of gas flows oxford. *Clarendon Press*, 199:9–1, 1994.

[12] Charles K Birdsall. Particle-in-cell charged-particle simulations, plus monte carlo collisions with neutral atoms, pic-mcc. *IEEE Transactions on plasma science*, 19(2):65–85, 1991.

[13] Romulus Breban, Raffaele Vardavas, and Sally Blower. Theory versus data: how to calculate r 0? *PLoS One*, 2(3):e282, 2007.

[14] Timothy C Germann, Kai Kadau, Ira M Longini, and Catherine A Macken. Mitigation strategies for pandemic influenza in the united states. *Proceedings of the National Academy of Sciences*, 103(15):5935–5940, 2006.

[15] Charles M Macal and Michael J North. Agent-based modeling and simulation. In *Proceedings of the 2009 Winter Simulation Conference (WSC)*, pages 86–98. IEEE, 2009.

[16] Siddhartha Verma, Manhar Dhanak, and John Frankenfield. Visualizing the effectiveness of face masks in obstructing respiratory jets. *Physics of Fluids*, 32(6):061708, 2020.

[17] Sheng Yi, Hamid Krim, and Larry K Norris. Human activity modeling as brownian motion on shape manifold. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 628–639. Springer, 2011.

DISTRIBUTION

Hardcopy—Internal

Number of Copies	Name	Org.	Mailstop
1	D. Chavez, LDRD Office	1911	0359

Email—Internal (encrypt for OUO)

Name	Org.	Sandia Email Address
Daniel Barnes	02585	dbarnes@sandia.gov
Michael Gallis	01513	magalli@sandia.gov
Ronald Manginell	01513	rpmangi@sandia.gov
Jeffrey Payne	01510	jlpayne@sandia.gov
Russell Hooper	01463	rhoope@sandia.gov
Zakari Eckert	01463	zeckert@sandia.gov
Melissa Finley	06820	mfinley@sandia.gov
Sally Jensen	01513	sjjense@sandia.gov
Edward Cole	05000	coleei@sandia.gov
Patrick Finley	08721	pdfinle@sandia.gov
Kathy Simonson	06000	kmsimon@sandia.gov
Keith Matzen	01000	mkmatze@sandia.gov
Walter Witkowski	01540	wrwitko@sandia.gov
Anne Grillet	01513	amgrill@sandia.gov
Stan Moore	01444	stamoor@sandia.gov
Thomas Hughes	2585	thughe@sandia.gov
Technical Library	01177	libref@sandia.gov
Technical Library	01177	libref@sandia.gov

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.