A Kinetic Monte Carlo (kMC) Model for the Dynamics of Infectious Diseases

Talla Babou*, Julianna Capece†, Uzochi Gideon‡, Juliana Hirt§, Justin Le¶, Tamanna Tamanna‖, Haoru Yang#, SangEun Youn††, and Xinyue Yu‡‡

Abstract. The SIR model is used to describe the dynamics of infectious diseases. With the availability of vaccines, we consider susceptible, infectious, recovered, and vaccinated individuals, and investigate the interplay between such states. We refine an existing SIR kMC model to simulate the dynamics between infectious states over time. In doing so, we implement vaccination into the current SIR model, incorporate additional parameters into the model that affect state transitions (such as age, mobility, social activity, etc.), and employ the cell-list method to compute interactions between individuals based on spatial locations. These three aims work in tandem to effectively simulate infectious disease progression by first building a base model with vaccination, extend this to weighted, faceted *person*-level dimensions, and consider the spatial configurations of individuals. This framework demonstrates a holistic approach to disease modeling as the SIRV and SIRV model with facets validate one another, while the incorporation of cell-listing speeds up the SIRV models.

Key words. kinetic Monte Carlo (kMC), simulation, computational modeling, cell-list

AMS subject classifications. 92D30, 60J65, 65C05

1. Introduction.

Infectious disease modeling plays an important role in understanding how diseases spread through a population and how public health interventions may affect that spread. Classical compartmental models, such as the SIR model, describe population-level transitions among susceptible, infected, and recovered individuals. While these models provide a useful mathematical foundation, they often rely on simplifying assumptions such as homogeneous mixing and averaged population behavior. In many real-world settings, however, disease transmission depends on individual-level differences, spatial movement, local interactions, and intervention mechanisms such as vaccination.

To incorporate these features, this study considers a kinetic Monte Carlo (kMC) framework for infectious disease dynamics. In this setting, individuals are represented as agents whose health states change according to probabilistic rules [14]. Rather than describing only the average behavior of the population, a kMC-based simulation allows us to model stochastic transitions among susceptible, infected, recovered, and vaccinated states at the level of individual agents. This makes the framework suitable for studying how local interactions, personal characteristics, and random events influence the overall disease dynamics.

Recent work in infectious disease modeling has increasingly emphasized agent-based and stochastic approaches, especially when individual heterogeneity, mobility, spatial structure, and intervention strategies are important. These approaches provide greater flexibility than purely deterministic compartmental models, but they also introduce new computational challenges. In particular, when

*Massachusetts Institute of Technology, tbabou@mit.edu
†Florida State University, jc22@fsu.edu
‡Rice University, ug6@rice.edu
§University of Wisconsin-Milwaukee, jhirt@uwm.edu
¶University of Texas at Austin, justinle@utexas.edu
‖New Jersey Institute of Technology, tl446@njit.edu
#University of Colorado Boulder, haoru.yang@colorado.edu
††KyungHee University, shfkfk179@khu.ac.kr
‡‡University of California, Los Angeles tracy@math.ucla.edu
This research was conducted at the 2026 Graduate Student Mathematical Modeling Camp.
All authors contributed equally.

infection probabilities depend on pairwise distances or local contacts among individuals, the cost of checking all possible interactions can become expensive for large populations. Therefore, improving both the biological realism and the computational efficiency of such models is an important direction.

Motivated by these considerations, our study is organized around three related research questions. First, we refine an existing kMC model by incorporating vaccination and additional transitions among susceptible, infected, recovered, and vaccinated states. Second, we investigate how weighted, faceted person-level dimensions can influence transition probabilities in a heterogeneous population. Third, we explore a more efficient method for determining the probability that a susceptible individual becomes infected in a restricted domain. These three directions are developed through three approaches. The first two approaches extend the structure of the SIRV model by modifying state transitions and individual-level representations. In contrast, the third approach focuses on the computational implementation of the infection process by introducing a cell-list method to reduce the cost of local interaction searches. Together, these approaches aim to improve the flexibility, realism, and efficiency of kMC-based infectious disease simulations.

- **RQ1:** How does the implementation of vaccination into the current kMC SIR model affect the interaction between susceptible, infectious, recovered, and vaccinated states?
- **RQ2:** How do weighted, faceted *person*-level dimensions influence transitions among susceptible, infected, recovered, and vaccinated states in a kMC SIRV model?
- **RQ3:** What is a more efficient method for determining the probability a susceptible individual becomes infected in a bounded domain?

2. Theory and Methods.

Building on the methodological gap outlined above, this study adopts a phase-based modeling approach organized into three stages: Approach 1 (A1), Approach 2 (A2), and Approach 3 (A3). Each stage extends the previous model by introducing additional complexity to the representation of individuals and their transitions among susceptible, infected, recovered, and vaccinated states. The following subsections describe the purpose, structure, and implementation of each approach.

2.1. Approach 1.

We describe the kMC methods used to transition from various states: susceptible, infected, recovered, vaccinated. In modifying the SIR model, we introduce a return from recovered to susceptible and a possible transition from susceptible to vaccinated, as depictied in Figure 1. We refine the currently existing kMC model by considering immunity levels and vaccination rates, and discuss the advantages of the kMC model over the ODE model.

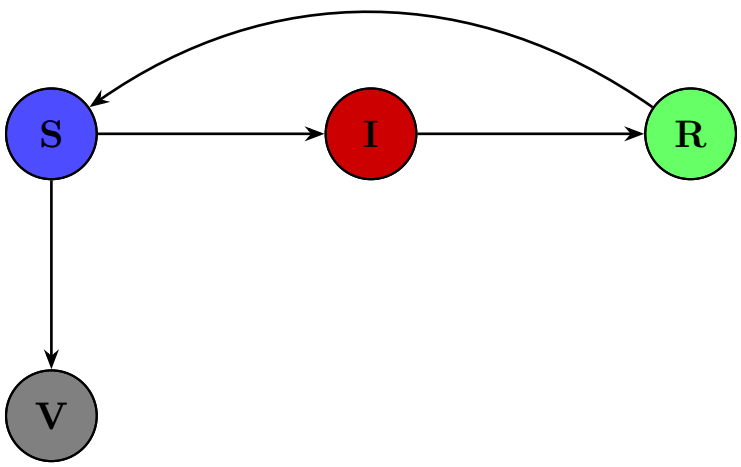

Figure 1: Standard SIRV model.

2.1.1. kMC Approach to the base SIRV Model. The population consists of $N = 1000$ individuals and observed over a period of $T = 365$ days. Each individual belongs to one of four states: Susceptible (S), Infected (I), Recovered (R), or Vaccinated (V). Individuals move continuously within a square domain of side length $L = \sqrt{N/\rho}$, where $\rho = 0.5$ is the population density per unit area.

At each time step, the infection exposure for a susceptible individual i is calculated as:

$$P_i = \sum_{j \in I} s_A(j) \exp\left(-\frac{r_{ij}^2}{2\sigma_{\text{Inf}}^2}\right),$$

where $s_A(j) \in [0, 1]$ is the social activity level of infected individual j,
$r_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$ is the distance between individuals i and j, and σ_{Inf} determines the spatial range of infection. The Gaussian kernel ensures that nearby infected individuals contribute more strongly to the infection risk than distant ones.

The model assumes some rules which are defined as follows. Individuals undergo Brownian motion at each time step with periodic boundary conditions. An infected individual recovers deterministically after 14 days of infection. Upon recovery, each individual receives immunity for some amount of time, which is chosen from a mixture of fast and slow-waning exponential distributions, stored as *ImmunityTime*. Once an individual's recovery period, *SARdays*, meets their assigned *ImmunityTime*, they return to the susceptible state and individuals transition from $R \to S$. At each time step, each susceptible S individual is independently vaccinated with probability p and moved from $S \to V$. Vaccinated individuals V are assumed to be fully immunized for this base model and do not transition back to susceptible.

2.1.2. kMC Approach to the Bidirectional SIRV Model. In this section, we extend the previously developed SIRV model by incorporating a reverse transition from the vaccinated compartment (V) to the susceptible compartment (S), thereby making it bidirectional as illustrated in Figure2. This extension accounts for the fact that vaccine-induced immunity is temporary and diminishes over time. We assume that vaccinated individuals are protected for a constant period, referred to as the Vaccine Immunity Duration (*VID*). After this period has elapsed, immunity wanes, causing individuals to re-enter the susceptible class and become eligible for infection.

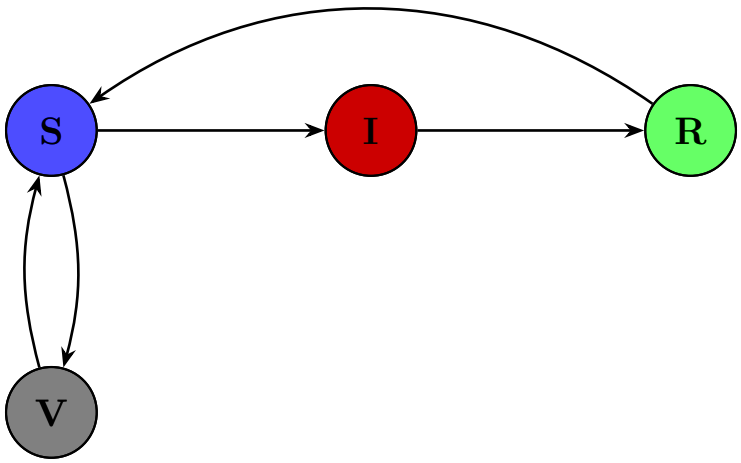

Figure 2: SIRV model with waning immunity and vaccination.

For this bidirectional model, we consider two different cases regarding the duration of naturally acquired immunity. In the first case, we assume that all recovered individuals possess the same constant natural immunity duration, denoted by ID. Consequently, individuals in the recovered

compartment return to the susceptible class after a fixed period. In the second case, we consider a more realistic scenario in which the duration of natural immunity varies among individuals. Some recovered individuals experience rapid waning of immunity, while others retain immunity for a longer period. This heterogeneous immunity structure better reflects what is observed in real-world populations.

Finally, we compare the outcomes of these two cases to investigate the impact of uniform versus heterogeneous immunity durations on disease transmission dynamics.

2.2. Approach 2. This section documents the theoretical armature that serves as the foundation for the Approach 2 (A2) model. Specifically, A2 seek to explore how weighted, faceted *person*-level dimensions shape transitions among susceptible, infected, recovered, and vaccinated states in a kMC SIRV model. This discussion begins in subsection 2.2.1 by situating susceptibility factors within Gärdenfors's theory of conceptual spaces [7], establishing how *person*-level dimensions can be represented as properties of susceptibility. It then turns to an overview of the solution package architecture in subsection 2.2.2, which outlines how these theoretical commitments are translated into a modular OOP computational workflow. Additionally, subsection 2.2.3 describes the synthetic data generation and population initialization process used to instantiate heterogeneous agents. Finally, subsection 2.2.4 details the simulation model design and rationale; with particular attention to how the A2 model extends the baseline SIRV structure through facet-informed individual representation.

The results and analysis for A2 are addressed in section 3.2.

2.2.1. Susceptibility Factors in Conceptual Space. The A2 model begins by developing a grounded, faceted understanding of a *person* as represented within the SIRV model. In S.R. Ranganathan's theory of faceted classification, complex entities can represented through analytically distinct facets; allowing a subject to be decomposed into meaningful semantic attributes of description [17]. This establishes a useful foundation for identifying attributes through which a synthetic individual may be represented, such as demographic, geographic, behavioral, and epidemiological characteristics. However, the extended A2 model requires more than the identification of facets as discrete descriptive attributes. A2 posits that susceptibility should thus be understood not as a fixed or binary condition, but as a variable property shaped by the configuration and relative intensity of individual attributes. For this reason, the A2 model extends faceted representation through the adoption of Gärdenfors's theory of conceptual spaces [7, 9]. As summarized in Figure 3, the translation of facets into orderable dimensions allows the A2 model to support graded differentiation between individuals. This dimensional structure enables the construction of a *person* concept: a synthetic individual representation in which susceptibility functions as the central property, defined through six underlying dimensions.

Gärdenfors's conceptual spaces model provides the theoretical structure for the A2 extension, organizing representation across three hierarchical levels: dimensions, properties, and concepts [7, 9]. Dimensions define the underlying axes of a conceptual space, allowing synthetic individuals to be positioned and compared according to measurable or orderable forms of variation. In A2, these axes are specified as *age*, *comorbidity*, *social activity*, *geography*, *mobility*, and *vaccine acceptance*. For example, *mobility* is encoded as immobile, assisted, or independent. Although initially identified through faceted representation, these values reflect ordered degrees of mobility and are therefore modeled as a dimension capable of supporting meaningful comparison within conceptual space.

From these dimensions, properties emerge as meaningful regions or configurations within the conceptual space. In the present model, *susceptibility* functions as an organizing property, orchestrating the six dimensions into a coherent basis for individual variation. The resulting *person concept* operates as the representational unit for each synthetic agent: a structured individual profile in which susceptibility is expressed through dimensional position rather than assigned as a uniform

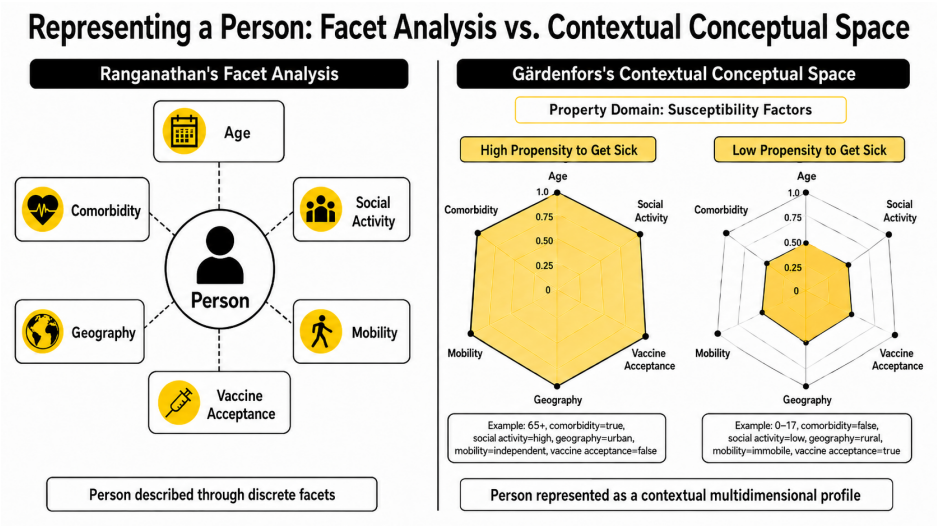

Figure 3: Representing *person*-level susceptibility factors as discrete facets and as contextual dimensions within a conceptual space [17, 7, 9].

population-level parameter.

Beyond the hierarchical relation between dimensions, properties, and concepts, two features of Gärdenfors's model are especially relevant to A2: weighting and dynamics. Weighting is particularly germane, as Gärdenfors states that "salience is context-dependent" [8, p. 38], with weights used to structure similarity judgments within conceptual space. The A2 model adopts this position epistemologically, applying weights according to the social and epidemiological conditions encoded in the simulation. This allows the same set of dimensions to contribute differently to susceptibility as risk, exposure, and vulnerability are defined within a given modeling context. Dynamics captures the possibility that conceptual representations may shift as new information is introduced. While A2 initializes susceptibility from static synthetic dimensions, this dynamic capacity provides a theoretical basis for future extensions in which *person*-level susceptibility may be updated across simulation executions.

2.2.2. Architecture of the Solution Package. The A2 solution package was implemented in Python 3.14.6 using a modular object-oriented programming (OOP) design, with the associated software package archived through Zenodo [2]. As shown in Figure 4, the architecture separates input configuration, synthetic data generation, simulation control, reaction logic, and analysis into distinct components. This structure supports maintainability and reproducibility by allowing each module to perform a specific function within the broader simulation workflow [22].

The full workflow is driven by YAML and JSON configuration files, which define the model parameters, population settings, and risk multipliers used during execution. This architecture decision removes the need to directly edit source code when adjusting simulation conditions. The synthetic data generator initializes the population, assigns *person*-level dimensions, identifies the initially infected population, and establishes starting positions for kMC movement. Further details on this generator module are presented in 2.2.3. The simulation module then controls the looping structure across time spans and epochs, while the reactions module contains the functions used to calculate and perform SIRV state transitions. These interactions are described in detail in 2.2.4. Finally, the analysis module provides standard SIRV outputs alongside dimension-based comparisons, allowing results to be examined both at the population level and through the weighted *person*-level dimensions introduced in A2.

2.2.3. Synthetic Data Generation and Population Initialization. The synthetic data generator module serves as the initialization point for the A2 simulation pipeline. Its purpose is to generate

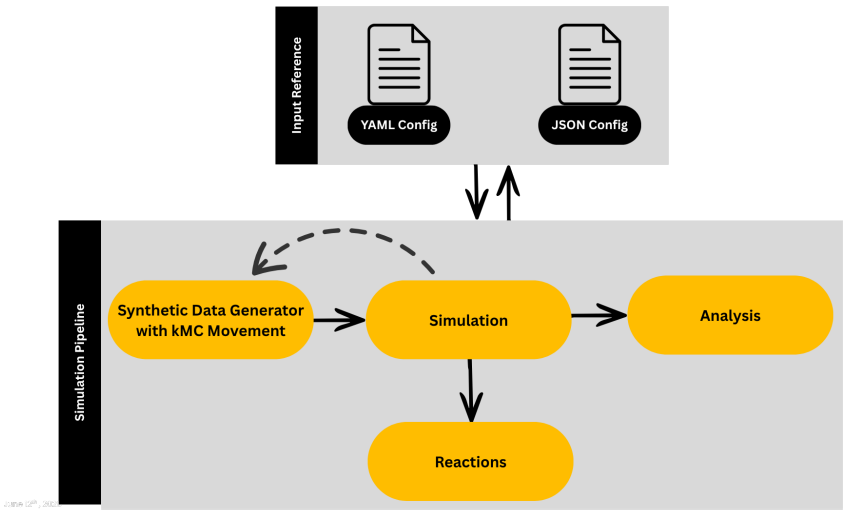

Figure 4: Architecture of the A2 solution package, showing the configuration-driven workflow connecting synthetic data generation, simulation control, reaction logic, and analysis.

a schema-aligned synthetic population of *people*, assign each *person* a unique GUID, and initialize both static *person*-level dimensions and dynamic SIRV-related states. The module is driven by the configuration files described in subsection 2.2.2, which specify population size, initial infection count, random seed, sampling weights, and infection risk multipliers. This configuration-driven structure allows the population generation process to be reproduced without direct modification to the Python source code. The configured sampling weights and infection risk multipliers used for A2 are summarized in Table 1.

Dimension	Encoded Value	Population Weight	Risk Multiplier
Age	0–17	0.22	0.50
Age	18–49	0.45	0.75
Age	50–64	0.20	0.75
Age	65+	0.13	1.00
Comorbidity	True	0.25	1.00
Comorbidity	False	0.75	0.50
Social Activity	High	0.25	1.00
Social Activity	Medium	0.50	0.75
Social Activity	Low	0.25	0.50
Geography	Urban	0.80	1.00
Geography	Rural	0.20	0.50
Mobility	Independent	0.85	1.00
Mobility	Assisted	0.12	0.75
Mobility	Immobile	0.03	0.50
Vaccine Acceptance	True	0.75	0.50
Vaccine Acceptance	False	0.25	1.00

Table 1: Encoded values, population weights, and risk multipliers for the six dimensions used in the A2 model.

Each generated agent is assigned six static dimensions: *age*, *comorbidity*, *social activity*, *geography*, *mobility*, and *vaccine acceptance*. *Age* represents the individual's age group and provides a demo-

graphic basis for differentiating susceptibility [3, 13]. *Comorbidity* indicates whether an individual has one or more underlying health conditions, allowing the model to represent medically relevant variation in infection risk [18, 21]. *Social activity* approximates the degree to which an individual participates in contact-generating activity [12], while *geography* distinguishes between urban and rural settings [16, 20]. *Mobility* captures whether an individual is independent, assisted, or immobile, providing an ordered dimension through which movement can be represented [10, 23]. Finally, *vaccine acceptance* indicates whether an individual is willing to accept vaccination, linking *person*-level attributes to vaccination-related state transitions [6, 11, 19].

For population generation, each dimension is assigned according to configuration-defined sampling weights. These weights determine the probability that a given encoded value will be assigned to a synthetic individual during generation. For example, *social activity* is sampled across high, medium, and low activity levels, with medium activity assigned the greatest probability ($high = 0.25$, $medium = 0.50$, $low = 0.25$). In this way, the generator controls the composition of the synthetic population prior to the application of infection risk multipliers.

Infection risk multipliers are distinct from these generation weights: the former modify how assigned dimensions contribute to infection risk during simulation, while the latter shape the composition of the synthetic population. The infection risk multipliers operationalize the conceptual spaces' salience-weighting logic introduced in subsection 2.2.1. In A2, these multipliers specify how strongly each dimension value contributes to transition behavior across susceptible, infected, recovered, and vaccinated states. This allows the same generated population to be interpreted through a context-sensitive risk structure, where *person*-level dimensions do not merely describe agents but modify their transition probabilities.

After static dimensions are assigned, the module initializes the dynamic SIRV states by randomly seeding the initially infected population from the generated agents. The remaining agents are initialized as susceptible, with recovered and vaccinated populations set according to the configuration. Finally, the generator assigns initial spatial coordinates and supports Brownian-motion-based position updates for the kMC simulation. The resulting synthetic population is then exported as the starting agent layer for the A2 simulation.

2.2.4. Model and Simulation Design. We perform a kinetic Monte Carlo (kMC) simulation of a stochastic agent-based SIRV model, where agents update their positions every day via Brownian motion. In our SIRV model, agents are grouped into four infection states: susceptible, infected, recovered, and vaccinated. Let $\text{Status}_i(t)$ be the status of individual i at time t. The probability of an individual changing from susceptible to infected is a function of their distance to other infected individuals and infection risk multipliers. We have that

$$\text{Status}_i(t+1) = \begin{cases} I, & \text{with probability } P_i \text{ if } \text{Status}_i(t) = S \\ \text{Status}_i(t), & \text{otherwise,} \end{cases}$$

where

$$P_i = \min\left\{1, \sum_{j\in\mathcal{I}} \left(\prod_{k\in\mathcal{R}} M_{jk}\right) \exp\left(-\frac{r_{ij}^2}{2\sigma_{\text{Inf}}^2}\right)\right\}.$$

Here $\mathcal{I}$ is the set of infected individuals, $\mathcal{R}$ is the set of risk factors, including age, comorbidity, social activity, geography, mobility, and vaccine acceptance, M_{jk} is the risk multiplier for infected individual $j \in \mathcal{I}$ and risk factor $k \in \mathcal{R}$, and r_{ij} is the Euclidean distance from susceptible individual i to infected individual $j \in \mathcal{I}$.

An individual recovers from an infection after an infectious period that is Geometric(p_{rcd})-distributed with mean $1/p_{\text{rcd}}$:

$$\text{Status}_i(t+1) = \begin{cases} R, & \text{with probability } p_{\text{rcd}} \text{ if Status}_i(t) = I \\ \text{Status}_i(t), & \text{otherwise,} \end{cases}$$

where $p_{\text{rcd}} = 1 - e^{-1/\tau_{\text{rcd}}}$. For large τ_{rcd}, we have $1/p_{\text{rcd}} = 1/(1 - e^{-1/\tau_{\text{rcd}}}) \approx \tau_{\text{rcd}}$, since $e^{-x} \approx 1 - x$ for small x, so the mean infectious period is τ_{rcd} days.

After an individual is recovered, they become susceptible after an immunity duration. The state changes from recovered to susceptible according to:

$$\text{Status}_i(t+1) = \begin{cases} S, & \text{with probability } p_{\text{sd}} \text{ if Status}_i(t) = R \\ \text{Status}_i(t), & \text{otherwise,} \end{cases}$$

where $p_{\text{sd}} = 1 - e^{-1/\tau_{\text{sd}}}$, so that the immunity duration is Geometric(p_{sd}) with mean $\approx \tau_{\text{sd}}$ days.

The probability that a susceptible individual becomes vaccinated is a function of their risk multipliers and a Poisson distribution, which was chosen to simulate true vaccination rates in the United States in 2020 [4]. In particular,

$$\text{Status}_i(t+1) = \begin{cases} V, & \text{with probability } p_{\text{vac}} \text{ if Status}_i(t) = S \\ \text{Status}_i(t), & \text{otherwise,} \end{cases}$$

where $p_{\text{vac}} = \min\left\{1, \max\left(0, \prod_{k \in \mathcal{R}} M_{ik} \cdot \text{Poisson}(0.008)\right)\right\}$.

Similar to the transitions above, a vaccinated individual loses vaccine-induced immunity and returns to susceptible with daily probability $p_{\text{vd}} = 1 - e^{-1/\tau_{\text{vd}}}$, i.e.

$$\text{Status}_i(t+1) = \begin{cases} S, & \text{with probability } p_{\text{vd}} \text{ if Status}_i(t) = V \\ \text{Status}_i(t), & \text{otherwise,} \end{cases}$$

where $p_{\text{vd}} = 1 - e^{-1/\tau_{\text{vd}}}$, so that the vaccine-induced immunity duration is Geometric(p_{vd}) with mean $\approx \tau_{\text{vd}}$ days.

2.3. Approach 3. The previous approaches introduced the overall SIRV framework. In this section, we focus on one specific component of this model: the infection transition from susceptible individuals to infected individuals, namely, $S \to I$. Since this transition requires evaluating local interactions between susceptible and infected individuals, its direct computation can become costly as the population size increases. Therefore, our goal is to compute this infection process more efficiently.

Approach 3 presents a numerical method for efficiently computing local interactions between individuals in the agent-based simulation. The method is based on the cell-list algorithm [1], which organizes susceptible and infected individuals according to their spatial locations to accelerate neighborhood searches.

Here, we consider three primary discussions. First, we introduce the cell-list method, highlighting its computational advantages and implementation. Second, we develop two heuristic strategies that extend the standard cell-list algorithm: one for selecting an appropriate cell side length and another for determining the cell-list update frequency. Finally, we evaluate the computational efficiency and modeling accuracy when incorporating the customized cell-list method into the kinetic Monte Carlo (kMC) simulation.

2.3.1. Introduction of the Cell-List Method. In the agent-based simulation, each individual is represented as a particle moving in a bounded two-dimensional square domain. To determine

possible local interactions, such as infection events, we need to identify individuals who are close to each other.

For this purpose, we use the cell-list method. The simulation domain is divided into square cells with side length l_{cell}. Each individual is assigned to a cell according to its current position. If the position of individual i is

$$\mathbf{x}_i = (x_i, y_i),$$

then the corresponding cell index is given by

$$c_x = \left\lfloor \frac{x_i}{l_{\text{cell}}} \right\rfloor, \qquad c_y = \left\lfloor \frac{y_i}{l_{\text{cell}}} \right\rfloor.$$

Let r_c denote the cutoff radius for local interactions. That is, two individuals separated by a distance larger than r_c are not considered to interact. The cell size is chosen so that [15]

$$l_{\text{cell}} \geq r_c.$$

With this choice, any two individuals whose distance is less than or equal to r_c are located either in the same cell or in neighboring cells. Therefore, the cell-list provides a systematic way to identify candidate pairs for local interactions.

When computing the infection probability for a susceptible individual q, we only compute distances for infected individuals in the same cell as q or in any of the 8 neighboring cells.

2.3.2. Advantage of the Cell-List Method. The main advantage of the cell-list method is the reduction of computational cost. Without the cell-list method, all possible pairs of susceptible and infected individuals must be checked. For a population size N, the number of pairwise distance calculations required by the direct approach will grow according to $O(N^2)$.

Using the cell-list method, however, each individual is compared only with individuals located in the same or neighboring cells. If the population is approximately uniformly distributed, the average number of individuals in each cell remains bounded as N increases. As a result, the total number of distance computations grows approximately linearly with N [1]. This reduction is important for large-scale simulations because it allows local interactions to be computed efficiently without checking every possible pair of individuals.

2.3.3. Local Interaction Search. Let C_α and C_β be two neighboring cells. For each pair of individuals

$$p_\alpha \in C_\alpha, \qquad p_\beta \in C_\beta,$$

we compute the squared distance between them:

$$r^2 = \left\| \mathbf{x}_{p_\alpha} - \mathbf{x}_{p_\beta} \right\|_2^2.$$

If $r^2 \leq r_c^2$, then the two individuals are considered to be within the interaction range. For example, in the infection model, this pair is used to compute the infection probability.

This cutoff check is still necessary because not all individuals in neighboring cells are necessarily within distance r_c. The cell-list method only provides candidate pairs, and the exact distance condition is checked afterward.

We impose periodic boundary conditions on the square simulation domain. Let the domain be

$$[0, L) \times [0, L).$$

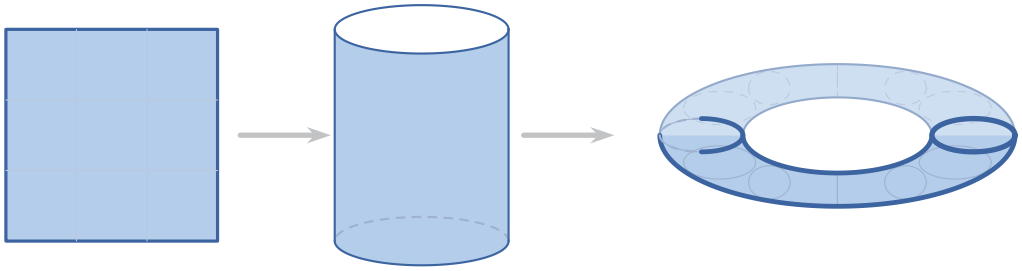

Figure 5: Schematic illustration of periodic boundary conditions on a two-dimensional square domain. Identifying one pair of opposite boundaries forms a cylinder, and identifying both pairs of opposite boundaries yields a torus. Thus, a particle leaving the domain through one boundary re-enters from the opposite boundary.

Under periodic boundary conditions, an individual leaving one side of the domain re-enters from the opposite side. After updating the position of individual i, we wrap the position back into the domain by

$$x_i \leftarrow x_i \bmod L, \qquad y_i \leftarrow y_i \bmod L.$$

This boundary condition removes artificial boundary effects and treats the simulation domain as a periodically repeated space.

When periodic boundary conditions are used, the distance between two individuals must be computed carefully. Two individuals may appear far apart in the original square domain, even though they are close across the periodic boundary.

To account for this, we use the minimum image convention. For two individuals located at

$$\mathbf{x}_i = (x_i, y_i), \qquad \mathbf{x}_j = (x_j, y_j),$$

we first compute the coordinate differences

$$\Delta x = x_i - x_j, \qquad \Delta y = y_i - y_j.$$

Then we apply the periodic correction:

$$\Delta x \leftarrow \Delta x - L \,\mathrm{round}\left(\frac{\Delta x}{L}\right),$$

$$\Delta y \leftarrow \Delta y - L \,\mathrm{round}\left(\frac{\Delta y}{L}\right).$$

The periodic distance between the two individuals is then computed as

$$r_{ij} = \sqrt{(\Delta x)^2 + (\Delta y)^2}.$$

This distance represents the shortest distance between the susceptible individual i and the infected individual j in the periodically repeated domain. Therefore, interactions across the boundary are treated consistently with interactions inside the domain.

2.3.4. Cell Side Length. The cell-list method divides the simulated domain into cells with side length greater than or equal to the cut-off radius of the interaction to be computed. In our simulation of an infectious disease, the simulated domain is a square of side length L, and the cut-off radius is related to the infectiousness of diseases. The cell side length l_{cell} represents the

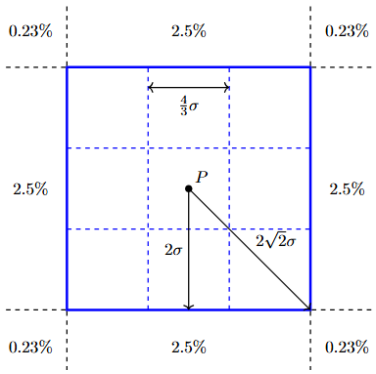

Figure 6: Geometric illustration of $l_{\text{cell}} = \frac{4}{3}\sigma_{\text{Inf}}$. The percentages indicate the probability of spreading the virus to the center of the neighborhood in all directions

maximal distance that an infected individual can transmit the virus to a susceptible individual. In an effective simulation, an appropriate l_{cell} should substantially reduce the number of individuals in the neighborhood of a selected susceptible individual, thereby lowering the computational cost. At the same time, it must ensure that any excluded infected individuals pose no risk of infecting the selected susceptible individual, thereby preserving the accuracy and reliability of the simulation. In the following, we are going to discuss how to choose an appropriate l_{cell} from two perspectives of our simulation.

Recall that we describe the infectiousness of the disease as a standard deviation σ_{Inf}. If a infected person is located at the origin, the virus will be spread in the x and y direction following a normal distribution $\mathcal{N}(0, \sigma^2_{\text{Inf}})$. Using properties of the normal distribution, we assume 95% of an infected person's infectiousness is within $2\sigma_{\text{Inf}}$ of them. In other words, for each susceptible person, we want to monitor the number of infected in a neighborhood of radius $2\sigma_{\text{Inf}}$ around them.

Assume that the susceptible individual is located at the center of a cell, and that its neighborhood consists of the cell itself together with the eight adjacent cells, as showed in Figure 6. Choosing $l_{\text{cell}} = \frac{4}{3}\sigma_{\text{Inf}}$ ensures that any infected individual outside this neighborhood is at least $2\sigma_{\text{Inf}}$ away from the susceptible individual. Under this assumption, an infected individual on one of the four edges of the neighborhood can transmit at most 2.5% of the virus concentration to the susceptible individual at the center, whereas an infected individual on a diagonal corner can transmit at most 0.23%. Since the transmission percentage decreases rapidly with distance, the contribution from infected individuals located beyond this neighborhood is negligible. Therefore, it is reasonable to assume that a susceptible individual is not at risk of infection from infected individuals outside this neighborhood. Consequently, $l_{\text{cell}} = \frac{4}{3}\sigma_{\text{Inf}}$ provides a suitable balance between computational efficiency and simulation accuracy.

Another way to find an acceptable l_{cell} is by examining the mathematical expression of the transmitting probability between the infected and susceptible individual, which is described by a Gaussian function of distance

$$P_i = \sum_j \exp\left(-\frac{r_{ij}^2}{2\sigma_{\text{Inf}}^2}\right) \times f_j,$$

where f_J is a factor that parametrizes the j-th infected person's conditions, r_i is the location of the i-th healthy person, and r_j is the location of the infected person. In the present model, we choose the quantity $r_i - r_j$ to associate with the spatial scale of infectiousness, that is $(r_i - r_j)^2 = c^2\sigma^2_{\text{inf}}$, where c is a constant. Implementing the cell-listing method, we want the infected people outside the cell that are negligible, that is, for the j's outside of the neighborhood

$$\exp\left[-\frac{r_{ij}^2}{2\sigma_{\text{Inf}}^2}\right] = \exp\left[-\frac{c^2\sigma^2_{\text{inf}}}{2\sigma_{\text{Inf}}^2}\right] = \exp\left[-\frac{c^2}{2}\right] \approx 0$$

And if the i-th susceptible person is at the center of its neighborhood, $r_{ij} \geq \frac{3}{2} l_{\text{cell}}$, then $l_{\text{cell}} = \frac{3}{2} c\sigma_{\text{Inf}}$. The parameter c provides a simple way to control l_{cell} in order to balance between accuracy and computational efficiency. The most direct mathematical bounding method controls the maximum structural error we are willing to accept by omitting interactions beyond the neighborhood. That is, for some tolerance $\epsilon > 0$, we have

$$\exp\left[-\frac{c^2}{2}\right] < \epsilon \Longrightarrow c \geq \sqrt{-2\ln(\epsilon)}.$$

For example, if we set $\epsilon = 10^{-5}$, $c \geq 5$. Usually, selecting $c \in [3, 5]$ captures between 99.73% ($3\sigma^2_{\text{Inf}}$) and 99.99% ($5\sigma^2_{\text{Inf}}$) of the spatial interaction probability density. Depending on the modeling setting, the user can choose relevant parameters to bound c. In the infectious disease modeling, ϵ can denote the population density of infected people outside of the cell neighborhood.

2.3.5. Frequency of Updating the Cell-List. To save on computational time, we may consider keeping the same cell-list for several time steps. Indeed, if the cell size greatly exceeds the expected change in position of each particle, we expect that most particles will remain in their current cell for several iterations. The number of time steps for which we can keep the same cell-list will depend on the mobility of the particles (i.e., the Brownian Motion variance σ_{BM}), and the size of each cell [15].

We propose updating the cell-list every

$$\max\left(\left\lfloor \frac{(l_{\text{cell}})^2}{2\pi\sigma^2_{BM}} \right\rfloor, 1\right) \tag{2.1}$$

time steps. Appendix A contains our derivation of this heuristic by considering escape times of a planar Brownian motion from a square domain. In particular, if the cell size is large and σ_{BM} is comparatively small (i.e., the population is less mobile than the disease is infectious), we can maintain the same cell-list for many iterations before updating it.

3. Results and Discussion.

3.1. Approach 1. Results are averaged over $n_{\text{sim}} = 500$ independent simulation runs. As shown in Figure 7, increasing the initial number of infected individuals leads to earlier outbreaks, reaches the infection peak sooner and with a higher value, and we notice that long-term dynamics of the population remain almost similar. This suggests that the initial infected count influences the timing and intensity of the outbreak. Figure 8 demonstrates if we reduce social activity, it lowers the infection peak and delays the transmission of the disease. Overall, we see that social activity is a key driver of disease transmission, and reducing it decreases the severity of the disease, and vaccination effectively suppresses infection.

The SIRV model presented here captures several important features of epidemic dynamics, including spatial heterogeneity in contact patterns, individual variation in social activity, and stochastic fluctuations in transmission events. The results suggest that the model reflects how mobility, social activity, temporary immunity, and vaccination jointly shape the spread and persistence of infection over time. When implementing vaccination into the currently existing SIR model, we consider both the number of initially infected individuals and social activity levels. The number of initially infected individuals affects how early the outbreak occurs and its intensity, with increasing infected individuals leading to higher infection peaks. Social activity is a major catalyst for disease transmission as lower levels of social activity lead to lower infection peaks and and delay disease progression. The bidirectional model considers the effect of waning immunity. After an individual is vaccinated, they eventually transition back into a susceptible state, leading to a rise in susceptible and decline in vaccinated individuals.

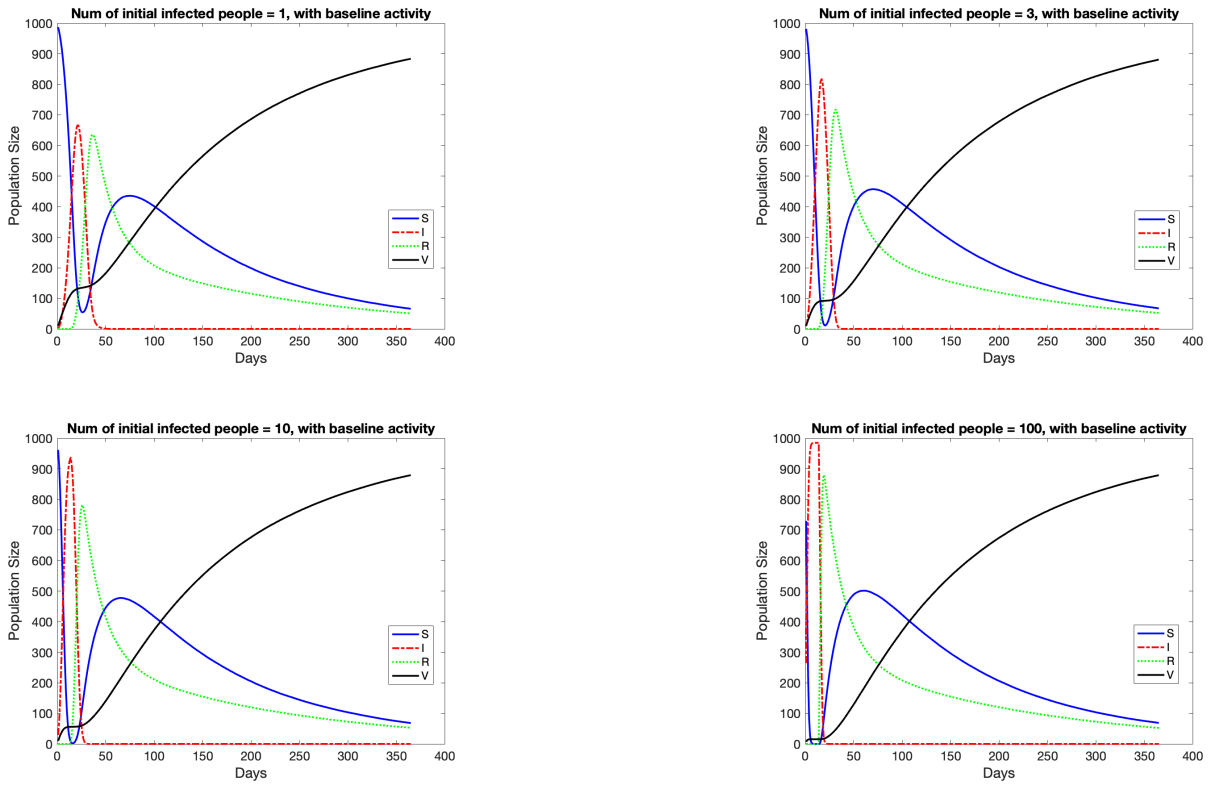

Figure 7: Varying initial infected individuals.

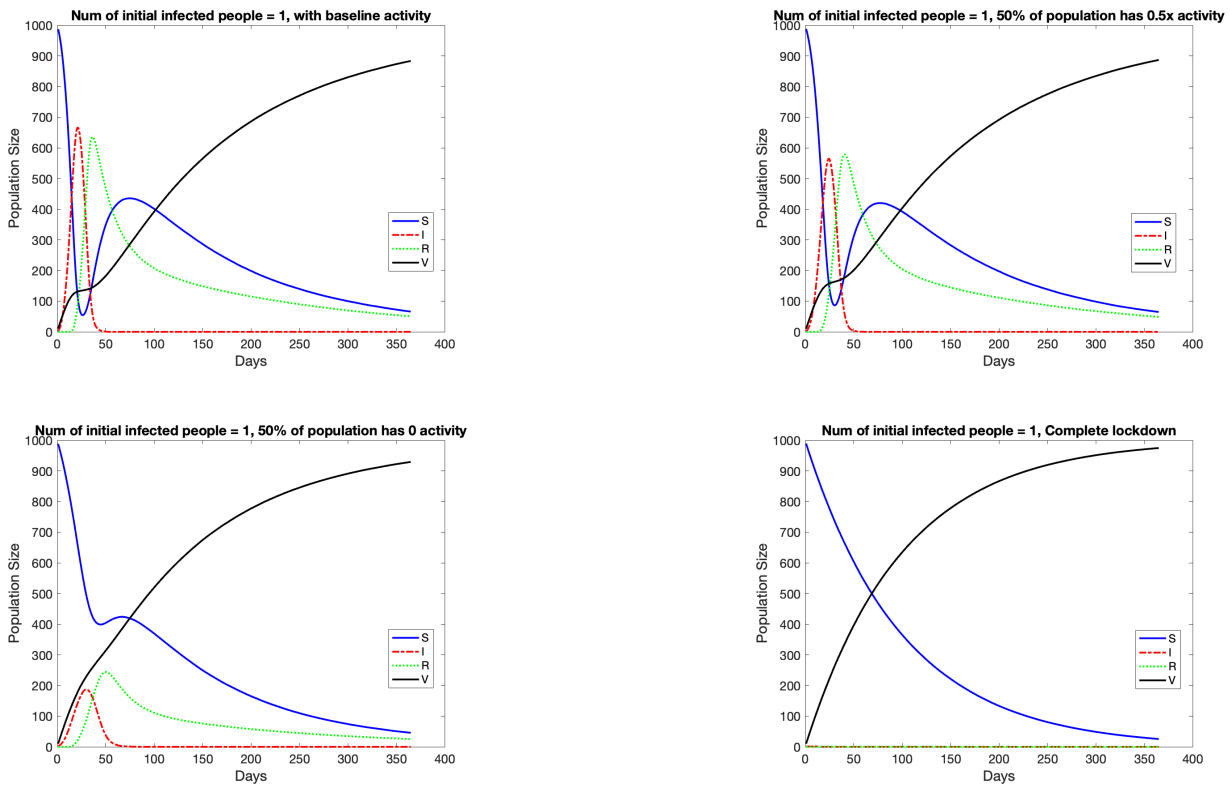

Figure 8: Effect of social activity.

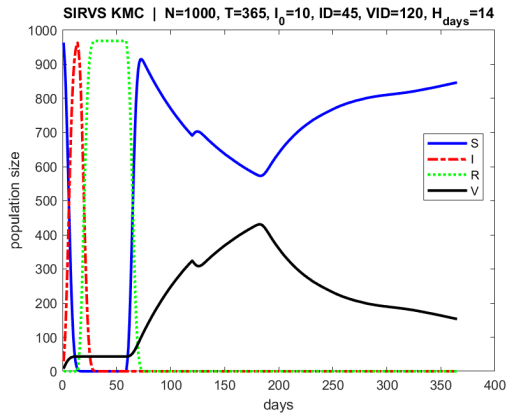

Figure 9: Bidirectional kMC for SIRV (case 1).

Figure 9 illustrates the results of our simulation over a period of 365 days for a population of 1,000 individuals. Initially, there are 10 infected individuals, each of whom remains infectious for 14 days before recovering. The natural immunity duration is assumed to be 45 days, while the vaccine immunity duration is taken to be 120 days.

From the figure, it can be observed that as the number of infected individuals increases, the susceptible population decreases accordingly. After the initial 14-day infectious period, the first group of infected individuals recovers, resulting in an increase in the recovered population, which remains zero during the first 14 days. As more individuals recover, the recovered compartment continues to grow and eventually reaches a peak. Thereafter, the number of recovered individuals begins to decline as their natural immunity expires and they return to the susceptible class.

Vaccination is introduced from the beginning of the simulation, causing the vaccinated population to increase gradually. However, since vaccine-induced immunity is temporary, vaccinated individuals eventually lose immunity and transition back to the susceptible class. This accounts for the increase in the susceptible population and the corresponding decrease in the vaccinated population observed toward the end of the simulation.

By the end of the 365-day period, all infected individuals have recovered, since each infected individual remains infectious for no longer than 14 days. Consequently, the infected compartment approaches zero. Similarly, the recovered compartment eventually declines to zero as individuals lose their natural immunity and re-enter the susceptible class. Overall, the long-term behavior of the model reflects the effects of waning natural and vaccine-induced immunity on the population dynamics.

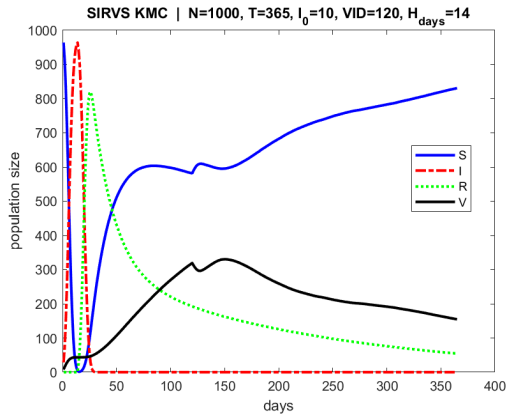

Figure 10: Bidirectional kMC for SIRV (case 2).

In Figure 10, the result for the second case is presented. A similar qualitative behavior is observed during the initial stages of the simulation. However, towards the end of the 365-day period, the recovered population does not decline to zero. This behavior reflects the presence of heterogeneous natural immunity durations within the population.

In this case, individuals do not share a uniform immunity period. Instead, some individuals experience relatively rapid waning of immunity and return to the susceptible class at earlier times, while others retain immunity for a longer duration before transitioning out of the recovered compartment. As a result, the recovered class persists over time, since a portion of the population continues to remain immune throughout the simulation period.

3.2. Approach 2.

3.2.1. Simulation Setup. As described in Section 2.2, we model the epidemic as a stochastic agent-based SIRV process, simulating N_{sim} epochs of a population of N individuals, with each epoch evolving all N individuals for T_{span} days. Each of N agents carries static risk attributes (age, comorbidity, social activity, geography, mobility, vaccine acceptance), each mapped to an infection-risk multiplier. The state advances in daily steps: agents undergo Brownian displacement (step standard deviation σ_{BM}) on a periodic square domain of side $L = \sqrt{N/\rho}$, then five stochastic reaction channels fire — $S \to I$ (spatial Gaussian transmission kernel of width σ_{Inf}, weighted by susceptible risk multipliers), $I \to R$, $R \to S$, $V \to S$ (exponential waning with means τ_{rcd}, τ_{sd}, τ_{vd}), and $S \to V$ (vaccination).

The relevant disease parameters for this suite simulations and default parameters are included in Table 2.

Table 2: Simulation Parameters and Baseline Defaults

Parameter	Meaning	Base default
N_{sim}	Number of stochastic runs aggregated	250
T_{Span}	Simulated days	365
N	Number of agents	1000
`initInfected`	Seed infections	1
`seed`	Base random-number-generator seed	12345
σ_{Inf}	Spatial transmission kernel width, in meters; larger values imply longer-range infection	2.4
τ_{rcd}	Mean infectious period, $I \to R$, in days	14
τ_{vd}	Vaccine effective duration, $V \to S$ waning, in days	180
τ_{sd}	Post-infection immunity duration, $R \to S$ waning, in days	90
ρ	Population density, measured in agents per m^2; sets domain size $L = \sqrt{N/\rho}$	0.1
σ_{BM}	Brownian daily movement step standard deviation	25.0

We perform sweeps overs four illustrative disease and population parameters, encompassing aspects of the distribution of population facets, as well as environment conditions. For each parameter that is being swept over, the remaining simulation parameters are set according to the default simulation values and population weights reported in Tables 2 and 1 respectively.

To explore how the disease dynamics change for different environments and populations, we compute two primary statistics relating to the rate of spread of the disease under a given set of initial conditions. For each epoch we estimate the **early exponential growth rate (daily)** $\hat{r}$, the doubling time $\ln 2/\hat{r}$, and the **takeoff probability** $P(\text{takeoff})$.

- $\hat{r}$ is estimated by fitting an exponential curve to the early growth period of an epoch,
- $P(\text{takeoff})$ is estimated as the proportion of epochs in which the infected population grows to at least 5 times the initial infected population.

For the purposes of this exploratory experiment we vary four input parameters

- Population Density ρ,
- Brownian Daily Movement Step Standard Deviation σ_{BM},
- Age Cohort Weights (fraction of the population in each of groups $[< 18, 18-49, 50-64, 65+]$),
- Vaccine Acceptance (proportion of the population whose encoded value is **True**).

3.2.2. Parameter Sweep Results. For each of the previously described parameter sweeps, we report the median early growth rate $\hat{r}$, doubling time, and proportion of runs that achieved takeoff, across 250 simulation runs. Estimated disease statistics are reported in Tables 3-6, as well as visualized in Figures 11- 14. The associated median SIRV curves with respect to time and confidence bands for each set of simulations can be found in Appendix B.

Population Density. We observe that the growth rate and probability of takeoff increase monotonically with ρ, achieving a 99% probability of takeoff at a density of 1 agent/m^2, while at a density of 0.05 agents/m^2 (1 agent/(20m^2)) only half of runs achieved takeoff rather than quickly fading out.

Brownian Daily Movement As the agents travel farther in each time step, the growth rate decreases (higher implied doubling time). This result may seem counterintuitive, but due to the spatial nature of our kMC model, disease spread is local and driven by spatial fronts, which do not develop as strongly due to the fact that susceptible and infected individuals have a lower likelihood of being co-located for extended periods of time. The probability of takeoff also decreases monotonically with mobility, implying that within this epidemic spread is reliant on extended proximity with infected individuals.

Age Cohort Weights We evaluate the the spread of disease for three representative distributions; younger, working age, and older. As described in Sections 2.2.3 and 2.2.4, older cohorts are modeled with both a higher infection risk, as well as higher probability of vaccine uptake, and we observe that the growth rate grows monotonically with the age of the population, as does the probability of takeoff, implying that even more aggressive vaccination policies may be necessary for elderly populations.

Vaccine Acceptance As described previously, Vaccine Acceptance is modeled according to a Bernoulli(p_{vac}) distribution. For those individuals where this variable is True, they are modeled as twice as likely to receive a vaccine and individuals where it is encoded as False. To provide a potential model for vaccine education policies, we consider three potential values for the population mean p_{vac}. We observe that monotonic increases in doubling and decreases in growth rate with p_{vac}, and at $p_{\text{vac}} = 0.9$, the probability of takeoff is about 70% within our sample.

3.2.3. Discussion. The simulation results showcase the potential of this modeling framework to allow finetuned and flexible experiments that can model specific populations, but also aid in forecasting the impact of potential policy decisions. While the data and parameter settings used in this synthetic experiment are not drawn from a representative population, initial simulation results are promising in their ability to capture expected behavior under different parameter choices, and highlight a potential opportunity for future refined models to be incorporated into population-aware policy decision frameworks.

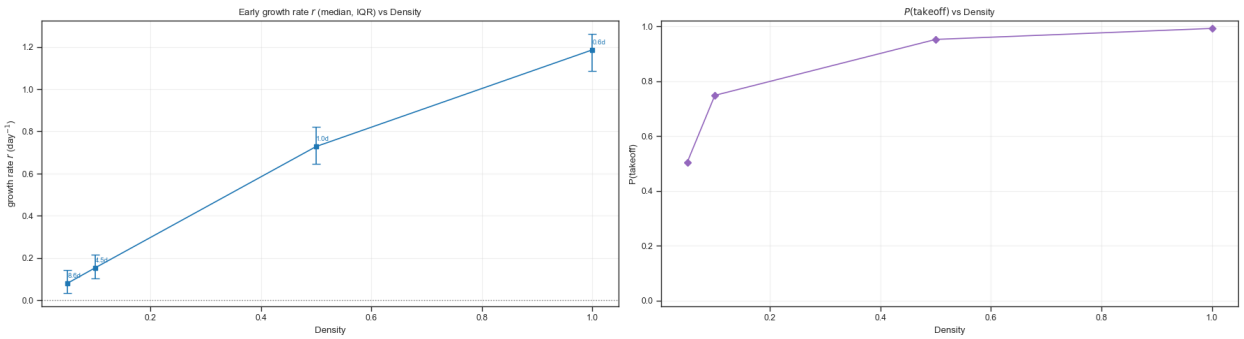

Figure 11: Population Density sweep summary panels vs. ρ: Median early growth rate (daily) $\hat{r}$ with IQR (doubling time annotated) and P(takeoff).

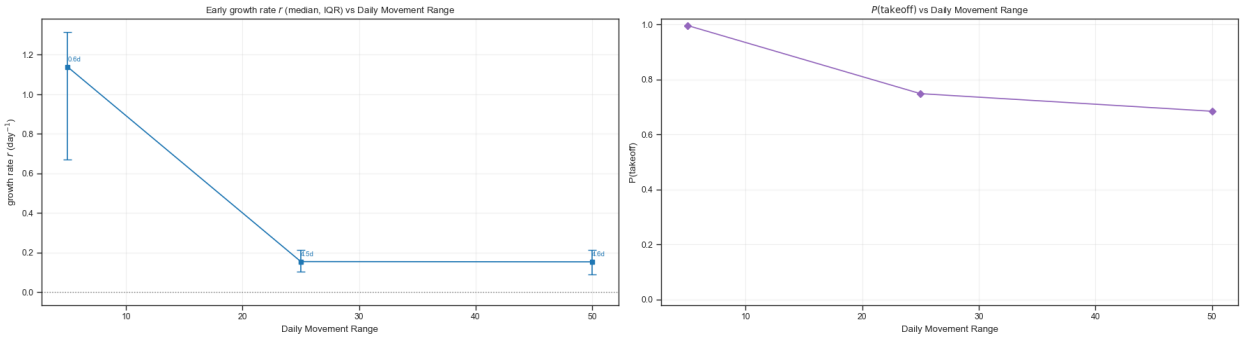

Figure 12: Movement Standard Deviation sweep summary panels: Median early growth rate (daily) $\hat{r}$ with IQR (doubling time annotated) and P(takeoff).

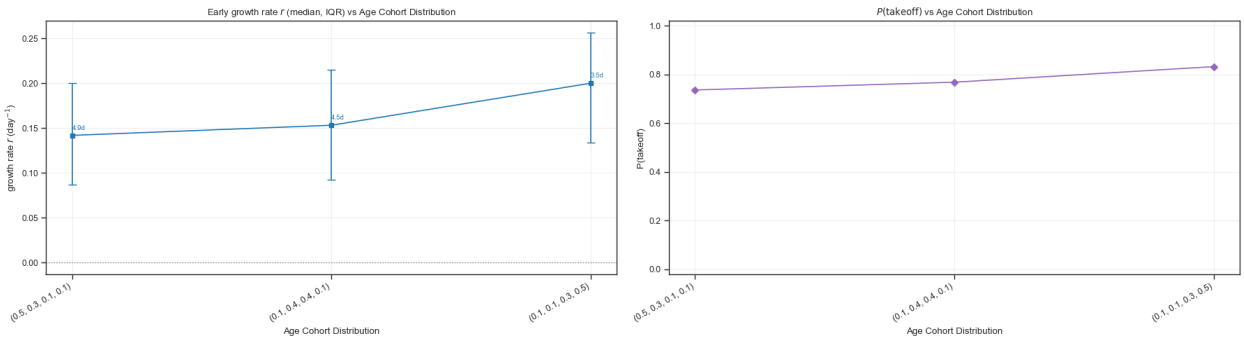

Figure 13: Age Cohort Weight sweep summary panels: Median early growth rate (daily) $\hat{r}$ with IQR (doubling time annotated) and P(takeoff).

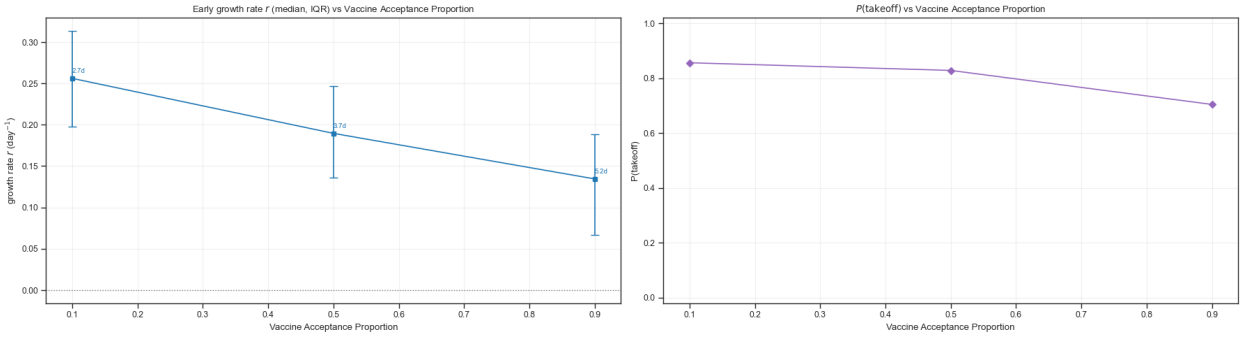

Figure 14: Vaccine Acceptance sweep summary panels: Median early growth rate (daily) $\hat{r}$ with IQR (doubling time annotated) and P(takeoff).

Table 3: Effect of population density ρ on epidemic intensity.

ρ (agents/m^2)	$\hat{r}$ (day^{-1})	Doubling Time (days)	P(takeoff)
0.05	0.081	8.6	0.50
0.1	0.15	4.5	0.75
0.5	0.73	0.95	0.95
1.0	1.2	0.58	0.99

Table 4: Effect of daily movement range σ_{BM} on epidemic intensity.

σ (m/day)	$\hat{r}$ (day^{-1})	Doubling Time (days)	P(takeoff)
5	1.1	0.61	1.0
25	0.15	4.5	0.75
50	0.15	4.55	0.68

3.3. Approach 3.

3.3.1. Computational Cost Reduction & Accuracy. In Figure 15, we see that the computational time of the naive approach for calculating P_i grows faster than the computational time for the cell-list method. Indeed, the slopes of this log-log plot tell us the polynomial order of complexity; the cell-list approach has an order of complexity close to $O(N)$, while the naive approach has an order of complexity closer to $O(N^2)$. Even though the computational time for this cell-listing method is faster, it should be noted that maintaining the cell-list has increased memory requirements [15].

Reducing the computational cost of the model is a great advantage of the cell-list method. Before incorporating the cell-list method into the kMC framework, however, it is essential to assess its impact on the accuracy of the simulation. We performed 100 independent simulations using both the brute-force kMC model and the cell-list-based kMC model, each with a population size of $(N = 10,000)$. All simulations were conducted under identical conditions, with fixed values of σ_{Inf}, σ_{BM}, simulation runtime, cell side length l_{cell}, cell-list update interval, population density, and all other relevant model parameters. This setup allows for a direct comparison of the two methods.

Figure 16 (Top) illustrates the simulation results obtained using cell-listing with those generated by the brute-force approach, which computes interactions exactly without any spatial restrictions. The two curves exhibit very similar behavior, indicating that the cell-listing method preserves the overall dynamics of disease transmission. The close agreement between the two curves demonstrates that restricting interaction calculations to local neighborhoods does not significantly alter the model's predictions. However, a small discrepancy can be observed: the equilibrium point of the cell-listing simulation occurs slightly later than that of the brute-force simulation, resulting in a minor rightward shift of the epidemic curve. This behavior is expected because the cell-list method neglects interactions originating from individuals located outside the selected neighborhood. By excluding these long-range contributions, the effective infection rate is slightly reduced, causing the spread of the disease to progress slightly slower than in the exact simulation.

In addition to the $N = 10,000$ case, simulations were conducted for $N = 1000$ (See Figure 19 in the Appendix). The results show that the cell-listing method successfully reproduces the same qualitative trends observed in the brute-force simulations. In particular, as the population size N increases, the epidemic reaches equilibrium earlier, producing the expected leftward shift of the infection curves. The ability of the cell-listing method to accurately capture this trend suggests that it preserves the key large-scale behavior of the model despite the computational approximations introduced.

To quantify the difference between the two approaches, we examined the relative error between the

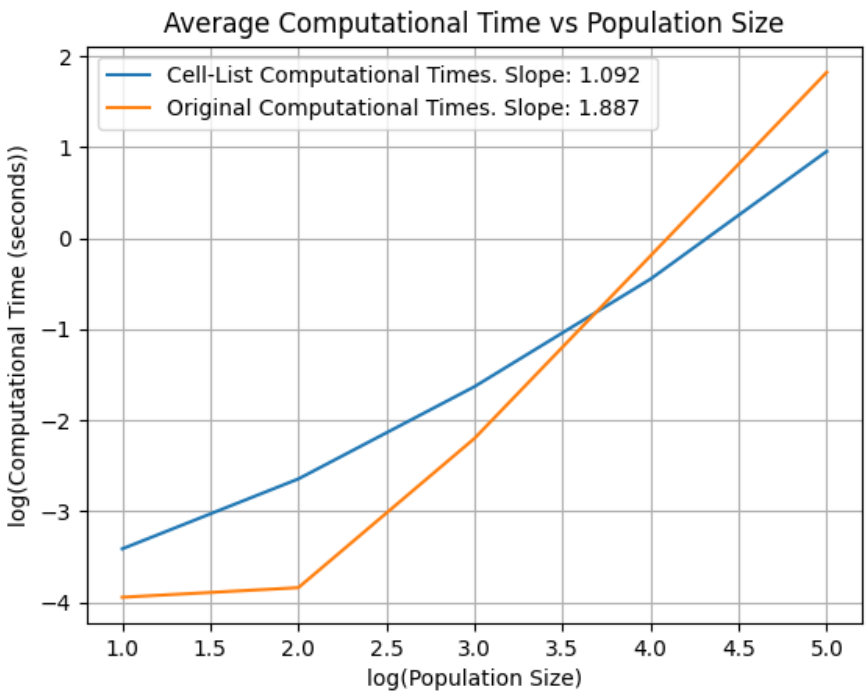

Figure 15: Log-log plot of computational times for the cell-list approach and the original, naive approach (i.e., computing all pairwise distances) for determining infection probability as population size increases. These computational times were averaged over 10 trials, maintaining conditions of $N/2$ susceptible and $N/2$ infected individuals. The average slope for the naive (original) method was taken after $N = 10^2$.

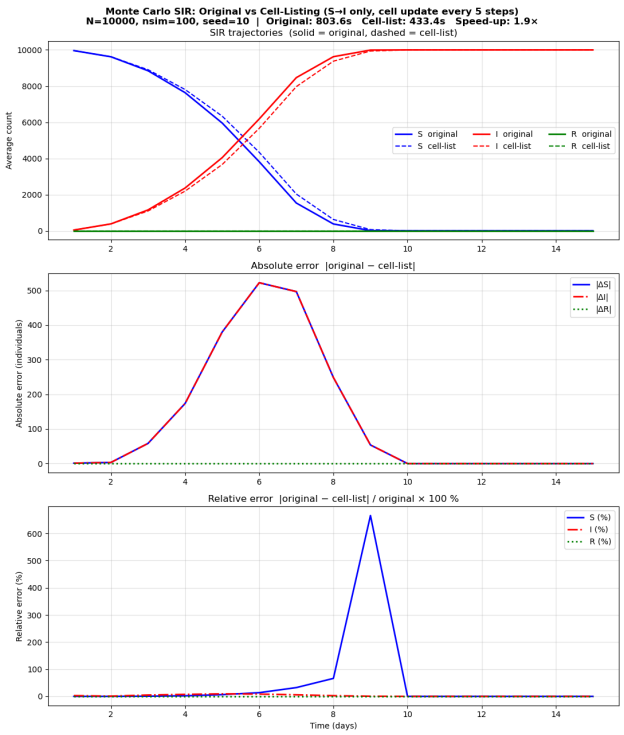

Figure 16: Top: Comparison of simulations with and without cell-list method with $N = 10^4$
Middle: The absolute error between two simulations
Bottom: The relative error between two simulations

Table 5: Effect of age structure on epidemic intensity.

age weights [<18, 18–49, 50–64, 65+]	$\hat{r}$ (day^{-1})	Doubling Time (days)	P(takeoff)
(0.5, 0.3, 0.1, 0.1) (younger)	0.14	4.9	0.74
(0.1, 0.4, 0.4, 0.1) (working age)	0.15	4.5	0.77
(0.1, 0.1, 0.3, 0.5) (older)	0.2	3.5	0.83

Table 6: Effect of vaccine-acceptance proportion on epidemic intensity.

acceptance	$\hat{r}$ (day^{-1})	Doubling Time (days)	P(takeoff)
0.1	0.26	2.7	0.86
0.5	0.19	3.7	0.83
0.9	0.13	5.2	0.70

corresponding simulation outputs (See Figure 16 Bottom). However, due to the highly symmetric nature of the epidemic trajectories, the relative error metric proved to be insufficient for evaluating accuracy. The absolute differences between the two curves (See Figure 16 Middle) are identical over time, which causes the relative error to provide limited insight into the actual discrepancy between the simulations. Consequently, relative error alone does not serve as a reliable measure of approximation quality for our model.

3.4. Holistic Approach. Within the scope of this project, we have outlined three approaches relevant to kMC modeling of infectious disease spread, and while they have been presented separately, the three approaches are intertwined as well as co-developed with the same broader framework. Each approach both informs and is informed by the other approaches, as illustrated in Fig. 17.

The SIRV model outlined in Section 2.1 provides both a base case and validation of the dynamics of the facet-augmented model outlined in Section 2.2, while the latter can also be used to probe how the degree of gain of additional complexity over the base model and assess the trade-off of ease and expressiveness. The cell-list approach described in Section 2.3 most directly can provide computational benefits to both the SIRV model as well as the facet-augmented SIRV model, but can also benefit from evaluating the performance of the cell-list in simulated populations with varying densities and mobilities. The flexibility of of the facet-augment SIRV model provides a convenient testbed to compare the theoretical performance of the approach against simulations with varying density, mobility, and disease characteristics.

Together these three approaches can be used in tandem, and with further refinement can work in concert to inform population-aware public policy, as well as develop and test new computational methods for large-scale simulations within a flexible framework.

4. Conclusion. Rather than treating SIRV modeling as a single computational problem, this project developed three approaches that progressively refine the representational and computational structure of infectious disease simulation. A1 established a baseline SIRV model capable of representing time-dependent vaccination efficacy degradation and variable recovery rates. A2 extended this structure by integrating weighted, faceted *person*-level dimensions through the decomposition of the *person* concept, allowing susceptibility to vary across synthetic individuals while supporting reproducibility and stability through configuration files. A3 further advanced the simulation framework through cell-list partitioning, providing a path toward improved computational performance in large-scale SIRV models. Taken together, these approaches demonstrate how theoretical representation, synthetic population design, and computational optimization can be integrated to support more flexible and applicable infectious disease simulation.

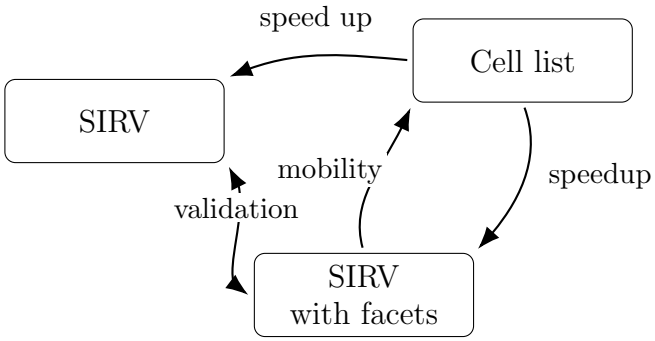

Figure 17: An illustration of the connections between the three approaches outlined in the scope of this project.

Future work will focus on refining both the epidemiological and computational dimensions of the model. First, additional testing is needed to determine whether SIRV state transitions should follow linear decay, Poisson-driven dynamics, or another transition structure. Second, open datasets should be identified to refine and validate the synthetic weights used during population generation. Third, a more rigorous analysis of the error between the cell-listing approach and the original pairwise approach (past our empirical findings in Section 3.3.1) may be possible. Finally, future implementations may explore the use of resilient distributed datasets (RDDs) to improve processing efficiency and scalability, particularly as the model is extended to larger synthetic populations and more complex transition dynamics.

Acknowledgments. We would like to thank the SIAM Graduate Student Mathematical Modeling Camp (GSMMC) directors, Pak-Wing Fok and Eirini Kilikian of the Department of Mathematical Sciences at the University of Delaware, for organizing and facilitating a productive and supportive modeling environment. We are especially grateful to our project mentor, Henry Boateng of the Department of Mathematics at San Francisco State University. His expertise in scientific computing, computational chemistry, and applied mathematics, along with his guidance, feedback, and encouragement, was invaluable throughout the development of this work. This support helped shape both our modeling approach and our understanding of the problem.

REFERENCES

[1] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Oxford University Press, second ed., 2017.

[2] T. Babou, J. Capece, U. Gideon, J. Hirt, J. Le, T. Tamanna, H. Yang, S. Youn, and X. Yu, *Software for kinetic Monte Carlo (kMC) simulation for modeling coronavirus spread*, 2026, https://doi.org/10.5281/zenodo.21043257, https://doi.org/10.5281/zenodo.21043257.

[3] N. G. Davies, P. Klepac, Y. Liu, K. Prem, M. Jit, and R. M. Eggo, *Age-dependent effects in the transmission and control of COVID-19 epidemics*, Nature Medicine, 26 (2020), pp. 1205–1211, https://doi.org/10.1038/s41591-020-0962-9.

[4] J. Diesel, N. Sterrett, S. Dasgupta, J. L. Kriss, V. Barry, K. Vanden Esschert, A. Whiteman, B. L. Cadwell, D. Weller, J. R. Qualters, L. Harris, A. Bhatt, C. Williams, L. M. Fox, D. Meaney-Delman, C. L. Black, and K. E. Barbour, *Covid-19 vaccination coverage among adults — united states, december 14, 2020–may 22, 2021*, Morbidity and Mortality Weekly Report, 70 (2021), https://doi.org/10.15585/mmwr.mm7025e1.

[5] R. Durrett, *Probability: theory and examples*, Duxbury Press, Belmont, CA, second ed., 1996.

[6] M. O. Elgendy and M. E. Abdelrahim, *Public awareness about coronavirus vaccine, vaccine acceptance, and hesitancy*, Journal of Medical Virology, 93 (2021), pp. 6535–6543.

[7] P. Gärdenfors, *Conceptual Spaces: The Geometry of Thought*, MIT Press, 2004.

[8] P. Gärdenfors, *The geometry and dynamics of meaning*, Topics in Cognitive Science, 17 (2025), pp. 34–56, https://doi.org/10.1111/tops.12767.

[9] P. Gärdenfors and M. Osta-Vélez, *Reasoning with Concepts: Conceptual Spaces as a Framework*, MIT press, 2026.

[10] T. Hu, S. Wang, B. She, M. Zhang, X. Huang, Y. Cui, J. Khuri, Y. Hu, X. Fu, X. Wang, and others, *Human mobility data in the COVID-19 pandemic: characteristics, applications, and challenges*, International Journal of Digital Earth, 14 (2021), pp. 1126–1147.

[11] J. V. Lazarus, K. Wyka, T. M. White, C. A. Picchio, L. O. Gostin, H. J. Larson, K. Rabin, S. C. Ratzan, A. Kamarulzaman, and A. El-Mohandes, *A survey of COVID-19 vaccine acceptance across 23 countries in 2022*, Nature Medicine, 29 (2023), pp. 366–375.

[12] C. Liu, J. Huang, S. Chen, D. Wang, L. Zhang, X. Liu, and X. Lian, *The impact of crowd gatherings on the spread of COVID-19*, Environmental Research, 213 (2022), p. 113604, https://doi.org/10.1016/j.envres.2022.113604.

[13] Y. Liu, B. Mao, S. Liang, J.-W. Yang, H.-W. Lu, Y.-H. Chai, L. Wang, L. Zhang, Q.-H. Li, L. Zhao, Y. He, X.-L. Gu, X.-B. Ji, L. Li, Z.-J. Jie, Q. Li, X.-Y. Li, H.-Z. Lu, W.-H. Zhang, Y.-L. Song, J.-M. Qu, and J.-F. Xu, *Association between age and clinical characteristics and outcomes of COVID-19*, European Respiratory Journal, 55 (2020), https://doi.org/10.1183/13993003.01112-2020.

[14] C. Macal and M. North, *Tutorial on agent-based modeling and simulation*, in Proceedings of the Winter Simulation Conference, 2005, https://doi.org/10.1109/WSC.2005.1574234.

[15] W. D. Mattson and B. M. Rice, *Near-neighbor calculations using a modified cell-linked list method*, Computer Physics Communications, 119 (1999), pp. 135–148, https://api.semanticscholar.org/CorpusID:7652497.

[16] D. J. Peters, *Community Susceptibility and Resiliency to COVID-19 Across the Rural-Urban Continuum in the United States*, The Journal of Rural Health, 36 (2020), pp. 446–456, https://doi.org/10.1111/jrh.12477.

[17] K. S. Raghavan, *Shiyali Ramamrita Ranganathan*, 2019, https://www.isko.org/cyclo/ranganathan.

[18] C. D. Russell, N. I. Lone, and J. K. Baillie, *Comorbidities, multimorbidity and COVID-19*, Nature Medicine, 29 (2023), pp. 334–343, https://doi.org/10.1038/s41591-022-02156-9.

[19] M. Sallam, *COVID-19 Vaccine Hesitancy Worldwide: A Concise Systematic Review of Vaccine Acceptance Rates*, Vaccines, 9 (2021), https://doi.org/10.3390/vaccines9020160.

[20] Y. Sun and S. M. Monnat, *Rural-urban and within-rural differences in COVID-19 vaccination rates*, The Journal of Rural Health, 38 (2022), pp. 916–922, https://doi.org/10.1111/jrh.12625.

[21] B. Wang, R. Li, Z. Lu, and Y. Huang, *Does comorbidity increase the risk of patients with COVID-19: evidence from meta-analysis*, Aging, 12 (2020), pp. 6049–6057, https://doi.org/10.18632/aging.103000.

[22] E. Westra, *Modular Programming with Python*, Packt Publishing Ltd, 2016.

[23] M. Zhang, S. Wang, T. Hu, X. Fu, X. Wang, Y. Hu, B. Halloran, Z. Li, Y. Cui, H. Liu, Z. Liu, and S. Bao, *Human mobility and COVID-19 transmission: a systematic review and future directions*, Annals of GIS, 28 (2022), pp. 501–514, https://doi.org/10.1080/19475683.2022.2041725.

Appendix A. A Heuristic Number of Time Steps before Updating the Cell-List. To determine an appropriate number of time steps for which we can use the same cell-list, we consider the expected time for a planar Brownian motion to exit a square-domain (i.e., a cell). Let $(B_t)_{t\geq 0}$ be a planar Brownian motion starting at 0, with X_t and Y_t its x and y-coordinates respectively. Given some length $L > 0$, we are interested in the expectation of the stopping time

$$\tau_{\text{sq}} := \inf\{t > 0 : \max(|X_t|, |Y_t|) \geq: L\},$$

which is the exit time from the square centered at 0 with side length $2L$. To estimate $\mathbb{E}[\tau_{\text{sq}}]$, we can consider the exit times from the inscribed and circumscribed circles of this square, namely,

$$\tau_{\text{in}} := \inf\{t > 0 : X_t^2 + Y_t^2 \geq L^2\},$$
$$\tau_{\text{ci}} := \inf\{t > 0 : X_t^2 + Y_t^2 \geq 2L^2\}.$$

It is clear that

$$\tau_{\text{in}} \leq \tau_{\text{sq}} \leq \tau_{\text{ci}}.$$

From here, we employ the following lemma:

Lemma A.1. *For $r > 0$, let $\tau_r := \inf\{t > 0 : X_t^2 + Y_t^2 \geq r^2\}$. Then, $\mathbb{E}[\tau_r] = \dfrac{r^2}{2}$.*

Proof. From properties of Brownian motion, we have that $(X_t^2 - t)_{t\geq 0}$ and $(Y_t^2 - t)_{t\geq 0}$ are both independent, identically distributed continuous martingales. By the optional stopping theorem [5],

$$\mathbb{E}[X^2_{\tau_r\wedge t} - (\tau_r \wedge t)] = \mathbb{E}[Y^2_{\tau_r\wedge t} - (\tau_r \wedge t)] = \mathbb{E}[X_0^2 - 0] = 0$$

for all $t > 0$. From this, we obtain

$$\frac{1}{2}\mathbb{E}[X^2_{\tau_r\wedge t} + Y^2_{\tau_r\wedge t}] = \mathbb{E}[\tau_r \wedge t]$$

for all t. By the monotone convergence theorem, the right hand side converges to $\mathbb{E}[\tau_r]$ as $t \to \infty$, and since $X^2_{\tau_r\wedge t} + Y^2_{\tau_r\wedge t} \leq r^2$ for all $t > 0$, the dominated convergence theorem tells us the left hand side converges to $\dfrac{1}{2}\mathbb{E}[X^2_{\tau_r} + Y^2_{\tau_r}]$. Taking these facts together gives us

$$\mathbb{E}[\tau_r] = \frac{1}{2}\mathbb{E}[X^2_{\tau_r} + Y^2_{\tau_r}] = \frac{r^2}{2},$$

since, by the continuity of Brownian motion, $X^2_{\tau_r} + Y^2_{\tau_r} = r^2$ almost surely. ∎

From Lemma A.1, we have $\mathbb{E}[\tau_{\text{in}}] = \dfrac{L^2}{2}$ and $\mathbb{E}[\tau_{\text{ci}}] = L^2$. In particular, $\dfrac{L^2}{2} \leq \mathbb{E}[\tau_{\text{sq}}] \leq L^2$.

Using these exit times, we now estimate $\mathbb{E}[\tau_{\text{sq}}]$ by considering the green and red areas in Figure 18. We see that the area of the green region is $4a^2 - \pi L^2$, while the area of the red region is $2\pi L^2 - 4L^2$. Taking the ratio of the green and red area gives us

$$\frac{4L^2 - \pi L^2}{2\pi L^2 - 4L^2} = \frac{4 - \pi}{2\pi - 4}.$$

From this, we estimate

$$\mathbb{E}[\tau_{\text{sq}}] \approx L^2\left(\frac{1}{2} + \frac{4-\pi}{2\pi - 4}\right) = L^2 \cdot \frac{2}{\pi} \approx 0.637L^2,$$

that is, we scale the known exit times of the balls according to the ratio of the areas of the green and red regions.

In practice, we use a scaled Brownian motion $(\sigma_{BM}B_t)_{t\geq 0}$. By repeating the above derivation with $\sigma_{BM}B_t$ in place of B_t, we have

$$\mathbb{E}[\tau_{\text{sq}}] \approx \frac{L^2}{\sigma_{BM}^2} \cdot \frac{2}{\pi}.$$

Finally, taking $L = \dfrac{l_{\text{cell}}}{2}$ gives us our heuristic number of time steps

$$\max\left(\left\lfloor \frac{(l_{\text{cell}})^2}{2\pi\sigma_{BM}^2} \right\rfloor, 1\right).$$

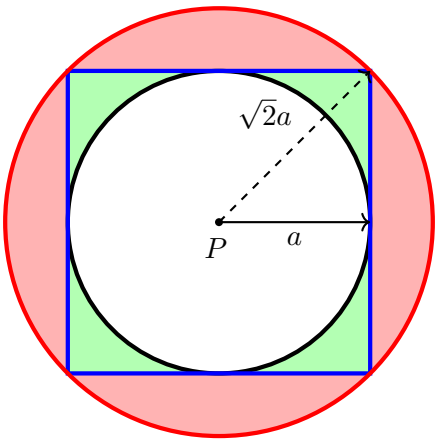

Figure 18: The expected exit time for a Brownian motion starting at center point P from the square (blue) is estimated using the exit times from the inscribed circle (black) and the circumscribed circle (red).

Appendix B. Approach 2 Results: SIRV Curves. Figures 20-23 display the SIRV evolution of simulated infectious disease spread under varying parameters, which have been discussed and summarized in 3.2.2. The median outcome over all N_{sim} runs, with shaded bands representing the $[2.5, 97.5]$ quartile intervals. Of note is that these intervals become large under certain sets of parameters, though is explained through the fact that when a a non-negligible portion of runs fade immediately while another proportion takes off, both the epidemic breakouts and fadeouts are included within this range. These confidence bands become tight when either nearly all runs take off or when most runs never achieve significant disease spread.

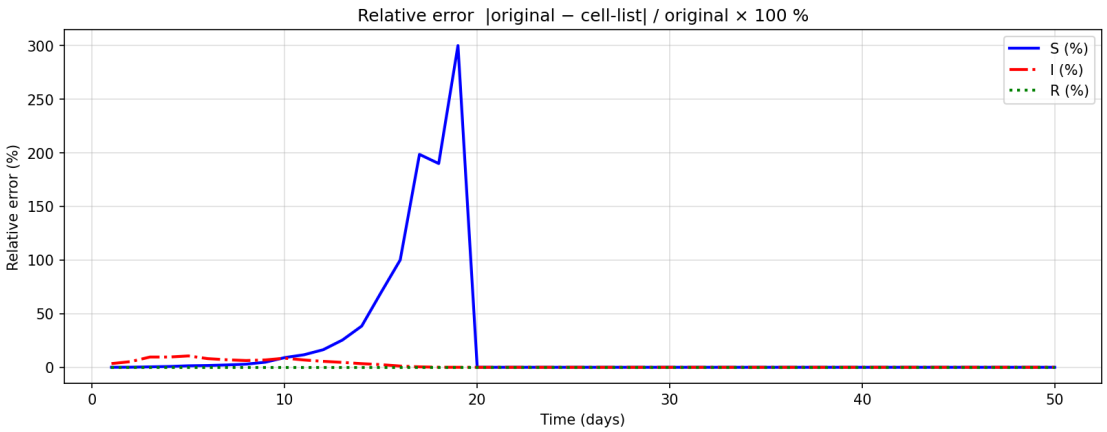

Figure 19: Comparison of simulations with and without cell-list method with $N = 10^3$.

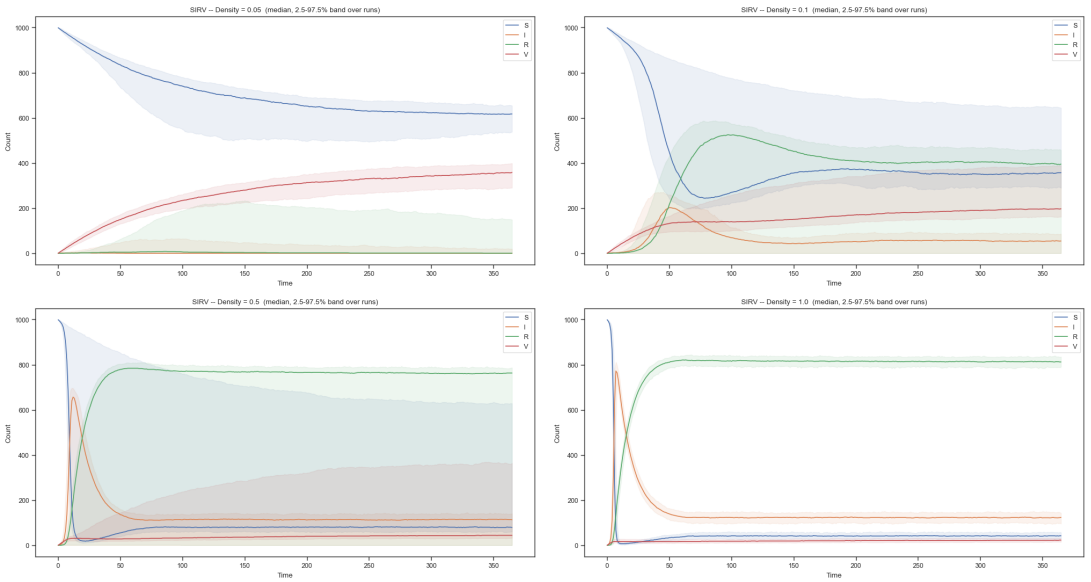

Figure 20: SIRV Curves for population densities $\rho \in \{0.05, 0.1, 0.5, 1.0\}$. Solid lines indicate the median outcomes over $N_{\text{sim}} = 250$ runs, while [2.5, 97.5] quartile bands are shaded.

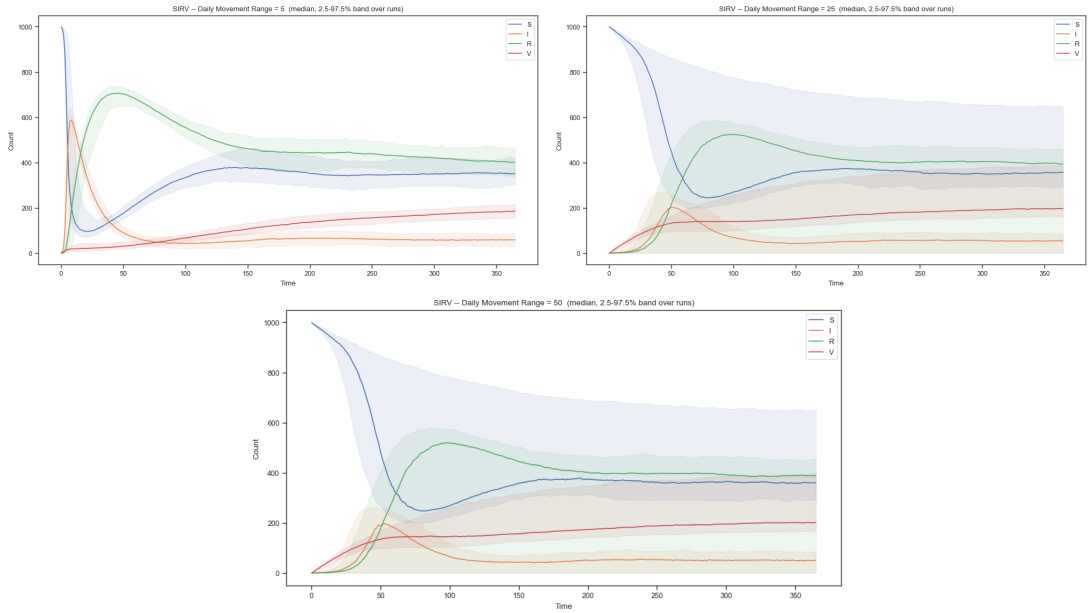

Figure 21: SIRV Curves for $\sigma_{\text{BM}} \in \{5, 25, 50\}$. Solid lines indicate the median outcomes over $N_{\text{sim}} = 250$ runs, while [2.5, 97.5] quartile bands are shaded.

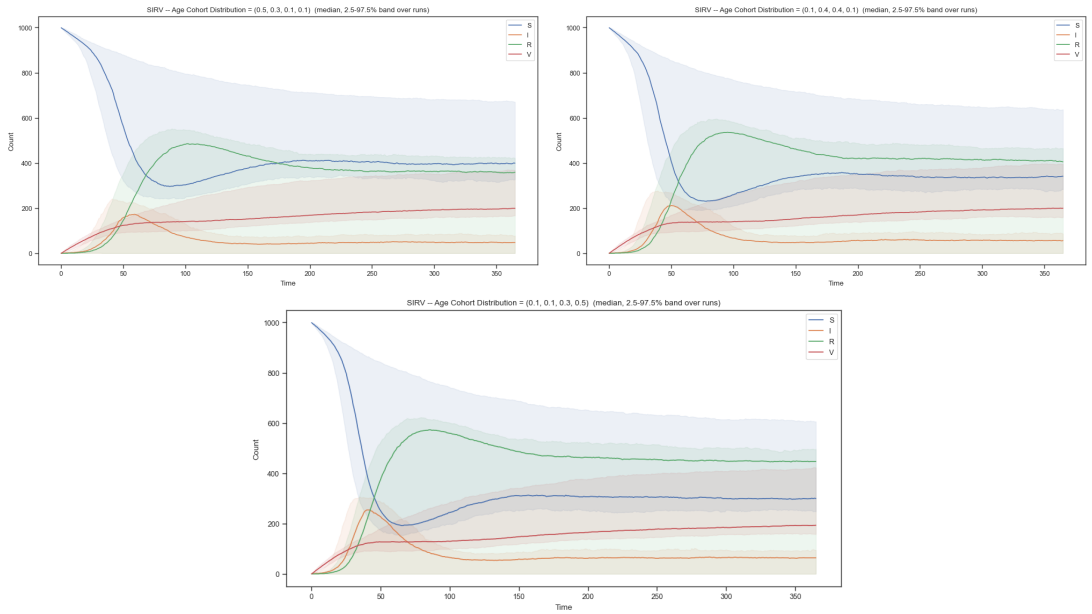

Figure 22: SIRV Curves for younger, working, and older age distributions. Solid lines indicate the median outcomes over $N_{\text{sim}} = 250$ runs, while [2.5, 97.5] quartile bands are shaded.

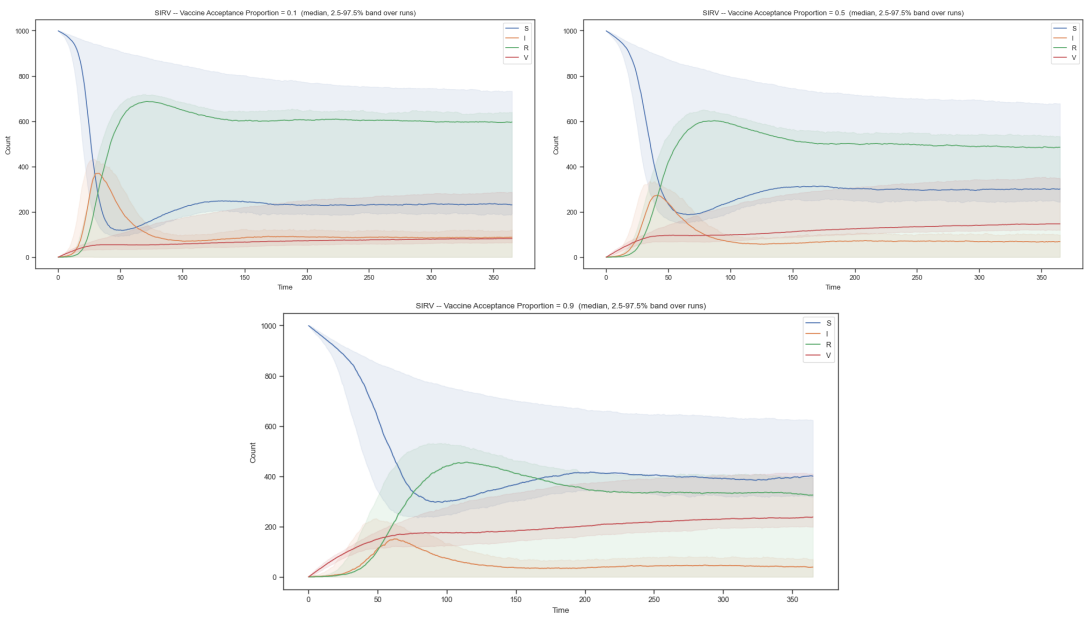

Figure 23: SIRV Curves for Vaccine Acceptance $\in \{0.1, 0.5, 0.9\}$. Solid lines indicate the median outcomes over $N_{\text{sim}} = 250$ runs, while [2.5, 97.5] quartile bands are shaded.