A COMPUTATIONAL APPROACH TO UNDERSTANDING SPATIAL AND TEMPORAL GRANULARITIES IN AGENT-BASED MODELING

BY

ERIC A. SHOOK

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Geography
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2013

Urbana, Illinois

Doctoral Committee:

Associate Professor Shaowen Wang, Chair
Emeritus Professor Bruce Hannon
Professor Laxmikant Kale
Professor Sara McLafferty

ABSTRACT

Epidemic agent-based models simulate individuals in artificial societies capable of moving, interacting, and transmitting disease amongst themselves. Due to limitations in data and computation, epidemic models simulating tens of millions of individuals often coarsen the finest representations of space and time–termed spatial and temporal granularities in this thesis. This dissertation examines and overcomes a set of computational challenges to investigate a fundamental problem in spatially explicit epidemic agent-based modeling. This research demonstrates that coarsening spatial and temporal granularities influence both computational tractability and epidemic ABM processes. By focusing on the nexus of space, time, and process my dissertation improves understanding of the interrelationships and trade-offs between space and time as they relate to spatial processes using an epidemic modeling case study.

To my wife Sara and our children Oliver, Parker, and Julia

ACKNOWLEDGMENTS

This dissertation would not have been possible without the support and guidance from family, friends, and advisors.

First, I thank my family that helped me endure this long, challenging, and rewarding process. The unending patience, love, and understanding of my wife Sara has been a constant source of inspiration. She always believed in me, even when I found it difficult to believe in myself. Her support never wavered while she single-handedly managed everything for our growing family while I worked late nights, early mornings, and over weekends. The boundless energy and love of our children Oliver, Parker, and Julia has always brought a smile to my face and a sense of determination in my heart. Watching them grow up throughout this process helped to shape my path and inspire me to become an educator. I am thankful for all the help and good times that we have been able to have with our siblings and their families. Finally, I owe many thanks to my parents and Sara's parents. While it was difficult for them, because we continued to move farther away for me to pursue a graduate education, they have always offered their unconditional love and support for my family and me, and for that, I am truly grateful.

I am fortunate to have had the opportunity to work with my advisor Shaowen Wang and committee members Bruce Hannon, Laxmikant Kale, and Sara McLafferty. Shaowen has provided unmeasureable amounts of guidance and support, both professionally and personally, during my time here at Illinois and at Iowa. His dedication to scholarship will continue to be a model for me to follow into the future. I also am grateful to Bruce, Sanjay, and Sara for their support and guidance during this process. Their suggestions and perspectives proved invaluable during the development and finishing stages of my dissertation.

Countless friends of Sara's and mine have helped us through these past few years, and I appreciate all their help and support. For all the people who brought us food, watched our children, and offered my family care and support while we were in the hospital, I cannot express how grateful I am. Especially Jim and Kari May and their children Andrew and Jami, who have become family

friends, and could always be counted upon whether at 2:00am the day after Christmas or just to get together to have a good time.

Many people at Illinois deserve my thanks and gratitude. To the many graduate students that listened to my ideas, debated with me, and helped me relax. Yanli Zhao, Trevor Fuller, and Michael Minn in particular were always good for a laugh or a debate. I would also like to thank all the members in the GIS Group who continually impressed me with their dedication and drive. I enjoyed all the conversations and coding sessions as they put up with my first trials as an advisor, and I especially wish to thank Varun Goel, Nate Banion, Julie Carlson, Brian Wilson, Jizhe Yang, and Ross Evans. I owe a great deal to Susan Etter, Matt Cohn, and Chris Wilcock for their help in navigating the university system. Thank you to all the students, postdocs, and staff in the CIGI lab. I have especially enjoyed working with Yan Liu and Anand Padmanabhan at both Illinois and Iowa, and I will miss their camaraderie. Outside of Illinois, I would like to thank Will Wheaton and RTI, International for providing the US Synthetic Population Dataset that enabled my research.

Last, but not least I would like to thank my undergraduate advisor Paul Gray. Simply put Paul's selfless dedication to research and teaching made this possible. He drove me and three other undergraduate students to our first supercomputing conference (SC03) in Phoenix, Arizona. He introduced me to many people whose friendship I now admire including Scott Lathrop, Charlie Peck, and Tom Murphy, and to the world of high-performance computing, which has served as the foundation of my research interests ever since.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Introduction

Spatially explicit agent-based models (ABMs) are a computational approach that have been widely developed and applied to study complex characteristics of coupled human-environmental systems (Batty et al., 2012; Torrens, 2012; An, 2012; An et al., 2005; Parker et al., 2003) at the level of an individual or agent (Epstein, 2009). "Agent-based models offer distinct explanatory power over previous approaches, in that observed phenomena emerge from interactions, rather than being imposed by the modelling framework and preconceived notions." (Wainwright and Mulligan (2004) (page 232) as cited in Smieszek (2010)) However, to achieve this explanatory power often demands rich data and massive amounts of computation for simulating millions of agents (Ajelli et al., 2010; Parker and Epstein, 2011). This dissertation overcomes a set of computational challenges for ABMs to systematically investigate the influence of variations of spatial and temporal representations on modeling epidemic dynamics.

While ABMs have been used to study the spread of disease at various spatial and temporal scales ranging from emergency room facilities to the world, over days, months, and years (Laskowski et al., 2011; Epstein, 2009), increasingly large-scale epidemic ABMs are used by state and national health officials to inform public policy in epidemic preparedness (Brown et al., 2011; Ferguson et al., 2006; Germann et al., 2006). These large-scale ABMs, simulating tens of millions of agents that are capable of moving, interacting, and transmitting disease amongst themselves, are used to help scientists and policy makers at state and national levels better understand how disease spreads through populations, space, and time (Smieszek et al., 2011; Lee et al., 2010; Chao et al., 2010). Unlike small-scale ABMs, large-scale ABMs are often challenged by limitations in data and computation due to simulating a large number of agents moving, interacting, and transmitting disease.

To overcome computational challenges, ABMs often exploit parallel and high-performance com-

puting (HPC) by decomposing and distributing a model to many computing elements to be solved simultaneously (Tang and Wang, 2009; Ramachandramurthi et al., 1997; Cornwell et al., 2001; Wang et al., 2006). However, there is growing impediment against the continued advancement of parallel ABM in scaling to high-end computers. Inter-processor communication is often required to coordinate many processor cores to generate a cohesive simulation, but is recognized as a major limiting factor in enabling parallel ABMs to effectively use large numbers of processor cores (Parry, 2009; Barrett et al., 2008; Parker, 2007). Chapter 2 examines and overcomes this computational bottleneck through the development of a communication-aware framework to enable the simulation of millions of agents.

Due in part to the data and computational demands, a discourse toward the usefulness of epidemic ABMs compared to other epidemic modeling approaches including metapopulation or equation-based models has emerged in the literature. Hupert et al. (2008) argue that representing complex spatial interactions in ABMs may not add sufficient value to the predictive capabilities for understanding disease spread. Yet, a recent literature review of epidemic modeling notes that epidemic models often employ simplifying assumptions to reduce computational burden, which may limit the types of intervention strategies that can be simulated (Prieto et al., 2012). This suggests that the ability to represent complex spatial interactions in ABMs, while computationally intensive, is advantageous when simulating intervention strategies. Balcan et al. (2010) lists detailed data and computational demands as barriers for the use of ABMs, particularly at global scales, but does acknowledge higher-level of individual level detail in ABMs as a distinct advantage compared to other epidemic modeling approaches. However, Parker and Epstein (2011) developed a global-scale epidemic ABM with relatively modest computational demands using a number of optimizations and simplifying assumptions. This suggests that large-scale epidemic ABMs can be computationally tractable even at global-scale. At the heart of this discourse lie the assumptions of contact or mixing patterns, which are central to simulating how disease spreads in epidemic models (Koopman, 2004; Riley, 2007; Bian, 2013). Mixing patterns may include homogeneous (i.e. perfectly mixed) or heterogeneous (Mishra et al., 2011). It is shown that the incorporation of structure whether spatial, network, or community often reduces homogeneous mixing and influences disease spread dynamics (Ajelli et al., 2010; Ferrari et al., 2011; Chao et al., 2010) and in the case of spatial structure

generally increases computational requirements, which may limit computational tractability in large-scale models (Riley, 2007).

The representations of space and time often vary between different epidemic models (see Table 4.1 in Chapter 4), and can be influenced by factors such as the size of a modeled population, computational requirements, and quality of population or disease data (Bian, 2013; Ajelli et al., 2010; Parker and Epstein, 2011). Epidemic ABM simulations often operate with time-step durations ranging from minutes (Bisset et al., 2012) to a day (Brown et al., 2011). Large-scale spatially explicit epidemic models often use transmission abstractions based on spatial representations such as patches with sizes ranging from sub-building (Stroud et al., 2007) to sub-city (Balcan et al., 2010), for example. To enable individuals to transmit disease within finer-grained spatial representations such as within a building or city, these epidemic models often rely on homogeneous mixing (Riley, 2007). Studies implicitly (Ajelli et al., 2010) or explicitly (Ajelli and Merler, 2008) comparing spatial representations reveal differences in disease spread dynamics, but fail to explore the underlying impact on ABM processes that cause these differences. A study comparing three ABMs finds differences in simulated effectiveness of intervention strategies (Halloran et al., 2008), but due to differences in modeling assumptions and data it is difficult to determine how much the variation of spatial and temporal representations contributes to these differences. Studies that investigate influence of altering synthetic network configurations show consequent changes, sometimes dramatic, of disease spread dynamics (Xu and Sui, 2009; Keeling et al., 2010; Keeling and Eames, 2005; Shirley and Rushton, 2005; Keeling, 1999). These studies often use the same model reducing the complexity of comparing multiple models with different assumptions, but network-based models rely on an implicit representation of geographic space (see Chapter 3) and it is often difficult to extrapolate this influence to real-world scenarios. Further, disease passes along network connections in network-based models minimizing homogeneous mixing amongst individuals, which as stated previously is often assumed in large-scale epidemic ABMs. While the literature provides evidence that variations in the representation of space and time may influence disease dynamics, previous studies have not systematically investigated how variations in the representations of space and time may influence disease spreading processes in epidemic ABMs. This dissertation fills a gap in the literature to improve understanding of the influence that variations in the represen-

tations of space and time have on ABM processes and computational tractability. An improved understanding of this influence will add new insights toward the effects of simplifying assumptions with respect to coarsening the representations of space and time (Prieto et al., 2012) and help to answer the question of whether computationally demanding fine-grained epidemic models produce similar results to spatially and temporally coarse-grained models and as a result may be superfluous (Hupert et al., 2008).

1.2 Research Problem

This dissertation investigates how the finest representations of space and time–termed spatial and temporal granularities (STGs) in this thesis–influence computational tractability and shape ABM processes. Computational tractability is critical for policy makers and public health experts that may rely on epidemic ABMs to make timely and informed decisions when faced with an epidemic (Epstein, 2009). Parallelizing ABMs to take advantage of HPC resources helps to achieve computational tractability while providing several benefits including enhanced capabilities for modelling more complex phenomena, exploration of a large number of alternative scenarios, improved sensitivity analysis, rigorous and comprehensive validation, and improved computational performance (Abbott et al., 1997; Crooks et al., 2008; Clarke, 2003; Lysenko and DSouza, 2008). However, many epidemic models rely on simplifying assumptions or approximations such as coarsening STGs to make simulations computationally tractable (Prieto et al., 2012; Stroud et al., 2007; Ajelli et al., 2010). My dissertation systematically investigates how varying STGs influences both computational tractability and ABM processes.

Processes in ABMs capture agent movements and interactions, which operate within the constraints of STGs. As spatial granularities in a model are coarsened, finer grained spatial units are aggregated into coarser (i.e. larger) spatial units, which is related to a well-known problem in geography literature–the Modifiable Areal Unit Problem (MAUP) (Openshaw, 1984). MAUP may occur if small spatial units (e.g. points or polygons) are aggregated into larger spatial units, whose configuration is modifiable, and where different configurations lead to different results (Openshaw, 1984). However, MAUP does not explicitly address spatial processes (Goodchild, 2004). This dissertation investigates the influence of not only coarsened spatial granularities, but also coarsened

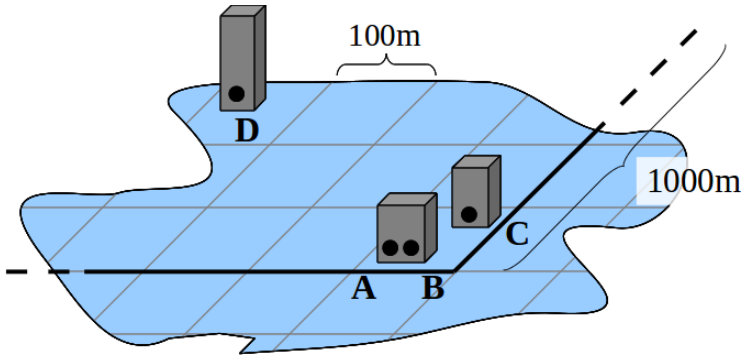


Figure 1.1: This figure is used to illustrate how coarsening STGs can influence a disease infection process where four agents are distributed across a geographic area.

temporal granularities, on epidemic processes. By focusing on the nexus of space, time, and process my dissertation improves understanding of the interrelationships and potential trade-offs between space and time as they relate to spatial-temporal processes using an epidemic modeling case study.

The following conceptual example illustrates the influence of varying STGs on model dynamics. Consider four agents (A-D) that are distributed across a geographic area (Figure 1.1).

- Agents A and B are located at opposite ends of the same small building.
- Agent C is located in a building on the same block as agents A and B.
- Agent D is located in a building several blocks away from agent A, B, and C.

Assume that agents B, C, and D are susceptible to infection if they come in contact with agent A who is infectious. For the illustrative purpose, let us assume agent A may walk in any direction at 3.6 kilometers an hour (i.e. 1 meter a second), but agents B, C, and D do not move. Chapter 3 details the case where all agents move, which is a straightforward extension of this example.

Initially agent A is too far away from the other agents to infect them, but as time progresses is able to walk further and potentially infect the other agents. Within seconds, agent A is able to walk to and potentially infect agent B, which resides in the same building. However, agent A cannot reach agents C or D, because they are too far away at this point in time. Within a couple of minutes, agent A is able to walk to agent C that resides in a building on the same block. Within a few more minutes, agent A is able to walk to agent D. Based on this conceptual example, the following ten scenarios suggest the influence of STGs on model dynamics.

The first three scenarios address the influence of temporal granularity on model dynamics (Table

1.1). It is straightforward to calculate distances between agent locations to determine when agent A may reach the other agents. If agent A is 15m, 80m, and 310m from agents B, C, and D, respectively, then the minimum time to walk to each agent is 15, 80, and 310 seconds, respectively. The temporal granularity of a model may affect the simulated times when agent A may come into contact and potentially infect the other agents. In the first scenario, a temporal granularity of one second is sufficiently fine to capture the exact times when agent A may reach the other agents. The second scenario coarsens temporal granularity to one minute. In this scenario the model cannot differentiate whether agent A reaches another agent in 2 seconds or 52 seconds; because time progresses in minutes, both cases result in agent A reaching another agent in 60 seconds adding a slight delay for agent A to reach the other agents. In the third scenario, it is possible for agent A to reach each agent in the first time step, but the time step duration is one hour, adding further delays. These scenarios show that as temporal granularity is coarsened, longer time step durations delay the time at which agent A may reach the other agents introducing a temporal lag.

	Temporal granularity	**Spatial granularity**	**Minimum time to reach agent B**	**Minimum time to reach agent C**	**Minimum time to reach agent D**
Scenario 1	1 second	Point	15 (15s)	80 (80s)	310 (310s)
Scenario 2	1 minute	Point	1 (60s)	2 (120s)	6 (360s)
Scenario 3	1 hour	Point	1 (3600s)	1 (3600s)	1 (3600s)

Table 1.1: This table summarizes three scenarios to illustrate the influence of temporal granularity. The time to reach each agent represents discrete-time steps, which are calculated based on agent A moving to another agent using the quickest route. Each time step is converted to seconds for reference (in parenthesis).

The three aforementioned scenarios also highlight an interrelationship between time and space. An interrelationship becomes clear when considering the possibility of agent A reaching agents B, C, or D in a single discrete-time step as temporal granularity is coarsened. Due to the distances between agents, temporal granularity affects the number of agents that agent A may come into contact in a single time step. For example, if the temporal granularity is 1 second, then agent A may come into contact with at most one other agent in any given time step in this example. If the temporal granularity is coarsened to 1 minute and agent A is located between agents B and C, then it could walk to and come into contact with agent B or agent C in that time, which was

not possible when temporal granularity was 1 second. If the temporal granularity is coarsened to 1 hour, then agent A may come into contact with all three agents in a single time step. The interplay between temporal lags and a potential increase in the number of agents that agent A may reach in a single time step indicates a potential space and time dependence. On one hand, a temporal lag may delay subsequent infections thus slowing the spread of disease; on the other hand, an increase in the number of agents that an infectious agent may come in contact with may speed up the spread of disease, which may vary based on the density of agents in a simulation.

The next set of scenarios coarsen spatial granularity from a spatially explicit point-based representation to a gridded patch-based representation where agents may be located anywhere in a patch by considering patch sizes of 10m×10m, 100m×100m, and 1000m×1000m (Table 1.2). Figure 1.1 illustrates 100m×100m patches (delimited with grey lines) and 1000m×1000m patches (delimited with black lines). Due to the lack of spatial information when coarsening spatial granularity, agents have a non-zero probability of being located next to each other and therefore may come into contact at time step 0, which was impossible in the previous scenarios (Table 1.1), where spatially explicit locations were used to calculate distances and determine whether they were too far apart. In the fourth scenario, agent A may infect agent B at time step 0 because it is possible that they both reside on the edges of their respective patches and come into contact, but agents C or D reside on patches that are too far away. In the fifth scenario, agent A may infect agents B and C at time step 0, because they all reside on the same patch. In the sixth scenario, agent A may infect all the other agents at time step 0, because they all reside on the same patch. Coarsening spatial granularity, similar to temporal granularity, potentially increases the number of agents that may be infected in a single time step.

	Spatial granularity	**Agent B reachable**	**Agent C reachable**	**Agent D reachable**
Scenario 4	10m	Yes	No	No
Scenario 5	100m	Yes	Yes	No
Scenario 6	1000m	Yes	Yes	Yes

Table 1.2: These three scenarios indicate whether agent A may reach agents B, C, or D in time step 0 based on three spatial granularities: 10m×10m, 100m×100m, and 1000m×1000m. An agent may be reached by agent A if it resides in the same patch or an adjacent or cornering patch, because it is possible that both agents are at the edges of their respective patches and may come into contact.

The final four scenarios explore combinations of STGs (Table 1.3). It has already been established that agent A may reach all the agents if temporal granularity is an hour or spatial granularity is 1000m (Table 1.3) so we eliminate these granularities from consideration. Due to coarsened spatial granularity, agent B is reachable in time step 0 for all combinations, so attention is focused on agents C and D in these scenarios. Notice coarsening spatial granularity shortens the minimum time to reach agents C and D. Coarsening temporal granularity increases the minimum time to reach agents C and D, but the delay is influenced by spatial granularity. For example, the spatially fine-grained scenario (e.g. 10m×10m) results in an 8× delay to reach agent D, while the spatially coarse-grained scenario (e.g. 100m×100m) results in a 60× delay when coarsening temporal granularity. Coarsening temporal granularity did not affect the number of time steps in the spatially coarsened scenarios (i.e. 9 and 10), because the patch size was greater than the distance agent A could move in a single time step (i.e. 1 or 60 meters). Whereas in the spatially fine-grained scenarios (i.e. 7 and 8) agent A could move across multiple patches if the temporal granularity was 1 minute.

	Temporal granularity	**Spatial granularity**	**Minimum time to reach agent B**	**Minimum time to reach agent C**	**Minimum time to reach agent D**
Scenario 7	1 second	10m	0 (0s)	7 (7s)	29 (29s)
Scenario 8	1 minute	10m	0 (0s)	2 (120s)	4 (240s)
Scenario 9	1 second	100m	0 (0s)	0 (0s)	2 (2s)
Scenario 10	1 minute	100m	0 (0s)	0 (0s)	2 (120s)

Table 1.3: These four scenarios capture the simulated time for agent A to reach agents B, C, and D based on a two temporal granularities and two spatial granularities. The time to reach each agent represents discrete-time steps, which are calculated based on agent A moving to the furthest patch possible every time step to reach another agent using the shortest path. The time step is converted to seconds in each instance by multiplying the time step by the time step duration (in parenthesis).

A number of observations can be made based on this conceptual example. First, coarsening temporal granularity introduces a temporal lag suggesting that coarsening temporal granularity may slow disease spread. This delay is amplified if spatial granularity is also coarsened suggesting an interrelationship between space and time. While coarsening temporal granularity introduces a temporal lag, it also increases the number of agents that agent A may reach in a single time

step, further indicating the interrelationship between space and time. Second, coarsening spatial granularity in all scenarios shortens the minimum time for agent A to reach and potentially infect the other agents suggesting that coarsening spatial granularity may speed disease spread in epidemic models. Taken in combination, this conceptual example also reveals a potential space and time trade-off, because coarsening temporal granularity increases the minimum time for agent A to reach the other agents, which is amplified as spatial granularity is coarsened, but coarsening spatial granularity shortens the minimum time for agent A to reach other agents. Further investigation is necessary beyond this conceptual example, which is limited to a particular spatial configuration of agents, to improve our understanding of the influence of STGs. Chapters 2 and 3 investigate the influence of STGs on ABM processes and disease spread dynamics in epidemic ABMs.

1.3 Organization

This dissertation consists of three papers addressing computational and modeling challenges culminating in an investigation of the influence of STGs on simulated disease spread amongst residents in the state of Ohio, USA. First, I examine and address a computational challenge for parallel ABMs to enable the simulation of millions of agents (Chapter 2). Second, I design a novel modeling approach that enables ABMs to adapt to different STGs enabling the investigation of their influence (Chapter 3). Third, I design an epidemic ABM based on the work established in Chapters 2 and 3 to demonstrate the influence of STGs on the speed, intensity, and spatial spread of a synthetic disease in an epidemic ABM simulating residents in the state of Ohio, USA (Chapter 4). Finally, I discuss the contributions and major findings of my dissertation (Chapter 5).

The second chapter titled "A Communication-Aware Framework for Parallel Spatially Explicit Agent-Based Models" formulates a conceptual design of the relationships between parallel ABMs and inter-processor communication and establishes a communication-aware framework based on the conceptual design. Inter-processor communication is widely recognized as a computational bottleneck for parallel ABMs (Parry, 2009; Tang and Wang, 2009; Barrett et al., 2008; Parker, 2007), which limits the use of large numbers of processing cores for large-scale ABMs. This chapter, published in the *International Journal of Geographic Information Science*, examines and addresses this challenge to enable parallel ABMs to effectively use high-performance computing resources.

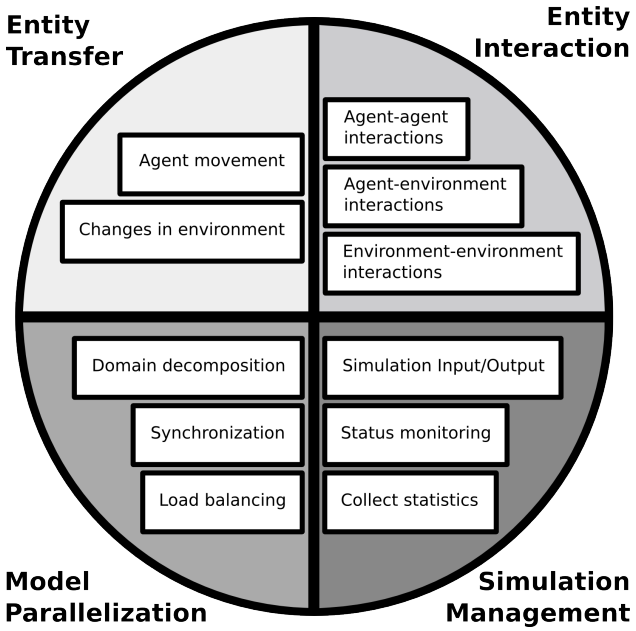


Figure 1.2: Conceptual design of inter-processor communication in a parallel ABM illustrating how different modeling functions such as agent movement or agent-environment interaction fit within one of four categories: entity interaction, entity transfer, simulation management, and model parallelization.

The second chapter formalizes the relationships between inter-processor communication and parallel ABMs in a conceptual design consisting of four categories: entity interaction, entity transfer, simulation management, and model parallelization (Figure 1.2). The conceptual design considers both agent and environment features and represent them generically as entities. Based on this conceptual design[1], the paper examines and resolves the challenge of inter-processor communication by establishing a generic framework for the management of inter-processor communication to enable parallel ABM to scale to high-performance parallel computers. The framework synthesizes four interrelated components: agent grouping, rectilinear domain decomposition, a communication-aware load-balancing strategy, and entity proxies. The results of a series of computational experiments based on a template ABM demonstrate that parallel computational efficiency diminishes as inter-processor communication increases, particularly when scaling a fixed-size ABM to thousands of processor cores. Therefore, effective communication management is crucial. The communication framework is shown to efficiently scale up to 2048 cores, demonstrating its ability to effectively

[1]The remainder of this paragraph is based on the abstract in the paper (Shook, et al., 2013) published in the *International Journal of Geographic Information Science*.

scale to thousands of processor cores to support the simulation of billions of agents. In a simulated scenario, the communication-aware load balancer reduced both overall simulation time and communication percentage improving overall computational efficiency. By examining and addressing inter-processor communication challenges, this research enables parallel SE-ABM to efficiently use high-performance computing resources.

The third chapter titled "Investigating the Influence of Spatial and Temporal Granularities in a Parsimonious Epidemic Agent-based Model" demonstrates that STGs affect speed, intensity, and spatial spread of a synthetic disease in a parsimonious ABM. The overarching goal of this chapter is to gain systematic understanding of the influence of STGs on epidemic ABM processes, which is enabled by the development of a novel modeling approach. Methodologically, this study is similar to those of network-based epidemic models (Xu and Sui, 2009; Keeling et al., 2010; Keeling and Eames, 2005; Shirley and Rushton, 2005; Keeling, 1999), but rather than altering network connections between agents in a synthetic model this study alters the representations of space and time to examine how variations of STGs influence ABM processes.

To enable the investigation of the influence of STGs, this chapter describes a novel approach, named "spatiotemporal process models," that contextualizes ABM processes explicitly within space and time and enables ABMs to adapt to different STGs. A parsimonious epidemic ABM is designed based on two spatiotemporal process models: agent movement and agent infection. In this way, the ABM is able to adapt to different STGs and facilitate the investigation of their influence without the drawback of comparing models with differing assumptions (Ajelli et al., 2010). In addition to spatial granularities, the modeling approach enables the investigation of temporal granularities, which until now has received little attention in the literature even though they may influence the ability of a model to capture changes in processes or model dynamics (Reitsma and Albrecht, 2005; Roche et al., 2011). The parsimonious ABM and spatiotemporal process modeling approach designed in this chapter serve as a basis for further inquiry in the fourth chapter.

The fourth chapter titled "The Influence of Spatial and Temporal Granularities in Agent-based Modeling of Disease Spread in Ohio, USA" integrates the communication-aware framework and spatiotemporal process modeling approach developed in Chapters 2 and 3 as part of an epidemic ABM that simulates every resident in the state of Ohio. The overarching goal of this chapter is to

study the influence of STGs on disease spread dynamics and computational tractability in a realistic scenario with varied population densities (i.e. urban versus rural), routine movement patterns (i.e. commuting), and tens of millions of heterogeneously distributed agents. Methodologically, this study is similar to studies comparing large-scale epidemic models (Ajelli et al., 2010; Ajelli and Merler, 2008; Halloran et al., 2008), but instead uses a single ABM to avoid drawbacks of comparing models with differing assumptions.

The epidemic ABM described in the fourth chapter is developed based on data from a number of sources including a synthetic United States population database developed for the Models of Infectious Disease Agent Study (MIDAS) (Wheaton et al., 2009), an US Census, an Annual School Construction Report, and an American Housing Survey. The ABM uses multiple movement processes to enable multiscale movement as agents move within and between home, school, and work buildings. A spatiotemporal process model calculates probabilities of infection for susceptible agents nearby infectious agents based on the STG of an ABM simulation. A series of experiments are designed to investigate the influence of STGs in a number of simulation scenarios. One experiment, in particular, examines if a commonly employed parameterization technique for epidemic models affects disease spread dynamics in the same way for spatially coarse- and fine-grained simulations. The findings of these experiments may have implications for the interpretation of epidemic model results including those used for informing state and national public policies.

The fifth chapter concludes the dissertation by discussing the contributions and synthesizing the major findings that revolve around the central theme of understanding the influence of STGs in agent-based modeling. Finally, I point out promising areas of future research.

References

Abbott, C., Berry, M., Comiskey, E., Gross, L., and Luh, H. (1997). Parallel individual-based modeling of Everglades deer ecology. *IEEE Computational Science & Engineering*, 4(4):60–72.

Ajelli, M., Gonçalves, B., Balcan, D., Colizza, V., Hu, H., Ramasco, J., Merler, S., and Vespignani, A. (2010). Comparing large-scale computational approaches to epidemic modeling: Agent-based versus structured metapopulation models. *BMC Infectious Diseases*, 10(1):190.

Ajelli, M. and Merler, S. (2008). The impact of the unstructured contacts component in influenza pandemic modeling. *PLoS One*, 3(1):e1519.

An, L. (2012). Modeling human decisions in coupled human and natural systems: Review of agent-based models. *Ecological Modelling*, 229:25–36.

An, L., Linderman, M., Qi, J., Shortridge, A., and Liu, J. (2005). Exploring complexity in a human–environment system: An agent-based spatial model for multidisciplinary and multiscale integration. *Annals of the Association of American Geographers*, 95(1):54–79.

Balcan, D., Gonccalves, B., Hu, H., Ramasco, J., Colizza, V., and Vespignani, A. (2010). Modeling the spatial spread of infectious diseases: The global epidemic and mobility computational model. *Journal of Computational Science*, 1(3):132–145.

Barrett, C., Bisset, K., Eubank, S., Feng, X., and Marathe, M. (2008). EpiSimdemics: An efficient algorithm for simulating the spread of infectious disease over large realistic social networks. In *SC '08: Proceedings of the 2008 ACM/IEEE conference on Supercomputing*, pages 1–12, Piscataway, NJ, USA. IEEE Press.

Batty, M., Crooks, A., See, L., and Heppenstall, A. (2012). Perspectives on agent-based models and geographical systems. In Heppenstall, A., Crooks, A., See, L., and Batty, M., editors, *Agent-Based Models of Geographical Systems*, pages 1–15. Springer Netherlands.

Bian, L. (2013). Spatial approaches to modeling dispersion of communicable diseases: A review. *Transactions in GIS*, 17(1):1–17.

Bisset, K., Aji, A., Marathe, M., and Feng, W. (2012). High-performance biocomputing for simulating the spread of contagion over large contact networks. *BMC Genomics*, 13(Suppl 2):S3.

Brown, S., Tai, J., Bailey, R., Cooley, P., Wheaton, W., Potter, M., Voorhees, R., LeJeune, M., Grefenstette, J., Burke, D., et al. (2011). Would school closure for the 2009 H1N1 influenza epidemic have been worth the cost?: A computational simulation of Pennsylvania. *BMC Public Health*, 11(1):353.

Chao, D., Halloran, M., Obenchain, V., and Longini, I. (2010). FluTE, a publicly available stochastic influenza epidemic simulation model. *PLoS Computational Biology*, 6(1):e1000656.

Clarke, K. (2003). Geocomputation's future at the extremes: High performance computing and nanoclients. *Parallel Computing*, 29(10):1281–1295.

Cornwell, C., Wille, L., Wu, Y., and Sklar, F. (2001). Parallelization of an ecological landscape model by functional decomposition. *Ecological Modelling*, 144(1):13–20.

Crooks, A., Castle, C., and Batty, M. (2008). Key challenges in agent-based modelling for geo-spatial simulation. *Computers, Environment and Urban Systems*, 32(6):417–430.

Epstein, J. (2009). Modelling to contain pandemics. *Nature*, 460(7256):687–687.

Ferguson, N., Cummings, D., Fraser, C., Cajka, J., Cooley, P., and Burke, D. (2006). Strategies for mitigating an influenza pandemic. *Nature*, 442(7101):448–452.

Ferrari, M., Perkins, S., Pomeroy, L., and Bjørnstad, O. (2011). Pathogens, social networks, and the paradox of transmission scaling. *Interdisciplinary Perspectives on Infectious Diseases*, 2011(267049):1–10.

Germann, T., Kadau, K., Longini, I., and Macken, C. (2006). Mitigation strategies for pandemic influenza in the United States. *Proceedings of the National Academy of Sciences of the United States of America*, 103(15):5935–5940.

Goodchild, M. (2004). GIScience, geography, form, and process. *Annals of the Association of American Geographers*, 94(4):709–714.

Halloran, M., Ferguson, N., Eubank, S., Longini, I., Cummings, D., Lewis, B., Xu, S., Fraser, C., Vullikanti, A., Germann, T., et al. (2008). Modeling targeted layered containment of an influenza pandemic in the United States. *Proceedings of the National Academy of Sciences*, 105(12):4639–4644.

Hupert, N., Xiong, W., Mushlin, A., et al. (2008). The virtue of virtuality: The promise of agent-based epidemic modeling. *Translational Research: The journal of Laboratory and Clinical Medicine*, 151(6):273–274.

Keeling, M. (1999). The effects of local spatial structure on epidemiological invasions. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 266(1421):859–867.

Keeling, M., Danon, L., Vernon, M., and House, T. (2010). Individual identity and movement networks for disease metapopulations. *Proceedings of the National Academy of Sciences*, 107(19):8866–8870.

Keeling, M. and Eames, K. (2005). Networks and epidemic models. *Journal of the Royal Society Interface*, 2(4):295–307.

Koopman, J. (2004). Modeling infection transmission. *Annual Review of Public Health*, 25:303–326.

Laskowski, M., Demianyk, B., Witt, J., Mukhi, S., Friesen, M., and McLeod, R. (2011). Agent-based modeling of the spread of influenza-like illness in an emergency department: A simulation study. *IEEE Transactions on Information Technology in Biomedicine: A publication of the IEEE Engineering in Medicine and Biology Society*, 15(6):877–889.

Lee, B., Brown, S., Korch, G., Cooley, P., Zimmerman, R., Wheaton, W., Zimmer, S., Grefenstette, J., Bailey, R., Assi, T., et al. (2010). A computer simulation of vaccine prioritization, allocation, and rationing during the 2009 H1N1 influenza pandemic. *Vaccine*, 28(31):4875–4879.

Lysenko, M. and D'Souza, R. (2008). A framework for megascale agent based model simulations on graphics processing units. *Journal of Artificial Societies and Social Simulation*, 11(4):10.

Mishra, S., Fisman, D., and Boily, M. (2011). The ABC of terms used in mathematical models of infectious diseases. *Journal of Epidemiology and Community Health*, 65(1):87–94.

Openshaw, S. (1984). *The modifiable areal unit problem*. Geo Books Norwich, UK.

Parker, D., Manson, S., Janssen, M., Hoffmann, M., and Deadman, P. (2003). Multi-agent systems for the simulation of land-use and land-cover change: A review. *Annals of the Association of American Geographers*, 93(2):314–337.

Parker, J. (2007). A flexible, large-scale, distributed agent based epidemic model. In *Proceedings of the 39th conference on Winter simulation (WSC)*, pages 1543–1547, Piscataway, NJ, USA. IEEE Press.

Parker, J. and Epstein, J. (2011). A distributed platform for global-scale agent-based models of disease transmission. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 22(1):2:1–2:25.

Parry, H. (2009). Agent based modeling, large scale simulations. In Meyers, R., editor, *Encyclopedia of Complexity and Systems Science*, pages 148–160. Springer New York.

Prieto, D., Das, T., Savachkin, A., Uribe, A., Izurieta, R., and Malavade, S. (2012). A systematic review to identify areas of enhancements of pandemic simulation models for operational use at provincial and local levels. *BMC Public Health*, 12(1):13.

Ramachandramurthi, S., Hallam, T., and Nichols, J. (1997). Parallel Simulation of Individual-Based, Physiologically Structured Population Models. *Mathematical and Computer Modelling*, 25(12):55–70.

Reitsma, F. and Albrecht, J. (2005). Implementing a new data model for simulating processes. *International Journal of Geographical Information Science*, 19(10):1073–1090.

Riley, S. (2007). Large-scale spatial-transmission models of infectious disease. *Science*, 316(5829):1298–1301.

Roche, B., Drake, J., and Rohani, P. (2011). An Agent-Based Model to study the epidemiological and evolutionary dynamics of Influenza viruses. *BMC Bioinformatics*, 12(1):87.

Shirley, M. and Rushton, S. (2005). The impacts of network topology on disease spread. *Ecological Complexity*, 2(3):287–299.

Smieszek, T. (2010). *Models of epidemics: How contact characteristics shape the spread of infectious diseases*. PhD thesis, Diss., Eidgenössische Technische Hochschule ETH Zürich, Nr. 18971, 2010.

Smieszek, T., Balmer, M., Hattendorf, J., Axhausen, K., Zinsstag, J., and Scholz, R. (2011). Reconstructing the 2003/2004 H3N2 influenza epidemic in Switzerland with a spatially explicit, individual-based model. *BMC Infectious Diseases*, 11(1):115.

Stroud, P., Del Valle, S., Sydoriak, S., Riese, J., and Mniszewski, S. (2007). Spatial dynamics of pandemic influenza in a massive artificial society. *Journal of Artificial Societies and Social Simulation*, 10(4):9.

Tang, W. and Wang, S. (2009). HPABM: A Hierarchical Parallel Simulation Framework for Spatially-explicit Agent-based Models. *Transactions in GIS*, 13(3):315–333.

Torrens, P. (2012). Moving agent pedestrians through space and time. *Annals of the Association of American Geographers*, 102(1):35–66.

Wainwright, J. and Mulligan, M. (2004). *Environmental Modelling: Finding Simplicity in Complexity*. John Wiley & Sons Inc., Wiley Chichester, UK.

Wang, D., Berry, M., Carr, E., and Gross, L. (2006). A parallel fish landscape model for ecosystem modeling. *Simulation*, 82(7):451–465.

Wheaton, W., Cajka, J., Chasteen, B., Wagener, D., Cooley, P., Ganapathi, L., Roberts, D., and Allpress, J. (2009). Synthesized population databases: A US geospatial database for agent-based models. RTI Press paper available at http://www.rti.org/pubs/mr-0010-0905-wheaton.pdf (Accessed: March, 2013).

Xu, Z. and Sui, D. (2009). Effect of small-world networks on epidemic propagation and intervention. *Geographical Analysis*, 41(3):263–282.

CHAPTER 2

A COMMUNICATION-AWARE FRAMEWORK FOR PARALLEL SPATIALLY EXPLICIT AGENT-BASED MODELS

2.1 Introduction

Spatially-explicit agent-based models (SE-ABM) have been widely used to study a variety of phenomena including, for example, pandemics, gentrification, and pedestrian movement (Epstein, 2009; O'Sullivan, 2002; Torrens, 2012). As the size of complex problems and sophistication of SE-ABM continue to increase, the computational limits of sequential computing approaches have become a major barrier to SE-ABM, and, thus, driven tremendous needs for exploiting parallel and high-performance computing (HPC) (Parker and Epstein, 2011; Tang and Wang, 2009). It is desirable for SE-ABM to gain straightforward access to not only massive computational resources, but also related spatial analysis methods and tools (Crooks and Castle, 2012), which can be accomplished through integration with CyberGIS, a new generation of geographic information systems (GIS) based on CyberInfrastructure (CI) (Wang, 2010).

To reap the benefits of CyberGIS, however, basic computational research is necessary to enable parallel SE-ABM to scale to larger numbers of processor cores for efficiently taking advantage of HPC resources, which allows, for example, enhanced capabilities for modeling larger and more complex problems at fine spatiotemporal granularities and improved computational performance (Parry and Bithell, 2012; Crooks et al., 2008). Within this context, an important research area is inter-processor communication that has been recognized as a major limiting factor in scaling parallel SE-ABM (Parry, 2009; Tang and Wang, 2009; Wang et al., 2006b; Barrett et al., 2008; Parker, 2007). This paper investigates the role of inter-processor communication (hereafter communication) in parallel SE-ABM and develops a communication framework that enables parallel SE-ABM to scale to a massive number of parallel processing cores (hereafter cores). Experiments based on a template agent-based model demonstrate that the framework is capable of supporting computationally intensive SE-ABM and scaling to thousands of cores.

The effective management of communication is important to the efficient use of HPC resources

especially for harnessing large numbers of cores (Kale et al., 1999). While many strategies and methods have been developed to manage and reduce communication in parallel SE-ABM (Parker, 2007; Scheutz and Schermerhorn, 2006; Wang et al., 2006; Parker and Epstein, 2011), efficient communication continues to challenge the scaling of large- and multi-scale spatial simulations. Therefore, this research focuses on a suite of methods for effective and efficient communication to drastically improve the scalability of parallel SE-ABM.

As a case study, we parallelized a Sugarscape model (Epstein and Axtell, 1996) to examine the effectiveness of the communication framework. Sugarscape is a widely used agent-based simulation testbed based on a set of agent interaction rules and environment specifications (Ginot et al., 2002; Lysenko and DSouza, 2008). Agents within the model represent people in an artificial society who interact within a two-dimensional lattice environment, with a renewable resource (sugar) that agents consume. Agents' behaviors are governed by rules that can be combined to construct sophisticated models. These models form a collective testbed to examine a variety of different phenomena including social structure, group dynamics, and the effects of combat, trade, or epidemiology (Epstein and Axtell, 1996).

To aid our investigation of the roles of communication in parallel SE-ABM, a conceptual design is established for parallel SE-ABM (Section 2.3). We then elaborate the framework in Section 2.4 encompassing a suite of four interrelated methods. Next, we assess the effectiveness of the methods by conducting computational performance experiments and extensively examining characteristics of communication within a simulated scenario (Section 3.4). The final section discusses our results and conclusions.

2.2 Background

This section provides an overview of SE-ABM and parallel SE-ABM. First, we detail the utility of SE-ABM and identify the computational challenges that face SE-ABM. Subsequently, we review the state of the art of parallel SE-ABM, and relationships between parallel SE-ABM and inter-processor communication.

2.2.1 Spatially-explicit Agent-based Models

ABM are a computational approach used to simulate dynamic phenomena at individual levels (Railsback et al., 2006), and have been applied in a variety of fields including for example epidemiology, ecology, geography, and social sciences (Epstein, 2009; Grimm et al., 2005; Parker et al., 2003; Heath et al., 2009). O'Sullivan and Haklay (2000) provide a definition of an agent as "... an autonomous, goal-directed software entity" (page 1410). Interactions among agents and their environment often produce complex dynamics with emergent properties (Crooks et al., 2008; O'Sullivan and Haklay, 2000) that may be difficult or infeasible to capture based on traditional analytical approaches (Goldstone and Janssen, 2005). SE-ABM represent a subset of ABM that explicitly represent space (e.g. field- or object-based) (Stanilov, 2012; Bian, 2003), to capture and simulate spatiotemporal dynamics. This paper focuses on SE-ABM simulating Lagrangian motion or agent movement across a landscape (Brown et al., 2005). A number of frameworks have been designed to facilitate such ABM development including for example NetLogo, Swarm, Repast, Ascape, and Mason (see reviews in Nikolai and Madey (2009); Crooks and Castle (2012)).

Computational challenges of large-scale ABM strain conventional sequential computing often forcing modelers to balance a trade-off between realistic representation and feasible computation (Crooks et al., 2008; Parry and Bithell, 2012). While some models may benefit from approaches such as agent aggregation, where an agent is modelled as a representative of a group of individuals also called super-individuals, which reduce computational requirements by modelling a smaller number of agents (Grimm, 1999; Hellweger, 2008). Other models are sensitive to spatial or individual variability (Shaman, 2007; Goldstone and Janssen, 2005), which aggregation approaches may fail to capture and as a result require computationally intensive simulations (see Parry and Bithell (2012) for discussion). Emergency management, for example, has a tremendous need for large-scale SE-ABM to simulate the responses and actions of millions or billions of individuals to hurricanes, wildfires, or disease epidemics (Hawe et al., 2012; Carley et al., 2006). Infectious disease ABM have grown from city and country to global scales (Parker and Epstein, 2011), and are now used to inform policy decisions in handling epidemics such as H5N1 in which models scaled to 6 billion agents enable simulations to capture disease transmission between agents and their individual behaviors under disease risk (Epstein, 2009). In these types of models, agent movement and interactions vary across

space and time creating computationally-intensive simulations with complex model dynamics that often require parallel and HPC approaches.

2.2.2 Parallel Spatially-explicit Agent-based Models

The computational requirements of a number of SE-ABM surpassed the limited capabilities of sequential computing necessitating the use of HPC (Tang and Wang, 2009), in which the computation of SE-ABM is decomposed into multiple sub-problems that are distributed to many cores to be solved simultaneously (Ramachandramurthi et al., 1997; Cornwell et al., 2001; Wang et al., 2006). Parallel and HPC can enable SE-ABM to significantly reduce overall simulation time and efficiently explore simulation provenance by effectively meeting significant computational requirements (Wang et al., 2006b; Abbott et al., 1997; Barrett et al., 2008; Bennett et al., 2011). A number of generic frameworks have been created to support parallel SE-ABM including ABM++ (https://www.epimodels.org/midas/pubabmplusplus.do), FLAME, SWAGES, Repast HPC, and HPABM (Deissenberg et al., 2008; Scheutz and Harris, 2011; Collier and North, 2011; Tang and Wang, 2009). Parallel ABM often use a discrete time-step or event-based representation of time in which time progresses in distinct stages (e.g. agent movement) or in response to individual events (e.g. agent X responds to agent Y) (Brown et al., 2005; Zeigler et al., 2000). The latter has benefited from active development in parallel discrete-event simulations that may reach massive-scale computation (Fujimoto, 1999; Perumalla and Seal, 2011).

Parallelization often requires inter-processor communication where a model is distributed among a number of cores, which must coordinate their actions to generate a cohesive simulation.[1] Communication in parallel SE-ABM can be implemented based on shared-memory and message-passing architectures. Parallel SE-ABM have been developed to exploit shared memory systems (Lysenko and DSouza, 2008; Wang et al., 2008), distributed memory systems (Barrett et al., 2008; Wang et al., 2006), and a collection of distributed systems referred to as grids (Tang et al., 2011; Wang et al., 2005). Communication has been understood as a bottleneck to the scalability of several models (akin to the number of cores that a model can effectively use) (Nichols et al., 2008; Tang and Wang,

[1] We differentiate inter-processor communication generated during a parallel agent-based modeling simulation from agent or environment communication within the ABM. For the remainder of this paper, we refer to inter-processor communication as communication, and communication among agents and environment as interaction.

2009; Barrett et al., 2008; Minson and Theodoropoulos, 2008; Wang et al., 2006). As a result, many modelers attempt to minimize and effectively manage communication using a number of strategies including for example load-balancing (Parker, 2007), bulk communication (Parker and Epstein, 2011), aggregation of agents (Lees et al., 2003), message brokers (Barrett et al., 2008), and ghost zones to improve efficiency of communications that simulate agent-agent and agent-environment interactions (Quinn et al., 2003; Wang et al., 2006; Nichols et al., 2008). Synchronization is commonly employed to maintain a cohesive simulation, but must be carefully managed to fully exploit parallelism. Barrett et al. (2008) broke a simulation into multiple phases to reduce the total number of synchronizations. Step-wise synchronization has been used by Scheutz and Schermerhorn (2006) to eliminate global synchronizations, while multiple synchronizations per time step can be engaged to maintain a lock-step model (Wang et al., 2006). Integrating communication-oriented strategies and methods into a cohesive framework to alleviate computational bottlenecks for parallel SE-ABM is the primary focus of our research.

2.3 Conceptual Design

Inter-processor communication links multiple cores to enable cohesive simulations in parallel SE-ABM. To help contextualize relationships between communication and parallel SE-ABM, our conceptual design of communication consists of four categories: entity interaction, entity transfer, simulation management, and model parallelization where an entity represents an agent or environment feature in SE-ABM.

Entity interaction coordinates communication generated by the interactions between entities (i.e. agent-agent, agent-environment, and environment-environment interactions) that reside on separate cores. Interactions among agents are often based on spatial proximity or shared network links (Qiu et al., 2008; Brown et al., 2005; Barrett et al., 2008). Agent clustering can occur if agents move into a common space or are connected to a network shared by a large number of agents, which can compound the amount of interactions between agents and their environment and subsequently the amount of communication. Environment-environment interactions are used to model environmental phenomena such as traffic or mudflows (Nagel and Rickert, 2001; Dattilo and Spezzano, 2003) and often require communication to coordinate information exchange between

environment features (e.g. number of cars on a highway segment or amount of material flowing downhill).

Entity transfer oversees the movement of entities (agents and environment features) across cores. Entities may be transferred due to changes in space, networks, or other contexts. The dynamic changes of agents and environments often result in significant fluctuations for entity transfer, which create communication imbalance among cores, because cores transfer different numbers of entities. Such communication imbalance may degrade computational performance if a few cores transferring a large number of entities slow down the majority of cores transferring only a few entities.

Simulation management coordinates processor activities. Functions within this category include: status monitoring, statistics gathering, and simulation input/output (I/O). Status monitoring enables the retrieval of status information about agents, environments, and simulations. Statistics gathering organizes simulation information (e.g. agent population characteristics) for simulation analysis. Simulation management in general engages extensive inter-processor communication to coordinate monitoring, gathering, and I/O of simulation information.

Model parallelization encompasses a set of parallel computing strategies for SE-ABM, including, but not limited to, domain decomposition, load-balancing, and model synchronization. Domain decomposition often has a large effect on communication, because a decomposition method defines how a parallel SE-ABM is partitioned which directly influences the amount of communication required by SE-ABM. Load-balancing strategies can help reduce communication (see Section 2.4.3 for details). Model synchronization is often engaged to maintain a cohesive simulation by assuring that all involved cores operate in unison. Communication is required to signal that all synchronizing cores have reached a synchronization point.

Our conceptual design provides insights toward understanding how various communication aspects influence computational performance by characterizing the relationships between communication and parallel SE-ABM into four categories. This design is applied to guide and evaluate our communication framework for parallel SE-ABM.

2.4 Communication Framework

Our communication framework bridges parallel SE-ABM and HPC by eliminating the need for SE-ABM to directly manage inter-processor communication, which can be understood as one core sending pieces of information to another core that correspondingly interprets the received information. Specifically, the framework needs to automatically identify (1) who (e.g. entities) needs to share information, (2) what information needs to be shared, and (3) with whom. To bridge this gap between identifying entities sharing information within an SE-ABM and the need to send, receive, and interpret the resultant information among cores in a parallel simulation we designed a communication framework composed of four interrelated components: group organization and operation, rectilinear domain decomposition, a communication-centric load-balancing method, and entity proxies. These four methods, described in detail as follows, address three important aspects based on the conceptual design, namely model parallelization (decomposition and load-balancing methods), entity transfer (group organization and operation), and entity interactions (entity proxies).

2.4.1 Groups

The group concept (e.g. scapes and swarms) is widely used in a number of ABM frameworks. In our research, groups are established as an abstraction for facilitating communication. Carefully constructed groups enable our framework to automatically identify each of the three elements for effective communication in SE-ABM, namely who is sharing, what information, and with whom. Groups are also used to achieve improved computational performance and reduced development complexity.

A *group* consists of a set of entities, which allows the communication framework to organize entities for supporting the separation of two conceptually different steps: group organization and group operation. Group organization identifies which entities to include in a particular group (e.g. group all entities within polygon P) addressing the first and third elements in determining who is sharing information and with whom. Group operation applies a procedure to a group (e.g. send group G to core X) addressing the second element in determining what information to share. The separation of these two steps allows group operations to be conceptualized and developed

independently of group organizations, which has a positive influence on modular development of the framework. For example, agent transfer, which often follows an agent movement step in parallel SE-ABM where agents may move off the subdomain of one core into the subdomain of another core, can be achieved by organizing a group of agents no longer residing in the local subdomain and transferring the group to its appropriate core using a send-group operation. The same send-group operation can be used by a load-balancing method to transfer a group of agents on an overloaded core to a lightly loaded core to improve workload balance. Additional examples of group organizations and operations can be found in Table 2.1.

Groups also help improve computational performance by reducing the number of communication functions while enabling bulk and asynchronous communication. If group organization and operation functions were combined, then the number of communication functions needed to support the various needs of SE-ABM would grow exponentially. Separating these two steps results in the need of fewer communication functions, which can be tuned and optimized for performance individually. One such optimization inherent in groups is bulk communication (Parker and Epstein, 2011), which is known to be effective for improving communication performance. Instead of sending N messages each containing one entity, one message is sent containing N entities of a group. If the framework knows when a particular piece of information is needed (e.g. a group) then it can use asynchronous communication, which overlaps communication and computation improving efficiency and performance (Gropp et al., 1999). For example, if a group is transferred from core X to core Y but is not needed immediately, the communication framework may allow the SE-ABM simulation to continue instead of waiting for the communication operation to finish. This flexibility can lead to improved performance for large-scale simulations (Barrett et al., 2008).

2.4.2 Rectilinear Domain Decomposition

In the case of parallel SE-ABM, domain decomposition partitions agents and their environments, so that they can be distributed to a number of processor cores to be executed in parallel (Parry and Bithell, 2012). Spatial domain decomposition exploits spatial characteristics to partition environments and agents into subdomains, and is widely used for SE-ABM (Parker and Epstein, 2011; Wang et al., 2006; Abbott et al., 1997). This research uses a two-dimensional regular grid

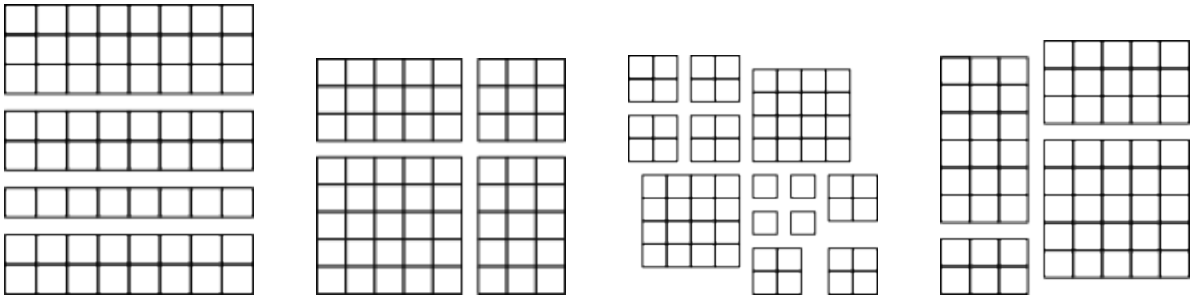
Organization
Entities residing in a rectangle region
Entities within radius R from point X, Y
Entities residing in subdomain S
Agents with $age = 10$
Environment cells with an attribute's value = 0
Agents residing on environment cells with elevation = 4 meters

(a)

Operation
Send a group to processor core P_1
Receive a group from processor core P_2
Count group size
Build histogram of agent age in a group
All entities in a group interact with agent A
Delete entities in a group

(b)

Table 2.1: Example group organizations and operations which can be reused in different combinations.



(a) Row (b) Grid (c) Quadtree (d) Recursive bisection

Figure 2.1: Illustrative decompositions for a 2D lattice that is partitioned for 4 cores.

for spatial domain decomposition. Such a grid-based field representation has been used by many SE-ABM (Linard et al., 2009; Tang et al., 2009; Wang et al., 2006; Abbott et al., 1997) to develop various decomposition strategies often based on row (1D) (Abbott et al., 1997; Wang et al., 2006b), grid (2D) (Quinn et al., 2003), quadtree, and recursive bisection decompositions (Figure 2.1).

Table 2.2 provides a qualitative comparison of the strengths and weaknesses among the aforementioned domain decomposition strategies. 1D decomposition (row/column) has at most two neighbors and thus high perimeter-to-area ratios that may cause excessive communication, because communication is often intense along the perimeters of subdomains (Wang et al., 2006b; Ding and Densham, 1996) due to entity transfer and interaction across cores. Grid decomposition does not over-decompose the environment where a core manages multiple subdomains while decompositions based on quadtree and recursive bisection achieve small perimeter-to-area ratios and better load-balancing flexibility. Quadtrees may result in over decomposition with each core managing a large number of fragmented subdomains of an environment (i.e. leaves in a quadtree), but can flexibly

Strategy	Max # neighbors	Constant neighbors	Overdecompose	Perimeter-to-area ratio	Load-balancing
Row/Column	2	yes	no	high	worst
Grid	8	yes	no	low-medium	good
Recursive Bisection	>10	no	no	low	better
Quadtree	>10	no	yes	low	best

Table 2.2: Characteristics of domain decomposition strategies for a 2D lattice environment, including a contiguity measure of the maximum number of surrounding neighbors of an environment section, whether these neighbors will remain constant if the environment is re-decomposed (e.g. for load-balancing), the possibility of overdecomposing the environment, perimeter-to-area ratio, and a rating of flexibility for load-balancing based on the number of possible decomposition configurations.

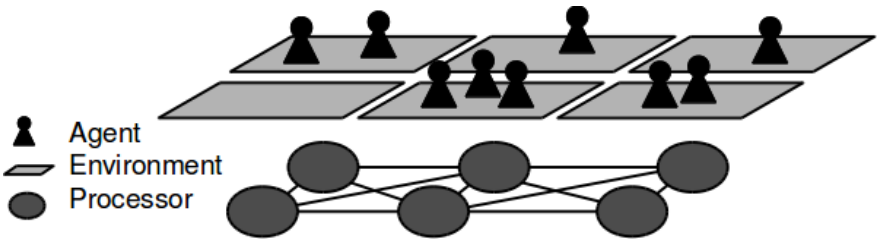


Figure 2.2: A 1-to-1 mapping between environment subdomains that share a common edge and a Cartesian grid of processing cores according to a rectilinear domain decomposition strategy.

support many small subdomains for load-balancing. Recursive bisection typically minimizes the perimeter-to-area ratio, but needs to manage the overhead of mapping irregular subdomains to the communication network of processing cores.

The communication framework adapts a 2D grid decomposition method named rectilinear domain decomposition (RDD) (Nicol, 1994). In our RDD implementation, each core owns a subdomain based on a 1-to-1 mapping between environment subdomains and cores. This 1-to-1 mapping arranges cores in a matching Cartesian grid to its corresponding decomposed spatial environment where any two cores responsible for subdomains that share a common edge or corner are neighbors (Figure 2.2). This decomposition strategy simplifies the management of communication operations, because cores have at most 8 neighbors including corners that are regularly arranged, and do not change throughout the simulation. Therefore, straightforward communication strategies such as those based on swapping information in North-South and East-West directions can be employed to implement communication that exchanges information among subdomains that share a common edge or corner (see Pinches et al. (1991)).

2.4.3 Distributed load-balancing strategy for rectilinear domain decomposition

Load-balancing improves computational performance by re-decomposing a model during a simulation to address changes of workload distribution among cores. Workload may change dynamically across processor cores, for example, due to movement of entities (e.g. agents move from one area to another area) or changes in environment features, which may cause more computation or communication workload for a core that was previously lightly loaded. Catalyurek et al. (2009) identified four primary objectives of load-balancing: 1. balanced loads; 2. low communication cost after the

application of load-balancing; 3. low cost to move data from an old to new workload distribution; and 4. short re-decomposition time.

Different load-balancing strategies emphasize different combinations of these objectives. Load-balancing based on RDD is proven to be an *NP*-hard problem called general block distribution of a matrix (Grigni and Manne, 1996). Therefore, near optimal approximations are acceptable to shorten re-decomposition time. Several approximation algorithms based on RDD have been developed (Nicol, 1994; Aspvall et al., 2001). These algorithms represent centralized strategies, which means that a single core calculates new workload partitions based on all workload information. These centralized RDD-based load-balancing algorithms do not satisfy the needs of our communication framework, because they do not take communication into account. While centralized strategies achieve more balanced loads, in general, they do not scale to a large number of cores because one core must balance the load of all cores increasing re-decomposition time as the number of cores increases. Distributed load-balancing strategies, on the other hand, use multiple cores, each core only requiring a limited amount of workload information, to calculate partitions (Catalyurek et al., 2009). Using multiple cores to calculate partitions enables de-centralized strategies to scale to a larger number of cores, because the cores involved in balancing workload can scale with the number of cores that need to be balanced rather than being limited to a single core.

We have developed a distributed load-balancing strategy that is designed to efficiently scale to a massive number of cores and improve computational performance by taking both computation and communication cost into account. Our strategy iterates over three primary steps to achieve the balance of computation and communication workload. One single iteration of execution of the strategy repartitions either rows or columns. Multiple executions can be used to partition rows and columns, referred to as iterative refinement (Nicol, 1994). Below, we detail the three steps, namely preparation, partition calculation, and redistribution.

The preparation step gathers and aggregates workload information representing the workload of the model domain. While our illustration is focused on the process for repartitioning columns of a rectilinear grid domain, the same steps can be applied to rows by changing the orientation. The process of load-balancing for our strategy begins with a RDD where each core is assigned with a rectangular subdomain. All of the subdomains are divided into fixed-width columns of blocks.

A block of a subdomain is the finest spatial unit for which our method estimates computation and communication workloads. Blocks can be spatially aggregated to represent the workloads of larger parts of a subdomain. Computation workloads are calculated based on the expected execution time needed to enable the simulation, transfer, and interaction of entities for an entire block. Communication workloads are calculated in the same way as computation workloads, but for edges of a block (Figure 2.3a). Transforming domain decomposition to workload information in this way can be understood as deriving a spatial computational domain to represent computation and communication intensity of subdomains (Wang and Armstrong, 2009).

The partition calculation step repartitions the blocks among subdomains to ensure no subdomain overloads any processor core. Similar to pairing neighboring white and black squares of a chessboard, this step pairs subdomains that share a column partition, then exhaustively searches for a partition that reduces the computation and communication workloads of overloaded subdomains. This is achieved through the use of blocks by testing the workload balance of every partitioning of the blocks. The load-balancing strategy exhausts all combinations and selects the partition that optimally balances the workload. Next, pairs of subdomains are alternated and the remaining half of the partitions are calculated in the same way.

Each core executes the partition calculation step by dividing its subdomain into blocks and calculating workloads for each block. Multiple strategies can be used to estimate workloads. One strategy to calculate computation workload is to count the number of agents (or environment cells) within a block and multiply it by a weight factor, such as an estimated time to simulate one agent for 100 iterations. An additional factor can be used to approximate clustered interactions, by raising the number by a small exponent based on the density of the agents within the block. Communication workload can be calculated in each direction (North, South, East, West) by counting the number of agents near the block perimeter and multiplying it by a weight factor, such as an estimated time to transfer one agent and interact with a remote entity 100 times each. Additional calculations can be included to account for modeling specific dynamics.

Then, each core sends its array of blocks to the top-most core of its column in the Cartesian grid of cores, analogous to the top squares of a chessboard (see the example Cartesian grid in Figure 2.2). These top-most cores organize the blocks of the subdomains by *row*, pair up, ex-

change and combine their respective arrays of blocks (arr), and then execute Algorithm 1, which iterates over all possible partitions of the blocks ($index$) while considering all subdomain workloads ($left\_$ *and* $right\_workload$ in each row) to find a partition that reduces the workload of overloaded subdomains ($minmax$). Figure 2.3 illustrates how workloads are calculated in a 4-core column-based example.

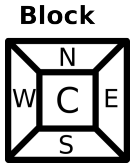
Finally, in the redistribution step, the new partitions are distributed to the remaining cores. According to the new partitions, the entities are grouped (see Groups in Section 2.4.1) and transferred, which finishes one iteration of executing the load-balancing strategy. Another load-balancing iteration could be executed alternating column-based with row-based partitioning to iteratively refine workload balance (Nicol, 1994).

Our distributed load-balancing strategy has both advantages and disadvantages. Partition indexes are constrained to the blocks within pairs of subdomains. Consequently, workload balancing might not achieve the best optimality compared to centralized strategies that have no such constraints on partitions (Catalyurek et al., 2009). However, centralized strategies cannot efficiently balance the workload of large-scale simulations that take advantage of massive numbers of cores. Our strategy can scale to a larger number of cores compared to centralized strategies, because partition calculations are distributed across a set of cores as opposed to a single core. Furthermore, our strategy incorporates communication workload to improve overall load-balancing performance.

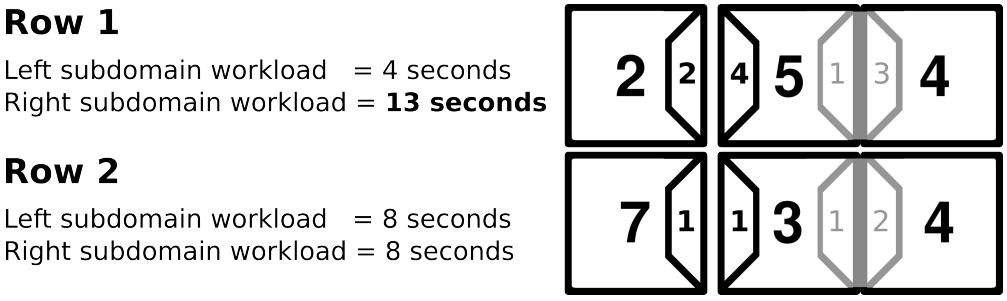
Algorithm 1 find_optimal_partition_index($arr, colsize, rowsize$)

Find the optimal column partition index ($minindex$) that minimizes the workload of the maximally loaded subdomain ($minmax$) based on workloads of the blocks (arr) by iterating over all partition indices ($index$)

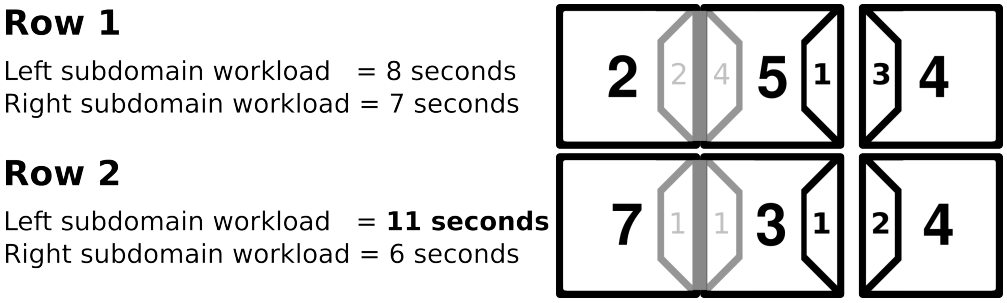
Require: block array (arr), width of block array ($colsize$), number of subdomains per column ($rowsize$)
  $minmax$=∞
  for $index = 2$ to $colsize - 1$ **do**
    $current\_max$=find_max_subdomain($arr, index, colsize, rowsize$)
    if $current\_max < minmax$ **then**
      $minmax = current\_max$
      $minindex = index$
    end if
  end for
  return $minindex$



(a) Block



(b) Block array, column partition index=1



(c) Block array, column partition index=2

Figure 2.3: A block contains workload information for computation (C), and communication in four directions (N, S, E, W). Figures 2.3b and 2.3c illustrate the workloads calculated by the *find_max_subdomain* operation for two partition indexes of a block array of size 2×3. Communication workloads are included only along an edge of a subdomain where communication occurs, otherwise are grayed out. The workload for the maximally loaded subdomain for each index is highlighted in bold. The partition index returned by the *find_optimal_partition_index* operation, minimizing the maximally loaded subdomain is 2.

Algorithm 2 find_max_subdomain($arr, index, colsize, rowsize$)

Find the workload of the maximally loaded subdomain ($max\_workload$) given a partition index ($index$) and workloads of the blocks (arr) by iterating over (row) left ($left\_workload$) and right ($right\_workload$) paired subdomains in a column

Require: block array (arr), partition index ($index$), width of block array ($colsize$), number of subdomains per column ($rowsize$)

```
max_workload=−∞
for row = 1 to rowsize do
  define left_workload as the workload of the left subdomain for row row
  calculated by aggregating blocks from (row,1)-(row,index) in arr
  define right_workload as the workload of the right subdomain for row row
  calculated by aggregating blocks from (row,index + 1)-(row,colsize) in arr
  if left_workload >= max_workload then
    max_workload=left_workload
  end if
  if right_workload > max_workload then
    max_workload=right_workload
  end if
end for
return max_workload
```

2.4.4 Entity Proxies

Ghost (or halo) zones are a parallel computing technique used to keep a local copy of data that reside in remote cores, and have also been employed in parallel SE-ABM (Quinn et al., 2003; Wang et al., 2006; Nichols et al., 2008; Tang and Wang, 2009). Generally, ghost zones surround a local computing environment serving as static placeholders of entities from remote cores. A ghost zones size often depends on entity interaction distances (e.g., agent vision or sphere of influence). We adapt the capabilities of ghost zones that often serve as repositories merely storing information from neighboring cores, to an active proxy of remote information that helps automatically manage communication. Specifically, we design agent and environment proxies that reside in ghost zones to represent corresponding entities on remote cores that forward information to their corresponding remote entity and receive its response. Entity proxies enable entity interaction among entities that reside on different cores, because a proxy can be treated as a local entity, compared to the case of ghost zones where an entity merely acknowledges the existence of a remote entity and must initiate inter-processor communication themselves to interact.

Automatic entity interaction is made possible through the use of group operations. When one

group interacts with another that contains entity proxies, those proxy interactions are captured and forwarded to the corresponding remote entity for each proxy using a send operation. Responses from remote entities are received and given to the local entities completing the group interaction. Again, groups facilitate bulk communication by representing multiple interactions among entities, enabling the framework to improve computational performance by sending multiple interactions using one send operation.

Entity proxies, similar to ghost zones, need to be carefully designed to avoid stale information, and reduce memory and communication costs. In our framework, remote entity proxies send and receive information to and from their corresponding local entities thereby eliminating issues related to incoherency and stale information, which can occur when a local copy is altered but a static remote proxy is not updated (Protic et al., 1996). Duplicated information among cores requires both additional memory and communication, because each processor is required to store its local entities as well as send, receive, and store entity proxies. Memory and communication costs may increase as the number of cores, number and size of agents or environments, or interaction distances increase. Therefore these costs are model dependent. For example, a model that limits agent interactions to 10 feet will use smaller ghost zones and likely fewer entity proxies reducing memory and communication costs, compared to a model that allows interactions up to 20 feet. The effect of overdecomposition is especially pronounced in this comparison. Consider dividing a 100x100 foot square in half using a ghost zone size of 10 or 20 feet resulting in 2,000 or 4,000 square feet of ghost zone coverage, respectively. Instead, divide the square into quarters assigning 2 sections per core (e.g. overdecomposition in the case of quadtree, for example). The resulting ghost zone coverage becomes 4,400 and 9,600 square feet (overlapping corners), effectively more than doubling the memory and communication costs. The interrelated components of our framework are designed to complement each other, minimizing such overheads as overdecomposition on entity proxies, to ensure effective and efficient communication management.

2.5 Experiments

A set of computational experiments was designed to evaluate the communication framework. The experiments aim to characterize the computational performance of the framework by measuring

speedup, efficiency, and communication-to-computation ratio. In addition, a simulated scenario is assessed to address the effectiveness of the conceptual design described in Section 2.3.

2.5.1 Sugarscape Parallelization

We developed a parallel Sugarscape model (Epstein and Axtell, 1996) based on the communication framework. The agents in our model share an identical set of traits. Specifically, agents begin with an identical amount of sugar (20 units); share the same vision trait (2 cells); and metabolic rate (0.5 units per cell). During each iteration, they search for and attempt to collect sugar from the environment and consume sugar while they move. Our Sugarscape model uses the following 3 rules: (M)ovement; (R)eplacement; and (G)rowback.

- **Agent movement (M)** look in all 4 lattice directions (North, South, East, West) and rank the cells according to sugar amount; randomly select and move to a cell with a higher probability given to cells with higher sugar amounts; and collect 50% of the sugar.
- **Agent replacement (R)** if an agent dies by consuming all its sugar, replace it with a new agent in a new randomly selected location within the dead agent's subdomain.
- **Environment growback (G)** sugar resources in each environment cell grow back at a rate ($\alpha * \{cell's\ sugar\ capacity\}$) up to the capacity of the cell. $\alpha = 75\%$[2].

The parallelized Sugarscape model is implemented in C programming language and uses the Message Passing Interface (MPI) (Gropp et al., 1999). Synchronization during a communication phase is maintained by each core 1. sending at least one message (even if empty) to each neighboring core, 2. responding to appropriate messages, and 3. maintaining a list of expected responses (e.g. entity interactions) from sent messages. In this way, each core has comprehensive knowledge of communication (e.g. information sent and expected to receive) that guarantees all cores receive all messages and that communication can continue as long as necessary to resolve any conflicts or enduring interactions. No further synchronization (e.g. barrier) is necessary, because each core has sent and received all messages. The model based on the associated communication framework

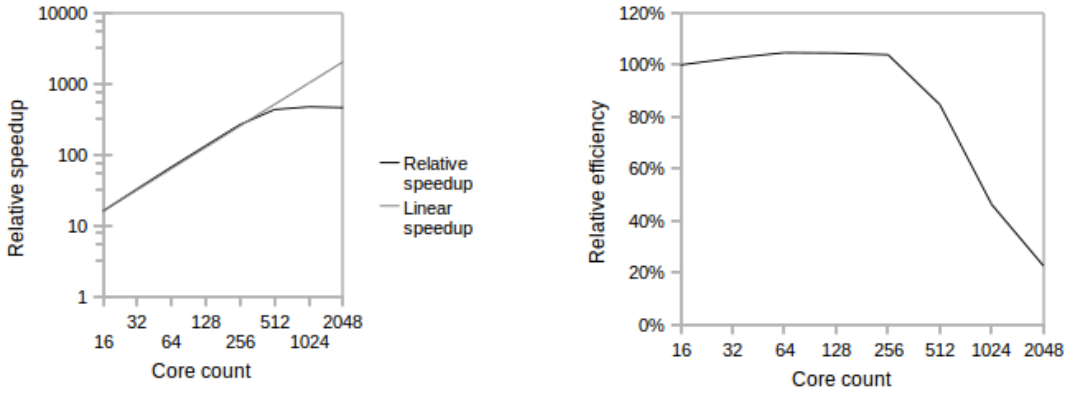
[2]The value 75% is selected based on experimental testing in order to support a large number of agents per cell such that the model is not resource bound and avoids eliminating agent-environment interactions entirely.

was executed on the Ranger supercomputer (supported by the National Science Foundation) that is comprised of 15,744 16-way AMD Opteron processors connected using InfiniBand with 123 TB memory and 1.7 PB of disk space. Performance information was collected using the Integrated Performance Monitoring (IPM) tool (Skinner, 2005).

2.5.2 Performance Analysis

We designed two experiments to measure the performance characteristics of the parallel Sugarscape model. Both experiments distribute a model among an increasing number of cores ranging from 16 to 2048. One experiment used a model of fixed size such that the model is decomposed into increasingly smaller subdomains that are mapped to an increasing number of cores. The model size in the other experiment increases linearly with the number of cores such that the subdomain size for each core remains fixed. The first experiment (hereafter referred to as fixed) is designed to gauge the ability of the communication framework to parallelize a model of fixed size to an increasing number of cores. The second experiment (hereafter referred to as scaled) is designed to test the scalability of the communication framework to handle increasingly larger models. To measure the effectiveness of the communication framework in the two experiments we use the speedup and scaled speedup metrics and their corresponding efficiencies. We compare the communication to computation ratios between the two experiments to understand the important role that communication plays in parallel SE-ABM. In these experiments the agents and environment cells are equally divided among the cores with the agents being distributed randomly within their subdomains. The load-balancing strategy is not applied, because the even distribution strategy results in little load imbalance.

Speedup (S_p) is commonly used to assess the performance of a parallel computing application (Amdahl, 1967) and is often employed for strong scaling evaluation where an increasing number of cores are used to solve a problem of fixed size (Barrett et al., 2008). Speedup is defined as the execution time on a single core (T_1) over the execution time on p cores (T_p) (Amdahl, 1967). Linear or ideal speedup is reached when $S_p = p$. Efficiency (E) is correlated with speedup and is used to evaluate the effective utilization of parallel computing resources (Kumar and Gupta, 1994). It is the ratio of speedup (S_p) divided by the number of cores (p). Efficiency typically ranges from 0 to



(a) Relative speedup

(b) Relative efficiency

Figure 2.4: Fixed experiment: speedup and efficiency relative to 16-core simulation times (measurements averaged over total time of executing the simulation 10 times with 100 iterations per simulation).

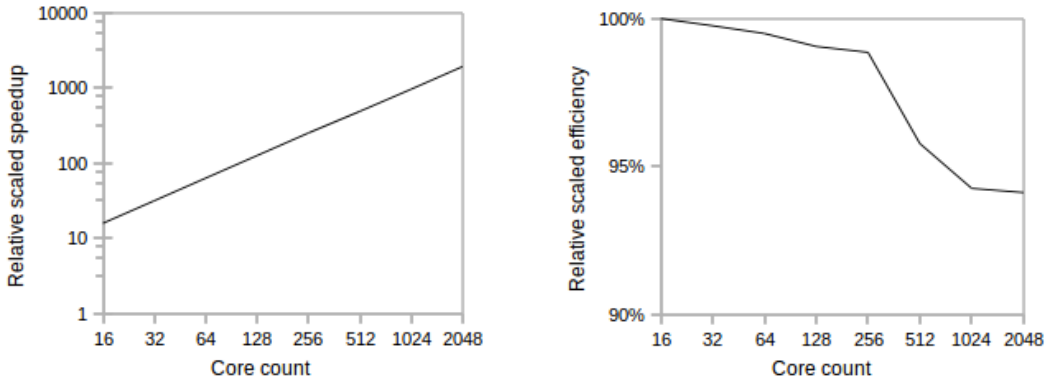
1 with linear speedup having efficiency of 100%.

$$S_p = \frac{T_1}{T_p} \tag{2.1}$$

$$E = \frac{S_p}{p} = \frac{\frac{T_1}{T_p}}{p} \tag{2.2}$$

The fixed experiment measures the speedup and efficiency of the communication framework (Figure 2.4) by increasing the number of cores from 16 to 2,048 while fixing the model size to $16,777,216$ agents with an environment size of $8,192 \times 8,192$ cells. The fixed experiment results exhibit super linear speedup, in which efficiency exceeds 100% (Gustafson, 1990). This can be explained by the following reasons: 1. memory bottlenecks increase higher latency memory accesses in a non-uniform memory architecture (NUMA) (Bolosky et al., 1991) at 16 cores (i.e. each core was responsible for processing too much data) that are alleviated by distributing the ABM across more cores, and 2. low parallel overhead resulting from effective computational management including communication across a larger number of cores. Due to these memory bottlenecks and Ranger's node size (16 cores), both experiments are relative to 16 cores instead of single core execution, consistent with similar large-scale simulations (Barrett et al., 2008; Perumalla and Seal, 2011).

Another metric used to evaluate performance is scaled speedup (Gustafson, 1988; Kumar and Gupta, 1994). We define scaled speedup as the ratio of execution time on a single core (T_1) times the



(a) Relative scaled speedup

(b) Relative scaled efficiency

Figure 2.5: Scaled experiment: scaled-speedup and efficiency relative to 16-core simulation times (measurements averaged over total time of executing the simulations 10 times with 100 iterations per simulation).

number of cores (p) over the time to execute on p cores (T_p). This metric is often employed for weak scaling experiments, and has been used to measure computational performance of parallel SE-ABM (Barrett et al., 2008). Scaled efficiency is used to evaluate the utilization of parallel computing resources when linearly scaling the problem size to the number of cores and is calculated by dividing the scaled speedup (SS_p) by the number of cores (p).

$$SS_p = \frac{p * T_1}{T_p} \tag{2.3}$$

$$SE = \frac{SS_p}{p} = \frac{\frac{p*T_1}{T_p}}{p} \tag{2.4}$$

Figure 2.5 shows the scaled experiment results (e.g. scaled speedup and efficiency). In this experiment the number of agents (2^{22} or $4,194,304$) and the environment size ($2,048x2,048$) are fixed per core and thus will increase with the number of cores. These parameters differ from the first experiment, because the number of entities grows linearly with the number of cores (from 16 to 2,048). We note that four outliers were removed from these results, because their simulation times dramatically altered the results (e.g. over 2% change in average simulation time when included), which were likely caused by system noise on the shared supercomputer Ranger which has been shown to be significant (Bhatelé et al., 2010).

The difference in efficiencies between the two experiments can be explained by the effect of

over decomposition. Over decomposition is the result of a model being decomposed among too many cores such that each core has little work, often resulting in a change in the communication to computation (C2C) ratio. C2C ratio is defined as the time for communication divided by the time for computation (Barrett et al., 2008; Wang et al., 2006b).

$$C2C = \frac{T_{comm}}{T_{comp}} \tag{2.5}$$

We compare the C2C ratios between these two experiments. Results with outliers removed, again likely caused by system noise on Ranger (Bhatelé et al., 2010), show that C2C ratio is higher for the fixed experiment compared to the scaled experiment (Figure 2.6). We noticed a slight dip in the 2048-core experiments. This can be explained by the situation where other users on the supercomputer were contending for the network infrastructure at the time of experiments, but in the case of the 2048-core experiments a significant portion of the supercomputer, including the network infrastructure, was being used to run the large experiments consequently reducing network usage (i.e. contention) from other users. It is important to note that efficiency diminishes as communication becomes a dominant factor. In fact Kale et al. (1999) showed that the C2C ratio must remain constant to maintain the same efficiency as the number of cores increases. This implies that effective management of communication is a key factor in achieving the scalability of a parallel application and thus has significant impact on the performance of parallel SE-ABM. The results of these experiments show that our communication framework maintains a low C2C ratio, not only in the scaled experiment but also in the fixed experiment, which is a key factor in the excellent efficiencies in both experiments (Figures 2.4 and 2.5).

Our performance analysis shows the important role that inter-processor communication plays in parallel SE-ABM and the effectiveness of our communication framework. The fixed experiment demonstrates that as the C2C ratio increases the efficiency of a simulation decreases, in particular the subtle dip in efficiency in the 1024-core fixed experiment coincides with the peak in the 1024-core C2C ratio further illustrating the importance of communication to performance. The scaled experiment demonstrates that our communication framework, if given a large enough model, maintains an exceptionally low C2C ratio to achieve excellent efficiency even as simulations execute on more than two thousand cores. Even in the fixed experiment, our communication framework

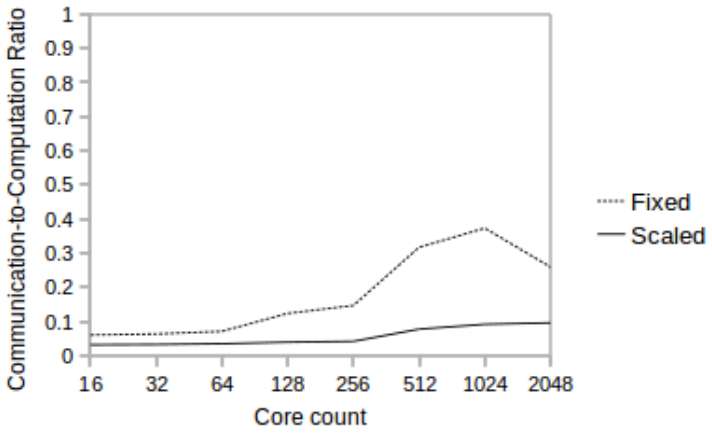


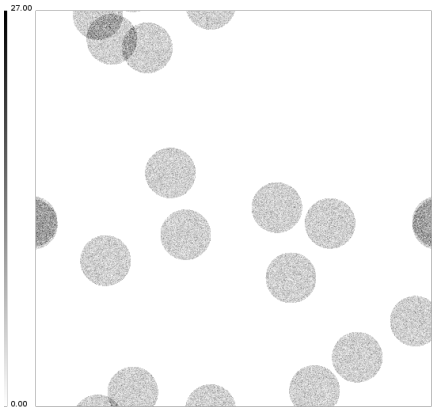
Figure 2.6: Communication-to-computation ratios for fixed and scaled experiments.

maintained a low C2C ratio and excellent efficiency until the effects of over decomposition reduced efficiency beyond 256 cores.

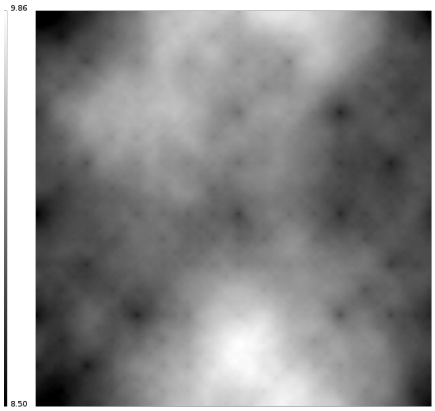
2.5.3 Simulated Scenario

This experiment is designed to emulate a real-world scenario by combining an initial distribution of clustered agents with varied environmental resources (see Figure 2.7 for representative agent distributions and environments). Specifically, we generated a 2,048x2,048 fractal landscape that corresponds to the amount of sugar in a Sugarscape environment. The fractal landscape is generated using a diamond-square method based on the equation c^{-l*H} (Fournier et al., 1982). The initial agent distribution is controlled using a parent-daughter clustering method (a modified Neyman-Scott cluster process (Neyman and Scott, 1958)). In the parent-daughter clustering method a number of daughter points (i.e. agent locations) are randomly distributed within a given radius to a parent point. Specifically, 16 parent points are randomly distributed within the landscape; and for each parent point, $1,048,576$ daughter points are randomly assigned within radius=128 from the parent point.

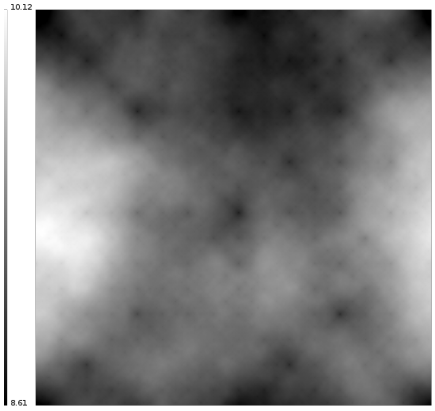
First, the distributed load-balancing strategy described in Section 2.4.3 is evaluated. This strategy is designed to address both computation and communication in decomposition calculations and, thus, should reduce both simulation time and the C2C ratio. We ran 10 simulations, each for 300 iterations using 256 cores with the load-balancer disabled and another 10 simulations with



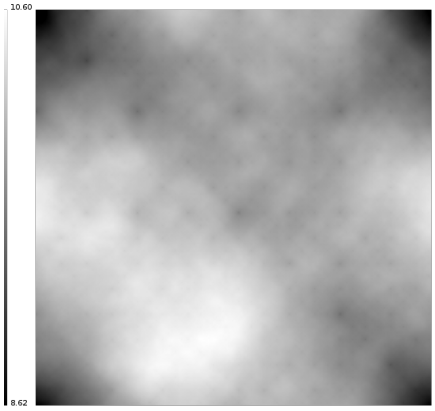
(a) A scenario of agent distribution



(b) Fractal environment: scenario 1



(c) Fractal environment: scenario 2



(d) Fractal environment: scenario 3

Figure 2.7: Representative scenarios of agent distributions (radius=128) and fractal environments (H=0.7).

Communication	Average time	Minimum time	Maximum time
Agent transfer	26.364	3.181	46.421
Environment transfer	18.004	0.135	42.920
Load-balancer	2.102	0.598	3.444
Simulation management	0.310	0.061	0.652

Table 2.3: Average, minimum, and maximum communication time (seconds) for agent and environment transfer, load-balancer, and simulation management for the simulation.

the load-balancer enabled. On average, the load-balancer reduced execution time by 20.2% (e.g. 69.58 seconds with the load-balancer disabled and 57.89 seconds with it enabled) and reduced communication percentage by 7.84% (e.g. a communication percentage of 85.35% with the load-balancer disabled and 79.14% with it enabled) even while taking into account the communication generated by the load-balancer itself to move workload data.

Next, we dissected the communication operations within the model to better understand the characteristics of inter-processor communication. We selected the simulation with the median simulation time out of the 10 simulations with the load-balancer enabled as a sample for further examination. The time of executing the simulation is 58.61 seconds with a communication percentage of 79.82%. We look at the change in communication throughout the simulation by comparing the average, minimum, and maximum communication time among processors for agent transfer, environment transfer, load-balancer, and simulation management throughout our sample simulation (see Table 2.3). Agent transfer dominated the communication time at 26.364 seconds on average with a change between minimum and maximum time of 43.24 seconds. The change in times provides evidence that communication imbalance can be significant for parallel SE-ABM and emphasizes the importance of effective communication management. When applying the conceptual design in Section 2.3 to this simulation, we see that entity transfer represents 94.83% of communication followed by parallelization at 4.49% with simulation management less than 1%. The breakdown of communication cost shows that our communication framework adds approximately 5% communication as overhead (i.e. simulation management, and redistributing workload through load balancing). However, the load-balancer reduced communication percentage by 7.84%. Therefore, our load-balancing strategy reduces overall inter-processor communication including accounting for its own overhead.

2.6 Concluding Discussion

SE-ABM have become an important approach for the study of complex spatial dynamics. These models have outpaced the computational capabilities of sequential computing approaches, demanding the research and development of parallel SE-ABM enabled by HPC. While CyberGIS enables computationally intensive spatial modeling based on straightforward access to HPC resources, parallel SE-ABM must overcome a major computational bottleneck, namely inter-processor communication. Past research on parallel SE-ABM has not sufficiently examined the role of inter-processor communication. This paper argues the importance of such communication in parallel ABM with a spatially explicit lens, and has established a communication framework to enable scalable parallel SE-ABM for efficiently using cutting-edge HPC resources.

Inter-processor communication is conceptualized into four categories: entity interaction, entity transfer, simulation management, and model parallelization, which form our conceptual design of communication. This design allows inter-processor communication to be explicitly taken into account in the development of methods for parallel SE-ABM. Specifically, the framework implements four interrelated methods that address different communication aspects involved in parallel SE-ABM: groups (entity transfer and interaction), rectilinear domain decomposition (model parallelization), a load-balancing strategy (model parallelization), and entity proxies (entity interaction). The framework supports dynamic decomposition based on load-balancing to enable efficient and scalable use of HPC resources. Groups and entity proxies allow the communication framework to efficiently organize, operate on, and manage, for example, agent-environment interactions based on massively parallel computer architectures. These methods specifically address the spatial characteristics of ABM by incorporating, for example, spatial group organization functions and using spatial domain decomposition and load-balancing strategies to adapt to varying spatial configurations of model entities.

Computational performance evaluation based on experiments has demonstrated that inter-processor communication plays a vital role in achieving desirable computational performance of parallel SE-ABM, particularly when a model is computed using a large number of processor cores on a tightly integrated high-performance computer. Furthermore, the experiments show that our communication framework scales to thousands of cores to support the simulation of billions of

agents, and has the ability to significantly improve overall computational performance. Lastly, the experiments show that our communication framework is able to adapt to different spatial configurations (e.g. agent clustering), particularly enabled through an effective load-balancing strategy.

In sum, the major contributions of this paper include a conceptual design and its associated implementation for an inter-processor communication framework enabling large-scale parallel SE-ABM on state-of-the-art high-performance computers. An implementation of the framework is available for download at http://cybergis.org/seabm. We look to extend this communication framework in two future directions. One focuses on investigating the synergistic integration between parallel SE-ABM and CyberGIS for improved pre- and post-simulation analysis and visualization (Tang et al., 2011). The other will adapt the framework to enable an epidemic SE-ABM to scale to massive computational resources simulating the movement and infection of millions of agents to help better understand spatial characteristics of disease spread.

Acknowledgements

This paper is based upon work supported by the National Science Foundation (NSF) under Grant Numbers: OCI-1047916 and BCS-0846655. Computational resources were provided through a supercomputing resource allocation award - SES070004 - by the NSF Extreme Science and Engineering Discovery Environment program. The authors greatly appreciate the insightful comments from Prof. May Yuan and three anonymous reviewers on the early version of this manuscript.

References

Abbott, C., Berry, M., Comiskey, E., Gross, L., and Luh, H. (1997). Parallel individual-based modeling of everglades deer ecology. *IEEE Computational Science & Engineering*, pages 60–72.

Amdahl, G. (1967). Validity of the single processor approach to achieving large scale computing capabilities. In *AFIPS '67 (Spring): Proceedings of the April 18-20, 1967, spring joint computer conference*, pages 483–485, New York, NY, USA. ACM.

Aspvall, B., Halldrsson, M., and Manne, F. (2001). Approximations for the general block distribution of a matrix. *Theoretical Computer Science*, 262(1–2):145–160.

Barrett, C., Bisset, K., Eubank, S., Feng, X., and Marathe, M. (2008). EpiSimdemics: an efficient algorithm for simulating the spread of infectious disease over large realistic social networks. In *SC '08: Proceedings of the 2008 ACM/IEEE conference on Supercomputing*, pages 1–12. IEEE Press Piscataway, NJ, USA, IEEE Press.

Bennett, D., Tang, W., and Wang, S. (2011). Toward an understanding of provenance in complex land use dynamics. *Journal of Land Use Science*, 6(2-3):211–230.

Bhatelé, A., Wesolowski, L., Bohm, E., Solomonik, E., and Kalé, L. (2010). Understanding application performance via micro-benchmarks on three large supercomputers: Intrepid, Ranger and Jaguar. *International Journal of High Performance Computing Applications*, 24(4):411–427.

Bian, L. (2003). The representation of the environment in the context of individual-based modeling. *Ecological Modelling*, 159(2-3):279–296.

Bolosky, W., Scott, M., Fitzgerald, R., Fowler, R., and Cox, A. (1991). Numa policies and their relation to memory architecture. In *ACM SIGARCH Computer Architecture News*, volume 19, pages 212–221. ACM.

Brown, D., Riolo, R., Robinson, D., North, M., and Rand, W. (2005). Spatial process and data models: Toward integration of agent-based models and GIS. *Journal of Geographical Systems*, 7(1):25–47.

Carley, K., Fridsma, D., Casman, E., Yahja, A., Altman, N., Chen, L., Kaminsky, B., and Nave, D. (2006). BioWar: scalable agent-based model of bioattacks. *IEEE Transactions on Systems, Man and Cybernetics, Part A*, 36(2):252–265.

Catalyurek, U., Boman, E., Devine, K., Bozdağ, D., Heaphy, R., and Riesen, L. (2009). A repartitioning hypergraph model for dynamic load balancing. *Journal of Parallel and Distributed Computing*, 69(8):711–724.

Collier, N. and North, M. (2011). Repast HPC: A platform for large-scale agent-based modeling. In Dubitzky, W., Kurowski, K., and Schott, B., editors, *Large-Scale Computing Techniques for Complex System Simulations*, pages 81–110. Wiley-IEEE Computer Society Press.

Cornwell, C., Wille, L., Wu, Y., and Sklar, F. (2001). Parallelization of an ecological landscape model by functional decomposition. *Ecological Modelling*, 144(1):13–20.

Crooks, A. and Castle, C. (2012). The integration of agent-based modelling and geographical information for geospatial simulation. In Heppenstall, A., Crooks, A., See, L., and Batty, M., editors, *Agent-Based Models of Geographical Systems*, pages 219–251. Springer Netherlands.

Crooks, A., Castle, C., and Batty, M. (2008). Key challenges in agent-based modelling for geo-spatial simulation. *Computers, Environment and Urban Systems*, 32(6):417–430.

Dattilo, G. and Spezzano, G. (2003). Simulation of a cellular landslide model with CAMELOT on high performance computers. *Parallel Computing*, 29(10):1403–1418.

Deissenberg, C., Van Der Hoog, S., and Dawid, H. (2008). EURACE: A massively parallel agent-based model of the european economy. *Applied Mathematics and Computation*, 204(2):541–552.

Ding, Y. and Densham, P. (1996). Spatial strategies for parallel spatial modelling. *International Journal of Geographical Information Systems*, 10(6):669–698.

Epstein, J. (2009). Modelling to contain pandemics. *Nature*, 460(7256):687–687.

Epstein, J. and Axtell, R. (1996). *Growing Artificial Societies: Social Science from the Bottom Up*. MIT Press.

Fournier, A., Fussell, D., and Carpenter, L. (1982). Computer rendering of stochastic models. *Communications of the ACM*, 25(6):371–384.

Fujimoto, R. (1999). Parallel and distributed simulation. In *Simulation Conference Proceedings, 1999 Winter*, volume 1, pages 122–131. IEEE.

Ginot, V., Le Page, C., and Souissi, S. (2002). A multi-agents architecture to enhance end-user individual-based modelling. *Ecological Modelling*, 157(1):23–41.

Goldstone, R. and Janssen, M. (2005). Computational models of collective behavior. *Trends in Cognitive Sciences*, 9(9):424–430.

Grigni, M. and Manne, F. (1996). On the complexity of the generalized block distribution. *Lecture Notes in Computer Science*, 1117:319–326.

Grimm, V. (1999). Ten years of individual-based modelling in ecology: what have we learned and what could we learn in the future? *Ecological modelling*, 115(2):129–148.

Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W., Railsback, S., Thulke, H., Weiner, J., Wiegand, T., and DeAngelis, D. (2005). Pattern-oriented modeling of agent-based complex systems: lessons from ecology. *Science*, 310(5750):987–991.

Gropp, W., Lusk, E., and Skjellum, A. (1999). *Using MPI: portable parallel programming with the message-passing interface*. MIT Press.

Gustafson, J. (1988). Reevaluating Amdahl's law. *Communications of the ACM*, 31(5):532–533.

Gustafson, J. (1990). Fixed time, tiered memory, and superlinear speedup. In *Distributed Memory Computing Conference, 1990., Proceedings of the Fifth*, volume 2, pages 1255–1260. IEEE.

Hawe, G., Coates, G., Wilson, D., and Crouch, R. (2012). Agent-based simulation for large-scale emergency response: A survey of usage and implementation. Available at http://www.dur.ac.uk/resources/profiles/8266/CompSurvReviewpaper.pdf (Accessed on July 4, 2012).

Heath, B., Hill, R., and Ciarallo, F. (2009). A survey of agent-based modeling practices (january 1998 to july 2008). *Journal of Artificial Societies and Social Simulation*, 12(4):9.

Hellweger, F. (2008). Spatially explicit individual-based modeling using a fixed super-individual density. *Computers and Geosciences*, 34(2):144–152.

Kale, L., Skeel, R., Bhandarkar, M., Brunner, R., Gursoy, A., Krawetz, N., Phillips, J., Shinozaki, A., Varadarajan, K., and Schulten, K. (1999). NAMD2: Greater scalability for parallel molecular dynamics. *Journal of Computational Physics*, 151:283–312.

Kumar, V. and Gupta, A. (1994). Analyzing scalability of parallel algorithms and architectures. *Journal of Parallel and Distributed Computing*, 22(3):379–391.

Lees, M., Logan, B., Oguara, T., and Theodoropoulos, G. (2003). Simulating agent-based systems with HLA: The case of SIM_AGENT – part II. In *Proceedings of the 2003 European Simulation Interoperability Workshop*, pages 285–293. European Office of Aerospace R&D, Simulation Interoperability Standards Organisation and Society for Computer Simulation International.

Linard, C., Ponçon, N., Fontenille, D., and Lambin, E. (2009). A multi-agent simulation to assess the risk of malaria re-emergence in southern France. *Ecological Modelling*, 220(2):160–174.

Lysenko, M. and DSouza, R. (2008). A framework for megascale agent based model simulations on graphics processing units. *Journal of Artificial Societies and Social Simulation*, 11(4):10.

Minson, R. and Theodoropoulos, G. (2008). Distributing RePast agent-based simulations with HLA. *Concurrency and Computation: Practice and Experience*, 20(10).

Nagel, K. and Rickert, M. (2001). Parallel implementation of the TRANSIMS micro-simulation. *Parallel Computing*, 27(12):1611–1639.

Neyman, J. and Scott, E. (1958). Statistical approach to problems of cosmology. *Journal of the Royal Statistical Society. Series B (Methodological)*, 20(1):1–43.

Nichols, J., Hallam, T., and Dimitrov, D. (2008). Parallel simulation of ecological structured communities: Computational needs, hardware capabilities, and nonlinear applications. *Nonlinear Analysis*, 69(3):832–842.

Nicol, D. (1994). Rectilinear partitioning of irregular data parallel computations. *Journal of Parallel and Distributed Computing*, 23(2):119–134.

Nikolai, C. and Madey, G. (2009). Tools of the trade: A survey of various agent based modeling platforms. *Journal of Artificial Societies and Social Simulation*, 12(2):2.

O'Sullivan, D. (2002). Toward micro-scale spatial modeling of gentrification. *Journal of Geographical Systems*, 4(3):251–274.

O'Sullivan, D. and Haklay, M. (2000). Agent-based models and individualism: is the world agent-based? *Environment and Planning A*, 32:1409–1425.

Parker, D., Manson, S., Janssen, M., Hoffmann, M., and Deadman, P. (2003). Multi-agent systems for the simulation of land-use and land-cover change: a review. *Annals of the Association of American Geographers*, 93(2):314–337.

Parker, J. (2007). A flexible, large-scale, distributed agent based epidemic model. In *WSC '07: Proceedings of the 39th conference on Winter simulation*, pages 1543–1547. IEEE Press, IEEE Press.

Parker, J. and Epstein, J. (2011). A distributed platform for global-scale agent-based models of disease transmission. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 22(1):2:1–2:25.

Parry, H. (2009). Agent based modeling, large scale simulations. In Meyers, R., editor, *Encyclopedia of Complexity and Systems Science*, pages 148–160. Springer New York.

Parry, H. and Bithell, M. (2012). Large scale agent-based modelling: a review and guidelines for model scaling. In Heppenstall, A., Crooks, A., See, L., and Batty, M., editors, *Agent-Based Models of Geographical Systems*, pages 271–308. Springer Netherlands.

Perumalla, K. and Seal, S. (2011). Discrete event modeling and massively parallel execution of epidemic outbreak phenomena. *Simulation*.

Pinches, M., Tildesley, D., and Smith, W. (1991). Large scale molecular dynamics on parallel computers using the link-cell algorithm. *Molecular Simulation*, 6(1):51–87.

Protic, J., Tomasevic, M., and Milutinovic, V. (1996). Distributed shared memory: Concepts and systems. *IEEE Parallel & Distributed Technology: Systems & Applications*, 4(2):63–79.

Qiu, F., Li, B., Chastain, B., and Alfarhan, M. (2008). A GIS based spatially explicit model of dispersal agent behavior. *Forest Ecology and Management*, 254(3):524–537.

Quinn, M., Metoyer, R., and Hunter-zaworski, K. (2003). Parallel implementation of the social forces model. In *Proceedings of the Second International Conference in Pedestrian and Evacuation Dynamics*, pages 63–74. CMS Press, University of Greenwich.

Railsback, S., Lytinen, S., and Jackson, S. (2006). Agent-based simulation platforms: Review and development recommendations. *Simulation*, 82(9):609–623.

Ramachandramurthi, S., Hallam, T., and Nichols, J. (1997). Parallel simulation of individual-based, physiologically structured population models. *Mathematical and Computer Modelling*, 25(12):55–70.

Scheutz, M. and Harris, J. (2011). An overview of the simworld agent-based grid experimentation system. In Dubitzky, W., Kurowski, K., and Schott, B., editors, *Large-Scale Computing Techniques for Complex System Simulations*, pages 59–80. Wiley-IEEE Computer Society Press.

Scheutz, M. and Schermerhorn, P. (2006). Adaptive algorithms for the dynamic distribution and parallel execution of agent-based models. *Journal of Parallel and Distributed Computing*, 66(8):1037–1051.

Shaman, J. (2007). Amplification due to spatial clustering in an individual-based model of mosquito-avian arbovirus transmission. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 101(5):469–483.

Skinner, D. (2005). Performance monitoring of parallel scientific applications. Technical report LBNL/PUB–5503, Lawrence Berkeley National Laboratory, Berkeley, CA.

Stanilov, K. (2012). Space in agent-based models. In Heppenstall, A., Crooks, A., See, L., and Batty, M., editors, *Agent-Based Models of Geographical Systems*, pages 253–269. Springer Netherlands.

Tang, W., Malanson, G., and Entwisle, B. (2009). Simulated village locations in Thailand: a multi-scale model including a neural network approach. *Landscape Ecology*, 24(4):557–575.

Tang, W. and Wang, S. (2009). HPABM: a hierarchical parallel simulation framework for spatially-explicit agent-based models. *Transactions in GIS*, 13(3):315–333.

Tang, W., Wang, S., Bennett, D., and Liu, Y. (2011). Agent-based modeling within a cyber-infrastructure environment: a service-oriented computing approach. *International Journal of Geographical Information Science*, 25(9):1323–1346.

Torrens, P. (2012). Moving agent pedestrians through space and time. *Annals of the Association of American Geographers*, 102(1):35–66.

Wang, D., Berry, M., Carr, E., and Gross, L. (2006a). A parallel fish landscape model for ecosystem modeling. *Simulation*, 82(7):451–465.

Wang, D., Berry, M., and Gross, L. (2006b). On parallelization of a spatially-explicit structured ecological model for integrated ecosystem simulation. *International Journal of High Performance Computing Applications*, 20(4):571–581.

Wang, D., Berry, M., and Gross, L. (2008). A parallel structured ecological model for high end shared memory computers. *Lecture Notes in Computer Science*, 4315:107–118.

Wang, D., Carr, E., Gross, L., and Berry, M. (2005). Toward ecosystem modeling on computing grids. *Computing in Science & Engineering*, 7(5):44–52.

Wang, S. (2010). A CyberGIS framework for the synthesis of cyberinfrastructure, GIS, and spatial analysis. *Annals of the Association of American Geographers*, 100(3):535–557.

Wang, S. and Armstrong, M. (2009). A theoretical approach to the use of cyberinfrastructure in geographical analysis. *International Journal of Geographical Information Science*, 23(2):169–193.

Zeigler, B., Praehofer, H., and Kim, T. (2000). *Theory of modeling and simulation: Integrating discrete event and continuous complex dynamic systems*. Academic Press.

CHAPTER 3

INVESTIGATING THE INFLUENCE OF SPATIAL AND TEMPORAL GRANULARITIES IN A PARSIMONIOUS EPIDEMIC AGENT-BASED MODEL

3.1 Introduction

Computational models are formulated to better understand complex spatiotemporal dynamics by comparing computer simulations to observations and theories. While previous work in computational spatial modeling recognizes the importance of the embodiment of space and time (Batty et al., 2012; Manson et al., 2012; Bian, 2004), computational representations of space and time are often coarsened due to limitations in fine-scale understanding, data, and computational capabilities (Hagen-Zanker and Jin, 2012; Ajelli et al., 2010; Riley, 2007). This research investigates how the finest representations of space and time, here termed spatial and temporal granularities (STGs), shape model processes–capturing movements and interactions of agents and environments–that operate within the constraints of such representations. Specifically, we formulate a novel approach that enables disease models to adapt to different STGs. This approach is applied to a disease model for providing a new spatial and temporal lens to critically examine epidemic models as motivated by previous studies comparing epidemic models (Xu and Sui, 2009; Ajelli et al., 2010; Ajelli and Merler, 2008; Kaplan and Wein, 2003).

Disease spread is a complex, multi-scale process dispersing from an individual to a community, city, and country over days, weeks, and months. As a result, disease spread can be modeled and studied from a variety of spatial and temporal perspectives (Figure 3.1). Disease maps have long been used to understand disease spread (Snow, 1855; Beyer et al., 2012). Spatial analyses have been applied to studies of disease spread including for disease detection (Emch et al., 2012; Kulldorff and Nagarwalla, 1995). Multiple studies have conceptualized epidemic disease spread as a wave moving through space and time to investigate spatial and temporal characteristics of disease spread based on historical diseases (Cliff and Haggett, 2006; Cliff et al., 2008). Disease models, including dynamic, cellular-automata, and agent-based ones, rely on simplifying assumptions to simulate disease spread in a computational model using desktop or high-performance computing (Hethcote,

2000; Bian, 2013; Parker and Epstein, 2011).

Researchers and policy makers increasingly rely on epidemic agent-based model (ABM) simulations where individual agents move and interact passing disease amongst themselves to understand how communicable diseases such as influenza and smallpox may spread throughout a population (Epstein, 2009; Ferguson et al., 2003). Comparing the results of these models that often differ in assumptions and STGs (Bobashev et al., 2007; Ajelli et al., 2010; Halloran et al., 2008) may lead to opposing interpretations (Kaplan and Wein, 2003). Therefore, the influence of STGs on modeling results needs to be better understood in the context of epidemic modeling.

Our study systematically examines how variations of representations of both space and time affect the dynamics of a single epidemic ABM that adapts to different STGs. Methodologically, our approach is similar to the work on network-based epidemic models (Xu and Sui, 2009; Keeling et al., 2010; Keeling and Eames, 2005; Shirley and Rushton, 2005; Keeling, 1999) that seeks to understand how alterations in network connections between agents influence the spread of disease. In both cases, synthetic models are manipulated to understand how changes in disease spreading mechanisms (i.e., network connections or STGs) influence the spread of disease. We design an ABM that is capable of adapting to different STG's thus enabling our investigation into the influence of STGs on simulated disease spread while improving spatial and temporal understanding of modeling epidemic dynamics.

3.1.1 Space, Time, and Processes in Epidemic Models

A grand challenge for epidemic models is to accurately simulate disease spread in the absence of exact and complete data, including, for example, the locations of individuals and their homes or workplaces. Modelers face data limitations for both population and disease data, which is often coarsened due to concerns such as privacy (Maantay and McLafferty, 2011). Data used to assign persons to workplaces or schools is also often aggregated or largely unavailable (Wheaton et al., 2009). Further, data is often lacking to understand where people may come into contact outside the home or workplace, termed unstructured contacts (Ajelli and Merler, 2008). To compensate for limitations in data, epidemic models often rely on simplifying assumptions or abstractions, which may influence simulated disease dynamics including the locations of simulated individuals or with

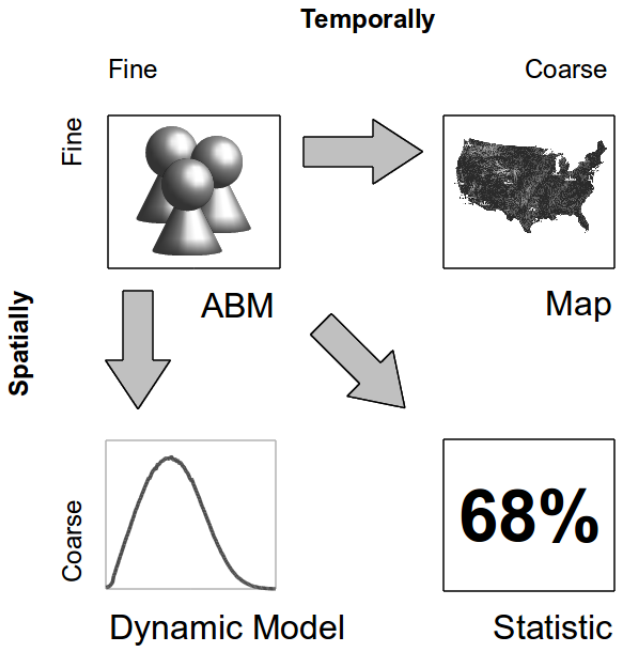


Figure 3.1: This figure illustrates a spectrum of spatial and temporal granularities situating fine-grained ABMs with dynamic models (i.e., no spatial representation), disease maps (i.e., no temporal representation), and disease statistics (i.e., no spatial or temporal representation). Different ABMs may lie at different points along this spatiotemporal spectrum.

whom they contact.

Interactions among agents form contact or mixing patterns that are central to understanding how a disease spreads through populations (Koopman, 2004). Mixing patterns include homogeneous (i.e., perfectly mixed) or heterogeneous, and are influenced by model assumptions and parameters (Mishra et al., 2011), which often differ amongst ABMs. Such differences include how agents move around and how a disease spreads from agent to agent (Atti et al., 2008; Germann et al., 2006; Ferguson et al., 2006). Changing spatial patterns of contact or mixing amongst agents is shown to influence simulated disease spread (Ajelli and Merler, 2008; Ajelli et al., 2010). While it is known that human mobility (Brockmann, 2009; Balcan et al., 2009), agent-to-agent contacts (Rahmandad and Sterman, 2010), and duration of such contacts (Smieszek, 2009) play an important role in agent-based modeling of disease spread, much less is known about how variations of the representations of space and time determine when and where agents come into contact or the duration of agent contacts.

Disease transmission modeling methods often adopted in spatial models can be categorized into

four common abstractions (Riley, 2007). We recast these within an agent-based modeling context for consistency within our discussion.

1. **Group**. Disease transmission among groups of agents can only occur if they share a connection. Transmission rates among groups can vary, but agents within a group share the same rate of infection.

2. **Network** models disaggregate groups into a network of agents in which transmission can only occur between connected agents.

3. **Patch**. Agents within a patch share the same rate of infection, but disease transmission to nearby patches may occur at different rates usually based on distance between patches.

4. **Distance** is calculated between susceptible and infectious agents to determine a rate of infection.

Embedded within each of the four abstractions lie distinctive assumptions about the representations of space for epidemic modeling that may influence disease spread. Patch and distance transmission methods *explicitly represent space* by relying on an agent's location. Whereas groups and networks *implicitly represent space* by relying on spatial characteristics such as agents' home, work, or school buildings to establish connections amongst groups or agents (Germann et al., 2006; Ferguson et al., 2006). These abstractions also specify different spatial granularities: *coarse- to fine-grained* (group and patch) and *very fine-grained* (network and distance).

Time representation is critical to ABMs (Liu and Andersson, 2004; Crooks and Castle, 2012). Epidemic ABMs commonly use two types of time representation, namely continuous and discrete. Continuous-time or event-based models operate on a temporally ordered sequence of events in which each event may change something in the model (e.g., an agent's location) or trigger a new event (Zeigler et al., 2000; Wainer and Mosterman, 2011). Discrete-time models operate based on a global clock that ticks at fixed-time intervals (e.g., 6 or 12 hours) and enforces the model to be updated simultaneously (Germann et al., 2006; Ferguson et al., 2006). More epidemic ABMs, particularly at the national or multi-national scale, use discrete representations of time (Chao et al., 2010; Atti et al., 2008; Germann et al., 2006; Ferguson et al., 2006; Longini et al., 2004).

3.2 A Parsimonious Epidemic Agent-based Model

Our study is novel in transforming a distance-based disease model to a patch-based one (Riley, 2007) to investigate the influence of STGs on simulated disease spread. Our parsimonious epidemic ABM facilitates simulated movement and infection of agents based on user-defined parameters including patch size (spatial granularity) and time step duration (temporal granularity) enabling our ABM to adapt to different STGs. This capability is distinctly different from other epidemic models and enables simulation results from the same ABM operating under different STGs to be compared to elucidate the influence of STGs rather than comparing results of different epidemic models that may be based on different assumptions or data (Ajelli et al., 2010; Halloran et al., 2008).

We sought parsimony in our ABM to provide clear interpretation of results aimed toward improving our understanding of the influence of STGs on ABM processes. Significant progress has been made in understanding the simulation of geographic processes (Reitsma and Albrecht, 2005; Albrecht, 2005), and we find similarities to coarsening STGs within rich literature of representing spatial-temporal dynamics in data models (Yuan, 2001; Peuquet, 1994; Goodchild and Glennon, 2008). As spatial granularities in a data model are coarsened, finer grained spatial units are aggregated into coarser (i.e., larger) spatial units, which is related to a well-known problem in geography literature–the Modifiable Areal Unit Problem (MAUP) (Openshaw, 1984). MAUP may occur if small spatial units (e.g., points or polygons) are aggregated into larger spatial units, whose configuration is modifiable, and where different configurations lead to different results (Openshaw, 1984). Hornsby and Egenhofer (2002) examine how coarsening the granularity of time affects geospatial lifelines, which capture an individual's movement in a data model composed of time-stamped records of locations. However, MAUP does not explicitly address spatial processes (Goodchild, 2004) and Hornsby and Egenhofer (2002) do not explicitly look at the influence of coarsening the granularity of time on the simulation of processes. Goodchild (2004) states that ABMs may be used to better understand processes by comparing a simulated world with the real world. This study uses simulations with the finest STGs as a proxy for the real world and compares these simulations with those of spatially and temporally coarsened simulations to improve understanding of the influence of STGs on ABM processes in the context of epidemic modeling.

The spatial landscape of our ABM consists of a grid of square patches whose edge length is controlled by a spatial granularity variable (SG). Several epidemic ABMs use gridded representations of space (Parker and Epstein, 2011; Smieszek et al., 2011; Laskowski et al., 2011). Our ABM uses regularly sized and shaped patches to enable the systematic study of a range of spatial granularities from fine- to coarse-grained without complications related to aggregating irregularly shaped patches. The duration of discrete-time step in our ABM is controlled by a temporal granularity variable (TG). Our ABM adopts a discrete-time representation due to its widespread use in epidemic modeling (Chao et al., 2010; Atti et al., 2008; Germann et al., 2006; Ferguson et al., 2006; Longini et al., 2004).

Spatial and temporal extent (SE and TE, respectively) control the size of a spatial landscape and temporal duration of a simulation, and can be any multiple of SG and TG, respectively. The other primary variables in our ABM consist of the number of agents (N), movement speed (S), infection radius (R), infection probability (P), infectious period (I), and number of simulations to execute in spatiotemporal process models (M), which are listed in Table 3.1. Movement speed controls how fast agents move across a spatial landscape. Infection radius controls how close two agents must be to potentially transmit disease. Infection probability controls the chance of disease transmission when a susceptible is within an infectious agent's infection radius.

Our ABM adopts a commonly used Susceptible, Infectious, Recovered (SIR) epidemiological model to capture primary stages of a disease (Mishra et al., 2011), which has been adopted or adapted by several epidemic modeling studies (Xu and Sui, 2009; Hethcote, 2000; Roche et al., 2011). Most agents start each simulation being susceptible to disease. Upon disease transmission, agents become infectious and transmit disease to nearby susceptible agents for a fixed period. After which agents permanently recover from disease and are no longer susceptible.

3.3 Spatiotemporal Process Models

Processes in ABMs operate within the constraints of space and time representations, capturing spatiotemporal dynamics both implicitly and explicitly (Batty et al., 2012). As STGs are coarsened, ABM processes may become increasingly challenged to explicitly capture dynamics within a simulation, instead relying simplifying assumptions to implicitly capture dynamics. Simplifying

Parameter	Name	Examples
SE	Spatial extent	16×16, 1024×1024
TE	Temporal extent	16, 64, 1024
SG	Spatial granularity	1, 4, 16, 64, 256, 1024
TG	Temporal granularity	1, 2, 8, 32, 128, 512
N	Number of agents	128, 1024
S	Movement speed	1, 2
R	Infection radius	1, 2
P	Infection probability	20% ,50% ,70%
I	Infectious period	20, 64, 128
M	Spatiotemporal process model executions	10,000 or 1,000,000

Table 3.1: Primary model variables

assumptions are often embedded in processes designed to model continuously occurring dynamics to compensate for actions or behaviors that cannot be explicitly captured (Riley, 2007; Grimm et al., 2006; Sattenspiel and Dietz, 1995). Two common examples include homogeneous mixing of infected individuals in epidemic models (Ajelli et al., 2010), and introducing stochasticity (e.g., randomness) in ABMs (Bonabeau, 2002). Within the context of spatially explicit epidemic ABMs, there is limited understanding of how commonly used simplifying assumptions (e.g., homogeneous mixing amongst agents) in ABM processes are influenced by STGs. This study formulates a novel approach—spatiotemporal process model—that helps to examine the influence of STGs on ABM processes.

A spatiotemporal process model contextualizes an ABM process within space and time. Defining a spatiotemporal process model enables modelers to specify how every variable, equation, and algorithm used in a process is situated in space and time. They ask themselves, for example, how an infection probability variable should change if the duration of a discrete-time step (e.g., temporal granularity) is coarsened from two to four hours. Similarly, they ask how an infection probability variable should change if spatial granularity of a lattice is coarsened from ten kilometers to twenty or fifty. These questions reveal implicitly captured dynamics embedded in model assumptions. The answers help to make these dynamics explicit, represented in a spatiotemporal process model. The undertaking of addressing these often-difficult questions deepens our knowledge of ABM processes by bringing to light their underpinning assumptions of space and time.

A spatiotemporal process model generates as output a set of probability matrices, each rep-

resented as a two-dimensional array P. Each cell in P represents the probability of a process occurring in a spatiotemporal region whose size corresponds to a STG of a simulation (e.g., the size of a patch and duration of a time step). These regions surround the relative location of an individual agent. For example, a probability matrix may define probabilities for an agent to move to a nearby patch or remain in the same patch. While spatiotemporal process models are not limited to probabilistic processes or two-dimensions, we use them as examples in this article. Each spatiotemporal process model may simulate a spatially and temporally explicit process such as agent-agent interactions, agent-environment interactions, or agent movements thousands or millions of times, to derive statistically sound probabilities of process occurrence. Multiple probability matrices may be constructed for complex processes such as capturing differing probabilities of interaction amongst agents in different age categories (Germann et al., 2006).

Spatiotemporal process models help frame a modeler's understanding of how a process is situated within STGs of an ABM and, through this spatial and temporal contextualization, help to expose simplifying assumptions and examine their influence. ABMs may use probability matrices constructed by spatiotemporal process models to help control their processes. In this way, ABMs become capable of adapting to different STGs by executing spatiotemporal process models based on the STGs of an ABM simulation, which if different may alter the spatiotemporal regions and constructed probability matrices. For example, the probabilities of an infectious agent transmitting disease to a susceptible agent residing in the same patch will likely be higher if the patch size is small, because the two agents will tend to be closer to each other providing more opportunities to transmit disease. If the patch size is very large, the agents will tend to be farther away reducing the chance of disease transmission.

Spatiotemporal process models do not eliminate simplifying assumptions; rather they provide a mechanism for the influence of those assumptions to be examined as STGs are coarsened in ABMs. In much the same way as spatial analysts appreciate the impact of spatial scales due to issues such as the MAUP (Wong, 2009), agent-based modelers appreciate the impact of spatial and temporal scales on their models (Batty et al., 2012; Manson et al., 2012; Bian, 2004). However, as previous stated, due to limitations in fine-scale understanding, data, and computational capabilities ABMs often coarsen STGs (Hagen-Zanker and Jin, 2012; Ajelli et al., 2010; Riley, 2007) and as a

result often employ simplifying assumptions. Spatiotemporal process models are designed to help modelers examine the influence that coarsening STGs may have on model processes. We provide two concrete examples of designing spatiotemporal process models as part of a parsimonious epidemic ABM in the following sections.

3.3.1 Agent Movement

Our ABM uses spatiotemporal process models to address a challenging question in modeling random agent movement in patch-based spatial representation: when should an agent move from one patch onto another? Previous approaches, including island, meta-population, and stepping-stone models, control movement largely through user-defined parameters, commuting or travel data (Kareiva et al., 1990; Keeling et al., 2010; Balcan et al., 2010; Balcan and Vespignani, 2012). In the absence of such data, particularly at spatial scales of a neighborhood or building complex, new approaches are needed for systematically determining agent movement across patches.

Agents in our ABM move around a spatial landscape, changing their direction randomly each step, referred to as a random walk (Pearson, 1905). Random walk, due to its jarring random changes in direction, distorts pedestrian movement at the finest spatial and temporal scales (Batty, 2003). However, epidemic ABMs are often designed for coarser spatial and temporal scales, and random walk has been shown to capture human movement patterns at national scales (Brockmann et al., 2006) as well as on street networks (Jiang et al., 2009). We adopt random walk as a mechanism for agents in our ABM to move at varying STGs, similar to other epidemic models' adoption of random movement (Epstein et al., 2008; Keeling et al., 2010).

Our ABM uses a spatiotemporal process model, hereafter referred to as a movement model, to construct a probability matrix, hereafter referred to as a *movement_matrix*, which defines the probabilities for an agent to remain in the same patch or move to a nearby patch each time step. The spatially and temporally explicit movement model simulates a random walk process to represent fine-grained agent movements. Initially, an agent is randomly assigned to a point-based location within an area the same size of a patch (e.g., spatial granularity) in an ABM. The agent takes a number of steps in a random walk, proportional to the temporal granularity of a simulation. The random walk process is repeated numerous times, and each time the final location

of the agent is recorded. Then, based on the final locations of each random walk and the STG of an ABM simulation, the movement model calculates the probabilities that agents remain within a same patch or move to a nearby patch. Probabilities derived from these simulations are used to construct a $movement\_matrix$, which is used by our ABM to probabilistically move agents during a simulation.

3.3.2 Agent Infection

Agent infection in epidemic ABMs is a complex process involving multiple real-world dynamics (Atti et al., 2008; Germann et al., 2006; Ferguson et al., 2006). In temporally coarse ABMs, infection processes often implicitly capture agent movement through parameters, equations, and probabilities, because the duration of a time step (e.g., 6 or 8 hours) is too long to explicitly capture finer-grained movements (Germann et al., 2006; Ferguson et al., 2006). In spatially coarse ABMs, infection processes often implicitly capture disease transmission by assuming homogeneous mixing amongst agents within coarsened spatial landscapes (Mishra et al., 2011). As STGs are refined, ABMs explicitly capture agent movement, and agent infection begins to match our understanding of disease transmission (Laskowski et al., 2011).

Our ABM uses a spatiotemporal process model, hereafter referred to as an infection model, to construct a probability matrix, hereafter referred to as an $infection\_matrix$, that facilitates agent infection. The infection model is used to determine the probabilities of infection for susceptible agents in an ABM simulation based on whether they share the same patch or a nearby patch with one or more infectious agents. Similar to the movement model, the infection model randomly places agents within areas the same size as a patch, and uses a random walk process to represent fine-grained agent movements. Unlike the movement model, the infection model simulates two agents, in which an infectious agent attempts to infect a susceptible agent if it is within an infection radius, R, after each step in a random walk. The movement model simulates numerous pairs of agents to calculate the probabilities of infection if a susceptible agent is located within a same patch or nearby patch of an infectious agent in our patch-based ABM. Using a concrete example, we illustrate the design and application of our ABM, movement model, and infection model in the following section.

3.3.3 Illustration of Models

We illustrate a straightforward example of our ABM consisting of five agents walking around a spatial landscape of size 8×8. We compare two cases: a spatial landscape is represented as a 4×4 grid of patches in one case and a 2×2 grid of patches in the second case (i.e., $SG = 2$ or $SG = 4$, respectively). The agents in our illustration–labeled A, B, C, D, E–are each randomly assigned to a patch and randomly walks around the landscape, transmitting disease with 90% probability to susceptible agents within a distance of 0.8 while infectious. One randomly selected agent is infected at the beginning of a simulation. The duration of a time step (e.g., temporal granularity) is 2 and walking speed for all agents is 0.3. We detail a 1. movement model, 2. infection model, and 3. parsimonious epidemic ABM for the $SG = 2$ case and then repeat the steps for the $SG = 4$ case demonstrating our approach and providing two concrete scenarios to compare.

Before a simulation begins, our patch-based ABM uses a spatially and temporally explicit movement model that simulates a fine-grained random walk process (Figure 3.2a) numerous times to derive statistically sound probabilities of agent movement across patches based on the spatial granularity, temporal granularity, and movement speed. In the $SG = 2$ case, the movement model calculates that agents remain in the same patch 77% of the time, move to an adjacent patch approximately 5% of the time or to a cornering patch 0.3% of the time based on one thousand random walks (Figure 3.2b). The movement model constructs a probability matrix named $movement\_matrix$ based on these probabilities.

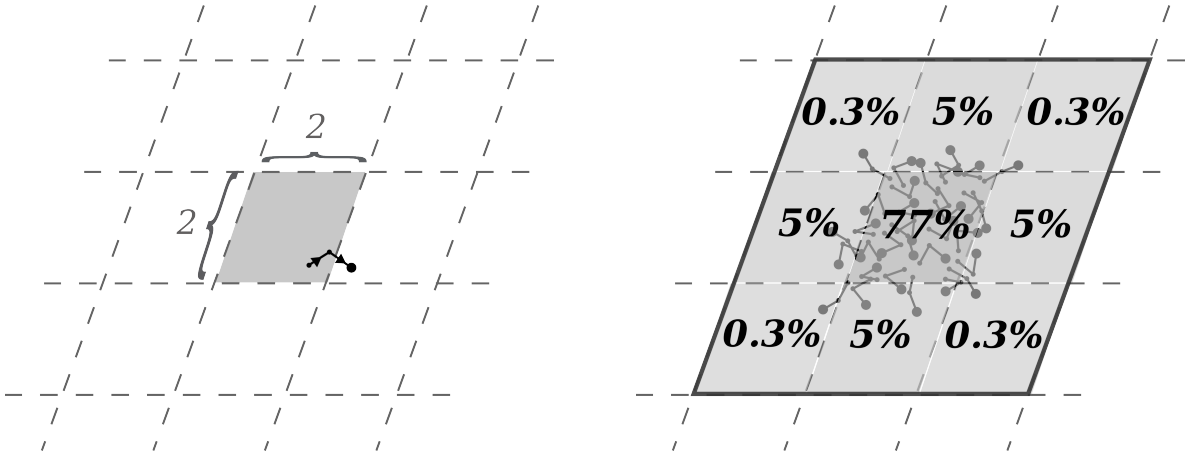
Similarly, our patch-based ABM uses a spatially and temporally explicit infection model that simulates two randomly walking agents–one infectious trying to infect the other that is susceptible (Figure 3.3a)–numerous times to derive statistically sound probabilities of infection for susceptible agents nearby infectious agents (Figure 3.3b). The infection model is repeatedly executed, each time initially locating a susceptible agent in a different nearby area corresponding to a patch in our ABM (Figure 3.3c). In the $SG = 2$ case, the infection model calculates that susceptible agents have a 51% probability to become infected if they share a patch with an infectious agent, approximately 6% if they are in an adjacent patch or 0.6% in a cornering patch to an infectious agent based on thousands of simulated infections. If a susceptible agent shares a patch with two infectious agents, for example, then the probabilities of it becoming infected are higher, because each infectious agent

has a 51% probability of infecting it. The infection model constructs a probability matrix named $infection\_matrix$ based on these probabilities.

Following the construction of probability matrices, $movement\_matrix$ and $infection\_matrix$, our parsimonious ABM begins executing a simulation. First, our ABM assigns each of the five agents to a randomly selected patch and infects a random agent (agent E in our example) (Figure 3.4a). Next, the ABM executes a series of iterations (TE/TG) that 1. use the $movement\_matrix$ to probabilistically move each agent (Figure 3.4b) and 2. use the $infection\_matrix$ to probabilistically infect susceptible agents near infectious agents (Figure 3.4c). Notice, in Figure 3.4c that agent E has 6% and 0.6% probability of infecting agents B and D, respectively, and that agents A and C cannot be infected, because they cannot be reached by agent E given walking speed, temporal granularity, infection radius, and their locations. In the next time step both agents E and B use the $infection\_matrix$ to try to infect nearby susceptible agents (e.g., agents A, C, and D). We repeated these steps, while coarsening the spatial granularity to $SG = 4$ and show an example of agents as being randomly assigned to one of the four patches in Figure 3.4d. The $movement\_matrix$ and $infection\_matrix$ for the $SG = 4$ case are shown in Figures 3.4e and 3.4f, respectively, for comparison. We draw the reader's attention to the exposure of all susceptible agents to infection from agent E due to coarsened spatial granularity, as well as different infection probabilities for B and D and movement probabilities for all agents. The remainder of this article elucidates how coarsening STGs influences model processes and disease spread dynamics.

3.4 Experimental Results

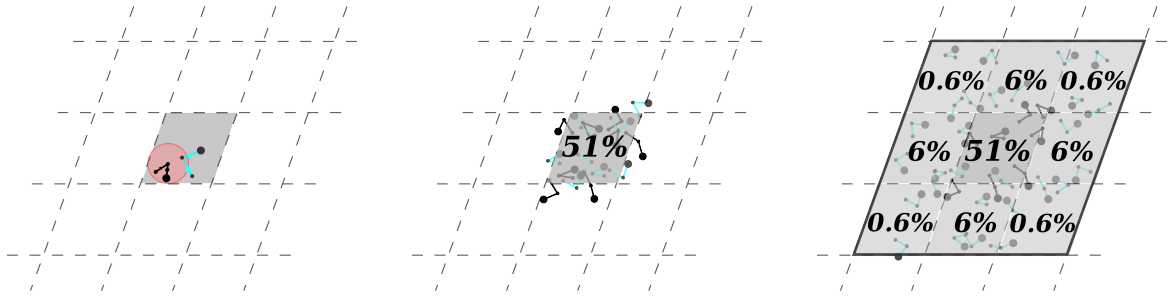
The goals of this study are to understand the influence of STGs on model processes and to build a foundation of spatial and temporal knowledge of epidemic simulations. Our goals are similar to those of network-based epidemic model studies that gain new insights into how disease spreads across networks of varying configurations, and whose results show that network configurations can significantly influence disease spread dynamics (Xu and Sui, 2009; Keeling et al., 2010; Keeling and Eames, 2005; Shirley and Rushton, 2005; Keeling, 1999). Our study assesses changes in speed, intensity, and spatial spread of a synthetic disease to investigate the influence of STGs on model processes in an ABM. Following the methodologies used in the network-based studies we sought



(a) Randomly select an explicit point-based location for an agent within a spatial area the size of a patch (i.e., 2×2). Randomly walk the agent, moving it distance $S = 0.3$ in a random direction $TG = 2$ times. The duration of a time step in our ABM determines the number of steps in a random walk.

(b) Repeat 3.2a $M = 1,000$ times, counting the number of agents that moved to nearby areas or remained within the initial spatial area to derive probabilities of agents moving to nearby patches or remaining within the same patch, which are used to construct a probability matrix named *movement_matrix* to be used by our ABM. Probabilities are truncated for brevity.

Figure 3.2: Illustration of a movement model that calculates probabilities of agents moving across patches based on the STGs of an ABM simulation using a random walk process. The parameter configurations in this illustration limit agent movements to adjacent and cornering patches, but certain parameter configurations may result in agents potentially walking to non-adjacent patches (e.g., two, three, or more patches away) and thus the probability matrix will increase in size from 3×3 to 5×5 or 7×7, for example.

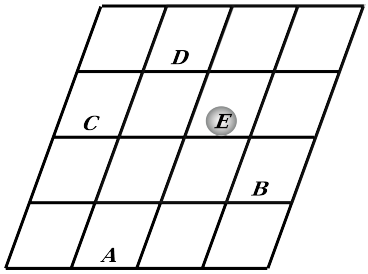


(a) Randomly select explicit point-based locations for one susceptible (blue) and one infectious agent (black) within a spatial area the size of a patch (i.e., 2×2). Randomly walk each agent, moving them distance $S = 0.3$ in a random direction $TG = 2$ times (i.e., the duration of a time step in our ABM). For each step, if the agents are within distance $R = 0.8$ (red area) then probabilistically transmit disease $P = 90\%$. Notice potential infection after first step (i.e., agents are within infectious radius R).

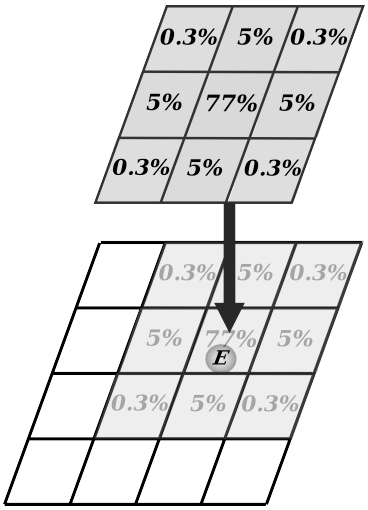
(b) Repeat 3.3a $M = 1,000$ times, counting the number of infections between the pairs to derive infection probabilities for susceptible agents occupying the same patch as an infectious agent.

(c) To calculate probabilities of infection for susceptible agents not sharing a patch with an infectious, initially locate the susceptible agent in a nearby area corresponding to an adjacent or cornering patch, for example, and repeat 3.3a and 3.3b. Execute the model for all nearby areas and calculate probabilities of infection for each area to construct a probability matrix named $infection\_matrix$ to be used by our ABM. Probabilities are truncated for brevity.

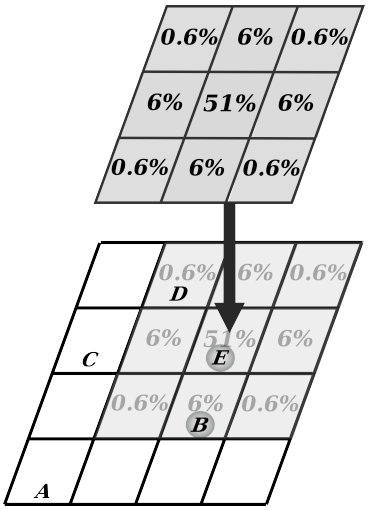
Figure 3.3: Illustration of an infection model that calculates probabilities of infection for susceptible agents occupying patches nearby infectious agents based on the STGs of an ABM simulation. Certain parameter configurations may result in calculating infection probabilities for larger numbers of patches (e.g., 5×5 or 7×7). Every time step during an ABM simulation the probabilities of infection are applied for each infectious agent to potentially infect any nearby susceptible agent.



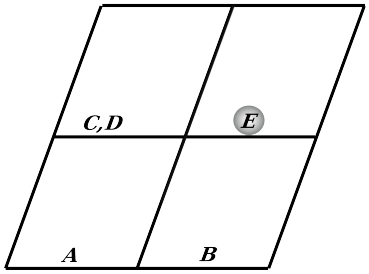
(a) Randomly assign each agent (A-E) to a patch. Randomly infect one agent (agent E in our example).



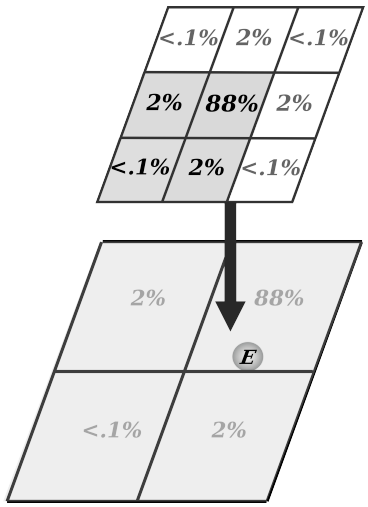
(b) Probabilistically move each agent based on the *movement_matrix* (Figure 3.2b). Movement probabilities to nearby patches for agent E are shown, but are also applied to all other agents each time step.



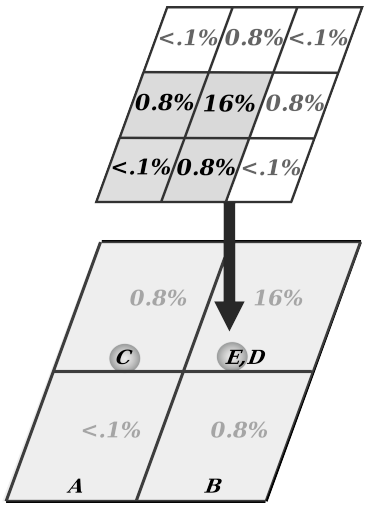
(c) For each infectious agent, probabilistically infect nearby susceptible agents based on $infection\_matrix$ (Figure 3.3c). Agent B becomes infected in this example.



(d) Randomly assign each agent (A-E) to a patch. Note that patch size has changed from 2×2 to 4×4 reducing the number of patches from 16 to 4.



(e) Probabilistically move each agent based on the *movement_matrix* constructed by simulating the movement model with a larger patch size (i.e., 4×4). If an agent attempts to walk beyond a landscape boundary, then retry random movement.



(f) For each infectious agent, probabilistically infect nearby susceptible agents based on $infection\_matrix$, which is constructed by simulating the infection model with a larger patch size (i.e., 4×4). Agent C becomes infected in this example.

Figure 3.4: Illustration of the steps in our patch-based ABM comparing $SG = 2$ (Figures 3.4a-3.4c) and $SG = 4$ (Figures 3.4d-3.4f). Movement and infection probabilities for agent E are shown in the steps for comparison.

parsimony in the design of a synthetic model, rather than focusing on matching population distributions, disease characteristics, and movement patterns (see for example Prieto et al. (2012) and Grassly and Fraser (2008)). Conceptually, a simulation of our synthetic model may represent disease spread at various spatial or temporal scales ranging from a national, state, county, city, or building scale spanning years, months, days, or hours based on the parameters used.

We use the following baseline parameters for our study and conduct sensitivity analysis altering these baseline parameters to understand their influence on the results of our study. We construct a scenario for conceptual exploration and use generic spatial and temporal units (SUs and TUs, respectively) rather than explicit spatial and temporal units, such as hectometers and hours. Our ABM simulates 1024 agents on a landscape with spatial extent 128×128 SUs and temporal extent of 2048 TUs. One randomly assigned agent is infected at the beginning of a simulation. Infectious agents infect susceptible agents within a radius of 0.7 SU with 70% probability of infection for 64 TUs (i.e., infectious period). Agents randomly walk at a speed of 0.7 SU per TU. The results are averaged over 10,000 simulations for each parameter configuration. Due to the large number of simulations per parameter configuration we re-use movement and infection matrices for 2,000 ABM simulations (M=1,000,000), which has little influence on simulation results. We examine a range of spatial granularities with edge lengths of size 2^s where s ranges from 0-5 and temporal granularities with time step durations of size 2^t where t ranges from 0-3. The finest-grained simulations with a spatial granularity=1 and temporal granularity=1 are used as a proxy for the real world to compare with spatially coarsened simulations.

3.4.1 Speed and Intensity of Disease

Attack rate curves plot new infections across the duration of an epidemic, capturing disease spread over time. They are often characterized by their peaks, termed peak attack rate (PAR), and when the peak occurred, termed time of peak attack rate (TPAR) (Ajelli et al., 2010; Xu and Sui, 2009; Ferguson et al., 2006). The PAR and TPAR help to quantify the intensity and speed of disease spread, describing the maximum number of individuals infected in a single period, at which point the disease has saturated the population, and how quickly the disease reaches saturation. The spatially and temporally finest-grained simulations—a proxy for the real world in

our study—result in a relatively low PAR infecting only 0.26% of the population at peak with a TPAR occurring at 240 TUs on average. In general, we find that coarsening spatial granularity results in higher PAR, infecting almost 8% of the population at peak, whereas coarsening temporal granularity results in lower PAR infecting around 4% of the population at peak (Figure 3.5). We also examine the percentage of agents that became infected during a simulation, termed cumulative attack rate (CAR), and find that coarsening spatial granularity in our simulations influences PAR, TPAR, and CAR more than coarsening temporal granularity overall (Figure 3.6). However, space and time remain interwoven based on our simulation results, and generally, TPAR in spatially fine-grained simulations occurs earlier if temporal granularity is more coarse, whereas TPAR in spatially coarse-grained simulations occurs earlier if temporal granularity is fine (Figure 3.6b). Interestingly, we found a non-linear relationship between PAR and TPAR when coarsening STGs (Figure 3.7) compared to a linear relationship found in reconfiguring links in small-world network-based epidemic simulations (Xu and Sui, 2009). Further studies may help to elucidate these differences, potentially adding value to recent work in studying contact networks that seek to bridge spatial proximity and networks (Salathé and Jones, 2010).

Understanding how coarsening STGs influence early disease spread (Ferguson et al., 2003) is important to understanding the chance that a disease becomes an epidemic or burns out (i.e., infecting a few individuals but not spreading to the larger population) early in a simulation (Rahmandad and Sterman, 2010). In general, we find that coarsening spatial granularity and fining temporal granularity reduce the chance of early disease burnout in simulations to less than 2% compared to more than 17% in spatially fine and temporally coarse simulations (Figure 3.8a). Meanwhile, coarsening spatial granularity speeds the initial spread of disease to 1% of the population from an average simulated time of 50 TUs compared to more than 92 TUs (Figure 3.8b). We modify the measure for early disease burnout in Rahmandad and Sterman (2010), from failing to infect 10% of the population or less to failing to infect 1% of the population or less, and remove the results from these simulations from further analyses. We use the time to infect 1% of the population as a measure of initial speed of disease spread similar as used in Danon et al. (2009). We found initial spread to 10% of the population to have similar trends as 1%. Our results are consistent with network-based modeling results (Keeling, 1999) in showing that increased spatial structure, in

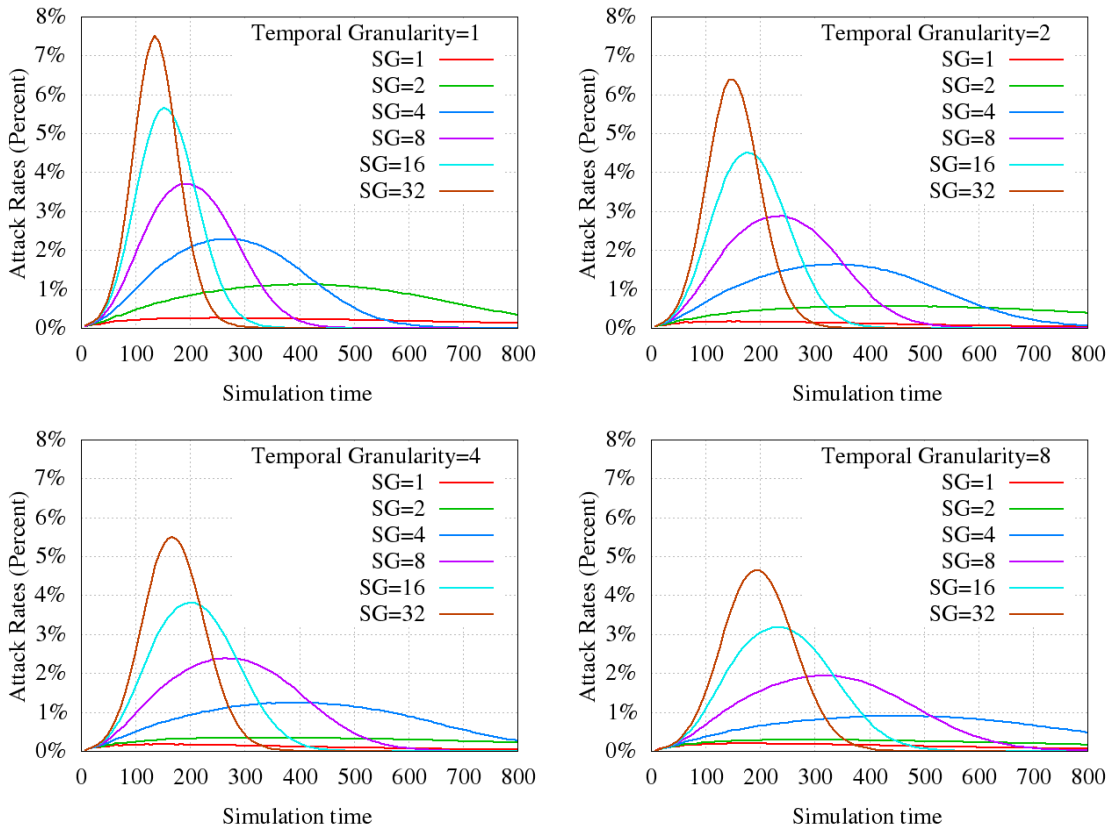
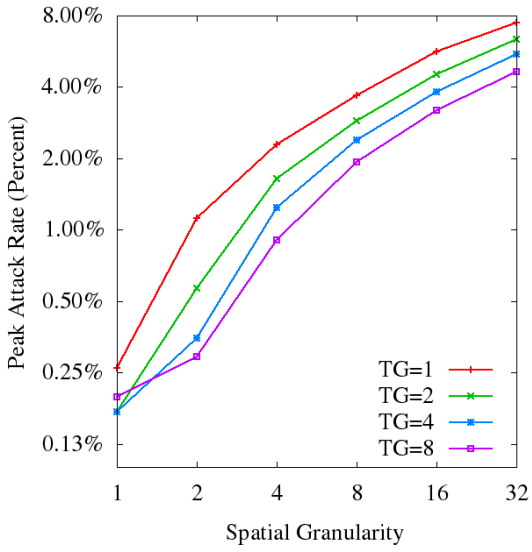


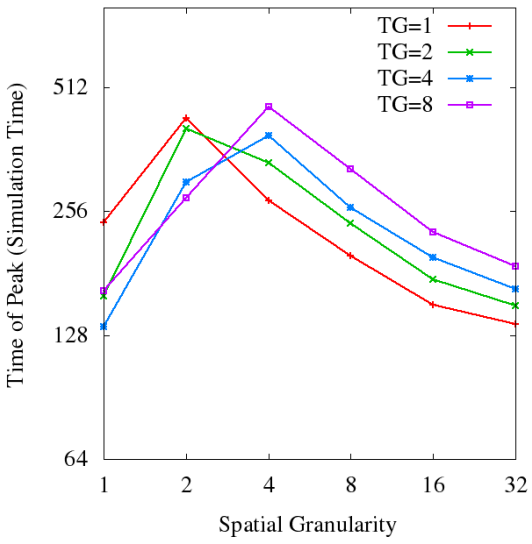
Figure 3.5: The four attack rate curves plot new infections per 8 units of time as a percentage of total population for different configurations of STG.

our case finer spatial granularities, and their case increased numbers of shared neighbors, play an important role in determining the chance of success or failure of an epidemic early in a simulation of disease spread.

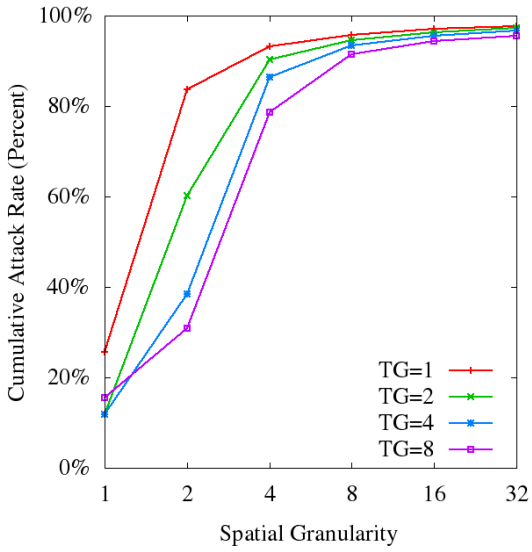
We conducted sensitivity analysis (Table 3.2) and show the relationship between PAR and TPAR for all parameter combinations (Figure 3.9). Sensitivity analysis shows that once PAR is below a certain threshold (e.g., approximately 1% of the population) the TPAR may occur earlier rather than later in a simulation, which is largely due to the peak being so small that the attack rate curve is almost flat as illustrated in the simulations with spatial granularity equal to one in Figure 3.5. When comparing maximum PAR against our baseline simulations (7.5%) we find that increasing walking speed (S) to 1.1 increases maximum PAR to 8.2%, infection radius (R) to 1.1 results in increases to 14.4%, and infection probability (P) to 90% increases maximum PAR to 9.3%. We reduced each of the three variables (S, R, and P) to 0.5, which resulted in maximum



(a) Peak Attack Rate



(b) Time of Peak Attack Rate



(c) Cumulative Attack Rate

Figure 3.6: Peak Attack Rate (PAR), Time to PAR (TPAR), and Cumulative Attack Rate (CAR) as spatial granularity is coarsened from 1 to 32 and for temporal granularity is coarsened from 1 to 8.

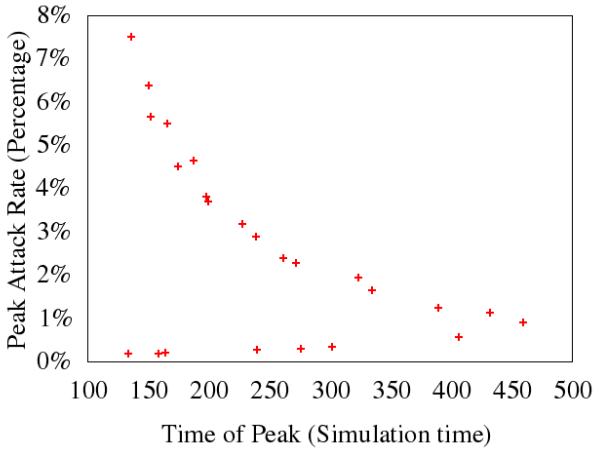
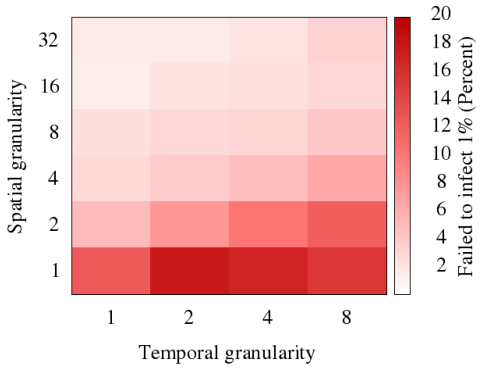
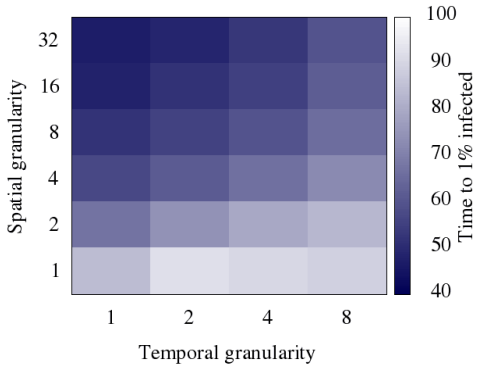


Figure 3.7: The relationship between average Peak Attack Rate (PAR) and Time to PAR (TPAR) for all configurations of STG tested. Each sample represents one STG configuration.



(a) Illustrates the percentage of simulations in which disease failed to infect at least 1% of a population. These simulations were removed from further analyses.



(b) Average simulation time to infect 1% of a population.

Figure 3.8: Spatially fine-grained and temporally coarse-grained simulations generally experience higher chances of disease burnout (infecting 1% or less of a population) and slower speeds of disease spread compared to spatially coarse-grained and temporally fine-grained simulations.

Parameter	Start value	End value	Increment
Infection probability (P)	0.5	0.9	0.2
Walking speed (S)	0.5	1.1	0.2
Infection radius (R)	0.5	1.1	0.2

Table 3.2: Sensitivity analysis parameters

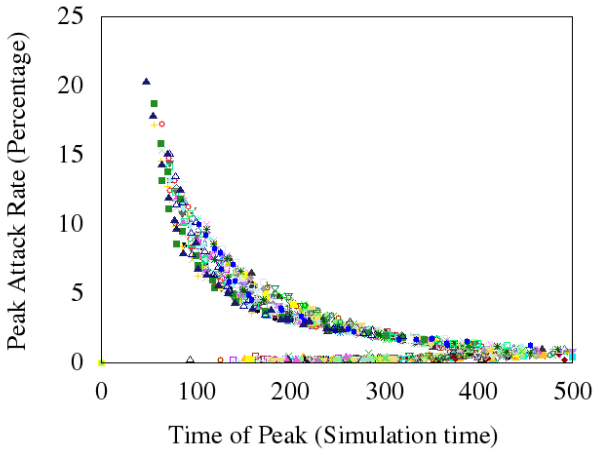


Figure 3.9: The relationship between Peak Attack Rate (PAR) and Time to PAR for results of sensitivity analysis. Each sample represents one parameter and STG configuration. Each parameter configuration shares the same symbol as STG is coarsened.

PAR of 6.3%, 2.9%, and 5.4%, respectively. We observed non-linear relationships in PAR and TPAR for all parameter combinations when varying STG (Figure 3.9).

While the influence of STG on the speed and intensity of disease is stark, it is not a complete picture. Attack rate curves are commonly used to capture disease spread in epidemic model simulations, but only provide a temporal view of disease spread. To fully understand disease dynamics and the underlying causes of peaks and valleys in these attack rate curves, we also investigated spatial characteristics of disease spread.

3.4.2 Spatial Spread of Disease

Disease spread unfolds in both time and space, and so we spatially contextualize our results using a swash-backwash model, which is designed to capture and analyze the leading and trailing edges of disease spread (Cliff and Haggett, 2006). A swash-backwash model divides a spatial area (e.g., simulated landscape) into subareas. For each subarea, a swash-backwash model records the time of the first infection occurrence (leading edge of disease spread) and last infection occurrence

(trailing edge). For each time period, the model subtracts the number of trailing edge subareas by leading edge subareas. Spatial spread of disease is visualized by plotting the difference in leading and trailing areas over time.

We apply a swash-backwash model to our simulation results, dividing the 128×128 simulated landscape into 8×8 sized subareas and plot the difference in leading and trailing areas (Figure 3.10). At the beginning of a simulation, new infections increase leading edge areas (positive values) as disease spatially spreads followed by agents recovering from disease increasing trailing edge areas (negative values). When spatial granularity of a simulation differed from 8×8 sized subareas, we considered all infections occurring in each subarea to calculate leading and trailing edges.

We use a swash-backwash model to understand why coarsening spatial granularity increases peak and cumulative attack rates (Figures 3.6a and 3.6c, respectively). The swash-backwash model results show that coarsening spatial granularity leads to a higher peak in leading edge area (Figure 3.10). Two factors contribute to higher peaks in leading edge area. Firstly, coarsening spatial granularity influences the movement process by obscuring fine-grained precision of an agent's location and as a result, agents are able to move longer distances for each time step. Our ABM acknowledges that agents may reside along an edge of a patch and, thus, each time step an agent may move to an adjacent patch. In 2 time steps, for example, agents may travel up to 2 adjacent patches whether spatial granularity equals to 1 or 32, however due to the increase in patch size the distance traveled is significantly different (i.e., 2 SUs compared to 64 SUs, respectively). Secondly, coarsening spatial granularity influences the infection process by exposing more susceptible agents to potential infection by infectious agents. This is evident in snapshots at the TPAR of several simulations that visualize the ratios of infected and recovered agents, which compare a range of spatially and temporally fine- and coarse-grained simulations (Figure 3.11). Returning to the illustration of our epidemic ABM in Section 3.3.3, agent C becomes infected in the SG=4 case (Figure 3.4f). Yet, in the SG=2 case agent C cannot be infected (Figure 3.4c), because the additional spatial information provided by fine spatial granularities establishes agent C as being too far away to be infected by agent E. The influence of coarsening spatial granularity on the two processes–movement and infection–results in faster spatial spread of disease as measured by higher peaks in leading-edge areas, which in turn exposes more susceptible agents to disease and increases

peak and cumulative attack rates.

We use a swash-backwash model to understand why coarsening temporal granularity generally increases the chance of disease burnout and slows down the speed of initial disease spread (Figures 3.8a and 3.8b, respectively). The swash-backwash model results show that coarsening temporal granularity for a given spatial granularity leads to slower growth in leading edge areas (Figure 3.10). Two factors contribute to this slower growth. Firstly, temporally fine simulations enable agents to potentially move further distances across a spatially coarsened landscape of patches. Acknowledging that it is possible for an agent to move to an adjacent patch every time step, an agent may move across 8 adjacent patches if TG=1 compared to a single patch in the same time if TG=8. The larger the patch size the greater distances an agent is able to travel in temporally fine-grained simulations compared to temporally coarse-grained simulations, which results in greater discrepancies in swash-backwash model results comparing peak leading edge areas for simulations with SG=1 to SG=32, for example (Figure 3.10). Secondly, coarsening temporal granularity introduces a temporal lag that affects subsequent infections as part of the infection process, which is analogous to compounding interest in a savings account. For example, a bank account with monthly compounding interest will earn less money over time compared to an account with daily compounding interest, because the earnings have fewer opportunities to compound on themselves each month compared to each day. Similarly, infectious agents have fewer opportunities to spread disease if temporal granularity is coarsened, and as a result, coarser temporal granularities in our experiments often exhibit slower initial disease spread as seen in cumulative attack curves that plot the accumulated number of infections across the duration of a simulation (Figure 3.12). The influence of coarsening temporal granularity on the two processes decreases the initial spatial spread as measured in leading edge area exposing fewer agents to disease, which slows the initial spread of disease (Figure 3.8b) and in turn increases the chance of disease burnout (Figure 3.8a). In summary, we find that coarsening STG plays a significant role in influencing model processes that alter disease spread in epidemic ABMs.

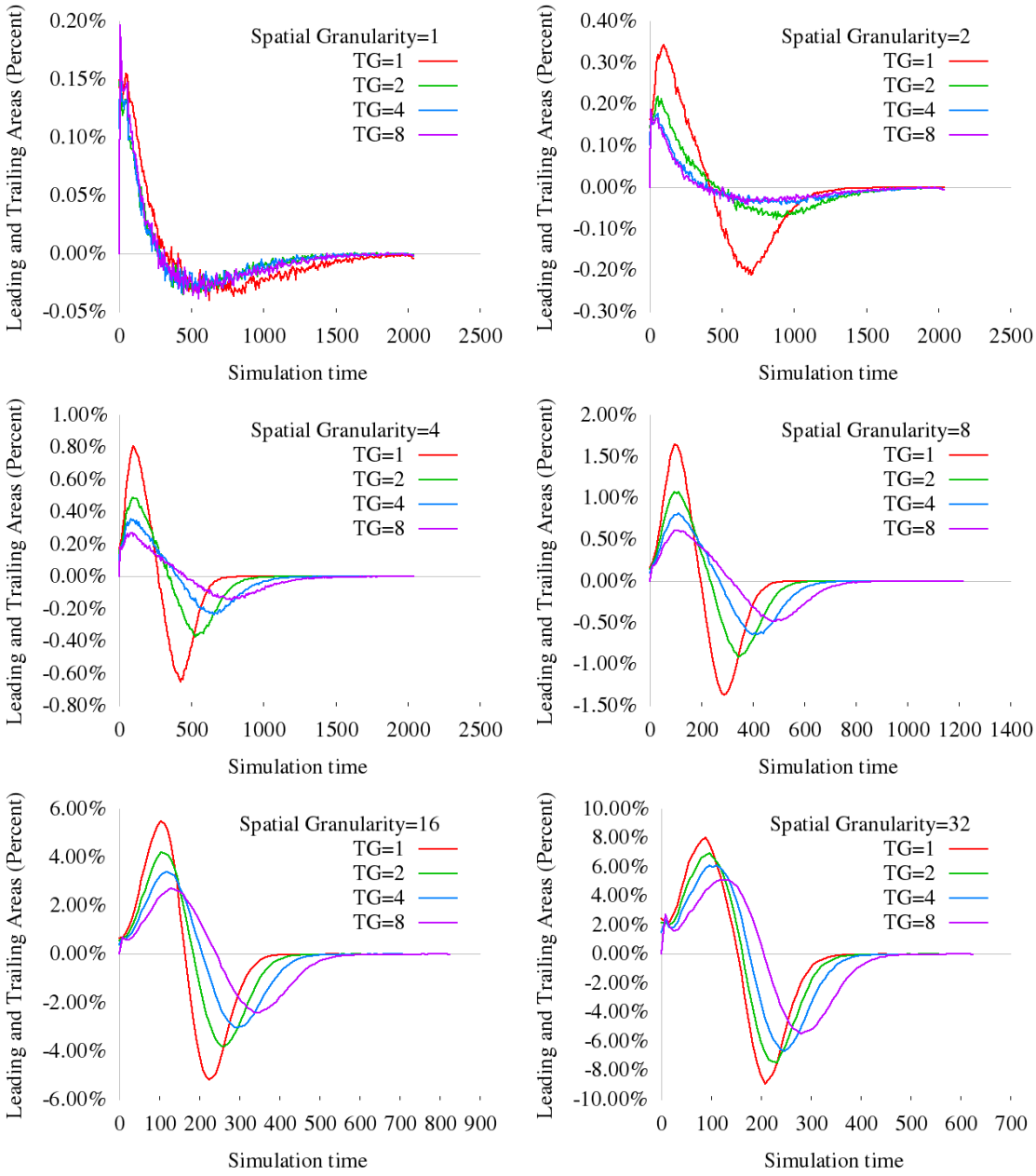
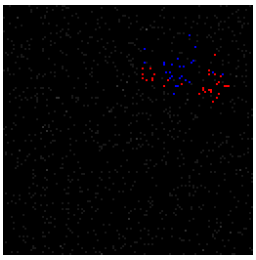
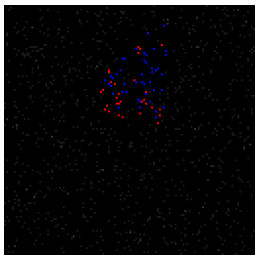


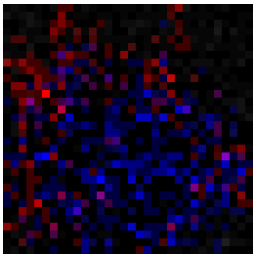
Figure 3.10: Swash-backwash model curves plotting increases in leading edge of disease spread as positive and increases in trailing edge of disease spread as negative for 8×8 sized subareas.



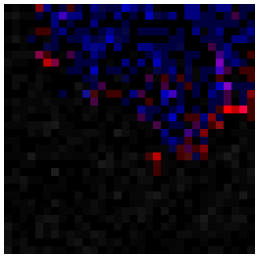
(a) Spatial granularity=1, Temporal granularity=1



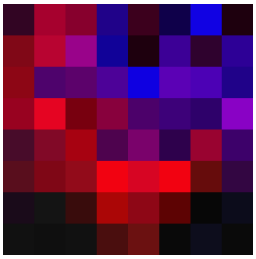
(b) Spatial granularity=1, Temporal granularity=8



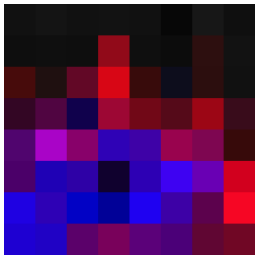
(c) Spatial granularity=4, Temporal granularity=1



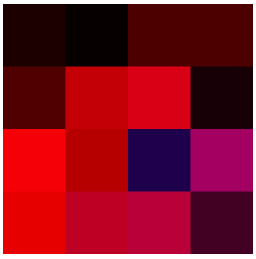
(d) Spatial granularity=4, Temporal granularity=8



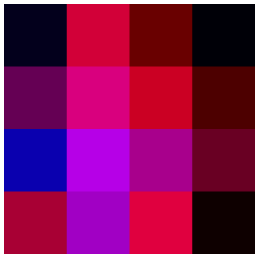
(e) Spatial granularity=16, Temporal granularity=1



(f) Spatial granularity=16, Temporal granularity=8



(g) Spatial granularity=32, Temporal granularity=1



(h) Spatial granularity=32, Temporal granularity=8

Figure 3.11: Maps of several sample simulations at the TPAR. The simulations are selected to represent various configurations of STGs for comparison. Each pixel in the map is drawn as a ratio of infected agents (red) to recovered agents (blue) with susceptible agents drawn as gray.

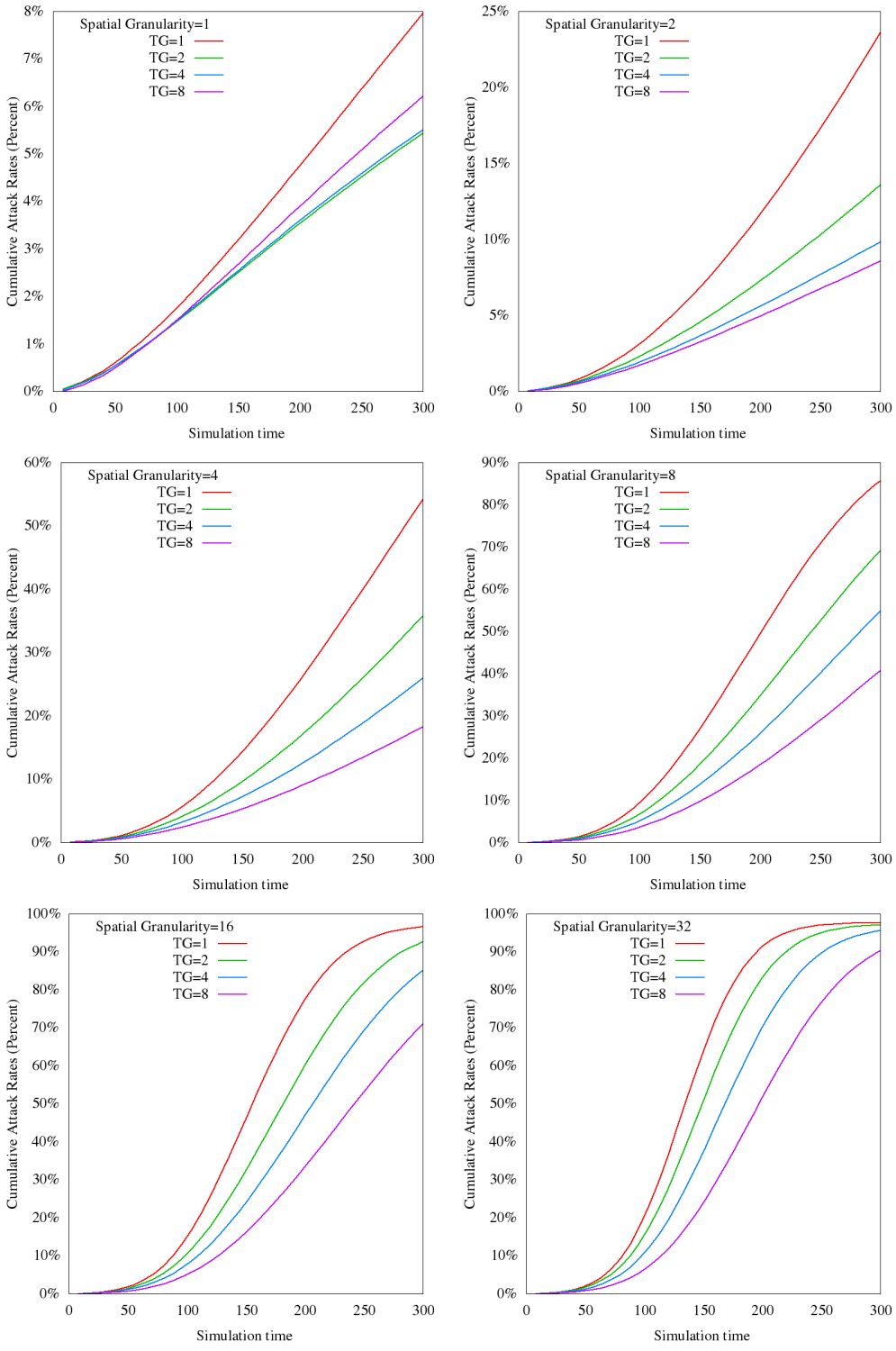


Figure 3.12: Cumulative attack rate curves plot accumulated infections as a percentage of total population. Simulation time 0-300 shown to illustrate initial increase in disease spread.

3.5 Concluding Discussion

Our study represents the first critical examination of how the representations of both space and time in a single epidemic ABM shape processes that operate within the constraints of such representations. We introduce spatiotemporal process models as an approach to contextualize ABM processes within space and time and to enable our ABM to adapt to different STGs. In the context of epidemic ABMs, our approach enables the assumption of homogenously mixed populations (Hethcote, 2000) to be relaxed by fining spatial granularities, similar to the way network-based epidemic models enable fully mixed populations to be relaxed by introducing network structure (Ferrari et al., 2011). This capability enables us to examine the results of simulations, ranging from coarse- to fine-grained, to understand the influence of not only spatial granularity, but also temporal granularity on model processes without the drawbacks of comparing models with differing assumptions (Ajelli et al., 2010). Our study used simple processes as examples, but spatiotemporal process models can be designed for more complex processes (e.g., spatial transmission kernels (Ajelli and Merler, 2008) or improved pedestrian movement (Kerridge et al., 2001)) or different spatial representations (e.g., polygons).

The results of our study show that coarsening STG influence model processes and subsequently disease spread in a parsimonious epidemic ABM. Specifically, we found that coarsening spatial granularities increase PAR by up to 7% and speeds TPAR by over 2×, which are consistent with related studies that observe increased attack rates with reduced spatial structure (Bobashev et al., 2007; Ajelli et al., 2010). Further, initial spread of disease is speeded up by almost 2× and chance of disease die-off (infecting less than 1% of a population) is reduced by more than 7×. We also found that coarsening spatial granularities in our ABM dramatically increases the speed of spatial spread of disease as measured by a swash-backwash model. On the other hand, we found that coarsening temporal granularities generally had the opposite effect resulting in decreased PAR, slowed down TPAR and initial spread of disease, and increased the chance of disease die-off. Interestingly, we found space and time to be interrelated as STGs are coarsened, and generally, TPAR in spatially fine-grained simulations occurs earlier if temporal granularity is coarse, whereas TPAR in spatially coarse-grained simulations occurs earlier if temporal granularity is fine.

Our study synergistically integrates both spatial and temporal perspectives to deepen under-

standing of the influence of STGs on epidemic ABM processes by spatially situating attack rate curves using the results of a swash-backwash model. This study is a first step in addressing a long-standing problem in epidemic modeling literature, namely how to systematically vary and study spatial structure (e.g., spatial granularities) in spatially explicit epidemic models while opening new opportunities to study temporal granularities, which until now have received little attention. As a first step, we sought parsimony in our ABM to gain systematic understanding of how changes in STGs impact disease spread similar to those works in network-based epidemic models (Rahmandad and Sterman, 2010; Xu and Sui, 2009; Keeling and Eames, 2005), but plan to build on this understanding in future studies. For example, our study found non-linear relationships between TPAR and PAR when varying STGs; whereas another study found linear relationships when altering network topologies (Xu and Sui, 2009). Understanding these differences may improve understanding in contact-networks based research (Emch et al., 2012; Cauchemez et al., 2011; Salathé and Jones, 2010). We also look to extend our ABM to a sub-national or national scale to investigate the influences of STGs when including multiscale mobility (Balcan et al., 2009), routine movements (e.g., commuting) (Danon et al., 2009; Keeling et al., 2010; Balcan and Vespignani, 2012), and heterogeneous population distributions.

Our future studies will also investigate the sensitivity of intervention strategies to STGs, elucidating any relationships between them (discussed with respect to simplifying assumptions in Prieto et al. (2012)). We find that the effects of fining STGs including decreasing and delaying peak attack rates share similar characteristics with those of intervention strategies. For example, Figure 3.5 in our article is similar to Figure 4c in Ferguson et al. (2006), which shows a decrease and delay in peak attack rate for different combinations of intervention strategies. Generally, intervention strategies whether medical or mobility (McLafferty, 2010) limit the potential distance of infection (e.g., travel restrictions) or the number of potential susceptibles (e.g., mass or targeted immunizations). Similarly, coarsening temporal granularity affects our movement process by reducing the number of patches infectious agents may move to, and fining spatial granularity affects our infection process by dividing larger spatial areas into many small areas limiting the number of potentially nearby susceptibles. It is therefore possible that simulations executed with certain STGs may over- or under-estimate the effectiveness of an intervention due to their influence on model processes (e.g.,

discussed with respect to homogeneous mixing in Ferguson et al. (2003)), which helps to explain why three spatially structured ABMs respond differently to various intervention strategies (Halloran et al., 2008). An improved understanding of the influence of STGs on simulated intervention effectiveness could provide new insights for scholars and policy makers when faced with a future epidemic.

References

Ajelli, M., Gonçalves, B., Balcan, D., Colizza, V., Hu, H., Ramasco, J., Merler, S., and Vespignani, A. (2010). Comparing large-scale computational approaches to epidemic modeling: Agent-based versus structured metapopulation models. *BMC Infectious Diseases*, 10(1):190.

Ajelli, M. and Merler, S. (2008). The impact of the unstructured contacts component in influenza pandemic modeling. *PLoS One*, 3(1):e1519.

Albrecht, J. (2005). A new age for geosimulation. *Transactions in GIS*, 9(4):451–454.

Atti, M., Merler, S., Rizzo, C., Ajelli, M., Massari, M., Manfredi, P., Furlanello, C., Tomba, G., and Iannelli, M. (2008). Mitigation measures for pandemic influenza in Italy: An individual based model considering different scenarios. *PLoS One*, 3(3):e1790.

Balcan, D., Colizza, V., Gonçalves, B., Hu, H., Ramasco, J., and Vespignani, A. (2009). Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences*, 106(51):21484–21489.

Balcan, D., Gonccalves, B., Hu, H., Ramasco, J., Colizza, V., and Vespignani, A. (2010). Modeling the spatial spread of infectious diseases: The GLobal Epidemic and Mobility computational model. *Journal of Computational Science*, 1(3):132–145.

Balcan, D. and Vespignani, A. (2012). Invasion threshold in structured populations with recurrent mobility patterns. *Journal of Theoretical Biology*, 293(0):87–100.

Batty, M. (2003). Agent-based pedestrian modelling. In Longley, P. and Batty, M., editors, *Advanced Spatial Analysis: The CASA Book of GIS*, pages 81–106. ESRI Press, Redlands, CA, USA.

Batty, M., Crooks, A., See, L., and Heppenstall, A. (2012). Perspectives on agent-based models and geographical systems. In Heppenstall, A., Crooks, A., See, L., and Batty, M., editors, *Agent-Based Models of Geographical Systems*, pages 1–15. Springer Netherlands.

Beyer, K., Tiwari, C., and Rushton, G. (2012). Five essential properties of disease maps. *Annals of the Association of American Geographers*, 102(5):1067–1075.

Bian, L. (2004). A conceptual framework for an individual-based spatially explicit epidemiological model. *Environment and Planning B*, 31(3):381–396.

Bian, L. (2013). Spatial approaches to modeling dispersion of communicable diseases: A review. *Transactions in GIS*, 17(1):1–17.

Bobashev, G., Goedecke, D., Yu, F., and Epstein, J. (2007). A hybrid epidemic model: Combining the advantages of agent-based and equation-based approaches. In *Proceedings of the 39th conference on Winter simulation (WSC)*, pages 1532–1537, Piscataway, NJ, USA. IEEE Press.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences of the United States of America*, 99(Suppl 3):7280–7287.

Brockmann, D. (2009). Human mobility and spatial disease dynamics. In Schuster, H., editor, *Reviews of Nonlinear Dynamics and Complexity*, pages 1–24. Wiley-VCH.

Brockmann, D., Hufnagel, L., and Geisel, T. (2006). The scaling laws of human travel. *Nature*, 439(7075):462–465.

Cauchemez, S., Bhattarai, A., Marchbanks, T., Fagan, R., Ostroff, S., Ferguson, N., Swerdlow, D., Sodha, S., Moll, M., Angulo, F., et al. (2011). Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza. *Proceedings of the National Academy of Sciences*, 108(7):2825–2830.

Chao, D., Halloran, M., Obenchain, V., and Longini, I. (2010). FluTE, a publicly available stochastic influenza epidemic simulation model. *PLoS Computational Biology*, 6(1):e1000656.

Cliff, A. and Haggett, P. (2006). A swash-backwash model of the single epidemic wave. *Journal of Geographical Systems*, 8(3):227–252.

Cliff, A., Haggett, P., and Smallman-Raynor, M. (2008). An exploratory method for estimating the changing speed of epidemic waves from historical data. *International Journal of Epidemiology*, 37(1):106–112.

Crooks, A. and Castle, C. (2012). The integration of agent-based modelling and geographical information for geospatial simulation. In Heppenstall, A., Crooks, A., See, L., and Batty, M., editors, *Agent-Based Models of Geographical Systems*, pages 219–251. Springer Netherlands.

Danon, L., House, T., and Keeling, M. (2009). The role of routine versus random movements on the spread of disease in Great Britain. *Epidemics*, 1(4):250–258.

Emch, M., Root, E., Giebultowicz, S., Ali, M., Perez-Heydrich, C., and Yunus, M. (2012). Integration of spatial and social network analysis in disease transmission studies. *Annals of the Association of American Geographers*, 102(5):1004–1015.

Epstein, J. (2009). Modelling to contain pandemics. *Nature*, 460(7256):687–687.

Epstein, J., Parker, J., Cummings, D., and Hammond, R. (2008). Coupled contagion dynamics of fear and disease: mathematical and computational explorations. *PLoS One*, 3(12):e3955.

Ferguson, N., Cummings, D., Fraser, C., Cajka, J., Cooley, P., and Burke, D. (2006). Strategies for mitigating an influenza pandemic. *Nature*, 442(7101):448–452.

Ferguson, N., Keeling, M., Edmunds, W., Gani, R., Grenfell, B., Anderson, R., and Leach, S. (2003). Planning for smallpox outbreaks. *Nature*, 425(6959):681–685.

Ferrari, M., Perkins, S., Pomeroy, L., and Bjørnstad, O. (2011). Pathogens, social networks, and the paradox of transmission scaling. *Interdisciplinary Perspectives on Infectious Diseases*, 2011(267049):1–10.

Germann, T., Kadau, K., Longini, I., and Macken, C. (2006). Mitigation strategies for pandemic influenza in the United States. *Proceedings of the National Academy of Sciences of the United States of America*, 103(15):5935–5940.

Goodchild, M. (2004). GIScience, geography, form, and process. *Annals of the Association of American Geographers*, 94(4):709–714.

Goodchild, M. and Glennon, A. (2008). Representation and computation of geographic dynamics. In Hornsby, K. and Yuan, M., editors, *Understanding dynamics of geographic domains*, pages 13–30. CRC Press.

Grassly, N. and Fraser, C. (2008). Mathematical models of infectious disease transmission. *Nature Reviews Microbiology*, 6(6):477–487.

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., Goss-Custard, J., Grand, T., Heinz, S., Huse, G., et al. (2006). A standard protocol for describing individual-based and agent-based models. *Ecological Modelling*, 198(1-2):115–126.

Hagen-Zanker, A. and Jin, Y. (2012). A new method of adaptive zoning for spatial interaction models. *Geographical Analysis*, 44(4):281–301.

Halloran, M., Ferguson, N., Eubank, S., Longini, I., Cummings, D., Lewis, B., Xu, S., Fraser, C., Vullikanti, A., Germann, T., et al. (2008). Modeling targeted layered containment of an influenza pandemic in the United States. *Proceedings of the National Academy of Sciences*, 105(12):4639–4644.

Hethcote, H. (2000). The mathematics of infectious diseases. *SIAM Review*, 42(4):599–653.

Hornsby, K. and Egenhofer, M. J. (2002). Modeling moving objects over multiple granularities. *Annals of Mathematics and Artificial Intelligence*, 36(1-2):177–194.

Jiang, B., Yin, J., and Zhao, S. (2009). Characterizing the human mobility pattern in a large street network. *Physical Review E*, 80(2):021136.

Kaplan, E. and Wein, L. (2003). Smallpox bioterror response. *Science*, 300(5625):1503–1504.

Kareiva, P., Mullen, A., and Southwood, R. (1990). Population dynamics in spatially complex environments: theory and data [and discussion]. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 330(1257):175–190.

Keeling, M. (1999). The effects of local spatial structure on epidemiological invasions. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 266(1421):859–867.

Keeling, M., Danon, L., Vernon, M., and House, T. (2010). Individual identity and movement networks for disease metapopulations. *Proceedings of the National Academy of Sciences*, 107(19):8866–8870.

Keeling, M. and Eames, K. (2005). Networks and epidemic models. *Journal of the Royal Society Interface*, 2(4):295–307.

Kerridge, J., Hine, J., and Wigan, M. (2001). Agent-based modelling of pedestrian movements: the questions that need to be asked and answered. *Environment and Planning B*, 28(3):327–342.

Koopman, J. (2004). Modeling infection transmission. *Annual Review of Public Health*, 25:303–326.

Kulldorff, M. and Nagarwalla, N. (1995). Spatial disease clusters: detection and inference. *Statistics in Medicine*, 14(8):799–810.

Laskowski, M., Demianyk, B., Witt, J., Mukhi, S., Friesen, M., and McLeod, R. (2011). Agent-based modeling of the spread of influenza-like illness in an emergency department: A simulation study. *IEEE Transactions on Information Technology in Biomedicine: A publication of the IEEE Engineering in Medicine and Biology Society*, 15(6):877–889.

Liu, X. and Andersson, C. (2004). Assessing the impact of temporal dynamics on land-use change modeling. *Computers, Environment and Urban Systems*, 28(1-2):107–124.

Longini, I., Halloran, M., Nizam, A., and Yang, Y. (2004). Containing pandemic influenza with antiviral agents. *American Journal of Epidemiology*, 159(7):623–633.

Maantay, J. and McLafferty, S. (2011). Environmental health and geospatial analysis: An overview. In *Geospatial Analysis of Environmental Health*, pages 3–37. Springer.

Manson, S., Sun, S., and Bonsal, D. (2012). Agent-based modeling and complexity. In Heppenstall, A., Crooks, A., See, L., and Batty, M., editors, *Agent-Based Models of Geographical Systems*, pages 125–139. Springer Netherlands.

McLafferty, S. (2010). Placing pandemics: geographical dimensions of vulnerability and spread. *Eurasian Geography and Economics*, 51(2):143–161.

Mishra, S., Fisman, D., and Boily, M. (2011). The ABC of terms used in mathematical models of infectious diseases. *Journal of Epidemiology and Community Health*, 65(1):87–94.

Openshaw, S. (1983). *The modifiable areal unit problem*. Geo Books Norwich, UK.

Parker, J. and Epstein, J. (2011). A distributed platform for global-scale agent-based models of disease transmission. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 22(1):2:1–2:25.

Pearson, K. (1905). The problem of the random walk. *Nature*, 72(1865):294.

Peuquet, D. (1994). It's about time: A conceptual framework for the representation of temporal dynamics in geographic information systems. *Annals of the Association of American Geographers*, 84(3):441–461.

Prieto, D., Das, T., Savachkin, A., Uribe, A., Izurieta, R., and Malavade, S. (2012). A systematic review to identify areas of enhancements of pandemic simulation models for operational use at provincial and local levels. *BMC Public Health*, 12(1):251.

Rahmandad, H. and Sterman, J. (2010). Heterogeneity and network structure in the dynamics of diffusion: Comparing agent-based and differential equation models. *Management Science*, 54(5):998–1014.

Reitsma, F. and Albrecht, J. (2005). Implementing a new data model for simulating processes. *International Journal of Geographical Information Science*, 19(10):1073–1090.

Riley, S. (2007). Large-scale spatial-transmission models of infectious disease. *Science*, 316(5829):1298–1301.

Roche, B., Drake, J., and Rohani, P. (2011). An agent-based model to study the epidemiological and evolutionary dynamics of influenza viruses. *BMC Bioinformatics*, 12(1):87.

Salathé, M. and Jones, J. (2010). Dynamics and control of diseases in networks with community structure. *PLoS Computational Biology*, 6(4):e1000736.

Sattenspiel, L. and Dietz, K. (1995). A structured epidemic model incorporating geographic mobility among regions. *Mathematical Biosciences*, 128(1-2):71–91.

Shirley, M. and Rushton, S. (2005). The impacts of network topology on disease spread. *Ecological Complexity*, 2(3):287–299.

Smieszek, T. (2009). A mechanistic model of infection: why duration and intensity of contacts should be included in models of disease spread. *Theoretical Biology and Medical Modelling*, 6(1):25.

Smieszek, T., Balmer, M., Hattendorf, J., Axhausen, K., Zinsstag, J., and Scholz, R. (2011). Reconstructing the 2003/2004 H3N2 influenza epidemic in Switzerland with a spatially explicit, individual-based model. *BMC Infectious Diseases*, 11(1):115.

Snow, J. (1855). *On the mode of communication of cholera.* John Churchill, 2nd ed. London (Reprinted 1936, New York).

Wainer, G. and Mosterman, P. (2011). *Discrete-event modeling and simulation: theory and applications.* CRC Press, Boca Raton, FL, USA.

Wheaton, W., Cajka, J., Chasteen, B., Wagener, D., Cooley, P., Ganapathi, L., Roberts, D., and Allpress, J. (2009). Synthesized population databases: A US geospatial database for agent-based models. RTI Press paper available at http://www.rti.org/pubs/mr-0010-0905-wheaton.pdf (Accessed: March, 2013).

Wong, D. (2009). The modifiable areal unit problem (MAUP). In Fotheringham, A. and Rogerson, P., editors, *The Sage Handbook of Spatial Analysis*, pages 105–124. SAGE Publications Limited.

Xu, Z. and Sui, D. (2009). Effect of small-world networks on epidemic propagation and intervention. *Geographical Analysis*, 41(3):263–282.

Yuan, M. (2001). Representing complex geographic phenomena in GIS. *Cartography and Geographic Information Science*, 28(2):83–96.

Zeigler, B., Praehofer, H., and Kim, T. (2000). *Theory of modeling and simulation: Integrating discrete event and continuous complex dynamic systems.* Academic Press.

CHAPTER 4

THE INFLUENCE OF SPATIAL AND TEMPORAL GRANULARITIES IN AGENT-BASED MODELING OF DISEASE SPREAD IN OHIO, USA

4.1 Introduction

Agent-based Models (ABMs) simulate dynamic phenomena at individual levels (Epstein, 2009) to study complex characteristics of coupled human-environmental systems (Batty et al., 2012; An et al., 2005; Parker et al., 2003) such as the spread of influenza-like illness (ILI) (Lee et al., 2010). However, a tension exists between the increasing ability of ABMs to represent complex spatial interactions across different geographic settings and the value that this additional complexity adds to the predictive capabilities for understanding disease spread (Cromley and McLafferty, 2011; Hupert et al., 2008; Ajelli et al., 2010). At the heart of this tension for spatially explicit epidemic models are the underlying assumptions of spatial disease transmission methods, which rely on differing representations of space (Chapter 3). This chapter investigates how coarsening the finest representations of space and time, termed spatial and temporal granularities (STGs), influence the spread of a synthetic ILI across the state of Ohio using a novel epidemic ABM and to add new insights toward the ongoing epidemic modeling discourse.

Spatial and temporal scales represent critical factors that often influence the results of spatial models and analysis methods (Manson et al., 2012; Openshaw, 1984). If spatial scales are coarsened to reduce computational cost (Stanilov, 2012) or in response to insufficiently detailed data (Ajelli et al., 2010), ABM processes change and models may become more speculative (Batty et al., 2012). If spatial scales are coarsened, finer grained spatial units are aggregated into coarser (i.e. larger) spatial units, which is related to a well-known problem in geography literature—the Modifiable Areal Unit Problem (MAUP). The MAUP may occur when small spatial units (e.g. points or polygons) are aggregated into larger spatial units, whose configuration is modifiable and where different configurations lead to different results (Openshaw, 1984). If temporal scales are coarsened to reduce computational cost, for example, model dynamics can be affected (Roche et al., 2011). However, understanding the influence of these critical factors is often difficult. Previous studies have

compared the results of differing disease models including differential equation, metapopulation, and agent-based ones (Rahmandad and Sterman, 2010; Ajelli et al., 2010; Halloran et al., 2008) to improve understanding of the influence of various factors including the introduction of heterogeneity and stochasticity, but differences in modeling assumptions and designs may result in opposing interpretations (Kaplan and Wein, 2003). This study uses a single epidemic ABM capable of adapting to different STGs to eliminate such differences.

While coarsening granularities weakens the agent-based modeling paradigm (Batty et al., 2012), ABMs are often motivated to coarsen STGs, which in turn decrease computational intensity (Roche et al., 2011) thus enabling computational tractability. Finer STGs and resultant fine-grained model processes may pose significant computational challenges, oftentimes resulting in longer execution times and significant increases in the amount of data produced and consumed (Ajelli et al., 2010; Torrens, 2012). To overcome these computational challenges, large-scale ABMs often adopt parallel and high-performance computing (HPC) approaches (Parry and Bithell, 2012; Parker and Epstein, 2011). In the case of large-scale epidemic ABMs that simulate ILIs, these approaches enable parallel ABMs to simulate millions or billions of humans (Epstein, 2009) commuting to and from work (Germann et al., 2006) often requiring multiple gigabytes of memory to store agent and model information (Parker and Epstein, 2011). Such large-scale epidemic models can inform public policy decisions (Lee et al., 2010) by helping policy makers and scientists better understand how disease spreads through populations by simulating past (Smieszek et al., 2011), present (Lee et al., 2010), and potential (Chao et al., 2010) diseases. However, large-scale epidemic models often operate at different STGs (Table 4.1), which has been shown to influence disease spread in a parsimonious epidemic ABM (Chapter 3). This study designs a novel parallel ABM that is capable of simulating more than ten million individuals and investigates the influence of STGs on disease spread dynamics in the state of Ohio.

4.2 Background

Epidemiological models capture primary stages of disease progression, which are often represented as a combination of four components: Susceptible, Exposed (Latent), Infectious (Contagious), and Recovered (SEIR) (Mishra et al., 2011). Many diseases including ILI have an incubation

Source	Model type	Primary transmission method	Spatial granularity	Temporal granularity	Study area	Processor core count
Bisset et al. (2012)	Agent-based	Network	Building	5 minutes	Delaware	8 Cores, 2 GPUs
Brown et al. (2011)	Agent-based	Group	Building, workgroups	24 hours	Pennsylvania	20 Cores
Chao et al. (2011)	Agent-based	Group	Household	12 hours	Los Angeles County and United States	Unspecified
Parker and Epstein (2011)	Agent-based	Patch, contact matrix	ModelBlock (grid)	∼15 minutes	Global and United States	32 Cores
Smieszek et al. (2011)	Agent-based	Patch	500×500m grid cells	24 hours	Switzerland	Unspecified
Balcan et al. (2010)	Metapopulation	Patch	Sub-City	Commuting 1/3 day	Global	Unspecified
Chao et al. (2010)	Agent-based	Group	Household	12 hours	United States	32 Cores
Bisset et al. (2009)	Agent-based	Network	Building	24 hours	California and Alabama	300 Cores
Atti et al. (2008)	Agent-based	Group	Building	6 hours	Italy	Unspecified
Barrett et al. (2008)	Agent-based	Network	Building	Minutes	Several US states	376 Cores
Stroud et al. (2007)	Agent-based	Group	Sub-building	Minutes	6 California counties	Hundreds of Cores
Ferguson et al. (2006)	Agent-based	Group	Building, workgroups	6 hours	United States and Great Britain	16 Cores
Germann et al. (2006)	Agent-based	Group	Building, workgroups	12 hours	United States	256+ Cores

Table 4.1: This table characterizes state-of-the-art large-scale spatial epidemic models simulating ILIs.

period that is represented as the exposed or latent time in which infected persons do not transmit disease immediately after infection. After the exposed period, a person becomes infectious transmitting disease to nearby susceptible persons. Finally, an infectious person becomes permanently recovered and is no longer susceptible to disease. Common variants of these stages include SIS, SIR, SIRS, SEIS, SEIR, and SEIRS (Liu et al., 1987; Hethcote, 2000). Other more complex stages may be included, for example, Contagious-Asymptomatic and Contagious-Symptomatic (Parker and Epstein, 2011). The duration of exposed and infectious periods may be selected to simulate past (Smieszek et al., 2011) or potential diseases (Epstein, 2009).

Commonly used abstractions for spatial disease transmission include group, network, patch, and distance (Riley, 2007). Diseases in group- and network-based models are transmitted amongst groups of individuals or individuals only if they share a connection. Diseases in patch- or distance-based models are often transmitted based on distance between patches or individuals. Transmission rates for individuals residing within a group or patch are often identical. Each of the four abstractions differs in their assumptions of representations of space ranging from coarse- (group and patch) to very fine-grained (network and distance), and implicitly (group and network) or explicitly (patch and distance) represented (Chapter 3). While many large-scale epidemic models simulating ILI adopt group-based transmission methods, network and patch have also been adopted in a number of models (Table 4.1).

Many large-scale spatial epidemic models operate at building-level spatial granularities, but range from sub-building (e.g. workgroups) to sub-city (Table 4.1). In addition to the finest-grained representations of space, several epidemic models enable disease to be transmitted at multiple spatial scales (e.g. neighborhoods, community) (Germann et al., 2006; Ferguson et al., 2006), which has been shown to influence disease spread at the scales of random community contacts (Ajelli and Merler, 2008). Models operate at temporal granularities ranging from minutes to a day (Table 4.1), which may influence the degree to which models may support human mobility (Brockmann, 2009; Balcan et al., 2009), or determine agent-to-agent contacts (Rahmandad and Sterman, 2010), or their durations (Smieszek, 2009) which are recognized as altering disease spread dynamics in epidemic modeling.

Previous studies comparing epidemic model results are insufficient to understand if coarsen-

ing STGs influence disease spread dynamics in a large-scale realistic scenario. Ajelli and Merler (2008) showed that different methods used to capture random contacts in the community, termed unstructured contacts, using the same epidemic ABM influence disease spread dynamics altering spatial diffusion and number of infected individuals. Ajelli et al. (2010) compares an agent-based and metapopulation model side-by-side. Metapopulation models are patch-based epidemic models where each patch is represented as a dynamic model that is linked via flows of people (Balcan et al., 2010). Ajelli et al. (2010) acknowledges a discrepancy between the two models with higher attack rates in the metapopulation model attributed to less detailed (e.g. spatial) information compared to the ABM. Halloran et al. (2008) compares the effectiveness of a number of intervention strategies simulated in three different epidemic ABMs simulating residents in Chicago, IL, however due to varying assumptions, methods, temporal granularities, and spatial granularities in the models it is difficult to understand why one strategy is more effective in one model compared to another or if coarsened STGs influence any of the three models. A number of studies have investigated how alterations in synthetic network configurations change disease spread dynamics (Xu and Sui, 2009; Keeling et al., 2010; Keeling and Eames, 2005; Shirley and Rushton, 2005; Keeling, 1999), but further work is needed to relate these results to realistic scenarios. Lastly, Chapter 3 investigated the influence of STG on model processes in a parsimonious epidemic ABM, but was limited to homogeneous population distributions and random movements. This study uses an epidemic ABM to investigate the influence of STGs on disease spread dynamics to systematically examine whether STGs affect large-scale epidemic models.

4.3 Agent-based Model

We design an epidemic ABM that simulates the spread of a synthetic ILI in the state of Ohio to investigate the influence of STGs on disease spread dynamics. The model consists of three primary entities: persons, buildings, and diseases and two primary processes: person movement and person infection. Buildings are subdivided into three types: home, work, and school.

Spatial and temporal granularities are defined parameters to enable the investigation of their influence. Our patch-based ABM uses a grid of square patches whose size is configurable, which controls the spatial granularity of a simulation. Buildings are located on one or more patches

and persons are located on a single patch. The probabilities of infection for a susceptible person are calculated based on the number of infectious agents residing on the same patch or on nearby patches. Our ABM uses discrete-time representation with time step durations in these experiments ranging from 8 hours to 1 hour, which controls the temporal granularity of a simulation. The use of discrete-time steps is consistent with other epidemic ABMs (Longini et al., 2004; Germann et al., 2006; Atti et al., 2008; Chao et al., 2010).

Our ABM uses data published by, for example, US Census Bureau and US Bureau of Transportation Statistics, and a Synthesized Population Database, hereafter referred to as the population database, produced by researchers at RTI, International for the Models of Infectious Disease Agent Study (MIDAS) (Wheaton et al., 2009), which has been used in other epidemic ABM studies (Brown et al., 2011). Our model uses generated individuals and buildings in the population database, which include assignment of individuals to households, work buildings, and schools. We summarize the usage of data in our study below and refer the reader to individual data sources listed in Table 4.2 for further details.

4.3.1 Persons and Buildings

Households in the population database are generated by statistically fitting households in the US Census Public Use Microdata Sample with household counts in the US Census data at the level of block groups (Wheaton et al., 2009). Households within block groups are then spatially distributed and assigned person occupancies to fit population data to the LandScan 90-meter gridded cells (Wheaton et al., 2009). In addition to synthesized household locations and occupancies from the population database, we derive household sizes by multiplying the total number of persons in the household by 673 square feet based on the ratio of median house size (1700 square feet) and average household membership (2.526) calculated based on data in an American Housing Survey of the United States.

Work buildings and person work assignments in the population database are generated based on data from the US Census Bureau and a commercial company named InfoUSA (Wheaton et al., 2009). The population database does not assign point-based locations for businesses, but does assign a work building to a census tract. We assign each work building to a random location

within their respective census tract. We derive the size of a work building by multiplying the total number of workers in the building by 193 square feet based on findings in a Workspace Utilization and Allocation Benchmark that determined the private sector has a median 193 Usable Square Feet (USF) per worker, which is comparable to a Federal benchmark of 190 USF per worker.

School buildings in the population database are generated based on data from National Center for Education Statistics and schoolinformation.com (Wheaton et al., 2009). Persons (up to age 17) are assigned to a nearby school based on the location of their household. In the case where a person is assigned to both a school and work building, our ABM assumes they attend school. We derive the size of a school building by multiplying the total number of students in the building by 154 square feet based on averaging the median square foot per student for Elementary, Middle, and High Schools built in 2012 (Annual School Construction Report).

4.3.2 Person Movement

Persons move based on their current activities including movement within a building (e.g. walking) and movement between home and work (e.g. commuting). Our ABM is designed with four activities: stay at household, study at school, work, and sleep. The first three activities represent a typical daily work or school schedule, and the last represents resting with no probability of transmission.

Activities determine movement between buildings and are similar to movement based on behavior (Carley et al., 2006), itineraries (Parker and Epstein, 2011), and time (Lee et al., 2008). Activities change based on three primary periods: all persons sleep from 12:00am until 7:59am, work or study at school from 8:00am until 3:59pm, and stay at household from 4:00pm until 11:59pm. The durations of sleeping and working correspond to American Time Use Survey Data and the remainder of the time is assigned to being at household interacting with cohabitants (e.g. family). These 8 hour periods ensure equivalent time in each activity as temporal granularity is fined from a duration of 8 hours to 4, 2, or 1 hour per discrete-time step. Movement between buildings is instantaneous with no consideration of a transportation network, which is consistent with the assumptions in many large-scale ABMs (Atti et al., 2008; Ferguson et al., 2005, 2006; Longini et al., 2005). To encourage persons to interact as in the real world, persons move within

buildings if their activity does not change between time steps by randomly selecting a new location within the building.

4.3.3 Person Infection

A disease in our ABM represents a single strain of an ILI, which are carried and transmitted by persons while they move within and between buildings. Our ABM uses a widely adopted epidemiological model of disease progression with four stages: Susceptible, Exposed, Infectious, and Recovered (SEIR) (Hethcote, 2000). The periods of each stage are adjustable variables in our ABM. Once an infectious person becomes recovered in our ABM and is no longer susceptible to disease it is removed from a simulation. For the experiments below, we used the following periods: an infected person is in the exposed stage for 2 days, after which it becomes infectious and remains so for 4 days, which is similar to periods used in other epidemic models (Longini et al., 2004; Germann et al., 2006).

Disease transmission may occur when a susceptible person falls within a 6 feet infectious radius of an infectious person. The infectious radius matches current understanding of the primary transmission mechanism of influenza through a cough or sneeze ejecting large-particle droplets that travel at most 6 feet (Fiore et al., 2011). Our ABM uses a spatiotemporal process model (see Chapter 3) that simulates and records how often persons may be within 6 feet to potentially transmit disease based on the STGs of a simulation and a distance-based disease transmission process.

We design a spatiotemporal process model, hereafter referred to as an infection model, which simulates an infectious person trying to infect a susceptible person numerous times to calculate probabilities of infection in our patch-based ABM. During the execution of the infection model, pairs of persons are each randomly assigned a spatially explicit point-based location within areas whose size corresponds to a patch in an ABM simulation. Each person has a 42% chance of moving every one minute discrete-time step, based on studies showing that individuals in the US spend approximately 55-58% of their waking hours in sedentary behaviors (Matthews et al., 2008; Owen et al., 2010). Each moving person walks in a random direction, termed random walk (Pearson, 1905), at a speed of 5 kilometers per hour to approximate average human walking speed (Cunningham et al., 1982). Every hour, all randomly walking agents are re-assigned to a spatially

explicit point-based location in the area to which they were originally located, a variant of random walk with restart (Tong et al., 2008) that uses fixed time periods rather than a probabilistic function to restart a random walk. The infection model restarts random walks to ensure fair comparison when coarsening temporal granularities from one-hour per time step to eight-hours per time step, because otherwise agents may walk further in temporally coarse simulations and influence the interpretation of simulation results. If a pair of agents in an infection model is within a 6 feet infectious radius, then the infectious person transmits disease to the susceptible person based on an infection probability variable. The infection model repeats this procedure numerous times recording the number of pairs resulting in infection to derive probabilities of infection for a susceptible agent residing in the same patch as an infectious agent. To account for susceptible agents occupying an adjacent patch, a cornering patch, or two or more patches away from an infectious agent, the infection model repeats the aforementioned procedure, but changes the area to which a susceptible agent is initially located. The probabilities are then used to construct an infection probability matrix as described in more detail in Chapter 3.

4.3.4 Model Parallelization

Our epidemic ABM implements a communication-aware framework (Chapter 2) to enable effective use of many processing cores. The spatial domain of the ABM is decomposed rectilinearly into a number of rectangular subdomains. Each subdomain is composed of a number of square patches and the number of subdomains corresponds to the number of cores used to simulate the parallel ABM. This approach differs from other scalable parallel systems such as Charm++ where a spatial domain is over decomposed into numerous subdomains and a runtime system dynamically assigns subdomains to cores to balance the workload (Kale and Zhen, 2010). Home, school, and work buildings are assigned to each subdomain based on their locations. Agents may move between buildings and within buildings. Agent transfers may occur after each agent movement step, which use inter-processor communication to send agents that moved off one subdomain onto another subdomain assigned to another core. This is achieved by grouping agents based on the subdomain to which they will move, and transferring each group of agents using a send group operation.

Our ABM uses agent proxies to encapsulate some of the complexities of parallelism and to facil-

itate disease spread across subdomain boundaries. For example, if a building spans across multiple subdomains then a building proxy is made for each subdomain that does not contain the lower left corner of the building, which represents the non-proxy building entity. Agent proxies enable interactions among agents that reside on different subdomains and thus require inter-processor communication to exchange information. Proxies reside on ghost zones that surround a subdomain and keep a local copy of data that reside in remote cores (Wang et al., 2006; Tang and Wang, 2009). The size of the ghost zone in our ABM is defined by the size of the infection probability matrix, which delimits the maximum distance that an infectious agent can infect a susceptible agent. Our ABM creates proxies of infectious agents rather than susceptibles, because we assume that, in general, there are fewer infectious agents compared to susceptible agents. Further, if a susceptible proxy were infected then it would be required to communicate this information back to the corresponding agent. Whereas, in the case of infectious agent proxies, no further inter-processor communication is necessary. This is because the status of an infectious agent does not change when infecting a susceptible agent, so an infectious agent proxy infecting a susceptible agent on another subdomain is not required to communicate this information back to the corresponding agent. The framework transfers proxy agents consisting of an agent ID and location, which reduces the amount of information that must be sent between cores each iteration. Proxies and agents are transferred asynchronously to improve computational performance. The ABM progresses in lock step each time step, but future implementations may employ multistep progression to reduce communication and improve performance as in (Barrett et al., 2008).

4.3.5 Limitations and Assumptions

The overarching design goal of our ABM is to investigate the influence of STGs on disease spread dynamics, therefore unlike other epidemic models it has not been tuned for one specific STG and has different assumptions compared to other models. Firstly, our ABM does not consider age, which has been shown to influence disease spread (Germann et al., 2006), future studies could use spatiotemporal process models to simulate probabilities of infections between agents of different age classes. Further, our ABM does not consider community structure or multi-scale transmission dynamics including, for example, work groups, classrooms, or neighborhoods (Smieszek et al.,

2011; Chao et al., 2010). Our ABM assumes agents randomly walk from a patch-based location in a building to infect nearby susceptible agents. While random walk has been shown to capture human movement (Brockmann et al., 2006; Jiang et al., 2009) and in part was chosen to approximate homogeneous mixing when comparing simulations of different spatial granularities, methods that are more sophisticated exist to capture pedestrian movement that would likely be more realistic (Torrens, 2012). Our ABM bases infection probabilities on relative locations of infectious and susceptible agents that are constrained by the buildings in which they reside and the STGs of a simulation. This enables our ABM to capture agents walking within buildings in spatially fine-grained simulations, rather than homogeneously mixing all agents residing in the same building (Chao et al., 2010).

4.4 Experimental Results

We investigate the influence of STGs on spatial and temporal dynamics of disease spread in the state of Ohio. The state of Ohio is a suitable study area, because it has varied population densities ranging from urban areas (e.g. Columbus, Cleveland, and Cincinnati) to rural areas as well as hosting multiple international airports, the largest being the Cleveland Hopkins International Airport. A report by the US Centers for Disease Control and Prevention (CDC) titled "Interim Pre-Pandemic Planning Guidance: Community Strategy for Pandemic Influenza Mitigation in the United States Early, Targeted, Layered Use of Nonpharmaceuticial Interventions" (Centers for Disease Control and Prevention, 2007) lists states, and in particular state health authorities such as the Ohio Department of Health, as being a fundamental unit to activate appropriate interventions and report illness in the case of a pandemic. Other epidemic studies have used the geographic unit of a state or densely populated counties (Bisset et al., 2012; Chao et al., 2011; Brown et al., 2011; Bisset et al., 2009; Barrett et al., 2008; Stroud et al., 2007) in simulating disease spread. There have also been several epidemic models designed to study disease spread at a national-level in European countries, which due to high amounts of travel amongst European countries and relatively small geographic size can be considered conceptually similar to a state when compared to this work (Atti et al., 2008; Smieszek et al., 2011). The following simulations focus on synthetic epidemic scenarios (Epstein, 2009; Chao et al., 2010; Xu and Sui, 2009) rather than reconstructing previous instances

Data	**Source**
Synthesized US Population Database	https://www.epimodels.org/midasdocs/infoMaterials/MIDAS_US_Syn_Pop_web.pdf
Workspace Utilization and Allocation Benchmark	http://www.gsa.gov/graphics/ogp/Workspace_Utilization_Banchmark_July_2012.pdf
American Household Survey of the United States	http://www.census.gov/housing/ahs/files/ahs11/National2011.xls
Annual School Construction Report	http://www.peterli.com/spm/pdfs/SchoolConstructionReport2013.pdf
American Housing Survey for the US	http://www.census.gov/housing/ahs/files/ahs11/National2011.xls
LandScan	http://http://www.ornl.gov/sci/landscan
American Time Use Survey	http://www.bls.gov/tus/tables/a2_2011.pdf

Table 4.2: This table details the data sources used to inform the design of our ABM.

of disease spread (Smieszek et al., 2011).

The ABM is implemented in C programming language and the Message Passing Interface (MPI) (Gropp et al., 1998). Random number generation is provided by a Mersenne twister method available in the GNU Scientific Library (GSL). The following experiments were executed on the supercomputer Stampede at the Texas Advanced Computing Center supported by the National Science Foundation-funded project Extreme Science and Engineering Discovery Environment (XSEDE). Stampede is comprised of 6,400 nodes that total 102,400 processing cores and 205 terabytes of memory. The Integrated Performance Monitoring (IPM) tool is used to collect simulation performance information (Skinner, 2005).

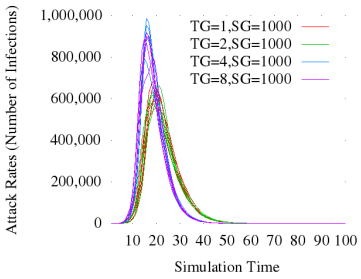
4.4.1 Influence of STG on Disease Spread Dynamics

We use the following baseline parameters for our study to gain insights toward the influence of STGs on simulated disease spread dynamics across the state of Ohio. Each parallel simulation uses 64 processing cores and is seeded by randomly infecting 100 agents on day 2, which is based on an assumption that 100 Ohio residents shared a flight with an infected person. Upon infection, agents are non-infectious (i.e. exposed) for 2 days and infectious for 4 days, which are similar periods as used in (Longini et al., 2004; Germann et al., 2006). Infectious agents infect susceptible agents within a 6 feet infectious radius with a probability of 0.17. The probability value is selected based on the finest-grained simulations being plausible for a reasonably infectious disease, which is described in further detail in the next section that discusses estimating disease infectiousness. Once an agent is recovered from disease, they are removed from a simulation, because they can no longer be infected by disease. These experiments simulate 170 days, unless disease dies out, in which case the simulation ends early. We examine temporal granularities of 1, 2, 4, and 8 hours per discrete-time step and spatial granularities of square shaped patches with edge lengths of 10, 100, 500, and 1000 meters. Our ABM uses a spatiotemporal process model to calculate probabilities of infection for all configurations of STGs as part of the experiments. The spatiotemporal process model simulates 500,000 pairs of agents for each cell in an infection probability matrix to calculate the probabilities of infection. We adopt a heuristic that assumes if five sequential cells in a single row of an infection probability matrix result in zero infections that the remainder of cells in the

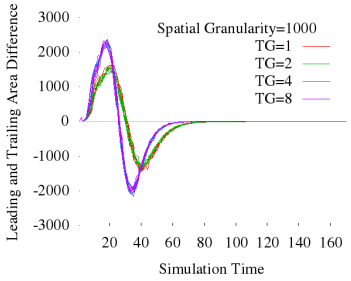
row will also result in zero infections. This heuristic has little to no influence on the results and was created to reduce the number of simulated pairs and the overall size of the probability matrix, which in the case of fine spatial granularities contains many zero probabilities for cells representing distance patches.

We execute 5 simulations for each STG configuration, and find coarsening STGs to alter daily attack rate curves, which plot the number of new infections per day across the duration of a simulation (Figures 4.1a, 4.1c, 4.1e, and 4.1g). Daily attack rate curves are often characterized by their peaks and time of peaks (Ajelli et al., 2010; Xu and Sui, 2009; Ferguson et al., 2006), and for the spatially and temporally finest-grained simulations we find that the average peak attack rate (PAR) is 1.7% and the time of peak attack rate (TPAR) is day 56. We find coarsening spatial granularities from 10m to 1000m increases PAR by 3× and speeds TPAR by 35 days on average (Figures 4.1a, 4.1c, 4.1e, and 4.1g). We find coarsening temporal granularities from 1 to 8 hours increases PAR by 1.4× and speeds TPAR by more than 5 days on average. In agreement with Chapter 3, we find a non-linear relationship between PAR and TPAR when coarsening STGs in these simulations that include multiscale movement and heterogeneous population distributions (Figure 4.2). A non-linear relationship differs from a linear relationship found when reconfiguring links in small-world network-based epidemic simulations (Xu and Sui, 2009).

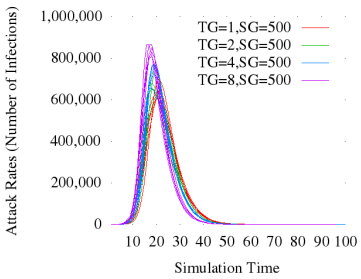
We apply a swash-backwash model to our simulation results to capture influence of STGs on spatial spread of disease (Cliff and Haggett, 2006). The state of Ohio is divided into 1000×1000 meter subareas and for each subarea the swash-backwash model records the time of the first infection (leading edge) and the time of the last infection (trailing edge). The swash-backwash model plots the difference between the number of leading and trailing edges each day to denote spatial change of disease spread. Positive values in the graph denote spatial growth, the swash portion of an epidemic wave, and negative values denote spatial shrinkage, the backwash portion of an epidemic wave. We found variations in spatial spread to broadly follow daily attack rate trends (Figures 4.1b, 4.1d, 4.1f, and 4.1h). Coarsening temporal granularities increases intensity and speed of spatial spread, measured by height and day of peak of the swash stage, similar to those changes observed in daily attack rates. Coarsening spatial granularities generally speeds spatial spread with some variation in the height of peak of the swash stage.



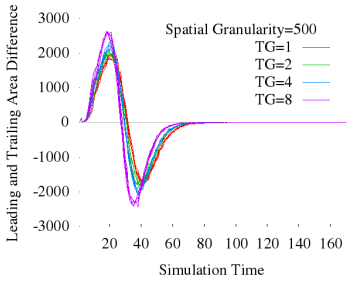
(a) Daily attack rates (Spatial granularity=1000m)



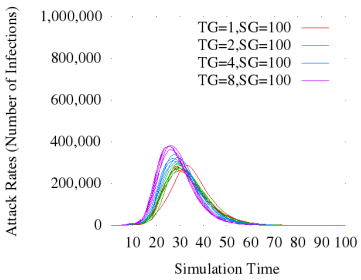
(b) Swash-backwash model results (Spatial granularity=1000m)



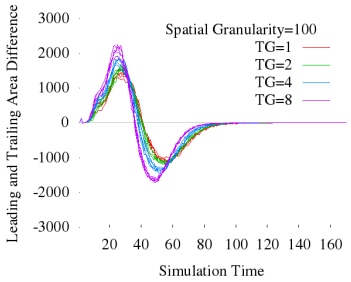
(c) Daily attack rates (Spatial granularity=500m)



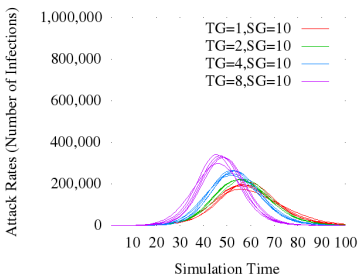
(d) Swash-backwash model results (Spatial granularity=500m)



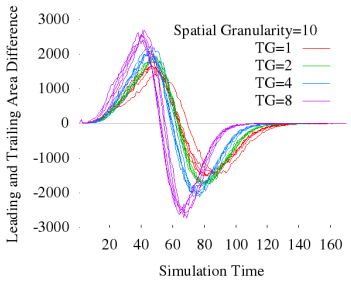
(e) Daily attack rates (Spatial granularity=100m)



(f) Swash-backwash model results (Spatial granularity=100m)



(g) Daily attack rates (Spatial granularity=10m)



(h) Swash-backwash model results (Spatial granularity=10m)

Figure 4.1: These curves plot daily attack rates (left) and the results of a swash-backwash model (right), which captures new infections and spatial spread of disease across the duration of a simulation, respectively. For each STG combination 5 simulations are executed and plotted with spatial granularities of 1000m, 500m, 100m, 10m and temporal granularities of 8h, 4h, 2h, and 1h considered. Simulation times 0-100 days are shown for daily attack rates.

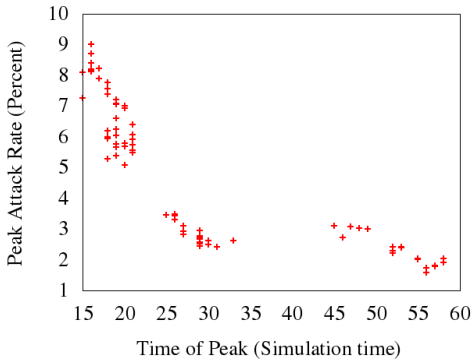


Figure 4.2: This figure illustrates a non-linear relationship between Peak Attack Rate (PAR) and Time of Peak Attack Rate (TPAR) when coarsening STGs.

It is important to observe that while daily attack rates experience significant changes when coarsening STGs there is less variation for spatial spread of disease when comparing STG configurations (i.e. swash period peaks around 1500-2500 subareas for all simulations). Observing broadly similar spatial spread helps to verify the spatiotemporal process model that follows our expectations. Recall an infection model is used by our ABM to construct infection probabilities based on the STGs of a simulation, which captures the probability of infection for a susceptible agent given it's relative location to an infectious agent. Since those probabilities reflect the STGs of a simulation, we expect that neither the patch size nor the duration of the time step should influence disease spreading amongst agents residing on different patches, which is reflected in similar trends in the spatial spread of disease. Yet, while spatial spread is similar, PAR generally increases as STGs are coarsened, which suggests that more agents are infected within similar sized spatial areas on average. This is not to say that the increase in infection is homogeneous across the state of Ohio, but is more likely to be influenced by factors such as number of persons within a household, workplace, or neighborhood, which is explored in further detail in the next experiment.

4.4.2 R_0

The basic reproductive number (R_0) is commonly used to estimate the infectiousness of a disease, which is defined as "the average number of secondary infections produced when one infected individual is introduced into a host population where everyone is susceptible" (Anderson and May, 1991) (pg. 17). For the following experiments we use random index case (Germann et al., 2006)

also referred to as first generation index (Ajelli and Merler, 2008) to estimate R_0 by executing numerous simulations with only one infectious agents in the state of Ohio and counting the number of secondary infections while ignoring tertiary infections. This method is likely to result in a slightly lower estimation than other methods of R_0 (Germann et al., 2006), but ensures that each STG combination is parameterized using the same method (Ajelli and Merler, 2008). We execute over 300 simulations for each parameter combination using a scaling method similar to (Chao et al., 2010) to scale the probabilities of infection until each parameter configuration resulted in an estimated R_0=1.6, which is within a common range tested in epidemic influenza studies (Chao et al., 2010; Ajelli et al., 2010; Ajelli and Merler, 2008; Ferguson et al., 2006). This is implemented by multiplying each entry in a probability matrix constructed in the previous experiment by a scaling factor so that the estimated R_0 value averaged 1.6. We ran sensitivity analyses changing the scaling factor used to estimate R_0 to gauge the influence of the scaling factor and found similar trends in daily attack rates and swash-backwash model results (Figure 4.3).

This experiment compares simulation results between two scenarios that are both parameterized to an estimated R_0 value of 1.6, but operate at different spatial granularities, namely 10m and 1000m. We execute 10 simulations for each scenario. The cumulative attack rate (CAR) or percentage of agents that became infected in each scenario is similar for both sets of simulations (57.7% (.01) for spatial granularity=10m and 61.2% (.02) for spatial granularity=1000m on average, standard deviation in parenthesis) and within a similar range in a study comparing three epidemic models (Halloran et al., 2008). Peak attack rates were also similar, differing by 0.42%. However, the TPAR occurred earlier in spatially coarsened simulations peaking before day 29 (1.5) whereas spatially fined simulations peaked after day 53 (1.8), on average (Figure 4.4a). While there is a delay in TPAR, other measures (i.e. PAR and CAR) generally align for these parameterized simulations.

Unlike PAR and CAR, spatial spread is significantly different when comparing these two scenarios (Figure 4.4b). Disease spatially spreads faster throughout Ohio in the spatially coarse-grained scenarios compared to the spatially fine-grained scenarios, with a peak in swash period generally corresponding to the TPAR in both scenarios. However, when comparing the two scenarios, coarsening spatial granularity limits the spatial spread of disease even while PAR and CAR are similar

overall. We find that infections tend to be concentrated around densely populated areas (i.e. urban areas) with fewer infections in rural areas in the spatially coarse-grained scenarios, whereas more infections occur in rural areas in spatially fine-grained scenarios. This is illustrated in a series of maps that capture cumulative infections (Figure 4.5).

To further illustrate the differences in spatial spread, we apply a disease risk function using kernel density estimation (KDE) to gauge the spatial distribution of risk to individuals throughout Ohio (Shi, 2010). KDE, represented as function d, estimates density values at location (x) over a two-dimensional raster space based on a number of sample points ($x_1 - x_n$), and for each location a kernel function K characterizes the contribution of each sample point within distance h, referred to as a kernel bandwidth (Silverman, 1986).

$$d(x) = \frac{1}{nh^2} \sum_{i=1}^{n} K\left(\frac{(x - x_i)}{h}\right)$$

In the context of disease mapping, disease risk (r) can be represented as a ratio of density of disease infection cases (c) and density of agent population (p). Disease infection cases for a simulation are output as spatially explicit points assigned to the lower-left corner of an agent's patch.

$$r(x) = \frac{c}{p}$$

We apply a disease risk function for a sample simulation for each scenario (Figure 4.6). In the spatially coarse-grained scenario disease risk is concentrated around the cities of Cincinnati (lower left), Columbus (center), and Cleveland (upper right) with disease risk radiating from these and other large population centers and little risk in other areas (Figure 4.6a). The spatially fine-grained scenario has slightly lower risk in the larger cities, but a noticeable increase in disease risk across the rural areas of the state (Figure 4.6b).

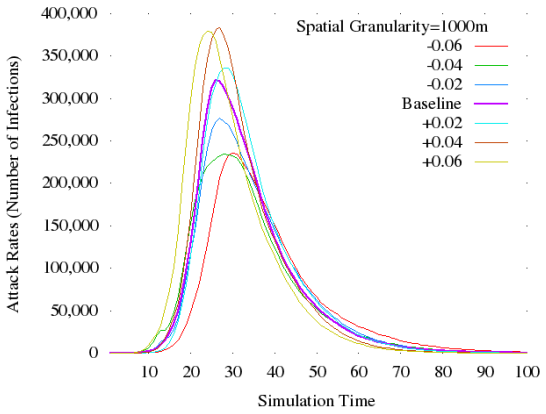
The differences in spatial spread are caused by a subtle, yet important change in probabilities of infection as spatial granularities are coarsened. In spatially fine-grained simulations, the probabilities of infection are very high for susceptible agents occupying a patch near an infectious agent, but quickly decrease with added distance. On the other hand, the chance of a susceptible

agents being next to an infectious agent in spatially coarse-grained simulations are much lower because the patch sizes are larger. As a result, the probabilities of infection for each patch are relatively low in spatially coarse-grained simulations. Due to these differences, there tends to be an insufficient number of agents in rural areas to lead to infection in spatially coarsened simulations compared to fine-grained simulations. However, in densely populated areas, the larger numbers of people occupying large patches leads to an increased outbreak of disease in spatially coarse-grained simulations, which is observed in the disease risk maps (Figure 4.6).

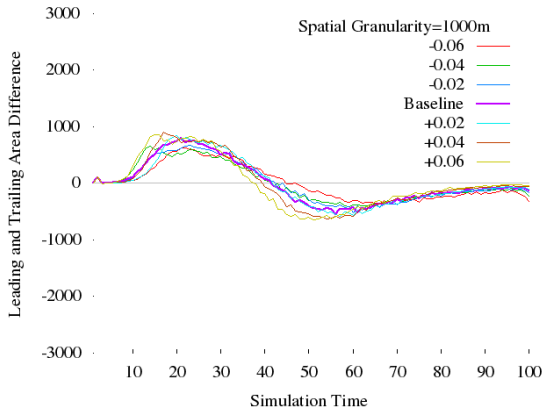
4.4.3 Computational Tractability

In general, coarsening STGs decreases overall simulation time in the first experiment (Table 4.3) with coarsening spatial granularities decreasing simulation time more than coarsening temporal granularities. Multiscale dynamics in our ABM affects a naïve expectation that when the duration of time steps are halved computation is expected to double, because it doubles the number of time steps. Recall, that agents move between buildings at 8:00am and 4:00pm and sleep from 12:00am-7:59am. During non-sleeping hours, agents also move within buildings and infectious agents try to infect nearby susceptible agents as part of an infection process. Computation does not increase linearly, because while fining temporal granularities increases the number of within building movement and infection process invocations, it does not increase invocations of between building movements. If between building movement and other processes were invoked each time step, then simulation times may increase close to linearly when fining temporal granularities. Simulations times are also influenced by other implementation details. For example, two optimizations used in our ABM is the removal of recovered agents that are no longer susceptible to disease and ending a simulation earlier when disease dies out. Coarsened spatial granularities result in higher PAR that peak earlier in the simulation so more agents are removed early from a simulation and it is more likely to end early, which results in significantly shorter simulation times.

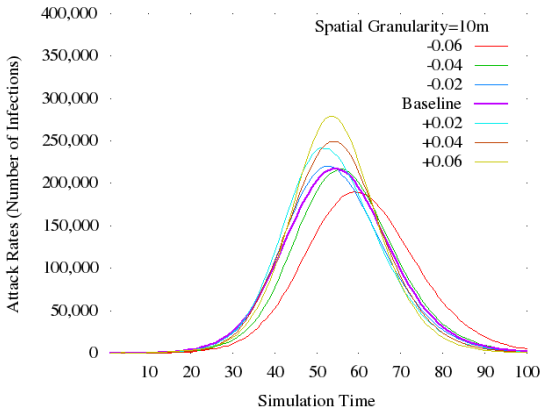
Simulation times in the R_0 experiments show approximately a $10\times$ reduction in simulation time when coarsening spatial granularity from 10m to 1000m, an average of 394 minutes and 42 minutes, respectively. Upon investigation, we find that the identification, grouping, and transfer of infectious agent proxies primarily contribute to the differences in simulation times between



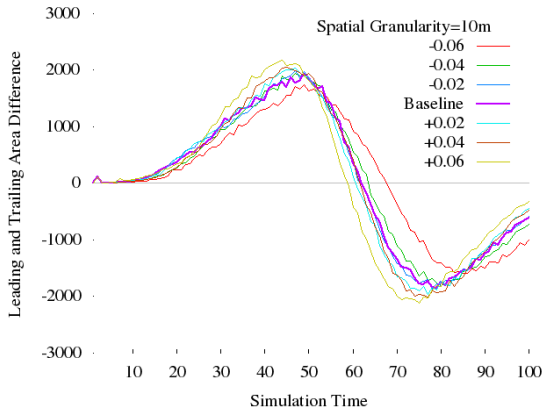
(a) Daily attack rates (Spatial granularity=1000m)



(b) Swash-backwash model results (Spatial granularity=1000m)



(c) Daily attack rates (Spatial granularity=10m)

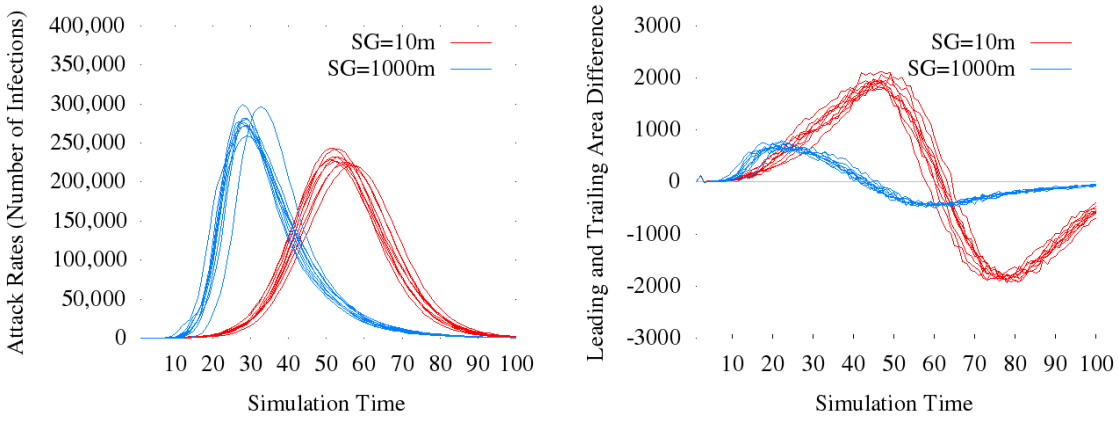


(d) Swash-backwash model results (Spatial granularity=10m)

Figure 4.3: Sensitivity analysis of a scale factor used to parameterize simulations to an estimated R_0 value of 1.6, which had a standard deviation of 0.39 and 0.22 for spatial granularities of 1000m and 10m, respectively. We change the scaling factor by increments of 0.02 and find similar trends with greater sensitivity in the 1000m scenario particularly for daily attack rates.

	1 hours	**2 hours**	**4 hours**	**8 hours**
10m	1330.5	852.3	579.2	439.4
100m	57.8	45.8	34.9	29.4
500m	21.8	17.3	15.1	12.4
1000m	27.0	21.2	13.4	12.6

Table 4.3: Average simulation execution times in minutes as STGs are coarsened.



(a) Daily attack rates

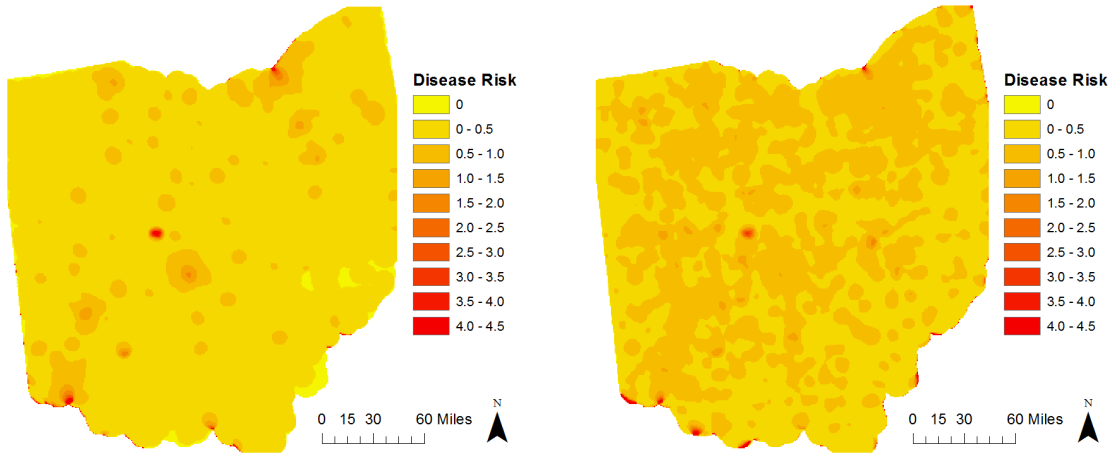
(b) Swash-backwash model

Figure 4.4: These figures compare daily attack rates and swash-backwash model results for simulations operating at spatial granularities of 1000m and 10m, both of which are parameterized to the same basic reproductive number, R_0=1.6.

the two scenarios. While further optimizations, including improvements in spatial hashing for example, may help to improve overall simulation execution times at finer spatial granularities, it is important to recognize the connections between spatial distributions of infection and infectious agent proxies. Recall, disease risk maps show concentrations of infections in highly populated areas in the spatially coarsened scenario compared to the spatially fined scenario (Figure 4.6). Further recall, the communication-aware framework used by our ABM decomposes a spatial landscape (e.g. the state of Ohio) rectilinearly with a goal of reducing perimeter-to-area ratio and thus inter-processor communication (Table 2.2). The concentration of infections in a limited number of small spatial areas results in few infectious agent proxies created and transferred across rectilinearly decomposed subdomains in the spatially coarse-grained scenario. Whereas infections in the spatially fine-grained scenario are spatially distributed resulting in significantly more infectious agent proxies needing to be created and transferred, which leads to an increase in communication and ultimately simulation time. These findings show the influence of STGs is not limited to simulated disease dynamics, but is also tightly linked to computational tractability.

(a) Day 20 (b) Day 40 (c) Day 60 (d) Day 80 (e) Day 100

Figure 4.5: Maps of cumulative infections across the duration of a simulation comparing spatial granularity=10m (top) and spatial granularity=1000m (bottom). Red areas denote presence of infection, grey areas denote home buildings to capture population distributions.



(a) Disease risk (Spatial granularity=1000m)

(b) Disease risk (Spatial granularity=10m)

Figure 4.6: Disease risk maps comparing a spatially coarse- and fine-grained simulation that are calculated as a ratio of disease case and population densities using a fixed bandwidth kernel estimation (BW=10km).

4.5 Concluding Discussion

This article simulated disease spread amongst more than ten million agents representing residents in the state of Ohio using an epidemic ABM to demonstrate the influence of STGs on spatial and temporal disease spread dynamics. An epidemic ABM is designed based on a spatiotemporal process model to enable it to adapt to different STGs. Experimental results show a clear influence of STGs on disease spread processes as STGs are coarsened from a patch size of 10m to 1000m and discrete-time step duration of 1 hour to 8 hours. Another experiment demonstrates the influence of spatial granularities on spatial spread of disease and simulation times. While both coarse- and fine-grained scenarios are parameterized to the same estimated basic reproductive number (R_0), disease risk is found to be concentrated in and radiating from densely populated urban areas in the spatially coarse-grained scenario. Whereas disease risk in the spatially fine-grained scenario was slightly lower in urban areas, but higher in rural areas. We found simulation time to be influenced by STGs in these simulations with a significant decrease in simulation time for coarse-grained simulations compared to fine-grained simulations, which are in part caused by spatial distribution of disease infections and consistent with previous comparisons of epidemic models operating at different spatial granularities (Ajelli et al., 2010).

Fining spatial granularities from 1000m to 10m is found to reduce PAR in our epidemic ABM, and our previous study using a parsimonious epidemic ABM based on a random walk process also shows a reduction in PAR when fining spatial granularities (Chapter 3). These results suggest that finer spatial granularities (e.g. 1m or finer) will further reduce PAR until ultimately disease transmission is distance-based relying on spatially explicit point-based representations. Extrapolating the results of these simulations would suggest that other epidemic ABMs that use spatially aggregated representations of space (i.e. group- or patch-based epidemic models) may also be influenced by their coarsened spatial granularities, which is consistent with other findings (Ajelli et al., 2010). However, epidemic models continue to be challenged by our limited understanding of human movement processes at fine spatial and temporal scales. Recent studies using tracking technologies such as GPS-enabled cellphones may provide a wealth of data to improve our understanding of fine-grained human movement (Isaacman et al., 2010), which could be used to inform the design of finer grained epidemic models.

This study also found differences in disease risk in different population densities. We speculate that population density may alter how STGs influence disease spread processes, because a previous study using a parsimonious epidemic ABM found a reduction in PAR as temporal granularity is coarsened (Chapter 3) whereas this study found an increase in PAR. Future studies may explore the effect of density-dependent transmission factors adopted in equation-based models (Mishra et al., 2011), but such approaches may not accurately reflect disease spreading processes and our experience leads us to suggest another avenue of exploration. A recent study using cellphone tracking found travel ranges to differ between residents in Manhattan, NY and Los Angeles, CA and these ranges tend to be inversely proportional to population density suggesting a potential role in human travel behavior (Isaacman et al., 2010). Commonly used spatial analysis techniques including adaptive or density-based bandwidth used in KDE (Silverman, 1986) could be used to limit the distance in which infectious agents can infect nearby susceptible agents, which is similar to those used in an epidemic modeling study by Ajelli and Merler (2008).

Contextualizing the results of both experiments indicates a potential trade-off in space and time. The first experiment used the same probabilities of disease transmission for all STG combinations, which resulted in similar trends in spatial spread, but significantly different daily attack rates. The

second experiment parameterized two scenarios to the same estimated R_0, one spatially fine-grained and the other coarse-grained, which resulted in similar trends in daily attack rates, but significantly different spatial spread. All simulations used the same epidemic ABM, indicating that alterations in spatial characteristics of disease spread may influence temporal characteristics and vice versa, which is supported by similar findings in Chapter 3 that used a parsimonious epidemic ABM, and helps to explain differences observed when comparing epidemic models operating at different spatial granularities (Ajelli et al., 2010). These findings have important ramifications for understanding simulated disease spread in epidemic models. In particular, our experiments demonstrate that a common method used to parameterize diseases to an estimated R_0 in epidemic models may align disease spread dynamics in the temporal dimension, but it may fundamentally alter spatial spread of disease.

While epidemic models differ in assumptions and approaches, conceptually we can make analogies between our patch-based epidemic ABM and other spatially aggregated epidemic models (i.e. patch- and group-based). For example, many epidemic ABMs couple fixed transmission probabilities with group-based disease transmission methods, with agents grouped at the level of a building (Table 4.1). Buildings with more employees or family members can be considered denser than buildings with fewer numbers of people, which is similar to urban areas compared to rural areas in our epidemic ABM. These models may over-represent disease infections in buildings with greater numbers of individuals compared to fewer, which is observed in a study showing transmission rates grow linearly with household size (Marathe et al., 2011) using a widely published model (Barrett et al., 2008; Eubank et al., 2004). Further studies are needed to validate whether these dynamics reflect the true nature of disease dynamics or if fundamental assumptions in epidemic models should be re-evaluated based on the new understanding that STGs influence processes in epidemic ABMs. This research also highlights a need for improved spatial thinking and adoption of new methods to spatially evaluate epidemic simulations such as those used in public health studies (Cromley and McLafferty, 2011) including, for example, swash-backwash models and disease risk maps.

References

Ajelli, M., Gonçalves, B., Balcan, D., Colizza, V., Hu, H., Ramasco, J., Merler, S., and Vespignani, A. (2010). Comparing large-scale computational approaches to epidemic modeling: Agent-based versus structured metapopulation models. *BMC Infectious Diseases*, 10(1):190.

Ajelli, M. and Merler, S. (2008). The impact of the unstructured contacts component in influenza pandemic modeling. *PLoS One*, 3(1):e1519.

An, L., Linderman, M., Qi, J., Shortridge, A., and Liu, J. (2005). Exploring complexity in a human–environment system: An agent-based spatial model for multidisciplinary and multiscale integration. *Annals of the Association of American Geographers*, 95(1):54–79.

Anderson, R. and May, R. (1991). *Infectious diseases of humans: dynamics and control*. Oxford University Press.

Atti, M., Merler, S., Rizzo, C., Ajelli, M., Massari, M., Manfredi, P., Furlanello, C., Tomba, G., and Iannelli, M. (2008). Mitigation measures for pandemic influenza in Italy: An individual based model considering different scenarios. *PLoS One*, 3(3):e1790.

Balcan, D., Colizza, V., Gonçalves, B., Hu, H., Ramasco, J., and Vespignani, A. (2009). Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences*, 106(51):21484–21489.

Balcan, D., Gonccalves, B., Hu, H., Ramasco, J., Colizza, V., and Vespignani, A. (2010). Modeling the spatial spread of infectious diseases: The global epidemic and mobility computational model. *Journal of Computational Science*, 1(3):132–145.

Barrett, C., Bisset, K., Eubank, S., Feng, X., and Marathe, M. (2008). EpiSimdemics: An efficient algorithm for simulating the spread of infectious disease over large realistic social networks. In *SC '08: Proceedings of the 2008 ACM/IEEE conference on Supercomputing*, pages 1–12, Piscataway, NJ, USA. IEEE Press.

Batty, M., Crooks, A., See, L., and Heppenstall, A. (2012). Perspectives on agent-based models and geographical systems. In Heppenstall, A., Crooks, A., See, L., and Batty, M., editors, *Agent-Based Models of Geographical Systems*, pages 1–15. Springer Netherlands.

Bisset, K., Aji, A., Marathe, M., and Feng, W.-C. (2012). High-performance biocomputing for simulating the spread of contagion over large contact networks. *BMC Genomics*, 13(Suppl 2):S3.

Bisset, K., F., X., Marathe, M., and Yardi, S. (2009). Modeling interaction between individuals, social networks and public policy to support public health epidemiology. In *Proceedings of the 2009 Winter Simulation Conference (WSC)*, pages 2020–2031.

Brockmann, D. (2009). Human mobility and spatial disease dynamics. In Schuster, H., editor, *Reviews of Nonlinear Dynamics and Complexity*, pages 1–24. Wiley-VCH.

Brockmann, D., Hufnagel, L., and Geisel, T. (2006). The scaling laws of human travel. *Nature*, 439(7075):462–465.

Brown, S., Tai, J., Bailey, R., Cooley, P., Wheaton, W., Potter, M., Voorhees, R., LeJeune, M., Grefenstette, J., Burke, D., et al. (2011). Would school closure for the 2009 H1N1 influenza epidemic have been worth the cost?: A computational simulation of Pennsylvania. *BMC Public Health*, 11(1):353.

Carley, K., Fridsma, D., Casman, E., Yahja, A., Altman, N., Chen, L., Kaminsky, B., and Nave, D. (2006). BioWar: Scalable agent-based model of bioattacks. *IEEE Transactions on Systems, Man and Cybernetics, Part A*, 36(2):252–265.

Centers for Disease Control and Prevention (2007). Interim pre-pandemic planning guidance: Community strategy for pandemic influenza mitigation in the United States early, targeted, layered use of nonpharmaceuticial interventions. http://www.flu.gov/planning-preparedness/community/community_mitigation.pdf (Accessed: March, 2013).

Chao, D., Halloran, M., Obenchain, V., and Longini, I. (2010). FluTE, a publicly available stochastic influenza epidemic simulation model. *PLoS Computational Biology*, 6(1):e1000656.

Chao, D., Matrajt, L., Basta, N., Sugimoto, J., Dean, B., Bagwell, D., Oiulfstad, B., Halloran, M., and Longini, I. (2011). Planning for the control of pandemic influenza A (H1N1) in Los Angeles County and the United States. *American Journal of Epidemiology*, 173(10):1121–1130.

Cliff, A. and Haggett, P. (2006). A swash-backwash model of the single epidemic wave. *Journal of Geographical Systems*, 8(3):227–252.

Cromley, E. and McLafferty, S. (2011). *GIS and Public Health*. The Guilford Press.

Cunningham, D. A., Rechnitzer, P. A., Pearce, M. E., and Donner, A. P. (1982). Determinants of self-selected walking pace across ages 19 to 66. *Journal of Gerontology*, 37(5):560–564.

Epstein, J. (2009). Modelling to contain pandemics. *Nature*, 460(7256):687–687.

Eubank, S., Guclu, H., Kumar, V., Marathe, M., Srinivasan, A., Toroczkai, Z., and Wang, N. (2004). Modelling disease outbreaks in realistic urban social networks. *Nature*, 429(6988):180–184.

Ferguson, N., Cummings, D., Cauchemez, S., Fraser, C., Riley, S., Meeyai, A., Iamsirithaworn, S., and Burke, D. (2005). Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature*, 437(7056):209–214.

Ferguson, N., Cummings, D., Fraser, C., Cajka, J., Cooley, P., and Burke, D. (2006). Strategies for mitigating an influenza pandemic. *Nature*, 442(7101):448–452.

Fiore, A., Fry, A., Shay, D., Gubareva, L., Bresee, J., and Uyeki, T. (2011). *Antiviral Agents for the Treatment and Chemoprophylaxis of Influenza*. Centers for Disease Control and Prevention.

Germann, T., Kadau, K., Longini, I., and Macken, C. (2006). Mitigation strategies for pandemic influenza in the United States. *Proceedings of the National Academy of Sciences of the United States of America*, 103(15):5935–5940.

Gropp, W., Huss-Lederman, S., Lumsdaine, A., Lusk, E., Nitzberg, B., Saphir, W., and Snir, M. (1998). *MPI-The Complete Reference: Volume 2, The MPI-2 Extensions*. MIT Press, Cambridge, MA.

Halloran, M., Ferguson, N., Eubank, S., Longini, I., Cummings, D., Lewis, B., Xu, S., Fraser, C., Vullikanti, A., Germann, T., et al. (2008). Modeling targeted layered containment of an influenza pandemic in the United States. *Proceedings of the National Academy of Sciences*, 105(12):4639–4644.

Hethcote, H. (2000). The mathematics of infectious diseases. *SIAM Review*, 42(4):599–653.

Hupert, N., Xiong, W., and Mushlin, A. (2008). The virtue of virtuality: the promise of agent-based epidemic modeling. *Translational Research: The Journal of Laboratory and Clinical Medicine*, 151(6):273–274.

Isaacman, S., Becker, R., Cáceres, R., Kobourov, S., Rowland, J., and Varshavsky, A. (2010). A tale of two cities. In *Proceedings of the Eleventh Workshop on Mobile Computing Systems & Applications*, HotMobile '10, pages 19–24, New York, NY, USA. ACM.

Jiang, B., Yin, J., and Zhao, S. (2009). Characterizing the human mobility pattern in a large street network. *Physical Review E*, 80(2):021136.

Kale, L. and Zhen, G. (2010). Charm++ and AMPI: Adaptive runtime strategies via migratable objects. In Parashar, M. and Li, X., editors, *Advanced Computational Infrastructures for Parallel and Distributed Adaptive Applications*, pages 265–282. Wiley Online Library.

Kaplan, E. and Wein, L. (2003). Smallpox bioterror response. *Science*, 300(5625):1503–1504.

Keeling, M. (1999). The effects of local spatial structure on epidemiological invasions. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 266(1421):859–867.

Keeling, M., Danon, L., Vernon, M., and House, T. (2010). Individual identity and movement networks for disease metapopulations. *Proceedings of the National Academy of Sciences*, 107(19):8866–8870.

Keeling, M. and Eames, K. (2005). Networks and epidemic models. *Journal of the Royal Society Interface*, 2(4):295–307.

Lee, B., Bedford, V., Roberts, M., and Carley, K. (2008). Virtual epidemic in a virtual city: Simulating the spread of influenza in a US metropolitan area. *Translational Research*, 151(6):275–287.

Lee, B., Brown, S., Korch, G., Cooley, P., Zimmerman, R., Wheaton, W., Zimmer, S., Grefenstette, J., Bailey, R., Assi, T., et al. (2010). A computer simulation of vaccine prioritization, allocation, and rationing during the 2009 H1N1 influenza pandemic. *Vaccine*, 28(31):4875–4879.

Liu, W., Hethcote, H., and Levin, S. (1987). Dynamical behavior of epidemiological models with nonlinear incidence rates. *Journal of Mathematical Biology*, 25(4):359–380.

Longini, I., Halloran, M., Nizam, A., and Yang, Y. (2004). Containing pandemic influenza with antiviral agents. *American Journal of Epidemiology*, 159(7):623–633.

Longini, I., Nizam, A., Xu, S., Ungchusak, K., Hanshaoworakul, W., Cummings, D., and Halloran, M. (2005). Containing pandemic influenza at the source. *Science*, 309(5737):1083–1087.

Manson, S., Sun, S., and Bonsal, D. (2012). Agent-based modeling and complexity. In Heppenstall, A., Crooks, A., See, L., and Batty, M., editors, *Agent-Based Models of Geographical Systems*, pages 125–139. Springer Netherlands.

Marathe, A., Lewis, B., Chen, J., and Eubank, S. (2011). Sensitivity of household transmission to household contact structure and size. *PLoS One*, 6(8):e22461.

Matthews, C., Chen, K., Freedson, P., Buchowski, M., Beech, B., Pate, R., and Troiano, R. (2008). Amount of time spent in sedentary behaviors in the United States, 2003–2004. *American Journal of Epidemiology*, 167(7):875–881.

Mishra, S., Fisman, D., and Boily, M. (2011). The ABC of terms used in mathematical models of infectious diseases. *Journal of Epidemiology and Community Health*, 65(1):87–94.

Openshaw, S. (1984). *The modifiable areal unit problem*. Geo Books Norwich, UK.

Owen, N., Sparling, P., Healy, G., Dunstan, D., and Matthews, C. (2010). Sedentary behavior: emerging evidence for a new health risk. In *Mayo Clinic Proceedings*, volume 85, pages 1138–1141. Mayo Foundation.

Parker, D., Manson, S., Janssen, M., Hoffmann, M., and Deadman, P. (2003). Multi-agent systems for the simulation of land-use and land-cover change: A review. *Annals of the Association of American Geographers*, 93(2):314–337.

Parker, J. and Epstein, J. (2011). A distributed platform for global-scale agent-based models of disease transmission. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 22(1):2:1–2:25.

Parry, H. and Bithell, M. (2012). Large scale agent-based modelling: A review and guidelines for model scaling. In Heppenstall, A., Crooks, A., See, L., and Batty, M., editors, *Agent-Based Models of Geographical Systems*, pages 271–308. Springer Netherlands.

Pearson, K. (1905). The problem of the random walk. *Nature*, 72(1865):294–294.

Rahmandad, H. and Sterman, J. (2010). Heterogeneity and network structure in the dynamics of diffusion: Comparing agent-based and differential equation models. *Management Science*, 54(5):998–1014.

Riley, S. (2007). Large-scale spatial-transmission models of infectious disease. *Science*, 316(5829):1298–1301.

Roche, B., Drake, J., and Rohani, P. (2011). An agent-based model to study the epidemiological and evolutionary dynamics of influenza viruses. *BMC Bioinformatics*, 12(1):87.

Shi, X. (2010). Selection of bandwidth type and adjustment side in kernel density estimation over inhomogeneous backgrounds. *International Journal of Geographical Information Science*, 24(5):643–660.

Shirley, M. and Rushton, S. (2005). The impacts of network topology on disease spread. *Ecological Complexity*, 2(3):287–299.

Silverman, B. (1986). *Density estimation for statistics and data analysis*, volume 26. Chapman & Hall/CRC.

Skinner, D. (2005). Performance monitoring of parallel scientific applications. Technical report LBNL/PUB–5503, Lawrence Berkeley National Laboratory, Berkeley, CA.

Smieszek, T. (2009). A mechanistic model of infection: Why duration and intensity of contacts should be included in models of disease spread. *Theoretical Biology and Medical Modelling*, 6(1):25.

Smieszek, T., Balmer, M., Hattendorf, J., Axhausen, K., Zinsstag, J., and Scholz, R. (2011). Reconstructing the 2003/2004 H3N2 influenza epidemic in Switzerland with a spatially explicit, individual-based model. *BMC Infectious Diseases*, 11(1):115.

Stanilov, K. (2012). Space in agent-based models. In Heppenstall, A., Crooks, A., See, L., and Batty, M., editors, *Agent-Based Models of Geographical Systems*, pages 253–269. Springer Netherlands.

Stroud, P., Del Valle, S., Sydoriak, S., Riese, J., and Mniszewski, S. (2007). Spatial dynamics of pandemic influenza in a massive artificial society. *Journal of Artificial Societies and Social Simulation*, 10(4):9.

Tang, W. and Wang, S. (2009). HPABM: A hierarchical parallel simulation framework for spatially-explicit agent-based models. *Transactions in GIS*, 13(3):315–333.

Tong, H., Faloutsos, C., and Pan, J.-Y. (2008). Random walk with restart: Fast solutions and applications. *Knowledge and Information Systems*, 14(3):327–346.

Torrens, P. (2012). Moving agent pedestrians through space and time. *Annals of the Association of American Geographers*, 102(1):35–66.

Wang, D., Berry, M., Carr, E., and Gross, L. (2006). A parallel fish landscape model for ecosystem modeling. *Simulation*, 82(7):451–465.

Wheaton, W., Cajka, J., Chasteen, B., Wagener, D., Cooley, P., Ganapathi, L., Roberts, D., and Allpress, J. (2009). Synthesized population databases: A US geospatial database for agent-based models. RTI Press paper available at http://www.rti.org/pubs/mr-0010-0905-wheaton.pdf (Accessed: March, 2013).

Xu, Z. and Sui, D. (2009). Effect of small-world networks on epidemic propagation and intervention. *Geographical Analysis*, 41(3):263–282.

CHAPTER 5

CONCLUDING DISCUSSION

This dissertation establishes STGs as influencing processes and computational tractability in epidemic ABMs. The key contributions of these three papers (Chapters 2-4) are summarized below followed by detailed discussions and promising areas of future research. Chapter 2 examines inter-processor communication, a widely recognized challenge in parallel ABMs, and formalizes the relationships between inter-processor communication and ABMs in a conceptual design. A communication-aware framework is developed to overcome the challenges of inter-processor communication, which is shown to enable a parallel ABM to efficiently scale to thousands of processor cores and simulate billions of agents. Chapter 3 formulates a novel modeling approach, named "spatiotemporal process models," to contextualize ABM processes in space and time. The modeling approach is applied to a patch-based epidemic ABM enabling it to adapt to different STGs. This chapter synergistically integrates both spatial and temporal perspectives to deepen understanding of the influence of STGs on epidemic ABM processes. Simulation results reveal coarsening spatial granularities speeds the spread of disease, while coarsening temporal granularities slows the spread of disease. Chapter 4 integrates the communication-aware framework and spatiotemporal process modeling approach as part of an epidemic ABM simulating more than ten million agents representing residents in the state of Ohio. Experimental results demonstrate varying STGs influence disease spread dynamics, and the existence of a space-time trade-off when parameterizing a synthetic disease. In addition to gaining systematic understanding of the interrelationships between space, time, and process, the results in the fourth chapter highlight an increased need to apply spatial thinking and analyses to epidemic modeling, which has largely been lacking in the literature for large-scale epidemic models that may be used to inform public policy decisions.

The primary contributions of the second chapter include a conceptual design to help agent-based modelers understand the role of inter-processor communication in parallel ABMs, and a suite of components organized into a communication-aware framework to overcome this computational bottleneck. Computational experiments based on an implementation of the framework

using the C programming language and the Message Passing Interface (MPI) (Gropp et al., 1998) demonstrate that the communication-aware framework is capable of efficiently executing on thousands of processing cores and support the simulation of billions of agents and environment features on state-of-the-art supercomputers.

The third chapter represents the first study to critically examine how variations of the representations of both space and time in a single epidemic ABM shape processes that operate within the constraints of such representations. This chapter has set the foundation for studying variations in spatial and temporal representations in a broad class of discrete-time spatial models using spatiotemporal process models. This study sought parsimony in an epidemic ABM to gain systematic understanding of the influence of STGs on model processes and to provide clear interpretation of results (Goldenfeld and Kadanoff, 1999). Simulation results of a synthetic disease demonstrate that, in general, coarsening spatial granularities, which effectively increases the amount of homogeneous mixing amongst agents: (1) increases peak attack rate, (2) speeds time of peak attack rate, (3) increases cumulative attack rate, (4) reduces chance of disease die-off, and (5) increases the speed of spatial spread. Whereas, in general, coarsening temporal granularities has the opposite effect. The spatial granularity portion of this study is consistent with previous studies that show that increased structure in the ways in which individuals interact, including, for example, network (Ferrari et al., 2011) and community (Chao et al., 2010) oftentimes reduces overall attack rates. The temporal granularity portion of this study is consistent with a previous study that stated dynamics differed as the duration of a time step is refined from one day to 0.1 days (Roche et al., 2011).

An epidemic ABM capable of adapting to different STGs is designed to investigate their influence in a realistic scenario simulating millions of agents in the fourth chapter. STGs are shown to significantly influence PAR and TPAR while having limited influence on spatial spread in an experiment that fixed the probabilities of infection. Another experiment using the same ABM demonstrated that parameterizing spatially fine- and coarse-grained simulations to the same estimated basic reproductive number (R_0) resulted in similar PAR and TPAR, but significantly different spatial spread. Specifically, disease spatially spread faster in spatially coarse-grained simulations, but exposed fewer agents to disease risk in rural areas compared to spatially fine-grained simula-

tions. The results of these experiments show that alterations in temporal characteristics of disease spread processes affect spatial characteristics and vice versa demonstrating a space-time trade-off in spatial processes. The practical implications of these findings, such as over- or under-estimating disease risk in rural versus urban areas, could impact understanding of epidemic modeling results, which are increasingly used to inform public policy decisions such as vaccine distribution strategies or estimating surge capacity for public health facilities in the case of an epidemic, for example (Epstein, 2009).

Broadly, this dissertation investigates how limitations in data and computation may influence spatial processes. In the context of epidemic modeling, current research suggests that finer STGs are necessary to capture detailed processes including, for example, mechanistic-based transmission and contact intensity (Smieszek, 2009); however, improvements in computational performance and understanding of human movement processes will be needed to achieve these granularities for large-scale models. This research improves understanding of how and why these fine-grained models may produce different results compared to coarse-grained models, which helps to resolve an ongoing discourse in epidemic modeling literature (Hupert et al., 2008). By systematically examining the influence of coarsening STGs on epidemic modeling processes, my dissertation demonstrates that space is intrinsically linked with time through spatial processes and that coarsening one will likely affect the other yielding a space-time trade-off. The remainder of this section describes promising future steps based on this research.

The communication framework developed in Chapter 2 provides a solid foundation for the design and development of a parallel computing platform for a CyberGIS–a cyberinfrastructure-based GIS (Wang, 2010). Such a platform could be used to create new computational methodologies that synergistically integrate spatial models and analyses with CyberGIS for efficient and effective development and sharing of geospatial knowledge. The advantages to such a platform are two-fold. Firstly, a platform built on a communication-aware framework for parallel ABMs would be capable of supporting spatial processes, which is a growing trend in GIScience research. Secondly, Prieto et al. (2012) states in a recent epidemic modeling review article that epidemic models should be housed in a cyber computing environment with an easy to use interface, and a CyberGIS-enabled platform would provide not only computational support enabled by cyberinfrastructure and a suite

of spatial tools, but also an easy-to-use interface for researchers and policy makers.

This study highlights a problem that is common in epidemic modeling, in which epidemic models are parameterized and evaluated almost exclusively using temporal methods. Commonly used methods to evaluate disease dynamics (i.e. daily and cumulative attack rates) (Prieto et al., 2012) may not be sufficient to fully understand how disease spreads through a simulated society in epidemic ABMs, because disease spread is a spatial-temporal phenomenon. Improved spatial understanding may help to explain why a study comparing intervention effectiveness in three epidemic ABMs found significantly different outcomes when simulating residents in Chicago, IL (Halloran et al., 2008), which could be due to overly concentrated disease infections in urban areas that commonly used methods failed to capture. These findings suggest further studies are needed to understand simulated spatial spread of disease in epidemic models used to inform public policy decisions (Barrett et al., 2011), such as estimating surge capacities (Epstein, 2009) or evaluating the effectiveness of school closures (Brown et al., 2011).

Recasting an ongoing discourse in epidemic modeling that questions the complexities of ABMs (Hupert et al., 2008) to a discussion focused on STGs may add new insights based on our findings. This dissertation research shows that coarsening STGs influences both ABM processes and disease spread dynamics, which may be sensitive to other factors such as population density. Therefore, fine-grained epidemic ABMs operating at the level of an individual may be appropriate for simulating epidemic ABMs, even if they demand massive amounts of data and computation. However, due to limitations in data and computation there may be a practical limits to refining STGs. Empirical studies are needed to help identify appropriate spatial and temporal scales for epidemics such as influenza, which may differ by strain, and are challenged by the increasingly recognized interrelationships between biological, behavioral, and group factors that influence disease (Galea et al., 2010). Combining new empirical studies with this dissertation research, which designs a novel computational approach to investigate the influence of STGs on epidemic ABMs, may ultimately improve understanding of spatial spread of epidemic diseases leading to more effective intervention strategies.

References

Barrett, C., Eubank, S., Marathe, A., Marathe, M., Pan, Z., and Swarup, S. (2011). Information integration to support model-based policy informatics. *The Innovation Journal: The Public Sector Innovation Journal*, 16(1).

Brown, S., Tai, J., Bailey, R., Cooley, P., Wheaton, W., Potter, M., Voorhees, R., LeJeune, M., Grefenstette, J., Burke, D., et al. (2011). Would school closure for the 2009 H1N1 influenza epidemic have been worth the cost?: A computational simulation of Pennsylvania. *BMC Public Health*, 11(1):353.

Chao, D., Halloran, M., Obenchain, V., and Longini, I. (2010). FluTE, a publicly available stochastic influenza epidemic simulation model. *PLoS Computational Biology*, 6(1):e1000656.

Epstein, J. (2009). Modelling to contain pandemics. *Nature*, 460(7256):687–687.

Ferrari, M., Perkins, S., Pomeroy, L., and Bjørnstad, O. (2011). Pathogens, social networks, and the paradox of transmission scaling. *Interdisciplinary Perspectives on Infectious Diseases*, 2011(267049):1–10.

Galea, S., Riddle, M., and Kaplan, G. A. (2010). Causal thinking and complex system approaches in epidemiology. *International Journal of Epidemiology*, 39(1):97–106.

Goldenfeld, N. and Kadanoff, L. (1999). Simple lessons from complexity. *Science*, 284(5411):87–89.

Gropp, W., Huss-Lederman, S., Lumsdaine, A., Lusk, E., Nitzberg, B., Saphir, W., and Snir, M. (1998). *MPI-The Complete Reference: Volume 2, The MPI-2 Extensions*. MIT Press, Cambridge, MA.

Halloran, M., Ferguson, N., Eubank, S., Longini, I., Cummings, D., Lewis, B., Xu, S., Fraser, C., Vullikanti, A., Germann, T., et al. (2008). Modeling targeted layered containment of an influenza pandemic in the United States. *Proceedings of the National Academy of Sciences*, 105(12):4639–4644.

Hupert, N., Xiong, W., Mushlin, A., et al. (2008). The virtue of virtuality: The promise of agent-based epidemic modeling. *Translational Research: The journal of Laboratory and Clinical Medicine*, 151(6):273–274.

Prieto, D., Das, T., Savachkin, A., Uribe, A., Izurieta, R., and Malavade, S. (2012). A systematic review to identify areas of enhancements of pandemic simulation models for operational use at provincial and local levels. *BMC Public Health*, 12(1):13.

Roche, B., Drake, J., and Rohani, P. (2011). An Agent-Based Model to study the epidemiological and evolutionary dynamics of Influenza viruses. *BMC Bioinformatics*, 12(1):87.

Smieszek, T. (2009). A mechanistic model of infection: Why duration and intensity of contacts should be included in models of disease spread. *Theoretical Biology and Medical Modelling*, 6(1):25.