A Multiscale Stochastic Cellular Automaton Model for Dispersion Process with Applications to Mountain Pine Beetle Infestations

by

Yanjun Liu

M.Sc., DePaul University, 2013
B.Sc., Southwestern University of Finance and Economics, 2011

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
Department of Statistics and Actuarial Science
Faculty of Science

SIMON FRASER UNIVERSITY
Summer 2024

Declaration of Committee

Name:	**Yanjun Liu**
Degree:	**Doctor of Philosophy**
Thesis title:	**A Multiscale Stochastic Cellular Automaton Model for Dispersion Process with Applications to Mountain Pine Beetle Infestations**
Committee:	**Chair:** X. Joan Hu Professor, Statistics and Actuarial Science
	Donald Estep Supervisor Professor, Statistics and Actuarial Science
	Owen Ward Committee Member Assistant Professor, Statistics and Actuarial Science
	Derek Bingham Examiner Professor, Statistics and Actuarial Science
	Michael Antolin External Examiner Professor, Biology Colorado State University

Abstract

Dispersion, the collective spatial movement of individuals in a population, often plays a prominent role in the dynamics of the population. However, dispersion is a multiscale phenomena bridging processes affecting individuals with processes affecting collections of individuals while the processes affecting collections of individuals are often unknown or difficult to quantify. This raises the need to develop models of dispersion.

In this thesis, we construct and investigate a multiscale model of dispersion. Dispersion of population density is modelled at a macroscale by a stochastic process inspired by cellular automata models. The dispersion model is coupled to microscale models of population that describe the local production of individuals in the population. Assuming that dispersion occurs on a slower time scale than the production of individuals, we construct a systematic approach to couple the macroscale dispersion model to the microscale population model. The resulting multiscale model uses relatively few parameters and those parameters have physical interpretations with respect to the population being modelled. The model is also flexible in that it can incorporate a wide range of microscale population models.

We carry out a thorough numerical investigation of the properties of the multiscale model for dispersion. The analysis shows that the multiscale dispersion model can yield a wide range of spatial and dynamic behaviours depending on the choice of parameters. We also demonstrate that the model parameters can be calibrated using data from specific statistics that characterize the spatial patterns of dispersion.

We apply the stochastic model of dispersion to mountain pine beetle infestations of forest in North America. Our results suggest that the model can produce patterns that are qualitatively similar to those infestations. We also calibrate the dispersion model using data consisting of images of an actual pine beetle infestation.

Keywords: Dispersion process; Multiscale modelling; Stochastic Cellular Automaton model; Calibration; Mountain Pine Beetle infestations

Dedication

To my family.

Acknowledgements

I would like to thank my advisor, Donald Estep, for his wisdom, patience, and uncompromising approach to research. I could not have asked for a better guide than Don through graduate school and the beginnings of my scientific career. I want to acknowledge Don's skill as a principled editor of manuscripts. Thank you for teaching me how to do science.

Table of Contents

List of Figures

Chapter 1

Introduction

1.1 Motivation

Dispersion processes, governing the movement of a population of individuals, play a vital role in science and engineering including biology, microbiology, ecology, physics, epidemiology, materials and geography. The study of dispersion processes has important applications to the structure and dynamics of populations of various types. Dispersion processes typically operate at a range of temporal and spatial scales. In practice, the dispersion of population often has a stochastic nature at the macroscale that is driven by microscale processes affecting the population of individuals at the microscale. By microscale, we mean that the underlying stochastic driving process occurs at many highly localized spatial points at a relatively rapid rate. Models of dispersion that incorporate information from individuals at the microscale are thus appealing.

Modelling dispersion is a challenging task because the mechanisms are often unknown or difficult to describe and the evolution of dispersion patterns for large-scale populations depends on local interactions of individual entities. To address this challenge, we construct a multiscale stochastic model for dispersion of the state of population that is described as a density. The population is assumed to be generated on a small, fast time scale at many localized regions while the dispersion takes place over a larger spatial scale at a slower rate. The hierarchical structure of the model includes: 1. The microscale describes the individual behaviours locally and can be well-modelled by existing models; 2. The macroscale captures the movements of spatially distributed communities and is modelled by novel stochastic cellular automata (SCA) approach. We use a sequential approach to bridge the two scales such that the microscale can provide input for the macroscale and the macroscale influences individual behaviours at the microscale.

The multiscale SCA model can be used for various applications. The primary application in this thesis is modelling Mountain Pine Beetle (MPB) infestation. MPB epidemics have had strong ecological and economic consequences. According to the biology and life dynamics of mountain pine beetles, local individual behaviour leads to changes in tree

infestation severity over large areas as the epidemic progresses. The microscale model is used to characterize the individual dynamics of beetles in one or a few neighbouring trees, while the macroscale SCA model is applied to the dispersion of infestation within a forest region. Nuclear radiation damage provides another application for the multiscale SCA model. Nuclear reactor materials are continuously bombarded with different kinds of energetic particles during service. Such kinds of particles interact with materials to produce high concentrations of defects such as vacancies and interstitials. These defects eventually induce microstructural changes in materials such as dislocation loops and voids, which drastically change the thermal and mechanical properties of materials and limit their performance. The microscale model characterizes the atomic displacement collision cascades that happen within nanoseconds and induce microstructure changes in the fuel lattice, microstructure changes over long-range space and time lead to radiation damage in materials which can be evaluated by phase field models. The macroscale model characterizes the level of structure change in materials where a large number of voids grow, shrink, and/or interact with each other.

In this thesis, we derive a multiscale stochastic process model. Two main contributions are the derivation of a novel stochastic process model for dispersion processes and a systematic method for coupling a macroscale stochastic process model to a microscale stochastic driving process. In addition, we conduct a thorough numerical investigation of the model properties, including numerically demonstrating that the model parameters can be calibrated. We explain how the model can be applied to study the dispersion of pine beetles in an infestation and carry out a Bayesian calibration of the model using photographic evidence of pine beetle damage.

1.2 Thesis Outline

In Chapter 2, we describe some of the social and economic implications of MPB outbreaks, which motivates our work in modelling the population spread and predicting future MPB epidemics. We also describe the biology of MPB and forest management plans, and summarize the available data for MPB research. In Chapter 3, we provide the mathematical background for constructing the multiscale SCA model and describe details for the SCA model setup and choice of microscale model. We also present the sequential approach to connect the models at different scales. We provide implementation details for the implementation of the stochastic process model in Chapter 4. We use R programming language (version 4.1.1) to build a simulator that takes an initial cell configuration and a set of parameter values for the SCA transition rules as input and outputs the cell configuration evolution at a sequence of times as output. We investigate the properties of the SCA model through numerical simulations in Chapter 5 under various situations. We present the mathematical background for solving the calibration problem and explain how the model can be

applied to study the dispersion of pine beetles in an infestation and carry out a Bayesian calibration of the model using photographic evidence of pine beetle damage in Chapter 6.

1.2.1 The Dispersion Model

A fundamental goal in the study of dispersion is to describe the patterns that is generated. These patterns describe how individuals are distributed in space, which effects how they interact with each other and their surrounding environment over time. Population density, which is the number of individuals per unit area or volume, driven by individual motions such as aggregation and repulsion, plays a crucial role in determining the evolution of dispersion patterns. Dispersion processes often display multiscale characteristics, in which population is generated on a fast temporal and short spatial scale while dispersion takes place over a larger scale at a slower rate. For this reason, we aim to build a multiscale framework to characterize the complex population dynamics generated from dispersion process at the community level while explicitly modelling individual behaviour.

There are two scales in our modelling framework, the microscale and the macroscale. At the macroscale, dispersion is considered as a dynamic force for the movements of spatially distributed communities described in terms of population density. It is described by an SCA model that operates on grid of polygons. The dynamics of dispersion are governed by probabilistic rules that reflect mortality, reproduction, immigration related to distance, and chance of successful colonization in new locations. The parameters in the SCA model are simple to adjust and combinations of assumptions can be handled. At the microscale, we use a model that produces detailed information about individuals that has an impact on dispersion patterns. In the application to MPB, we choose a simple birth-death process to model the individual behaviours. The microscale process model produces input population for the macroscale SCA. The model easily accommodates other microscale models.

The model characterizes the three stages of dispersion process that account for local individual behaviours and spatial heterogeneity. The model is built upon four fundamental assumptions about the dispersion process: 1. We assume that the model focus on simulating the dispersion process that are driven by several mechanisms including active dispersion, an individual controls its own locomotion and passive dispersion, and the individual movement depends on external forces such as gravity and wind. 2. We assume that the model operates within a reasonable expanse of magnitude. Physical scale is defined by computational capacity. Currently, computer simulations are too computational intensive if the model is performed over a large landscape area, for example, thousands of kilometres. 3. We assume that the geographic features of the grid are static as the model simulates evolution of patterns over time. 4. The model applies to the situation in which there are enough individuals to produce a viable density. We use a sequential approach to connect the two scales such that the microscale can provide subject inputs for the macroscale and the macroscale in

turn feed back into the individual behaviours at the microscale. We present more details about the model in Chapter 3.

1.2.2 Bridging the Microscale and the Macroscale Model

We use a sequential approach to couple the microscale and macroscale models. In sequential approaches, information is obtained at the microscale, and the results summarizing microscale behaviours are passed to macroscale models by an "upscaling" process. The macroscale model is evaluated and the results are processed to provide to the microscale model. In this approach, the models at each scale have to accurately produce the information to be passed to the other scale. We simulate the microscale model for multiple realizations in each cell and compute the expected value of population. We also use the expected value of limiting distribution of microscale model when the macroscale time step is sufficiently large.

Numerical experiments are used to decide the appropriate number of microscale simulations between two time steps at the macroscale. According to the multiscale strategy to link the two scales, we assume the dispersion process at the macroscale do not interact with population fluctuation at the microscale. Insufficient simulations at the microscale between two macroscale time steps can cause indistinguishable density updates at the macro scale, while excessive simulations at the microscale between two macroscale time steps can lead to significant population variation at the microscale and hence inadequate cell density update at the macroscale. Therefore, it is crucial to determine the appropriate microscale time steps between two neighbouring macroscale time steps to obtain reasonable patterns.

1.2.3 Numerical Investigations of Model Behaviour

The SCA model is highly non-linear and stochastic which complicates theoretical analysis, and techniques that apply to linear Markov chain models are not applicable. Therefore, we conduct numerical simulations to gain insight into model behaviour and qualitative features of the model. Though the fundamental components of the model are simple, the SCA model exhibit complex behaviour, The complexity arises from local interactions between simple components and the interaction rules which lead to interesting geometric dispersion patterns.

The simulations aim to explore the SCA model behaviour under various scenarios and parameter values. Through numerical simulations, we develop a through understanding of the dispersion process including hidden behaviours. We perform sensitivity analysis to identify the critical parameters that influence the behaviour of the model. In addition to presenting the dispersion patterns under various scenarios, we also derive some quantitative statistics of the patterns to quantify important features of the model, and identify what kind of model behaviour can be used to calibrate the model using experimental data.

The model performance depend on several factors including both the microscale model behaviour and the interactions of macroscale model parameters. Extensive simulation studies suggest that the macroscale model approaches three limiting situations when the microscale model generate enough individuals for the dispersion, fading away completely, settling into a fixed size, and growing without limit at a fixed speed. Nevertheless when the microscale model shows periodic behaviour, the macroscale model also shows periodic patterns. The experiments suggest that even though the underlying rules for the model are simple, the model can generate complex behaviour characteristic of dispersion processes found in nature.

The simulation studies focus on scenarios when the model produces final patterns growing without limit. We conduct 100 experiments and 50 realizations in each of the experiments to investigate the distribution of quantitative statistics including the average cell density, average number of affected cells, and average complexity of the interaction neighbourhood. We choose the appropriate grid size and time steps to run the model such that the model produces perceptible patterns that are sensitive to different parameter settings. Based on our exploration in model behaviour with different grid sizes and run time, we choose to simulate the model on a grid of 30×30 cells with 30 time steps to investigate the distributions of the quantitative measurements.

There are four phases in the simulation studies. Phase I is to conduct sensitivity analysis for each parameter in the model; Phase II is to investigate the interactions between different parameter regimes; Phase III is to investigate the impact of geographical properties of the domain, which is achieved by changing the attributes of certain cells to make them suitable for various physical conditions; Phase IV is to investigate the evolution patterns between interaction neighbourhoods.

There are seven parameters at the macroscale that determine the dispersion pattern. We focus on two probability parameters to characterize the stochastic behaviour of the model, the inclusion probability and the survival probability. Those two probability rules are connected with different stages of the dispersion process. The inclusion probability is linked to the emigration stage in which the potential cells for dispersion are selected, and the survival probability is linked to the immigration stage in which the individuals establish population in new cells. In the dispersion process, the amount of individuals provides a source of cell densities that determines the pattern. The cell density is jointly regulated by the SCA model at the macroscale and the birth-death process at the microscale. More details are provided in Chapter 5.

1.2.4 Model Calibration

The parameters in a physical process model of a complex system represent important physical characteristics of the system being modelled. Scientific inference using a model of a complex system involves evaluating the model for parameter values that are relevant

for the current physical state of the system. Unfortunately, it is often difficult to directly observe the parameters in a physical process model. Thus solving a form of inverse problem to determine parameter values from observed data is essential for using a model to forecast behaviour of the system being modelled. We consider two Bayesian approaches to this calibration problem, namely the Stochastic Inverse Problem (SIP) and Bayesian Inverse Problem (BIP) using both simulated data and real data consisting of graphical representations that indicate the progression of tree damage in a pine beetle infestation. We find that different quantities of interest (statistics) give distinct results for different parameters, which suggest an influence of the choice of appropriate quantities of interest. We provide more discussions about the calibration problem in Chapter 6.

1.3 Contribution of the Thesis

We have created an innovative multiscale SCA framekwork in modelling the dispersion of mountain pine beetles. We set up two scales in the model, the microscale model simulates the individual behavior to provide the source for the macroscale SCA model which simulates the dispersion of population density. We use sequential approach to couple the two scales to convert between the population size and population density. In fact, the multiscale framework provides the flexibility to be expanded to handle a wide range of scales depending on the resolution of dispersion patterns, and the sequential approach to couple different scales of the model allows us to use various microscale models.

The SCA model is a novel non-linear stochastic process model for the dispersion of population density. In the SCA, we use hexagonal cells to avoid the geometric sours of bias, and we use density as the cell state where the probability of dispering to a small region that depends on the local density in the surrounding neighbourhood. This is a stochastic version of game-of-life that acts on density rather than the states of 0/1. We introduce stochastic components to the transition rules in the SCA model and include a variety of factors that influence the dispersion, therefore the model is not only capable of capturing the uncertainties observed in dispersion, but reflects the beliefs of the known physics of the dispersion. We conduct a comprehensive investigation of the properties of the model that shows the temporal behaviors of the model. We show the model can be calibrated from observational data using a Bayesian approach, we carry out the calibration using a real image of the mountain pine beetle infestation pattern.

Chapter 2

Mountain Pine Beetle Description

In this chapter, we introduce the social and economic implications of mountain pine beetle (MPB) outbreaks, which motivates our work in modelling the population spread and predicting the future MPB epidemics. We also describe the biology of MPB and forest management plan, and summarize the available data for MPB research.

2.1 Introduction

The lodgepole pine forests provide an extensive range of advantages, such as habitats for wildlife, recreational areas, and raw materials for wood products. However, lodgepole pines are prone to natural disturbances - for example, the widespread MPB epidemics, which have killed millions of acres of forests in the Rocky Mountains in the United States and Canada. The mountain pine beetle, (MPB; *Dendroctonus ponderosae Hopkins*, Coleoptera: Curculionidae, Scolytinae), is an indigenous insect in lodgepole pine forests and is an integral component of the forest ecosystem throughout western North America. The beetle activities can either beneficial or detrimental depending on the forest management objectives [68]. During some years, the beetles cause little damage as they infest weakened, standing dead trees. But in other years, they attack living healthy lodgepole pines over extensive areas when their population periodically reaches the epidemic level. The MPB epidemics may stem from an alignment of factors, such as extensive forests of trees and climate change to support MPB reproduction [110], and human impacts, for instance, logging that lead to an imbalanced ecosystem.

Extensive areas of suitable hosts and climate changes are believed to foster the massive MPB epidemics. In the past, regular wildfires have limited the lodgepole pines at mature ages and ensured forest resilience from natural disturbances. With the efficient fire management program, fire suppression has been observed to make a remarkable change to forest composition; for example, forests dominated by large-diameter trees and dense stands with a high proportion of hosts [160]; and this enables the beetle attacks to be more effective. Climate changes, such as increased temperatures and drought conditions are also considered to

be major drivers of beetle epidemics. Increased temperatures support beetle growth and dispersion in the summer, and lead to high insect survival in the winter [74]. Besides, increased temperatures cause more precipitation falling as rain rather than snow, which consequently accelerate the decline of snowpack and drought conditions in Rocky Mountains. Decreasing surface waterflow has further exacerbated drought conditions, leaving more mature healthy trees water-stressed and reducing tree defences which makes the trees vulnerable to beetle attacks [203]. Therefore, successive years of favourable weather conditions combined with an abundance of vulnerable host trees that provide good habitat for MPB on the landscape have contributed to the unprecedented MPB epidemics occurred in the last decades.

The MPB epidemics have caused various ecological and economic consequences, and one of the most important concern is causing widespread tree mortality which could potentially affect appearance of forests and harm the aesthetic value of the area and tourism [3]. Besides, widespread beetle-killed trees could impact timer commodities, leaving the timber capacity unsustainable in the long run. Another frequently mentioned consequence of MPB epidemics is increase in wildfire threats and impairing postfire resilience [93]. However, existing research does not suggest a direct link between tree mortality due to MPB attacks and threat of wildfires in affected areas; since the MPB hosts - lodgepole pines - naturally burn in fires to kill mature large trees and produce favourable conditions for the seeds to germinate, which lead to a natural regeneration of the pine forests. There are other implications associated with beetle epidemics, such as altering forest structure and composition [97], modifying forest fuels [105], and threatening watersheds [156] and wildlife [161].

In general, MPB complete their life cycle in one year, and their life cycle mainly is composed of three stages: attacking new hosts, mating and brooding under the bark, emergence and dispersion to new hosts [157]. Successfully attacked trees usually display bole symptoms such as boring dust, small and numerous pitch tubes on the outer bark [68]. Later, foliage fades to yellowish green, to red and finally brown [164]. Infested trees initially appear in groups locally, and later develop into large infested patches that can be used for aerial assessment over large areas [164].

It is challenging to control and manage MPB epidemics. Direct control strategies that are used to decrease MPB populations and preventive methods that are used to reduce stand susceptibility are combined to form an efficient forest management plan [164]. Tracking the spatial and temporal patterns of MPB dynamics is crucial for assessing the risks and controlling the epidemics, nevertheless, how MPB populations spread into new territories and when the next epidemic occur can be incredibly challenging to predict based on current remotely sensed data since the analysis results are typically subject to accuracy assessment protocols [197]. Recently, unmanned aerial vehicles (UAVs) have become powerful tools for remote-sensing in forest management due to its ability of collecting data at low altitude and provide high quality data [169].

2.2 The Biology of Mountain Pine Beetle

MPB usually only infest a few weakened, stressed pines scattered throughout the forest. However, MPB population periodically erupt in epidemics and cause severe social and economic consequences. Thus, there are extensive studies to understand the biology of MPB and develop various models of their population dynamics.

2.2.1 Life Cycle

The MPB populations are normally univoltine (completing its life cycle in a single year), while their life cycle varies depending on the temperature. Occasionally, MPB may remerge and develop a second brood during warmer than usual summers [157]. MPB has four life stages: egg, larva, pupa and adult. All the life stages happen within the subcortical tissues of the host trees aside from the dispersion phase by mature adults [165]. Each year, young adult beetles leave the dead, yellow- to red- top trees and disperse to new hosts in the warm summer months. Upon successfully colonization, the female bores through the bark and commences gallery construction in the inner bark. The male enters the gallery soon after and mating occurs. Once mated, the female elongates the egg gallery vertically, small oval eggs are laid into the sides of the gallery and are covered with boring dust. Following egg hatch, larvae tunnel away from the egg galleries, feeding on the phloem tissue in the inner bark, cutting off nutrient translocation, thereby killing the host tree. MPB larvae require several months to develop to pupal stage, usually overwintering as larvae under the bark. Mature larvae cut chambers in the inner bark in the spring and transform into pupae in June and July, and the new adults emerge from mid-June through September. At first, new adults are pale and soft, and they may require one to two weeks to mature and harden before emergence. [157] [164] [13].

2.2.2 Dispersion Process

Most young adult MPB tend to disperse within short range throughout the forest for reproduction [165] [164]. In a release-recapture field study [165], the number of recaptured beetles was found to decline exponentially with respect to distance from the release point. While short-range dispersion may foster beetle population growth, there is evidence for beetles travelling long-distance flights to seek for suitable host trees. Recently, there are observations of MPB expanding their historical range and thriving in the north of British Columbia, where used to be considered as unfavourable for MPB [164]. Especially, with aid of convective winds, beetles could be carried to a few kilometres from the host where they emerge [163]. The long-distance dispersion could contribute to the MPB epidemic spread.

2.2.3 Host Selection

The initial selection scheme of a host tree is unclear to human observers. The selection could be spatially random and independent of population density and during initial selection, beetles may not respond to the mass-attack strategy [24]. Alternatively, MPB may prefer large-diameter trees with a thicker phloem to provide a good habitat [165]. Furthermore, pioneer females could adopt a combination strategy of random landing and attraction to subtle cues that is not apparent to human [138] [94].

2.2.4 Host Colonization

The colonization process of MPB is achieved in a complex scheme that is facilitated by pheromones (beetle-produced) and kairomones (host-produced) [19] [164] [151]. Pioneering females use kairomones to locate suitable hosts, and once they find the new host suitable for reproduction, they release the aggregation pheromones to attract more beetles to start the mass attack, after which females penetrate the bark and initiate gallery construction [151]. The mass attack is crucial as it enables beetles to overcome the tree defences.

2.2.5 Beetle Host Interactions

MPB normally prefer large-diameter trees with a thicker bark as they require a certain thickness of bark to excavate egg galleries. Furthermore, large trees can provide more resources and nutrients for beetle brood production. Therefore, small-diameter, young trees with thin phloem are seldomly attacked in practice. However, during the dispersion-colonization process, the trees are far from passive. Vigorous trees with large-diameter can often resist the attack and successfully defend themselves while they are the priority host selection for beetles. The trees release toxic resins as beetles bore through the bark [158] [14] which slows and even expels the attacking beetles. The phloem become saturated as the resin is released into the egg galleries, the beetles will be confined and killed in a lesion of dead tissues [164].

2.3 The Mountain Pine Beetle Epidemics

There are four major phases in the MPB population cycle: endemic, incipient-epidemic, epidemic and post-epidemic. The transition between phases are determined by beetle population size and spatial behavioural patterns. In endemic phases, the MPB populations have low number of individuals and they usually attack stressed host trees for their subsequent generation. As the populations develop to overcome the defences of vigorous and healthy trees, the population transitions to incipient-epidemic phase. Epidemics happen when there are large, contiguous areas of susceptible host trees and favourable weather conditions. The post-epidemic phase is characterized by the decline in MPB populations, which could

be a consequence of adverse weather or depletion of the host trees to maintain increasing populations at the landscape level [164] [151].

At the beginning of the incipient-epidemic phase, spread is slow and infestation is limited to single trees. As the population develop, infestation becomes patches with multiple trees, the MPB population enters epidemic phase. The epidemic stage could last several years with favourable weather conditions and sufficient susceptible host trees. There are three conceptual models of the host switching process of pine beetles during an epidemic outbreak [84]. The first model is termed as 'Passive Model' and it suggests when a tree is infested with pine beetles, it will provide signal for incoming beetles to switch host trees because of the lack of resin exudation [189]. The second model is termed as 'Threshold Model' because it assumes that a threshold level of trans-verbenol is required to simulate landing and attack on adjacent trees. When the pheromone concentration surrounding adjacent trees reaches a sufficient threshold, incoming beetles are thought to land on and infest a nearby tree to create a new host [84]. The third model is termed as the 'Repelling Model', and it proposes the switching of trans-verbenol and a 'repelling pheromone'. At the initial stage of the mass attack, large quantities of trans-verbenol are secreted attracting large number of beetles. However, at higher densities of infestation, the beetles emit a repelling pheromone that deters further infestation and therefore infest neighbouring trees [78]. In the collapse stage of an outbreak, i.e. the post-epidemic phase, the MPB population declines as the number of susceptible trees decreases and the beetles switch back to weaker and smaller trees.

Trees attacked by MPB display various symptoms such as external-bole signs and crown symptoms [164]. External-bole signs include pitch tubes on the trunk where beetle tunnelling begins, boring dust in bark crevices, and patches of bark removal by woodpeckers in search of larvae. Successfully attacked trees are usually killed and their crowns begin to fade from green to greenish-yellow. Later, the crowns fade to a uniform yellow, red and finally brown. First signs of fading usually occur in late spring of the year after attack. Most needles drop from trees after two years of the attack, leaving a dead grey snag. The attacked trees are also visible at the landscape level. During the onset of epidemics, infested trees appear in small groups. Then the small groups of infested trees will grow and coalesce to form larger patches. The colours of crowns are frequently used to assess the severity of epidemics over large areas by aerial surveys.

2.4 Social, Ecological and Economic Impacts of Mountain Pine Beetle Epidemics

As one of the most devastating forest insects in North America, MPB lead to various social, ecological and economic costs far beyond impacts on forest industry. Due to climate change, MPB have spread far outside its native range and the boreal forests are at risk to the

(a) (b)

Figure 2.1: Aerial view of trees infested by mountain pine beetles. (a) A small group of tree infestation in endemic stage. Photo is from [175]. (b) Large patches of infestation in the epidemic stage. Photo is from [164].

infestation of MPB. MPB affect recreation, property values, landscape aesthetics, wildlife habitat, biodiversity, watershed management and most importantly disrupt the long-term forest management plans [68].

2.4.1 Social Impacts

MPB have significant negative impacts on healthy forests, which many communities rely on for tourism, recreation, fishing and hunting. The dead trees, the discolouring of trunks and crowns have led to aesthetic decline of forests, which affects the attitudes and perceptions of tourism and outdoor recreation businesses [46]. Dead trees also increase deadwood fuel, leading to potential wildfire, and consequently poor air quality and the loss of infrastructure. Besides, due to the potential risk of falling, dead trees can present safety hazard to forest users and nearby communities, therefore make the forests unsuitable for tourism and recreational activities. Furthermore, the decrease in pine supply resulted from MPB epidemics could lead to mill closure and loss of forestry jobs.

2.4.2 Ecological Impacts

MPB have ravaged extensive areas throughout the forests with potential impacts on biochemical and biophysical processes for decades. One primary concern is the reduction in forest productivity. Tree mortality and defoliation caused by MPB could decrease the plant net carbon uptake, reduce carbon sink capacity of affected forests [111]. In addition, carbon released by wildfires resulted from massive dead trees to the atmosphere, could aggravates climate change, leading to further MPB epidemics. With MPB expanding its range into new territories, the boreal forest is at risk of being infested, which could alter the forest ecosystems and biodiversity [57], and reduce habitat for native wildlife.

Another major concern is about the regeneration of lodgepole pine forests following the MPB epidemics. The preference of MPB for large-diameter trees that produce most of the seeds needed for regeneration, allows new seedling establishment following MPB epidemics [112]. Due the release of resources, advance regeneration tend to form a new stand. However, the regeneration composition and abundance may have long-term effects on silvicultural treatments and ecological attributes such as hydrology and vegetation types [64].

2.4.3 Economic Impacts

There are several economic consequences of the extensive MPB infestation. The most obvious economic impacts are on timber supplies and residential property values. People who built homes in the forests face financial losses and safety threats when there is extensive dead trees caused by MPB epidemics [150]. Especially for those who try to sell their homes, few potential buyers are inclined to buy houses surrounded by dead trees, and thus the home prices are likely to be depressed. Besides, the dead trees increase the vulnerability of the homes to wildfires, which implies supplemental costs for homeowners.

MPB have direct impacts on the forest industry. Extensive infested forests have led to the excess supply of low quality and low priced timber and in order to maintain the market value of timber, the government has increased the annual allowable cut to salvage the dead timber, which results in a long-term shortage in timber supplies since it takes decades for the stand to regrow [53].

While MPB epidemics have increased the number of jobs needed to manage forests, they have also increased the costs for forest management such as costs of remote sensing and ground surveys for beetle monitoring and detection, the removal of hazard trees, and other tactics and supplies to control beetle populations.

2.5 Mountain Pine Beetle Epidemics Management and Prevention Plan

Generally, there are two types of strategies used to manage MPB epidemics. Direct control approaches are used to reduce the MPB populations, and preventive management approaches are used to increased stand vigour and reduce tree susceptibility to the MPB. Direct control aims at reducing MPB populations after the observation of tree mortality while preventive management is a proactive approach since it is done before the MPB epidemics exacerbate. In practice, the two approaches are combined to form an efficient forest management plan, and the adoption of direct control depends on an alignment of factors such as beetle population levels and resources available for implementation [164]. Both approaches can save huge losses when the MPB population is at an endemic level, but the management becomes more complicated when the population reaches epidemic level.

2.5.1 Direct Control

Direct control approaches are expensive, and there have been debates about their effectiveness in reducing losses [68]. However, in endemic stages, when the infestations are local, direct control approaches can be very effective to prohibit further development into epidemics. Treatments need to be prioritized on economic criteria, thus direct control is usually used to treat valuable stands. With timely detection and survey of the infested areas, once newly attacked trees are discovered, direct control such as sanitation cutting and controlled burning [165] should be implemented. And to maintain the control effectiveness, thorough treatment and follow-up surveillance [68] are needed. When MPB populations reach epidemic levels, direct control approaches are no longer feasible.

2.5.2 Preventive Management

Preventive management is highly effective when the MPB population is at an endemic level, a proactive thorough treatment of infested trees when beetle population is low can prevent the infestation developing into an epidemic. Once populations increase to epidemic level, the management of beetle populations becomes more complicated depending on real situations.

2.5.2.1 Hazard Rating in Stands

The main objectives of forest management are to predict which pine stands are most susceptible to MPB infestation, and how much loss will occur if the stands become infested. Forest managers can use a hazard and risk rating system rate the susceptibility of stands to MPB populations. Generally, hazard represents the possibility of an epidemic in a certain time period and is defined as a function of stand conditions. Risk represents expected loss and is determined by beetle populations [85].

There are several hazard rating systems that are developed based on influential factors on beetle survival and dispersion such as climatic and stand conditions. Stand susceptibility is determined by tree age, size, phloem thickness, and climatic conditions. Therefore, those factors are considered when assessing the risk in lodgepole pine stands and it is often more effective to incorporate data relative to beetle populations. The hazard rating system is the first step to determine the response for next stage.

2.5.2.2 Silviculture Methods to Reduce Risks in Stands

Silviculture methods that reduce infestation risks by clearcutting small- to moderate-sized blocks and creating a mosaic of age classes and tree species [85]. While there is no direct evidence that mixed stands resist MPB infestation better than pure stands, the age and size mosaics create a minimum area that is susceptible to the beetle, for example, having

young trees or mixing other species in high risk areas, which makes the forest resilient to MPB populations, and make the direct control action more feasible. This technique works best in even-aged lodgepole pine tree forests and the implementation requires extensive long-range planning.

Salvage and sanitation logging is favoured as salvage operations utilize timber and reduce food source for beetles [85]. Salvage cutting should be planned according to the dynamics of MPB, and for salvage cutting to be effective, timber should be harvested as quickly as possible after tree death and before beetle dispersion to prevent timber deterioration and protect adjacent trees.

In pure lodgepole pine stands, stocking control is an effective method to maintain tree vigour and moderate sized stands. By maintaining a well-spaced tree area relative to desired level of density, stocking control could favour trees that are not favourable to MPB, and hence increase stand susceptibility.

2.5.2.3 Semiochemicals and Baiting

Semiochemicals are natural pheromones produced by beetles and kairomones produced by hosts, which lead to mass attack in combination. Semiochemicals can be used to trap and remove beetle populations from infested stands [85]. Spot baiting is developed to remove small infestation patches and can be used in less than 30 trees. Forest managers choose two or three susceptible trees as bait before beetle dispersion, and remove the infested and dead trees after beetle flight. Pheromones and bait are effective short-term solutions to manage MPB populations until resources are available to apply other silviculture techniques.

2.5.2.4 Replacement Fires

When there is no commercial value in stands, but there is essential social and ecological function concerned, the fire management operations are more appropriate and ecologically desirable because the dead trees can cause potential wildfire, build up of fuel and plant succession. By utilizing prescribed fires, the forests are less susceptible to MPB since the beetle populations and their hosts are eliminated [165]. But this method is not favourable at landscape level especially when the forests are of great economic value.

2.5.2.5 Increase Sustainability and Forest Resiliency

To construct a more resilient and sustainable forest, forest managers need to develop a more climate adapted portfolio of species to replace the pure, even-aged pine stands in combination with proper silviculture techniques and detection methods [68]. The key to develop a resilient and sustainable forest is to increase forest species compositions, promote species diversity, manage water consumption [85], and use silviculture methods to prevent

beetle disturbances. A healthy vigorous forests that are well adapted to the climate can provide more economic values and ecologic benefits in the long run.

2.6 Detection and Monitoring of Mountain Pine Beetle on a Global Scale

Forest management decisions are determined by the location, size and potential consequences of beetle population [164]. And information regarding the location, size and consequences of beetle population is collected by various survey techniques, such as a tree-by-tree survey on the ground, aerial survey using airborne platforms or satellite-based sensors. The choice of survey method depends on the objective for a particular forest management. For example, investigation of beetle behavioural patterns and impacts require information collected at the landscape level. Local severity of MPB infestations are detected through systematic surveys conducted in a timely manner [164].

MPB detection is to identify and record locations of affected trees and potential locations of trees that are being attacked. Field crews use detection methods to conduct infestation assessments or to determine a mitigation plan. Mapping provides information regarding spatially estimates of the number of trees affected, or of volume affected for a management unit [164]. In order to control beetle populations, surveys need to locate the infestations as soon as possible. Monitoring helps forest managers to detect trends in beetle population dynamics and spatial pattern, and forecast potential risks resulted from infestation.

Aerial survey provide overall forest health status in broad-scale and gather information about red attack for strategic planning, which is usually the primary concern of the government. The forest health status help to identify areas that need further investigation, and the red attacked trees provide clues to potential green attack locations to facilitate forest management. Aerial sketch mapping is normally adopted by forestry services to review timber supply, manage resource and decide the annual allowable cut. The maps of stand infestations may also be used to form land use plans depending on management concerns. While in preparation of sanitation cutting, forest managers need detailed information about red attack locations and severity, and results from local area mapping of red attack are helpful to target the current green attacks.

Different management objectives require different survey techniques that provide appropriate level of details and extent. Aerial surveys are used to capture overall infestation extent and intensity; ground surveys are used for local infestation investigations such as confirmation of insect species, evaluation of dead timber and current green attacks; infestation trend assessment is used to determine the trend of affected forest by examining brood status and counting the number of infested trees [164].

2.6.1 Aerial Survey

Aerial surveys detect red attacked trees by observers in fixed wing or rotary aircraft, and good visibility is required to undertake an aerial survey [164]. Up-to-date aerial photos and custom drawn GIS maps help to orient the observers and improve the accuracy of insect polygon placement. Aerial photos show clear details about logging and burns and help observers to distinguish from infested stands, and custom drawn GIS maps highlight fixed geographical features such as roads, water bodies and other landmarks.

All aerial surveys note the location and identification of insects and its infestation severity, but the notes may differ depending on the agency. Multiple maps are combined to form digital infestations, but the information gathered from the air needs to be combined with results obtained from ground surveys previous to and after sketch mapping since correct identification of tree and insect species, and attack category are difficult to decide from the air. Observer knowledge of the local forest and insects is also important to improve the accuracy of the mapping. Aerial surveys results need to be combined with ground surveys assessments to estimate the severity of beetle infestation since it is impossible to determine the exact number of infested trees from the air [164]. Consequently, information collected by aerial survey maps need to be cross-validated by aerial sketch mapping, global positioning system (GPS) point data, air photo interpretation and other ground plot data [164].

(a) (b)

Figure 2.2: Aerial view of trees infested by mountain pine beetles. (a)A local scale tree infestation with high resolution. Photo is from [139]. (b) A large scale tree infestation with low resolution. Photo is from [43].

2.6.1.1 Sketch Mapping

The most general approach to detect red attacked trees is to sketch map from a fixed wing aircraft, sketch maps can provide quick, quality information during epidemics. Observers make notations on topographic maps at scales from 1:100000 to 1:250000 over millions of hectares and check flights that repeat over an area are used to check consistency between observers. With consistent mapping, cumulative mortality in a specific stand can be es-

timated by using mapping of successive years of infestation taking the scales of mapping into consideration [164]. Sketch mapping is used to characterize the disturbance patterns over large areas, usually at the regional level, but there could be accuracy issues regarding boundaries when incorporating with data represents small areas.

2.6.1.2 Global Positioning Systems

Once sketch maps have been obtained, infested landscapes undergo more detailed aerial surveys conducted from a helicopter. Red trees are visually detected, their locations are recorded with a GPS, and noted on topographic maps of 1:20 000 to 1:50 000 scale. The helicopter pilot hovers above the centre of a group of attacked trees, and another person will capture the GPS waypoint for the site. An estimate of the number of infested trees at that location and the type of insect is also noted. The purpose of the GPS survey is to accurately locate the beetle impacts to aid in rendering local or regional strategic decisions.

The likelihood of observers logging a non-existent red attack location is extremely low. However, not detecting red attack areas on the landscape depends on the survey effort covering an area. As mentioned above, knowledge of the population size is also required support information for helicopter GPS surveys. The density of the affected trees at a given point is also an issue. For a given survey point the trees identified may be dispersed or clustered, yet this is not captured in the survey. The use of GPS carries the errors associated with that technology and additional positional errors that are a function of the viewing platform.

2.6.1.3 Air Photo Interpretation

Aerial photography is the most common imagery used in a forest inventory to characterize forests and meet management objectives. Surveys that utilize aerial photography can be grouped into classes based upon the type of information collected: damage detection, damage location, damage amount and estimation of the relative size of the insect population and the capacity for future damage.

Aerial photography is not as well suited to initial damage detection as visual aerial survey methods, such as sketch mapping; information regarding insect populations and potential for causing future damage is best collected through field survey. However, aerial photography can address the information need for generating mortality estimates and precise locating of infested areas [164]. The interpreter's experience may affect the interpretation results.

2.6.2 Ground Survey

Ground surveys assess the population size, or the degree of forest infestation, within a local area. Sample plots are generally less than 1 ha. Population estimates indicate whether a local beetle population is increasing, static, or decreasing. Infestation estimates indicate

the impact of a particular beetle population. Both types of techniques are used to drive the selection of the most appropriate management response. Population assessments may be based upon field surveys or aerial surveys. Field surveys enable a brood assessment to be undertaken. Brood assessments are carried out in the late summer to fall and in the spring. Beginning in mid-July, population surveys with sketch mapping may be undertaken. These aerial surveys influence the placement of subsequent ground surveys.

Walkthroughs are used largely as an initial ground reconnaissance survey for determining the characteristics of attacked stands and to contribute toward determining information needs for more intensive surveys. Probes are systematic strip surveys that collect more detailed information than the walkthrough survey. Probe information is compiled on a polygon basis and includes attributes such as: location on map; size of beetles under bark; relative brood success; percentage of attack category; rate of spread; and stems per hectare. While useful, these survey techniques do not provide sufficient information for assessment of volume or area infested. Prism cruises, on the other hand, are used for detection and impact assessment, where the volume affected can be estimated on a stand basis. Line transects are also used for detection and impact assessment, and are more efficient than prism cruising. With these data, affected volume and area can be estimated from the survey, and can potentially be statistically extrapolated to represent larger areas. An additional means to characterize the population trend of a mountain pine beetle infestation is by calculating a green-to-red ratio. A green-to-red ratio is the estimated number of currently attacked trees compared to the number of red attacked trees. This ratio gives a rough indication of the population growth [164].

2.6.3 Digital Remote Sensing

Digital remote sensing uses sensors (mounted on either airborne or satellite platforms) that collect digital imagery of various spatial and spectral resolutions. One advantage to using digital data in mapping red attack stage trees is that they portray continuous data across the landscape. In this way, all areas in the image are examined for possible red attack, independent of accessibility or position in a watershed. Another advantage of mapping from digital sensing is the reduction or elimination of interpreter bias afforded by automated classification algorithms. The products have greater consistency and reliability between different areas or dates by avoiding visual interpretation, which increase reliability of results from the high positional accuracy of image data compared to aerial survey data. The results of analysis of remotely sensed data are typically subjected to accuracy assessment protocols. This is a unique element of remote sensing analyses in contrast to the more heuristic assessments of the aerial survey products. The accuracy of an attribute, such as red attack, and infested areas that are missed, may be characterized in relation to an independent validation dataset to improve characterization accuracy.

The key to employing digital data for mapping mountain pine beetle impacts is to match the information needs of managers with the image information content and resolution characteristics. For example, under endemic conditions, the information needs are for detection of single and small clusters of red attack trees. To produce this information, the imagery must have sufficiently high spatial and spectral resolution. In contrast, under epidemic conditions, the information needs are for quantifying the impact of large groups of red attack trees over large areas. Therefore, less expensive imagery with medium spatial resolution and moderate spectral resolution would be sufficient.

2.6.3.1 Airborne Platforms

Digital images may be collected from airborne platforms over areas identified as infested during aerial surveys. The key differences between airborne images and aerial or ground surveys are that the spectral characteristics of the entire forest are captured and the data can be re-examined if uncertainties occur. Airborne imagery includes traditional air photos that are scanned into a digital format, digital camera images, videography, multispectral scanners, and imaging spectrometers. Airborne images may be used to map the location of small clusters or scattered red trees. The results are used to direct ground surveys or to dispatch ground crews for sanitation treatment. The airborne digital imagery may be subjected to enhancements that highlight the locations of red attacked trees.

Aerial videography provides some operational advantages over air photos including: lower cost, no delay for photographic development, option for including audio commentary, and high light sensitivity. Additionally, camera settings can be adjusted during data collection in response to viewing the imagery as it is acquired. Similar to most airborne sensors, the disadvantages are primarily image extent and resolution characteristics. Otherwise, similar planning and processing options are available as they are with the digital camera systems described above.

2.6.3.2 Satellite Platforms

Satellite images may be collected to map infested areas over a range of scales. Satellite imagery is similar to airborne imagery in that the data is continuous across the extent of the sampled area. In general, the comparatively high orbits of satellite systems result in more favourable viewing geometry when compared to those of airborne systems. Airborne systems often generate data that requires sophisticated processing to compensate for aircraft motion, view angles, and variable illumination conditions over the acquisition period. Satellite images are available over a range of spatial, spectral, and temporal scales. Therefore, they can be used to address a variety of strategic and tactical planning decisions. The large image extents of satellite imagery enable economies of scale.

2.6.3.3 Unmanned Aerial Vehicles

Unmanned aerial vehicle (UAV), or simply drone, has emerged over the past few decades. Flexible, low-cost, and high-resolution remote sensing systems that use drones as platforms are important to fill data gaps and supplement the capabilities of aircraft and satellite remote sensing systems. The advantages of drone remote sensing include low material and operational costs, flexible control of spatial and temporal resolution, high-intensity data collection, and the absence of risk to crews. The current forestry applications of drone remote sensing are still at an experimental stage, but they are expected to expand rapidly. To better guide the development of drone remote sensing for sustainable forestry, it is important to systematically and continuously conduct comparative studies to determine the appropriate drone remote sensing technologies for various forest conditions and/or forestry applications [181].

2.7 Existing Models of Mountain Pine Beetle

Ecological entities are usually so complex that it is impossible to choose more than a fraction of all factors to study, and the success of the models depend on whether those factors included are really essential to explanation. The factors chosen should be able to improve understanding of the relations under study and provide leads for subsequent investigations. The study of only a few factors for a short period of time can lead to misinterpretations. Conceptual or theoretical models give clues to modes of behavioural interactions and may point toward important population characteristics. Empirical models give an accounting of these facets. In either case, if the model is to be used as a measurement tool, the nature or mode of operation of the ecological unit must be thoroughly understood to correctly interpret these components in terms of population dynamics.

Both conceptual and predictive models are found in MPB literature. Models chosen to be discussed in this section have to do with interactions of trees and beetle populations and are presented in an order that is consistent with the life cycle of the beetle.

2.7.1 Beetle Life Table Models

The life table is one of the oldest, most useful and best known expressions in the field of population studies. A life table is a concise summary of certain vital statistics of a population. The life table approach was used to determine the principal factors affecting natality and mortality of MPB. The life tables were based on intensive population sampling from annually established plots within both high and low level populations of the beetle. These data have been supplemented with detailed studies in both field and laboratory. Life tables can be used to demonstrate population trend.

2.7.2 Beetle Life Stage Models

The beetle life stage models are used to describe smoothed beetle density trends over years by life stage and size of infested tree using data obtained by sampling for life tables. The data include survival of various beetle stages with respect to lodgepole pine size, and plot brood densities for each life stage over the infested tree and observed year of the infestation. The model is derived by fitting through the plots using least deviations approach.

2.7.3 Beetle Dispersion and Aggregation Models

The beetle dispersion and aggregation models are developed based on several assumptions: 1. Pioneer beetles attack randomly over the available trees; 2. There is a threshold of beetle population to induce aggregation on a tree; 3. Any tree that beetle aggregate on will be mass attacked and killed. Usually, the surface area of a tree is measured in equivalent units. Then the pioneer beetle attack density per unit could be estimated using the Poisson distribution function, and the probability of having one or more pioneer beetles attack could be calculated. The threshold of aggregation was estimated based on the number of square units of the tree that must be attacked by one or more pioneer beetles to induce aggregation. Thus, the probability of the tree becoming an aggregator could be calculated [94] [163] [68].

The mortality prediction model assumed that the threshold of aggregation is uniform across both diameter and years within an epidemic. Thus, surface killed was estimated for the stand. However, for each year, the predicted surface area mortality was forced to be the same as that observed by adjusting the pioneer beetle attack density. Models show that in the early part of the epidemic the pioneer density is low and the comparative probability of large tree mortality versus small tree mortality is large. As the epidemic progresses, pioneer density rises and smaller trees are attacked because of this and the depletion of larger trees. Toward the end of the epidemic, pioneer beetle density decreases because emergence from infested trees declines and, as a result, the epidemic collapses. However, data taken from many epidemics show that gallery starts (the positive results of attack densities) are greater in the early stages of an epidemic, decreasing in the midyears, and increasing again in the latter years.

Other efforts have developed a mathematical model to both describe beetle aggregation and predict the relation of tree susceptibility and switching to changes in beetle density. The term “switching” is defined as the phenomenon of incoming beetles attacking an adjacent recipient tree after the first tree had been attacked by the pioneering beetles. This recipient tree is quickly mass attacked and killed more readily than the focus tree. When switching occurs, the beetles usually attack trees with a diameter greater than the average in the stand. This then enhances beetle survival because larger trees usually have thick phloem.

2.7.4 Stand-Beetle Models

Interactions of the beetle with the infested trees, and tree losses to the beetles within stands, follow a predictable pattern. A large proportion of large-diameter trees provides the threshold food supply necessary for an epidemic because of the thick phloem found in such trees. The epidemic potential exists primarily under optimal temperatures for beetle development. Losses of lodgepole pine over the main epidemic years are proportionately much greater in the large diameter classes. Most tree losses occurred during a 6-year period. Cumulative losses show that most large diameter trees were killed during the infestation. Losses shown are typical for stands of trees of similar size and distribution at similar elevations and latitudes.

2.7.5 Rate of Tree Loss Model

The rate of loss model refines existing risk-rating systems and provides a method for stratifying lodgepole pine stands and predicting tree and volume loss for habitat types. The model is provided to assist land managers in projecting tree mortality over time, and can be linked to the model for use in forest planning. Most models for epidemic processes have dealt with the continuous-infection model for treating epidemic processes in a fully stochastic manner, and most of these processes dealt with diseases. Continuous infection assumes that an individual host tree can be infectious from the moment it receives the infection (i.e. the beetle) until it dies, recovers, or is removed. This clearly is not the case with the MPB. The beetles have a discrete generation and discrete stages of growth, and its epidemic behaviour does not fit the continuous-infection assumptions.

There is an alternative to the continuous-infection assumption, and the model was postulated that the period of infectiousness is comparatively short. This alternative assumption best fits the epidemic behaviour of the mountain pine beetle and amount of tree loss. The period of attacking a tree is fairly short, a beetle epidemic can begin with one or several simultaneously infested trees within a stand. The infestation then will spread in a series of stages by each new generation of adult beetles attacking living green trees. The probability of a tree becoming infested at any generation depends upon the numbers of infested trees and susceptible green trees during the preceding generation of beetles.

Chapter 3

Multiscale Modelling Framework

In this chapter, we describe the construction of a multiscale model based on an extension of the concept of cellular automata (CA) that uses probabilistic rules to govern dispersion process while a microscale physical process drives the evolution. The complex interrelated processes at microscale and macroscale are connected via a hierarchical multiscale approach, and structural mechanisms are constructed to drive phenomena at the macroscale using collective behaviour at the microscale. In this coupled multiscale approach, computations at different scales are linked, in which the simulation schemes at the macroscale employ the detailed information obtained from microscale method and macroscale information initiates the microscale model. Motivated by an application to the Mountain Pine Beetle infestation, we focus on applications to dispersion processes in ecology and epidemiology, but the model framework is of general interest and can be applied in other fields, such as void dynamics in irradiated materials. We use a general birth-death process to represent individual behaviours at the microscale, and leverage the stochastic CA model to address the dispersion patterns at the macroscale. The model easily accommodates other microscale models.

3.1 Introduction

Dispersion, which is the movement of individual objects from one area to another, is a fundamental process that typically operates at multiple temporal and spatial scales. The definition of dispersion may vary within different contexts, as well as the scale considered. The study of dispersion is ubiquitous in numerous fields such as biology, microbiology, ecology, mathematics, physics, epidemiology, engineering and geography. For example, dispersion has profound implications for both the reproduction of individuals and the structure and dynamics of populations [142]. Modelling dispersion is a challenging task because the mechanism of dispersion is often unknown or difficult to describe in nature and the evolution of dispersion patterns for large-scale populations emerge from local interactions of individual objects.

In population dynamics, a process may be viewed as a three-stage system: emigration, transfer and immigration [130, 155]. Emigration is the stage when individuals leave their source area. The transfer stage refers to exploratory phase where individuals explore and travel through their surrounding areas. This stage can involve a high risk of mortality associated with natural forces. The immigration stage refers to entering a new area or joining a neighbouring group. In this case, there is a risk of mortality associated with individuals adapting to the new environment and establishment of population in the new areas.

A fundamental goal in the study of dispersion is to describe the patterns that it generates. Those patterns describe how individuals are distributed in space, and how they interact with each other in the population and their surrounding environment over time. For example, researchers study and measure the dispersion patterns in a certain species to make general spatial recommendations for conservation [183]. In nature, individuals tend to be influenced by the presence of their surrounding neighbours during dispersion, for example, motile animals tend to aggregate to create a favourable environment for reproduction and survival purposes. Therefore, population density, which is the number of individuals per unit area or volume, driven by individual motions such as aggregation and repulsion, plays a crucial role in determining the evolution of dispersion patterns. Dispersion processes often display multiscale characteristics, in which population is generated on a fast temporal and short spatial scale while dispersion takes place over a larger scale at a slower rate. For this reason, we aim to use a multiscale approach to explore the complex population dynamics generated from dispersion process at the community level while explicitly modelling individual behaviour.

There are two scales in our modelling framework, which we call microscale and macroscale. At the macroscale, dispersion is considered as a dynamic force for the movements of spatially distributed communities described in terms of population density. It is modelled by a stochastic process that takes place on a grid of polygons. The dynamics are covered by probabilistic rules for mortality, reproduction, immigration related to distance, and chance of successful colonization. The parameters in the stochastic process model simple to adjust and combinations of assumptions can be handled. Because of the inherent uncertainty in dispersion processes, we use a model that includes stochasticity in the dynamics. At the microscale, we use a model that produces detailed information about individuals that has an impact on dispersion patterns. For our applications, we choose the birth-death process to capture the individual behaviours locally such as reproduction, mortality, or competition for resources. The microscale process describes the input properties of the macroscale subjects. The model easily accommodates other microscale models.

The model characterizes the three stages of dispersion process that account for local individual behaviours and spatial heterogeneity. The model is built upon four fundamental assumptions about the dispersion process: 1. We assume that the model focuses on simulat-

ing the dispersion process that are driven by several mechanisms including active dispersion, an individual controls its own locomotion and passive dispersion, and the individual movement depends on external forces such as gravity and wind [130]. 2. We assume that the model operates within a reasonable expanse of magnitude. Physical scale is defined by computational capacity. Currently, computer simulations are too computational intensive if the model is performed over a large landscape area, for example, thousands of kilometres. 3. We assume that the geographic features of the grid are static as the model simulates evolution of patterns over time. 4. The model applies to the situation in which there are enough individuals to produce a viable density. We use a sequential approach to connect the two scales such that the microscale can provide subject inputs for the macroscale and the macroscale in turn feed back into the individual behaviours at the microscale.

3.2 Multiscale Strategy for Bridging Scales

The ultimate goal of multiscale modelling is to simulate the population dispersion patterns at larger scales by leveraging information obtained from smaller scales. Multiscale modelling takes advantage of combining models specialized to different length and time scales to obtain a model for describing the population behaviours over large regions. From a computational standpoint, combining models at different scales is a demanding task, and this challenge is further compounded by finding a well-founded strategy to bridge the scales to balance accuracy and efficiency.

We use a sequential multiscale approach. In sequential approaches, information is obtained at the microscale, and the results summarizing microscale behaviours are passed to macroscale models by an "upscaling" process. The macroscale model is evaluated and the results are processed to provide to the microscale model. In this approach, the models at each scale have to accurately produce the information to be passed to the other scale.

Both the micro and macroscale models are stochastic. We simulate the microscale model for multiple realizations in each cell and compute the expected value of population. We also use the expected value of limiting distribution of microscale model when the macroscale time step is sufficiently large. For example, in the application to MPB infestation, we consider a population model at the microscale and a population density model at the macroscale. When modelling the dispersion patterns generated by MPB epidemics, each cell represents a small patch of trees with similar levels of infestation. We repeat the birth-death process with the same parameter value settings for multiple realizations and compute the average number of individuals obtained from those realizations, which is used as expected population size in each cell, i.e., the expected number of beetles in each patch of trees. To use a birth-death process, we determine the carrying capacity of each cell, i.e., the maximum number of beetles that could survive in each patch of trees, from relevant literature or using the field experiment data. We convert the population size obtained from microscale model into a cell

density by using the ratio between population size to the carrying capacity, which is used as the initial cell configuration for macroscale model. After simulating the dispersion process at the macroscale, we acquire an updated cell density and convert it back to population size in each cell. The updated population size is used as the new starting value for microscale models to achieve the cell density for the next time step at macroscale. We alternate the microscale and macroscale model for a sufficient number of time steps to investigate the different patterns formed at the macroscale.

Numerical experiments are used to decide the appropriate number of microscale simulations between two time steps at the macroscale. According to the multiscale strategy to link the two scales, we assume the dispersion process at the macroscale do not interact with population fluctuation at the microscale. Insufficient simulations at the microscale between two macroscale time steps can cause indistinguishable density updates at the macro scale, while excessive simulations at the microscale between two macroscale time steps can lead to significant population variation at the microscale and hence inadequate cell density update at the macroscale. Therefore, we need to determine the appropriate microscale time steps between two neighbouring macroscale time steps to obtain reasonable patterns. We provide more details below.

3.3 Modelling Schemes at Different Scales

We employ our macroscale dispersion model on a grid structure. The cell states are represented by population density, and the state transitions depend on state of single cells and other cells in a local region where those cells interact with each other and a set of rules. We illustrate properties of the model with simple examples in this section. The macroscale model can accommodate any population model at the microscale. We choose the birth-death process to describe individual behaviours at the microscale because it is a widely used model for population dynamics with a banded maximum population.

3.3.1 Macroscale Process and Model

We assume the model for the microscale system provides a detailed picture of the dynamic evolution of individual organisms. On the other hand, the macroscale model aims to characterize the interactions between different regions that include a collection of communities. It is challenging and expensive to coordinate interactions between the microscale behaviours in macroscale models.

At the macroscale, population behaviours such as colonization, migration and formation, emerge from the interactions of individuals in combination with local physical properties, which typically display complex spatial patterns, and therefore requires a model that can generate complex and realistic configurations that match the spatial patterns.

We develop an analogue to a CA model called a stochastic cellular automaton (SCA). SCA are rule-based discrete dynamical systems that are comprised of a grid cells. Each cell has a state from a finite set, and the cells synchronously change their state according to a global update rule that determines the new state based on the old states of the cell and its immediate neighbours. All cells use the same update rule, and the update is repeated over and over at discrete time steps. An SCA approach is suitable for modelling dispersion process and can provide critical insights to investigate the dynamic behaviours of the dispersive system. The SCA is a variation of a Markov process, with a transition matrix that depends on the state properties at each time step, rendering the transition probability matrix nonautonomous. The properties depend on the state, so it is also a nonlinear process. The model uses relatively few parameters but exhibits many kinds of behaviours, from simple predictable behaviours to complex patterns. The nonlinear nature presents serious challenges for theoretical analysis.

CA have been studied extensively since they were first invented in the late 1940s by John von Neumann to build self-duplicating machines given proper instructions [83]. CA were demonstrated to model the complex patterns of natural phenomena with simple components [193] and later the general theoretical foundations for CA were constructed [194], CA were defined as discrete dynamic systems such that global changes in space and time could be generated by local interactions among simple components. CA are famous for their conceptual advances and computational power. Conceptual advances have led to the flexibility of CA such that CA can easily adapt to different real-world situations with special physical properties when the update rule is chosen properly, while computational power has made it possible for CA to perform complicated arbitrary computation tasks. The best-known example is John Conway's Game-of-Life, which resembles real-life processes given its analogies with alterations of a society of living organisms [82]. CA can be mainly classified into four categories based on their behaviours: CA that generate homogeneous stable patterns, CA that evolve into oscillating structures, CA that develop chaotic fashion over long times, and CA display complex patterns. The advantage of computational universality, homogeneity and massive parallelism makes CA very popular in biological and epidemiological field, such as interspecific competition for space [173], forest fire spreading [108], epidemic propagation [174], growth and inhibition processes [177], debris flows [60], and spatial dynamics of corals [120].

The SCA is well suited for modelling dispersion processes. We modify the original structure of CA in order to obtain a model that allows for more complexity. The motivation for developing such a model is that various external forces affect dispersion in a manner that can be described stochastically. The result is a simple, flexible, intuitive and transparent modelling framework to capture the complex patterns resulted from dispersion processes on the plane. The model can be integrated with other models and provides a linking bridge

between different scales to generate similar geometries that are characteristic of real-world phenomena. The model specifications are as follows:

(1) **Domain**: We use hexagonal cells that are less affected by geometric sources of bias to obtain a homogeneous pattern for evolution of dispersion.

(2) **Cell states**: Unlike conventional CA cell states that take binary values 1/0 or live/death, we let cell states take quantitative values that represent population density, therefore we have a continuum of values in $[0, 1]$ in each cell.

(3) **Non-homogeneous rules**: Conventional CA have homogeneous rules for dynamics across the domain. In the SCA, we allow the attributes of cells to vary across the domain. We present the design and implementation of this idea in model simulation section.

(4) **Transition rules**: We consider a variety of factors that influence cell state transitions. We design the transition rules by introducing stochastic components to model the uncertainties associated with ecological processes and including the weights and difference between population density of neighbouring cells to reflect dispersion theories.

(5) **Multiscale model**: we use a microscale population model to drive the macroscale model.

The model can be formally defined as follows.

3.3.1.1 Domain

We describe dispersion process on a tessellation of the domain using hexagonal cells. This yields some practical benefits: 1. The shape of a hexagonal grid gives advantages for representing curves of geographic features, such as rivers and roads more naturally than squares; 2. The distance between centroids is the same in all six directions with hexagons; 3. Finding neighbours of each cell is more straightforward with a hexagonal grid; 4. Hexagons have a competitive edge in determining the transition rules in cellular automata as all neighbouring hexagons have the same spatial relationship with the central hexagon; 5. A hexagonal grid suffers less distortion due to the curvature of the earth than square grids; 6. Compared to square grids that restrict movement to four directions, hexagonal grids let movement occur in six directions.

The model depends on identifying neighbourhoods of cells. We choose the Moore neighbourhood which means that each cell is surrounded by six neighbouring cells. Since we build the SCA model on a 2-dimensional space, we let Z^2 denote the two-dimensional lattice field which comprises a disjoint collection of $m \times n$ uniformly sized cells organized in m rows and n columns.

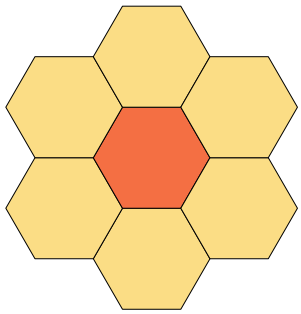


Figure 3.1: An example of a Moore neighbourhood. The central dark shaded cell is an affected cell, the six light shaded cells are its neighbouring cells, the population dispersion can occur from the central cell to the neighbouring cells.

3.3.1.2 Cell States

Cell states are represented by population densities that take values in an interval $\Pi = [0, 1]$. The configuration space is $Q = \Pi^{Z^2}$, then at any given time, the automaton configuration that specifies the states of cells is a mapping

$$c : Z^2 \rightarrow \Pi^{Z^2}$$

3.3.1.3 Interaction Neighbourhood

In the model, a cell only potentially interacts with its neighbours in Moore neighbourhood. We group all cells that potentially interact in one "interaction neighbourhood", which comprises a group of affected cells and their neighbouring cells. An interaction neighbourhood contains directly adjacent affected cells and potentially affected cells in regions that represent the spatial distance that the phenomena can spread in one time step during the dispersion process. The modelling framework is built upon observations of population movements, such as MPB infestations in lodgepole pines and radiation damage in nuclear fuel rods, which prefers short-distance dispersion because there is often a high cost associated with the dispersion process on longer scales [155]. Therefore, we consider directly adjacent neighbours when setting up the interaction neighbourhood in the model, with each affected cell having six neighbours in a Moore neighbourhood with the specific affected cell in the centre. A group of affected cells that are not separated by more than two nonaffected (zero density) cells along with all immediate neighbours formulate an interaction neighbourhood.

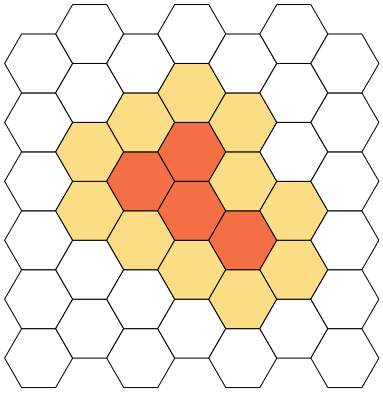


Figure 3.2: An example of an interaction neighbourhood. The dark shaded cells (affected cells) and their adjacent light shaded cells form an interaction neighbourhood in a portion of a 2D hexagonal cellular space, the population dispersion can only occur within each interaction neighbourhood.

3.3.1.4 Transition Rule

Transition rules specify how cells change their states over time. The SCA is simulated at discrete time steps, one step comprises an application of the rule to all cells in the grid. The transition rule is only applied once after all the cell states have been updated synchronously, the next state of each cell depends on current state of itself and its neighbouring cells according to the local transition rule F. For example, depending on whether the column j is even or odd, the state of the cell $(i, j) \in Z^2$ at $t + 1$ is:

If j is an even positive integer, then we have

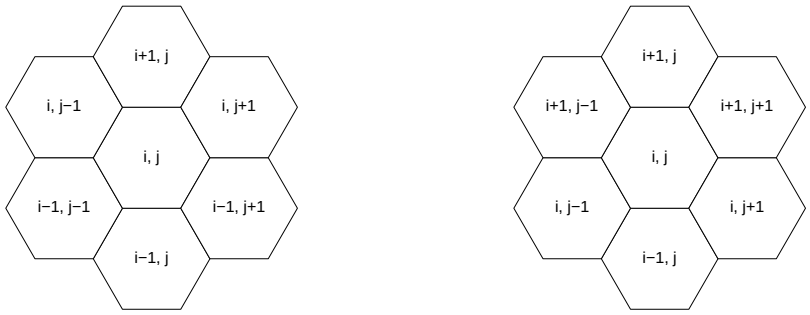
$$\pi_{ij}^{(t+1)} = F\left(\pi_{ij}^{(t)}, \pi_{i-1,j}^{(t)}, \pi_{i,j-1}^{(t)}, \pi_{i,j+1}^{(t)}, \pi_{i+1,j-1}^{(t)}, \pi_{i+1,j}^{(t)}, \pi_{i+1,j+1}^{(t)}\right) \tag{3.3.1}$$

If j is an odd positive integer, then we have

$$\pi_{ij}^{(t+1)} = F\left(\pi_{ij}^{(t)}, \pi_{i-1,j-1}^{(t)}, \pi_{i-1,j}^{(t)}, \pi_{i-1,j+1}^{(t)}, \pi_{i,j-1}^{(t)}, \pi_{i,j+1}^{(t)}, \pi_{i+1,j}^{(t)}\right) \tag{3.3.2}$$

We design the transition rules in a way such that they can provide flexibility for adapting to different situations. For example, the transition rules can simulate both positive density-dependent dispersion and negative density-dependent dispersion. In this modelling framework, we focus on simulating positive density-dependent dispersion situations in which increased population density drives individual movements. The transition rules are simple while producing a model that can represent complex relationships.

In real life, dispersion is nondirectional. In this section, we design the transition rule to simulate the nondirectional movement such that individuals can move in any direction



(a) (b)

Figure 3.3: The formal indexing for two configurations of the hexagonal cell neighbourhoods. (a) The indexing for adjacent cells when the column index j is odd. (b) The indexing for adjacent cells when the column index j is even.

without a specific destination in mind. However, we also recognize that there are influential factors, such as winds and geography for MPB, that can affect the dispersion direction. We illustrate in the simulation chapter reference. In the MPB application, the transition rule mainly explains the exploratory and immigrating stage of dispersion where individuals seek for new habitats and establish populations in the new habitats.

We now detail the model.

a. Inclusion probability

To appropriately address the random movement in dispersion, we use a probabilistic rule to characterize one stochastic feature of the model: at each time step, the outcome of local cell transitions is specified by a probability of updating to a new state. Specifically, the probabilistic rule is the probability of including a cell (i, j) in the interaction process at a given time step t. This probability is called the *inclusion probability*, denoted by $p_{I,ij}^{(t)}$, determined as

$$p_{I,ij}^{(t)} = \left[\frac{\left(w_{ij}^{(t)}\right)^{\alpha_{ij}^{(t)}}}{1 + \left(w_{ij}^{(t)}\right)^{\alpha_{ij}^{(t)}}} d_{ij}^{(t)} \right]^{\beta_{ij}^{(t)}} \tag{3.3.3}$$

where $w_{ij}^{(t)}$ is defined as the inclusion weight for cells and $d_{ij}^{(t)}$ is the maximum of relative density difference for neighbouring cells.

Inclusion weight $w_{ij}^{(t)}$. The model for dispersion is governed by population density between different areas. As a consequence, we assume that affected cells and non-affected cells that are surrounded by more affected cells are more likely to be included in the cell

interactions within each interaction neighbourhood. We define a dynamic inclusion weight parameter, denoted by $w_{ij}^{(t)}$, to determine the likelihood of cell interaction inclusion for each cell (i, j) at time t. For non-affected cells, the inclusion weight $w_{ij}^{(t)}$ is calculated as the number of adjacent affected neighbours for non-affected cells at each time step. For affected cells, as they are more likely to be involved in dispersion process, and individuals tend to leave high density areas and join their surrounding low density groups, so we add an extra weight w^* to the inclusion weight of affected cells to distinguish the likelihood of affected cells involved in dispersion processes. The inclusion weight $w_{ij}^{(t)}$ for affected cells is the number of its adjacent affected neighbours plus the extra weight w^*. In our model, we set $w^* = 1$.

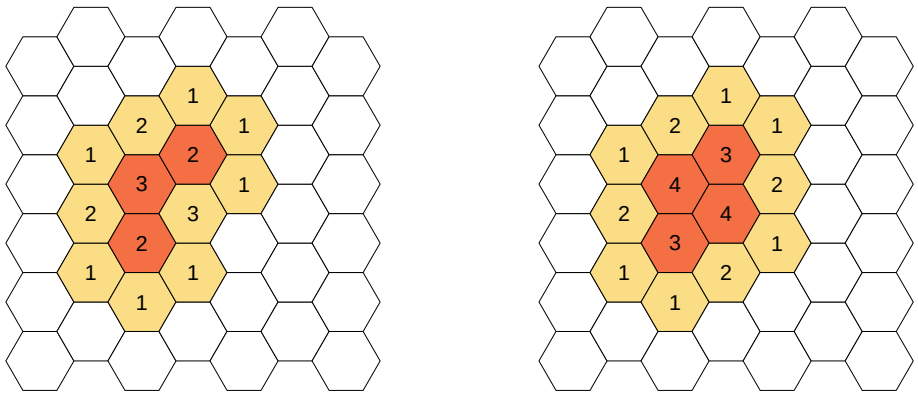


Figure 3.4: An example of inclusion weight $w_{ij}^{(t)}$ for each cell in an interaction neighbourhood. The number in each cell represents its inclusion weight used in the cell state transition process.(a) The cell states and their inclusion weights for an interaction neighbourhood when there are three infested cells at a specific time step t^*. (b) The updated interaction neighbourhood with new cell states and their corresponding inclusion weights at the next time step $t^* + 1$ when we assume there is one more cell that becomes infested during the transition process from time t^* to time $t^* + 1$.

Maximum of relative density difference for neighbouring cells $d_{ij}^{(t)}$. We assume that the relative density difference has an impact on the population density dispersion between neighbouring cells and dispersion is less likely to happen between neighbouring cells when their densities are close to each other. We use a parameter $d_{ij}^{(t)}$ equate to the maximum value of the density difference between the cell (i, j) and its six neighbouring cells to describe the influence of cell density difference on population density dispersion.

The parameters $\alpha_{ij}^{(t)}$ and $\beta_{ij}^{(t)}$ are used to adjust the inclusion probability under special considerations. For example, they are used to emphasize the difference in inclusion probability between high density cells and low density cells as the exact cell selection mechanism

is unknown. The parameters can be varied for each cell (i, j) at each time step t. $\alpha_{ij}^{(t)}$ is an integer, it can be varied to change the impact of inclusion weights on inclusion probabilities.

At each time step, for each cell (i, j), we associate a Bernoulli random variable $B_{I,ij}^{(t)}$ with inclusion probability $p_{I,ij}^{(t)}$ for taking the value of 1 and $1 - p_{I,ij}^{(t)}$ for taking the value of 0. We only include cells that have $B_{I,ij}^{(t)}$ of value 1 when sampled.

The inclusion probabilities tend to increase as $\alpha_{ij}^{(t)}$ increases, so that a cell with a specific weight is more likely to be included in the dispersion process when $\alpha_{ij}^{(t)}$ is large. $\beta_{ij}^{(t)}$ can also be used to adjust the inclusion probability, we set $\beta_{ij}^{(t)} \in (0, 1)$ to increase the inclusion probability for including a specific cell in the dispersion process, for example, regions where there are favourable factors for population dispersion. On the other hand, we set $\beta_{ij}^{(t)} \geq 1$ to decrease the inclusion probability for regions where the population dispersion is less likely to happen.

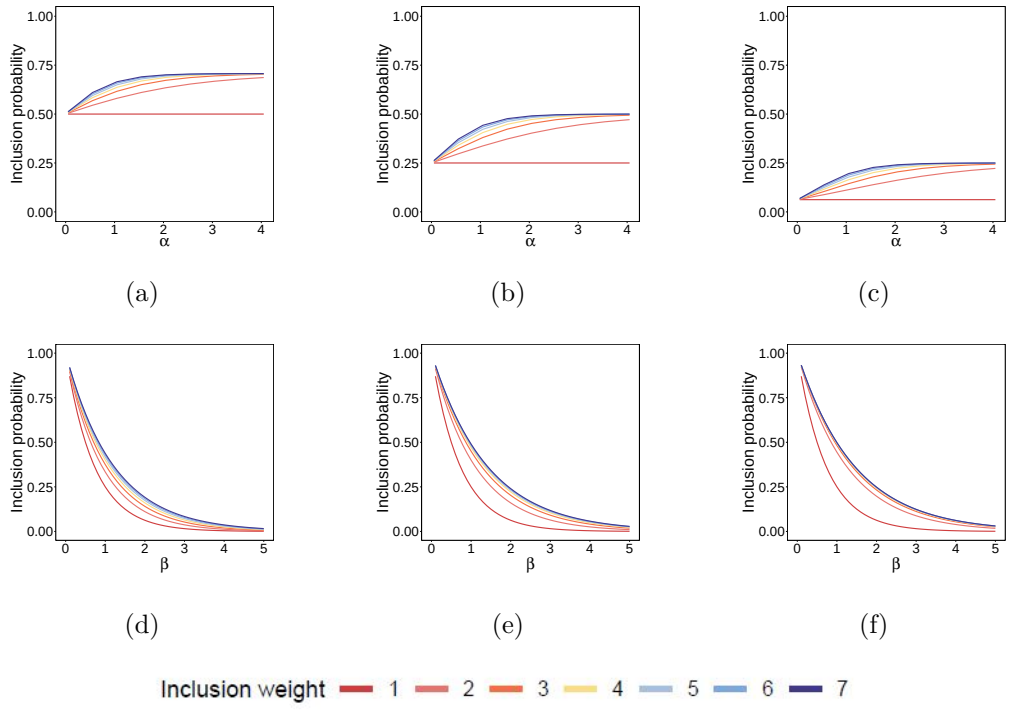


Figure 3.5: Inclusion probabilities for different α and β values. The inclusion weight can vary from 1 to 7 in the hexagonal cell neighbourhood. We set the maximum of relative density difference $d = 0.5$. The first row displays the inclusion probabilities for different α values when β is fixed. (a) $\beta = 0.5$. (b) $\beta = 1$. (c) $\beta = 2$. The second row displays the inclusion probabilities for different β values when α is fixed. (d) $\alpha = 1$. (e) $\alpha = 2$. (f) $\alpha = 3$.

We see in Figure 3.5, the inclusion probability increases as α increases and decreases as β increases. However, when the inclusion weight w and the relative density different d are fixed, α and β values do not have a radical influence on the inclusion probabilities unless when they take extreme values. In practical situations, the parameter values depend

on the physical properties of different interaction neighbourhoods. Therefore, $\alpha_{ij}^{(t)}$ and $\beta_{ij}^{(t)}$ values may be distinct for each cell at different time steps in the interaction neighbourhood. However, in the simplest form of the model, we set $\alpha_{ij}^{(t)}$ and $\beta_{ij}^{(t)}$ both to be a constant for each cell at each time step t. In Chapter 5, we present detailed study of the impact of the inclusion probability on model behaviour by varying α and β values.

b. Dispersion process

We determine a model for population dispersion assuming that populations tend to disperse into areas with a low population density. At each step of cell interaction in a given neighbourhood, we divide the included cells into two groups, the contributing cells and the non-contributing cells, where the contributing cell densities are higher than the mean density of all the included cells in the same interaction neighbourhood. We denote the group of contributing cells by $\{\pi_{i_h}^{(t)}\}_{h=1}^{H^{(t)}}$, where $\pi_{i_h}^{(t)} \geq \bar{\pi}^{(t)}$, $\bar{\pi}^{(t)}$ is the average density of the inclusive cells in each interaction neighbourhood and $H^{(t)}$ is the number of contributing cells at time step t. The rest of the cells in the interaction neighbourhood are considered to be non-contributing cells $\{\pi_{i_k}^{(t)}\}_{k=1}^{K^{(t)}}$, where $K^{(t)}$ represents the number of non-contributing cells at time t. Then the total amount of population to be redistributed is,

$$\tilde{\pi}^{(t)} = \sum_{h=1}^{H^{(t)}} \gamma_1(\pi_{i_h}^{(t)} - \bar{\pi}^{(t)})$$

where the parameter γ_1 determines the percentage of population to be redistributed.

To redistribute the population among the cells within the same interaction neighbourhood, considering a balance between the loss in the contributing cells and the gain in others, we allow the contributing cells to lose a proportion of population that is above the average of the interaction neighbourhood, thus the temporary state for contributing cells becomes,

$$\pi_{i_h}^{(t)} = \pi_{i_h}^{(t)} - \gamma_1(\pi_{i_h}^{(t)} - \bar{\pi}^{(t)}), \quad h = 1, 2, \ldots, H^{(t)}.$$

We assume the population survival rate is γ_2, and thus there is a background mortality rate $1 - \gamma_2$ that defines the population loss during the dispersion process. Assuming that the population uniformly spread within the same interaction neighbourhood, the state for contributing cells after receiving population is,

$$\pi_{i_h}^{(t+1)} = \pi_{i_h}^{(t)} - \gamma_1(\pi_{i_h}^{(t)} - \bar{\pi}^{(t)}) + \frac{1}{H^{(t)} + K^{(t)}}\gamma_2\tilde{\pi}^{(t)}, \quad h = 1, 2, \ldots, H^{(t)}.$$

The state for non-contributing cells becomes,

$$\pi_{i_k}^{(t+1)} = \pi_{i_k}^{(t)} + \frac{1}{H^{(t)} + K^{(t)}}\gamma_2\tilde{\pi}^{(t)}, \quad k = 1, 2, \ldots, K^{(t)}.$$

The dispersion process is performed for all interaction neighbourhoods simultaneously.

c. Survival probability

During the emigrating stage of dispersion, when individuals emigrate to new habitats, there is a risk of mortality associated with population establishment in the area [155]. We model this with a probabilistic rule as the probability of successful population dispersion to low density cells. The setup of survival probability is inspired by the field experiments conducted to investigate the dispersion of MPB described in [165]. The survival probability is mainly comprised of two parts. For one part, we use a constant ϕ to describe the natural conditions and geographical properties of the non-contributing cells in a particular region, e.g. wind, light and temperature. For the other part, we choose the continuous arctan function with two parameters ρ and ξ to describe the density-related effect during population establishment in the new area. The survival probability is determined as,

$$p_{S,ij}^{(t)} = \phi_{ij}^{(t)} + \rho_{ij}^{(t)} \arctan(\xi_{ij}^{(t)} \pi_{ij}^{(t)}), \tag{3.3.4}$$

where the shape parameters $\phi_{ij}^{(t)}$, $\rho_{ij}^{(t)}$ and $\xi_{ij}^{(t)}$ are used to adjust the survival probability for each non-contributing cell at each time step t.

At each time step, we also associate a Bernoulli random variable $B_{S,ij}^{(t)}$ with survival probability $p_{S,ij}$ of taking the value of 1 and $1 - p_{S,ij}$ of taking the value of 0 for each non-contributing cell. We only let the non-contributing cells that have $B_{S,ij}^{(t)}$ of value 1 successfully receive the population, and other non-contributing cells with a $B_{S,ij}^{(t)}$ of value 0 do not receive the population successfully, and thus their density becomes zero.

The survival probabilities tend to increase as the population density increases, suggesting that a given non-contributing cell is more likely to be colonized if its density is large enough. Similar to the inclusion probability, the parameter values in survival probability vary in principle for each cell at different time steps in the same interaction neighbourhood. Nevertheless, for the simplest form of the model, we let $\phi_{ij}^{(t)}$, $\rho_{ij}^{(t)}$ and $\xi_{ij}^{(t)}$ have constant values for all non-contributing cells at each time step in the same interaction neighbourhood. In Chapter 5, we also investigate the impact of the survival probability on model behaviours by varying the parameter values.

We transform the $m \times n$ configuration array into a $mn \times 1$ configuration vector for one particular interaction neighbourhood. Let $Q^{(t)}$ denote the system configuration for one interaction neighbourhood at time t, then

$$I : Q_{Z^2}^{(t)} \to Q_{Z^1}^{(t)},$$

where

$$Q_{Z^2}^{(t)} = \begin{bmatrix} \pi_{11}^{(t)} & \pi_{12}^{(t)} & \dots & \pi_{1n}^{(t)} \\ \vdots & \vdots & \vdots & \vdots \\ \pi_{m1}^{(t)} & \pi_{m2}^{(t)} & \dots & \pi_{mn}^{(t)} \end{bmatrix} \rightarrow Q_{Z^1}^{(t)} = \begin{bmatrix} \pi_{11}^{(t)} & \pi_{12}^{(t)} & \dots & \pi_{mn}^{(t)} \end{bmatrix}^T.$$

Using the probability transition matrix $P^{(t)}$, we obtain the cell update equation,

$$Q_{Z^1}^{(t+1)} = P^{(t)} Q_{Z^1}^{(t)}. \tag{3.3.5}$$

In the SCA model, the probability transition matrix $P^{(t)}$ is the product of the inclusion probability matrix and survival probability matrix, i.e.

$$P^{(t)} = P_S^{(t)} P_I^{(t)}. \tag{3.3.6}$$

5. Boundary Conditions

In the formal definition of a CA model, the lattice is assumed to be infinite in all dimensions. However, it is impossible to implement an infinite lattice on a computer, thus we need to prescribe boundary conditions on the boundary of a finite domain. The formulation of interaction neighbourhoods are different at a boundary as well. Three kinds of boundary conditions are discussed here: periodic, reflective and absorbing boundary conditions.

Periodic boundary conditions come closest to simulating an infinite lattice, where the top and bottom cells are connected, the left and right edges are connected, leading to the grid surrounded by a copy of one row/column from the opposite edge. Period boundary can avoid effects of border conditions by making the lattice invariant by translation, but are expensive to use in computational cost. Reflective boundary conditions are obtained by reflecting the grid at the boundary. They are suited for a situation in which the systems have physical boundaries but the value of the physical variables is not fixed. Absorbing boundary conditions are achieved by simulating "external" cells for the border which have the same updating rule as other cells in the lattice. Absorbing boundary addresses the border-effect, but it can avoid the reflective effects.

In the SCA model, we impose absorbing boundary conditions to handle boundary cell behaviour. We create auxiliary adjacent cells for those boundary cells that have no neighbours. We assume the auxiliary cells can only receive a proportion of population from the boundary cells according to the state transition rule, but they cannot send population to other cells in the interaction neighbourhood.

Using the same scheme, we derive a general form for cell updates of a community that comprises numerous cells and multiple interaction neighbourhoods. Suppose there are several interaction neighbourhoods in a local community defined by the lattice field $\mathbb{Z}^2$, the field comprises $M \times N$ nonoverlapping cells, the configuration array is represented by $\mathcal{Q}_{\mathbb{Z}^2}^{(t)}$

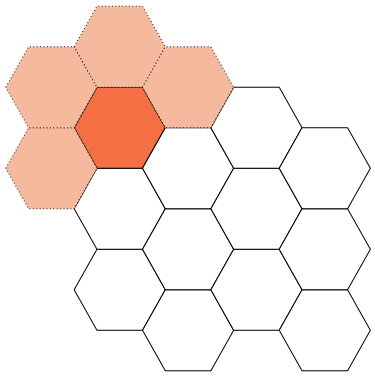


Figure 3.6: An example of auxiliary adjacent cells, the dark shaded cell represents a boundary affected cell, and the four light orange cells represent the corresponding auxiliary adjacent cells.

at time t. We adopt the same transformation scheme to obtain a new configuration vector $\mathcal{Q}_{\mathbb{Z}^1}^{(t)}$, the mapping is

$$I : \mathcal{Q}_{\mathbb{Z}^2}^{(t)} \to \mathcal{Q}_{\mathbb{Z}^1}^{(t)},$$

where

$$\mathcal{Q}_{\mathbb{Z}^2}^{(t)} = \begin{bmatrix} \pi_{11}^{(t)} & \pi_{12}^{(t)} & \cdots & \cdots & \cdots & \pi_{1N}^{(t)} \\ \pi_{21}^{(t)} & \pi_{22}^{(t)} & \cdots & \cdots & \cdots & \pi_{2N}^{(t)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \pi_{I1}^{(t)} & \pi_{I2}^{(t)} & \cdots & \cdots & \cdots & \pi_{IN}^{(t)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \pi_{M1}^{(t)} & \pi_{M2}^{(t)} & \cdots & \cdots & \cdots & \pi_{MN}^{(t)} \end{bmatrix}$$

$$\mathcal{Q}_{\mathbb{Z}^1}^{(t)} = \begin{bmatrix} \pi_{11}^{(t)} & \pi_{12}^{(t)} & \cdots & \cdots & \cdots & \pi_{MN}^{(t)} \end{bmatrix}^T.$$

With the probability matrix $\mathbb{P}^{(t)}$, we have,

$$\mathcal{Q}_{\mathbb{Z}^1}^{(t+1)} = \mathbb{P}^{(t)} \mathcal{Q}_{\mathbb{Z}^1}^{(t)}. \tag{3.3.7}$$

For computational efficiency, we do not attempt to deal with the entire matrix. Instead, we construct a new cell configuration vector and its corresponding transition matrix by assembling the cell configuration vector and transition matrix for each interaction neighbourhood into the full cell vector $\mathcal{Q}_{\mathbb{Z}^1}^{(t)}$ and the transition matrix $\mathbb{P}^{(t)}$.

Since there are many zero density cells and only a few interaction neighbourhoods in the local community, we can rearrange the zero cells into one arbitrary interaction neighbourhood and reposition the cell configuration $\mathcal{Q}_{\mathbb{Z}^1}^{(t)}$ and its corresponding matrix $\mathbb{P}^{(t)}$. For

example, suppose there are three active interaction neighbourhoods $Q_1^{(t)}$, $Q_2^{(t)}$ and $Q_3^{(t)}$ at time t in a local community with the corresponding transition matrix $P_1^{(t)}$, $P_2^{(t)}$ and $P_3^{(t)}$, we aggregate the rest of the cells in the community into one inactive neighbourhood $Q_0^{(t)}$, where all the cells in $Q_0^{(t)}$ have zero density. We rewrite the cell configuration as,

$$\mathcal{Q}_{\mathbb{Z}^1}^{(t)} = \begin{bmatrix} Q_1^{(t)} & Q_2^{(t)} & Q_3^{(t)} & Q_0^{(t)} \end{bmatrix}^T.$$

Since the cells do not interact with each other across different interaction neighbourhoods, the transition matrix for the community can be expressed in the form

$$\mathbb{P}^{(t)} = \begin{bmatrix} P_1^{(t)} & 0 & 0 & 0 \\ 0 & P_2^{(t)} & 0 & 0 \\ 0 & 0 & P_3^{(t)} & 0 \\ 0 & 0 & 0 & P_0^{(t)} \end{bmatrix},$$

where the submatrix $P_0^{(t)}$ is a diagonal matrix that has the same dimension as the cell configuration vector $Q_0^{(t)}$. The cell updates for the community is in Equation 3.3.7.

We provide two illustrative examples. We start with an example of one affected cell. At a given time t, suppose there is only one affected cell, we create the interaction neighbourhood by including the affected cell and its six neighbouring cells. We roll the dice on the six neighbouring cells with probability $\frac{1}{2^\beta}$, where β is the scale factor to adjust the inclusion probability. We only include cells with the coin flipping outcome of 1, and assume there are four selected non-contributing cells in the dispersion process.

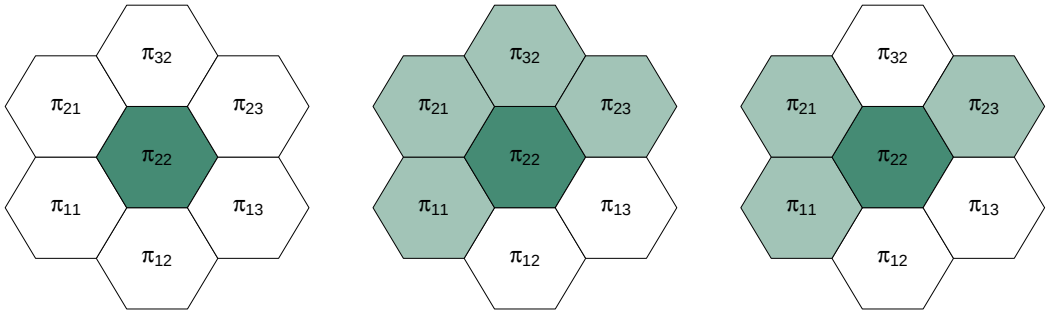
In the first example, the number of contributing cells $H = 1$ and the number of non-contributing cells $K = 4$ in the dispersion process. Let $\bar{\pi}^{(t)}$ be the mean density of the cells included in the dispersion process at each iteration, then

$$\bar{\pi} = \frac{1}{H+K}(\pi_{11} + \pi_{21} + \pi_{22} + \pi_{23} + \pi_{32}) = \frac{\pi_{22}}{5}.$$

The total amount of population to be redistributed is $\tilde{\pi} = \frac{4\gamma_1\pi_{22}}{5}$, and γ_1 determines the percentage of population to be redistributed. For the four non-contributing cells, the density becomes $\frac{4\gamma_1\gamma_2\pi_{22}}{25}$, γ_2 is the background population loss during the dispersion process; whereas for the contributing cell $(2,2)$, the density becomes $\pi_{22} - \frac{4\gamma_1\pi_{22}}{5} + \frac{4\gamma_1\gamma_2\pi_{22}}{25}$.

We then roll the dice on the four non-contributing cells with probability $\phi + \rho\arctan(\xi\frac{4\gamma_1\gamma_2\pi_{22}}{25})$, and suppose we obtain a zero only for the non-contributing cell $(3,2)$. Then, the cell density for $(3,2)$ becomes zero, and remains $\frac{4\gamma_1\gamma_2\pi_{22}}{25}$ for the other three non-contributing cells.

We obtain the inclusion probability matrix P_I and the survival probability matrix P_S, where



(a) (b) (c)

Figure 3.7: An example of dispersion process with one affected cell. (a) One particular interaction neighbourhood with the central cell affected and its six neighbouring cells that could potentially receive population from the central cell before the start of one specific dispersion process. (b) The same interaction neighbourhood with central cell affected and selected four neighbouring cells in light green that are considered as non-contributing cells during one specific dispersion process. (c) The same interaction neighbourhood with the central cell affected and selected three neighbouring cells in light green that successfully received the population at the end of one specific dispersion process.

$$P_I = \begin{bmatrix} 1 & 0 & 0 & 0 & \frac{4\gamma_1\gamma_2}{25} & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \frac{4\gamma_1\gamma_2}{25} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 - \frac{4\gamma_1}{5} + \frac{4\gamma_1\gamma_2}{25} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{4\gamma_1\gamma_2}{25} & 1 & 0 \\ 0 & 0 & 0 & 0 & \frac{4\gamma_1\gamma_2}{25} & 0 & 1 \end{bmatrix}$$

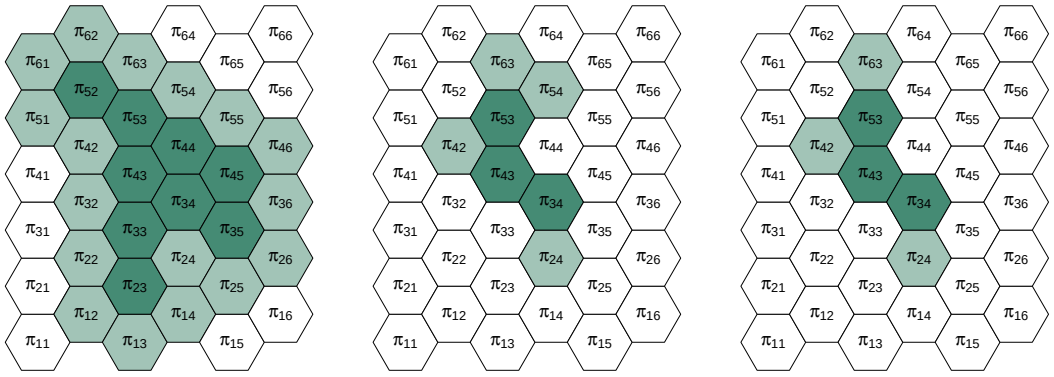
and

$$P_S = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

According to the cell states update Equation 3.3.5 and Equation 3.3.6, we obtain

$$\begin{bmatrix} \pi_{11}^{(t+1)} \\ \pi_{12}^{(t+1)} \\ \pi_{13}^{(t+1)} \\ \pi_{21}^{(t+1)} \\ \pi_{22}^{(t+1)} \\ \pi_{23}^{(t+1)} \\ \pi_{32}^{(t+1)} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & \frac{4\gamma_1\gamma_2}{25} & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \frac{4\gamma_1\gamma_2}{25} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 - \frac{4\gamma_1}{5} + \frac{4\gamma_1\gamma_2}{25} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{4\gamma_1\gamma_2}{25} & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \pi_{11}^{(t)} \\ \pi_{12}^{(t)} \\ \pi_{13}^{(t)} \\ \pi_{21}^{(t)} \\ \pi_{22}^{(t)} \\ \pi_{23}^{(t)} \\ \pi_{32}^{(t)} \end{bmatrix}.$$

We discuss another illustrative example with multiple affected cells. Suppose there are several affected cells at time step t, we create a minimum rectangular area that could envelop the interaction neighbourhood of the affected cells. We employ the same coin flipping scheme for each cell with respect to its inclusion probability (defined in the model) to decide which cells to be included in the dispersion process, we only include cells with the coin flipping outcome of 1. We assume at this iteration there are seven cells that are included in the dispersion process, and three of which are contributing cells in dark green.



(a) (b) (c)

Figure 3.8: An example of dispersion process with several affected cells. (a) One particular interaction neighbourhood with a group of infested cells in dark green and their neighbouring cells in light green, which are potential contributing cells and non-contributing cells before the start of one specific dispersion process. (b) The same interaction neighbourhood with the selected three contributing cells in dark green and four non-contributing cells in light green during one specific dispersion process.(c) The same interaction neighbourhood with the selected three contributing cells in dark green and three non-contributing cells in light green that successfully receive the population at the end of one specific dispersion process.

In this case, the number of contributing cells $H = 3$, and the number of non-contributing cells in the dispersion process $K = 4$. The mean cell density at time step t is:

$$\bar{\pi} = \frac{1}{7}(\pi_{24} + \pi_{34} + \pi_{42} + \pi_{43} + \pi_{53} + \pi_{54} + \pi_{63}).$$

The population to be redistributed at the given time is:

$$\tilde{\pi} = \sum_{h=1}^{3} \gamma_1(\pi_{i_h} - \bar{\pi}).$$

Subtracting the allocated population from contributing cells, the temporary density becomes

$$\pi_{i_h} = \pi_{i_h} - \gamma_1(\pi_{i_h} - \bar{\pi}) \quad h = 1, 2, 3.$$

The density for the cells that are included in the specific dispersion process at next time step becomes

$$\pi_l^{(t+1)} = \pi_l + \frac{1}{7}\gamma_2\tilde{\pi} \quad l = 1, 2, \ldots, 7.$$

Then the density at the next time step $t + 1$ is

$$\pi_{ij}^{(t+1)} = \begin{cases} \pi_{ij} - \gamma_1(\pi_{ij} - \bar{\pi}) + \frac{1}{7}\gamma_2\tilde{\pi} & \text{for contributing cells,} \\ \pi_{ij} + \frac{1}{7}\gamma_2\tilde{\pi} & \text{for non-contributing cells.} \end{cases}$$

We can obtain the inclusion probability matrix P_I and the survival probability matrix P_S, where

$$P_I = \begin{bmatrix}
1 & 0 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & 0 \\
\vdots & \ddots & & & & & & & & & & & & & \\
0 & \ldots & 1 & \ldots & C_{(34)} & \ldots & \ldots & C_{(43)} & \ldots & C_{(53)} & \ldots & \ldots & \ldots & \ldots & 0 \\
\vdots & & & \ddots & & & & & & & & & & & \\
0 & \ldots & \ldots & \ldots & C_{34} & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & 0 \\
\vdots & & & & & \ddots & & & & & & & & & \\
0 & \ldots & \ldots & \ldots & C_{(34)} & \ldots & 1 & C_{(43)} & \ldots & C_{(53)} & \ldots & \ldots & \ldots & \ldots & 0 \\
0 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & C_{43} & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & 0 \\
\vdots & & & & & & & & \ddots & & & & & & \\
0 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & C_{53} & \ldots & \ldots & \ldots & \ldots & 0 \\
0 & \ldots & \ldots & \ldots & C_{(34)} & \ldots & \ldots & C_{(43)} & \ldots & C_{(53)} & 1 & \ldots & \ldots & \ldots & 0 \\
\vdots & & & & & & & & & & & \ddots & & & \\
0 & \ldots & \ldots & \ldots & C_{(34)} & \ldots & \ldots & C_{(43)} & \ldots & C_{(53)} & \ldots & \ldots & 1 & \ldots & 0 \\
\vdots & & & & & & & & & & & & & \ddots & \\
0 & 0 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & 1
\end{bmatrix}$$

and P_S is a square matrix with entries on the main diagonal equal to 0 for cell $(5,4)$ and 1 for all the other cells, and the entries outside the main diagonal are zero.
Therefore, with the cell update equation 3.3.5 and 3.3.6,

$$\begin{aligned}\left[\pi_{11}^{(t+1)}\dots\pi_{24}^{(t+1)}\dots\pi_{34}^{(t+1)}\dots\pi_{42}^{(t+1)}\pi_{43}^{(t+1)}\dots\pi_{53}^{(t+1)}\pi_{54}^{(t+1)}\dots\pi_{63}^{(t+1)}\dots\pi_{66}^{(t+1)}\right]^T \\ = P_{36\times 36}\left[\pi_{11}^{(t)}\dots\pi_{24}^{(t)}\dots\pi_{34}^{(t)}\dots\pi_{42}^{(t)}\pi_{43}^{(t)}\dots\pi_{53}^{(t)}\pi_{54}^{(t)}\dots\pi_{63}^{(t)}\dots\pi_{66}^{(t)}\right]^{T'}\end{aligned}$$

where the transition matrix

$$P^{(t)} = \begin{bmatrix}
1 & 0 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 0\\
\vdots & \ddots & & & & & & & & & & & & & \\
0 & \dots & 1 & \dots & C_{(34)} & \dots & \dots & C_{(43)} & \dots & C_{(53)} & \dots & \dots & \dots & \dots & 0\\
\vdots & & & \ddots & & & & & & & & & & & \\
0 & \dots & \dots & \dots & C_{34} & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 0\\
\vdots & & & & & \ddots & & & & & & & & & \\
0 & \dots & \dots & \dots & C_{(34)} & \dots & 1 & C_{(43)} & \dots & C_{(53)} & \dots & \dots & \dots & \dots & 0\\
0 & \dots & \dots & \dots & \dots & \dots & \dots & C_{43} & \dots & \dots & \dots & \dots & \dots & \dots & 0\\
\vdots & & & & & & & & \ddots & & & & & & \\
0 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & C_{53} & \dots & \dots & \dots & \dots & 0\\
0 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 0 & \dots & \dots & \dots & 0\\
\vdots & & & & & & & & & & & \ddots & & & \\
0 & \dots & \dots & \dots & C_{(34)} & \dots & \dots & C_{(43)} & \dots & C_{(53)} & \dots & \dots & 1 & \dots & 0\\
\vdots & & & & & & & & & & & & & \ddots & \\
0 & 0 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 1
\end{bmatrix}$$

with

$$\begin{aligned}C_{(34)} &= \frac{\gamma_2\tilde{\pi}}{21\pi_{34}},\\ C_{(43)} &= \frac{\gamma_2\tilde{\pi}}{21\pi_{43}},\\ C_{(53)} &= \frac{\gamma_2\tilde{\pi}}{21\pi_{53}},\\ C_{ij} &= 1-\gamma_1\left(1-\frac{\bar{\pi}}{\pi_{ij}}\right)+\frac{\gamma_2\tilde{\pi}}{7\pi_{ij}}.\end{aligned}$$

From these examples, we conclude that the transition matrix for each interaction neighbourhood has different dimensions in different regions, and the matrix elements depend on

the corresponding cell states at time t. The dimensions for transition matrix changes conditional on whether the size of each interaction neighbourhood changes after each dispersion iteration.

The above multiscale stochastic process model describes dispersion process in each interaction region. The size of a complete model for an application depends on the size of the domain where the phenomena is modelled and the computational resources available for simulation. In the ideal case when there are unlimited computational resources, the stochastic process model can theoretically be operated over an infinite region. However, this is often not true in reality, i.e., the number of cells that are able to be performed by the model is limited. For example, in the study of forest infestation patterns resulted from MPB epidemics, the physical operational area of the stochastic process model depends on the definition of the microscale model when the computational capacity is fixed. In the case when the microscale model describes beetle behaviours on one single tree, the stochastic process model describes beetle density dynamics from tree to tree and the model can cover a large area when the trees are spaced far apart from each other. Under this circumstance, each cell represents one single tree, and the cell state describes the infestation severity on each tree. We compute the percentage of infestation for each tree as the macroscale input. But when the trees are spaced close to each other, we tend to use the microscale model to describe beetle behaviours on a patch of trees such that the macroscale stochastic process model can cover a relatively large area. In this situation, each cell represents a small patch of trees with similar infestation severity in the forest and the cell state refers to the MPB density of each patch, the microscale model is run multiple times and the cell state is taken as the average of the microscale model output for the macroscale model. Therefore, owning to the number of cells that are able to be performed by the model, the physical operational scale of the model depends on the tree density in a regional area.

3.3.2 Microscale Process and Model

The microscale is integrated with the macroscale model in order to provide the source of affected population. We assume a physics/biology based model for the microscale that can describe the dynamics of affected population locally. The macroscale model can employ any microscale model, whose output can be converted into a density. In the simplest framework, the microscale model is simulated for each macroscale step. But that is very computationally intensive so we may consider strategies to increase efficiency. For example, the microscale model has limiting behaviour for large macroscale time steps, we may be able to use that to speed computation.

We illustrate with a birth-death process. A birth-death process is a stochastic process that provides a useful theoretical framework for modelling population growth while taking into account stochastic demographic factors. In real world situations, there can be a different

(a) (b)

Figure 3.9: Photos of mountain pine beetles on a tree and trees infestations caused by mountain pine beetles. (a) Individual mountain pine beetles on a tree, which is described by the microscale model. Photo is from [1]. (b) Large patches of tree infestation caused by mountain pine beetles, which is described by the macroscale model. Photo was taken by Dezene Huber, University of Northern British Columbia. Source of the photo: https://www.flickr.com/photos/sfupamr/8621706223/. Licensed under a CC BY license.

modelling framework with different parameter values in each cell, we choose the birth-death process to describe all individual activities in each cell and we determine the parameter values are uniform regardless of the cell properties.

A birth-death process is a continuous-time Markov chain that counts the number of individuals in the population over time. Let X_τ denote the population size at time τ, and X_τ can take values in the finite state space $\{0, 1, 2, \ldots, C\}$, where C is the maximum population size (capacity). In a general process with n current individuals, the rate of births and deaths at any time depends on the number of extant individuals, assume a new individual is born with instantaneous rate b_n and an individual dies with instantaneous rate d_n, where b_n and d_n can be a function of n.

There are multiple forms of birth death processes with respect to different kinds of birth and death rates, among which, the simplest process has linear rates $b_n = nb$ and $d_n = nd$, where the constants b and d represent the per-individual birth and death rates with no density effects. However, in realistic situations, populations that occupy a finite space may be subject to various constraints on the population size over time. When there are too many individuals competing for the living space and food supply, the birth rate may decline when the population size is approaching the carrying capacity. We investigate two kinds of birth rates at the microscale model, one is $b_n = nb(1 - n/C)$ which is inspired by the logistic growth model [180] that can reflect the forementioned density-dependent constraints; and the other is $b_n = nb(1 - n/C) + b_0 nb(1 - n/C)sin(2t\pi/m)$ to reflect both the density-dependent constraint and the periodic effect, where b_0 is a constant reflecting the size of periodic effect, t is the macroscale time point and m is a constant used to determine the time period of the birth rate.

Given n individuals at time τ, consider a very short time period $\Delta\tau$, the probability of a birth in the interval $(\tau, \tau + \Delta\tau)$ is

$$Pr(X_{\tau+\Delta\tau} = n + 1 | X_\tau = n) = b_n \Delta\tau + o(\Delta\tau), \tag{3.3.8}$$

where $o(\Delta\tau)$ represents terms that have smaller order than $\Delta\tau$, i.e.,

$$\lim_{\Delta\tau \to 0} \frac{o(\Delta\tau)}{\Delta\tau} = 0.$$

The probability of a death in $(\tau, \tau + \Delta\tau)$ is

$$Pr(X_{\tau+\Delta\tau} = n - 1 | X_\tau = n) = d_n \Delta\tau + o(\Delta\tau). \tag{3.3.9}$$

At the micro scale, individuals usually have a much shorter life cycle than the system they belong to. Thus we can assume the time interval $\tau \to \tau + 1$ is small, and at most one event, either a birth or a death, can occur during this time interval. We formulate the transition probabilities as

$$Pr(X_{\tau+1} = j | X_\tau = i) = \begin{cases} b_i, & \text{if } j = i + 1, \\ d_i, & \text{if } j = i - 1, \\ 1 - b_i - d_i, & \text{if } j = i, \\ 0, & \text{otherwise.} \end{cases} \tag{3.3.10}$$

The transition matrix P has the form,

$$P = \begin{bmatrix} 1 & d_1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 - (b_1 + d_1) & d_2 & 0 & \dots & 0 & 0 \\ 0 & b_1 & 1 - (b_2 + d_2) & d_3 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 - (b_{N-1} + d_{N-1}) & d_N \\ 0 & 0 & 0 & 0 & \dots & b_{N-1} & 1 - d_N \end{bmatrix} \tag{3.3.11}$$

There are many microscale time steps for each macroscale time step. Suppose there are N_t, where N_t is a positive integer, microscale time steps in each macroscale time step. Then we have $\Delta\tau = 1/N_t$, and the transition probabilities at each microscale time step can be formulated as

$$Pr(X_{\tau+\Delta\tau} = n + 1 | X_\tau = n) = \frac{b_n}{N_t}$$

and

$$Pr(X_{\tau+\Delta\tau} = n-1 | X_\tau = n) = \frac{d_n}{N_t}.$$

The limiting population size is converted to a density by dividing by the carrying capacity of the birth-death process which can be used as the cell state at the macroscale.

3.4 Applications

The above multiscale modelling framework is suitable for dispersion process within a variety of ecological and physical applications where the systems can be organized in a hierarchical structure. We provide three examples, the MPB infestations, nuclear radiation damage, and disease transmission patterns. In these examples, we introduce two scales of the systems, the microscale and the macroscale, and explain how to connect different scales of hierarchy.

3.4.1 Mountain Pine Beetle Infestations

Forest infestation patterns provide a good application for the SCA model as MPB individual behaviours lead to changes in tree infestations severity - the state of density at the regional level. Simulation results from the model reflect the process of MPB infestations that are observed in the field and described in the literature, and exhibit non-linear dynamics expected in dispersion process. This modelling can provide useful information for forest management.

3.4.1.1 Macroscale Process and Model

We apply the SCA to MPB community dynamics to illustrate how local interactions can influence the overall community structure. The model is built upon two assumptions. The first assumption is that mountain pine beetles could only disperse to adjacent neighbouring cells whereas beetles could actually conduct exceptionally long flights ranging from 2.12 to 5.95 km [70] during flight seasons. The second assumption is that the model does not incorporate a separate model for host conditions. Most MPBs have a univoltine lifecycle, each summer young beetles emerge from current brood and seek for new hosts, this successful offspring production usually leads to the death of host trees. Since we do not consider any cell as "dead", the model could present an intensive dispersion pattern of the beetles.

The model is applied to the incipient-epidemic phase and the beginning of epidemic phase of the MPB infestation within a local forest region. Because from the biology perspective of mountain pine beetles, the onset of an outbreak usually starts from a localized infestation, and the beetle dispersion is slow and infestation is limited to small patches of trees at the beginning of an outbreak. In addition, during those phases, there are enough beetles to produce a viable density in each patch of trees. We represent the local forest

region by a regular grid composed of hexagonal cells with each cell consisting of a patch of trees with similar infestation severity. The MPB population density in each patch represents the cell state, and is computed as the ratio of beetle population size to the carrying capacity in each cell. In midsummer, young adult beetles emerge from their brood trees and seek for new host trees, usually nearby trees, for the next generation of MPBs. Their dispersion activities often happen in a small window and dispersion probabilities decrease with distance [44]. Thus we assume the beetle dispersion is restricted to adjacent neighbouring trees, and they cannot fly to trees that are separated by more than two non-infested patches of trees from their hosts. As defined in the model, each interaction neighbourhood should be distinguished by at least two non-infested patches of tree areas. We adopt the absorbing boundary conditions. There is a percentage of population loss in MPBs when the affected region reaches the boundary. The state transition rule describes the MPB population dispersion between neighbouring cells. It follows the dispersion scheme of pine beetles and also depends on the cell states (percentage of infestation in each cell). We assume uniform dispersion from high density cells to low density cells following a probabilistic rule within the same interaction neighbourhood. While the exact mechanisms of selecting perspective host trees is not completely understood, the selection of new host trees could be spatially random and independent of population density [14], we believe the states of neighbours and the relative population density between the neighbouring cells are crucial to beetle dispersion.

In practice, trees that are surrounded by more infested trees are more likely to be infested during the pine beetle dispersion process. By using the inclusion weight for each tree, we could characterize the impact of infestation severity of neighbouring trees on a specific tree. We use the parameter maximum of relative density difference for neighbouring cells to characterize the dynamic interaction between neighbouring trees. When neighbouring trees are lightly infested, pine beetles are less likely to disperse between the trees since there is enough resources on the current brood tree to support the reproduction of beetle population. On the other hand, when neighbouring trees are severely infested, MPBs tend to avoid those highly infested trees and seek colonization on less infested trees even though we do not know the exact chemical biology behind this scheme. Therefore, it is reasonable to propose the maximum of relative density difference for neighbouring cells as a determining parameter in the inclusion probability to decide whether there is population dispersion between neighbouring trees.

Besides the inclusion weights and maximum of relative density difference for neighbouring cells, there are other physical properties to determine the probability of including certain cells in dispersion process. We use parameters $\alpha_{(i,j)}^{(t)}$ and $\beta_{(i,j)}^{(t)}$ to distinguish the inclusion probability between high density and low density cells, and they can be varied for each tree at each time step, but considering computational efficiency, we set both of the parameter values to 1 since they do not have a radical influence on the inclusion probabilities.

Young beetles usually seek for new host trees when there are not sufficient resources for supporting next generation. When the female beetles initiate new attacks, they bore into the bark and release trans-verbenol that can attract both sexes; attracted males produce exo-brevicomin which primarily attracts females at low concentrations and inhibit aggregation at higher concentrations. As the attacks progresses, the tree becomes more vulnerable, the antiaggregation semiochemical, verbenone, is produced. The beetles seek other potential hosts [124]. We assume the pine beetle population dispersion follows the process as we described previously, the parameter γ_1 is used to characterize the local physical features to determine the amount of beetle population to disperse to neighbouring trees. The parameter γ_2 is used to determine the percentage of beetle population that can potentially enter the new host trees.

The attacks of MPBs are not always successful, and many trees do not undergo aggregation after being entered by beetles. The pitch tubes formed by the attacking beetles are filled with tree resin and beetles have to clear the resin from their galleries. If the tree resin seals the entrance of the galleries, the beetles either leave or die in the pitch tubes [124]. We use the survival probability to represent the stochastic feature of this process where the landing and colonization of beetles may not be successful.

3.4.1.2 Microscale Process and Model

According to the biology of mountain pine beetles, spatial and temporal dynamics vary with the level of aggressiveness. Host trees generally exhibit formidable defences that protect them from beetle attacks when population levels are low, but aggregation of large number of mountain pine beetles could successfully overwhelm the defences and colonize healthy trees for reproduction [15]. Thus, there exists a critical population threshold that separates endemic from epidemic phases. For the mountain pine beetles, they persist during endemic periods when stressors within host trees provide a constraint on population reproduction, and when appropriate conditions arise, they switch to an epidemic phase and there is a population explosion. The threshold serves as the minimum aggregation population size that is needed to successfully overwhelm tree defences and colonize a new host at the individual tree level.

The birth-death-process can capture the Allee effect of MPB population on a particular tree such that when there is no disturbance for pine beetles, their growth is positive below the carrying capacity C, and the growth rate reaches a maximum at $C/2$. Once the population level rises close the capacity, the growth rate begins to decrease. The population reproduction commonly happens in a small window of time during the summer. Therefore, we consider the time period $\Delta\tau$ is sufficiently small compared to the overall population dynamics at the landscape level.

3.4.1.3 Sequential Multiscale Approach to Link the Scales

We simulate the birth-death process at the microscale, use the percentage of infested trees in each cell as a proxy for initial cell configuration of SCA models at the macroscale. We update the cell density after simulating the dispersion process with SCA and convert the updated density to a population size. We alternate between the microscale and macroscale models and observe the resulting pattern at the macroscale.

3.4.2 Nuclear Radiation Damage

Nuclear reactor materials are continuously bombarded with different kinds of energetic particles during service. Such kinds of particles interact with materials to produce high concentrations of defects such as vacancies and interstitials. These defects eventually induce microstructural changes in materials such as dislocation loops and voids, which drastically change the thermal and mechanical properties of materials and limit their performance [5, 159, 176].

Due to the inherent multiscale nature of radiation effects in materials, the multiscale modelling scheme can be employed to simulate the resulting microstructure processes. Atomic displacement collision cascades that happen within nanoseconds induce microstructure changes in the fuel lattice, microstructure changes over long-range space and time lead to radiation damage in materials which can be evaluated by phase field models. There exists an intermediate level of structure change in materials where a large number of voids grow, shrink, and/ or interact with each other. [190] provides extensive discussion regarding determining the appropriate scales in structure changes in fuels. The main objective is to provide a brief description for our model application to these physical processes.

3.4.2.1 Macroscale Process and Model

The atomic displacement cascades, which occur in an uncorrelated fashion, produce an increase in the concentration of point defects such as vacancies and interstitials. Owning to dislocation bias - dislocations tend to absorb interstitials at faster rates than vacancies, the absorption of interstitials removes more interstitials than vacancies from the system and thus makes more vacancies available to cluster together to form voids in irradiated materials. Void formation results from the supersaturation of vacancies in irradiated materials. With interstitials having the tendency to form dislocation loops, the presence of gas atoms enhances the void nucleation process. These features have a significant impact on structural changes in materials.

At the macroscale, we use the SCA model to describe these point defects behaviours including annihilation, recombination, and diffusion which lead to void nucleation in irradiated materials. The cell space is represented by a regular grid composed of square cells. Cell

states are represented by the point defect percentage in each cell. We employ the stochastic transition rule to introduce uncertainties in modelling the spatial patterns in materials.

3.4.2.2 Microscale Process and Model

An energetic particle knocks an atom from its lattice, which displaces many other atoms through a cascade of atomic collisions and results in many vacancies and interstitials. These binary collision cascade events at the microscale are well-modelled by a group of computer programs called Stopping and Range of Ions in Matter (SRIM) which calculate interactions between ions and materials. The core of SRIM is a program called Transport of Ions in Matter (TRIM) [190, 202]. SRIM is used to generate ion implantation in any desired solid lattice. However, they don't take account of the structural or dynamic composition changes in materials.

3.4.3 Infectious Disease Transmission

There is direct evidence for the spread of infectious disease in places where people gather at high densities. Despite the fact that infectious disease transmission events are well documented in public health literature, the characteristics of disease transport patterns are not completely understood. Consequently, disease transmission to potential regions may not be recognized and corresponding precautions against the disease spread may not be taken. The SCA modelling scheme we build is applicable for multiple diseases. Through this multiscale framework, we incorporate pedestrian dynamics such as infected and susceptible individual interactions into a stochastic infection dynamics pattern, which can be used to assist in better understanding the transmission characteristics of infectious diseases, and provide guidelines for disease control and prevention.

3.4.3.1 Macroscale Process and Model

The macroscale represents the between-local community level disease transmission. The community can either represents geographical components such as village and district, or represent administrative units such as workplace, school, and hospital. For example, Q fever can transmit from livestock farms to their surrounding communities [121], as well as Legionnaires' disease can lead to community-level breaks [22]. The SCA model is used to simulate the emergence, evolution, spread patterns of infectious disease in the state of percentage of infectious or susceptible individuals in each community by varying the input parameters related to transmission mechanisms.

3.4.3.2 Microscale Process and Model

The microscale model describes the general behaviours between susceptible and infectious individuals within-local community, it is used to estimate the number of infectious

and susceptible individuals in the community. Each individual is treated as a mobile pedestrian, and the model captures the evolution of pedestrians and their interactions with other pedestrians, and the formation of groups of pedestrians in each community.

3.5 Summary

In this chapter, we have described the details for setting up the SCA model for dispersion processes. We are inspired by the concept of CA, and we extend this concept by relaxing some of the model settings to allow for more complexity which is characteristics of dispersion processes. The SCA model is a stochastic process, but the transition matrix depends on the state properties at each time step rendering the transition probability matrix nonautonomous, which presents serious challenges for theoretical analysis. Therefore, we investigate the model behaviour through numerical simulations. The model is intended to run within a finite domain and a finite time step at the macroscale as the resulting patterns exhibit similar behaviour when the model is run on an infinite domain and infinite time step, details can be found in Chapter 5.

Chapter 4

Implementation

In this chapter, we provide implementation details for the implementation of the stochastic process model defined in Chapter 3. We use R programming language (version 4.1.1) to build a simulator that takes an initial cell configuration and a set of parameter values for the CA transition rules as input and outputs the cell configuration evolution at a sequence of times as output. The simulator is designed such that the input can be varied. The microscale model that is embedded in the simulator can also be varied to reflect the choice of individual dynamic model.

4.1 Simulating the Dispersion Process

4.1.1 Domain

The simulation results for the stochastic cellular automaton model for dispersion presented in Chapter 3 that makes use of a two-dimensional tessellation, which generates evolutionary patterns at specified times. The two-dimensional tessellation suffices for the modelling purposes presented in the thesis. Using these images, we observe geometric invariant properties of dispersion patterns with different parameter settings. During an initial exploration process, we use one-dimensional tessellation to study the behaviour of cells during a dispersion process and the effects of boundary conditions. In general, the behaviour on one dimensional tessellations if of limited interest, and we do not present detailed simulation results based on one-dimensional tessellation.

We set up a domain that is composed of $m \times n$ non-overlapping cells using index i and j to identify each cell in the domain, where index i and j represent the row and column position of each cell, respectively. In the simulation studies, we focus on the scenarios when the model produces final patterns that grow without limit to capture the essential features of dispersion phenomena, and we use complexity of patterns and velocity of dispersion to provide further information about the final patterns. Therefore, in experiments to study the distribution of quantitative measurements including number of affected cells, average cell

density and complexity of dispersion patterns, we choose $m = 30$ and $n = 30$ as the domain size to obtain a "rough" pattern that is sensitive to the parameter values because the model tend to behave in a similar way in terms of complexity of patterns when the grid size is large and when the model is run long enough. In experiments to investigate the evolution patterns generated by the model, we choose $m = 50$ and $n = 50$ as the domain size to provide enough information about the dispersion pattern details and analyse the impact of parameters on the model behaviours as most of the cells on the grid are affected. Details for determining the domain size can be found in Chapter 5.

4.1.2 Interaction Neighbourhoods

In the modelling framework, the dispersion happens within interaction neighbourhoods. It is crucial to correctly set up the interaction neighbourhood for the affected cells. When there is only one affected cell in the domain, its interaction neighbourhood contains all cells that surround the particular affected cell. In a two-dimensional domain, common cell shapes are squares. However, we simulate dispersion process on a tessellation of the domain using hexagonal cells because of the forementioned practical benefits in Chapter 3. Since we use a tessellation based on hexagonal cells, the index varies according to the evenness and oddness of the column index. When the column index j is odd, the adjacent neighbouring cells include cells $(i-1, j-1), (i-1, j), (i-1, j+1), (i, j-1), (i, j+1), (i+1, j)$; and when j is even, the adjacent neighbouring cells include cells $(i, j-1), (i-1, j), (i, j+1), (i+1, j-1), (i+1, j+1), (i+1, j)$. When there is a group of adjacent affected cells, then the interaction neighbourhood contains the smallest collection of cells that envelopes these affected cells.

When there are several disconnected groups of affected cells, we create interaction neighbourhoods for each group. The distinct interaction neighbourhoods must be separated by at least two nonaffected adjacent cells. In this case, distinct affected cells that overlap in immediate neighbours must be grouped into one interaction neighbourhood. It turns out that the grouping of affected cells is incredibly expensive with even moderate domain size. Thus, we adopt a grouping scheme built on a grid-based clustering approach that has a faster processing time.

For cells on the boundary of the domain, we create auxiliary adjacent neighbours outside the domain that could receive population from the affected cells. We remove the auxiliary cells at the end of each dispersion iteration. The auxiliary cells only receive population and never send populations to other cells in the interaction neighbourhood.

To detect the simulation code, we verify the population redistribution mechanisms manually to make sure the code performs as intended.

Algorithm 1 Algorithm for creating interaction neighbourhoods

1: Determine all affected cells in the domain.
2: **for** each affected cell in the domain **do**
3: Randomly pick an affected cell as the starting point, where the cell has not been traversed before.
4: Create a two-cell radius shell around the picked cell, check if there is any affected cell(s) in the shell. If yes, include the affected cell(s) with the picked cell to form one interaction neighbourhood; otherwise, the starting point cell forms one separate interaction neighbourhood.
5: Repeat step 3 and 4 until there are no affected cells in the two-cell radius shell.
6: **end for**
7: All cells will be associated with some interaction neighbourhood.

4.1.3 Initial Cell Configuration

We create multiple initial cell configurations according to the study purposes.

1. We set up distinct initial cell configurations that display various shapes in the domain in order to study the effect of initial cell configuration on dispersion patterns.

2. We use the same initial cell configuration that is composed of one group of adjacent affected cells to perform sensitivity analysis and study the parameter interactions in the model. The cell densities are randomly taken from the interval $(0, 1]$.

3. We create two groups of affected cells that are far away from each other to study the evolutional patterns of two interaction neighbourhoods.

4.1.4 Boundary Conditions

Boundary conditions affect the implementation and the simulation outcome. Therefore, the information for boundary conditions is important. During some exploration, we used periodic boundary conditions to simulate an infinite domain with a finite number of cells. For the two-dimensional simulations, we use absorbing boundary conditions primarily because periodic boundary conditions are computationally expensive.

4.1.5 Simulating Birth-Death Process in Each Cell

We choose computational parameters for the microscale birth-death process to balance computational time versus predictability of the microscale model behaviour. We set the population capacity $C = 100$. Since the birth rate and death rate of the microscale birth-death process model depends on the time scale, we need the number of microscale time steps τ such that the generated individuals do not exceed C. In consideration of computational efficiency, we also choose $N_t = 100$. Given a specified initial cell density π^*, we can obtain

the population size n^* by multiplying the density π^* by the capacity C in the cell. If the population size n^* is smaller than 1, then we set the population size to be 0.

For each birth-death process within the affected cells, we generate random integers X_τ taking values -1, 0 or 1 for microscale time step $\tau = 1, \ldots, 100$. The probabilities for generating each value depends on the cell density and the predefined birth and death rate:

$$X_\tau = \begin{cases} 1, & \text{with probability } b_{ij,\tau}, \\ -1, & \text{with probability } d_{ij,\tau}, \\ 0, & \text{with probability } 1 - b_{ij,\tau} - d_{ij,\tau}. \end{cases}$$

where b and d represent the predefined birth and death rate, respectively.

In the mountain pine beetle application, we assume there are multiple trees in each cell, and each tree comes with a different density and birth-death process. Therefore, we simulate the birth-death process multiple times in each cell, and take the average of those processes to represent the average behaviour of those trees in the cell. In the simulation studies, we simulate the birth-death process for 20 realizations in each cell, and take the average of the final population size to obtain the updated cell density.

4.1.6 Evolution Patterns for Two Interaction Neighbourhoods

We create two disconnected groups of affected cells with different parameter settings to observe the evolution of two interaction neighbourhoods that eventually join. One interaction neighbourhood has a higher chance of dispersing populations. Once the two interaction neighbourhoods join together, we take the average of the parameter settings for the combined interaction neighbourhood.

4.1.7 Sequential Approach to Link the Scales

We use a sequential method to exchange cell densities between the microscale and the macroscale cellular automata models. We obtain the population size from the microscale birth-death process model, we transform it into the cell density and simulate the dispersion process at the macroscale to obtain the updated cell density. Then we transform the updated cell density to a population size that is passed to the birth-death process for simulation over the microscale steps until the macroscale time step is reached. We march the microscale and macroscale simulations over a sufficiently long time to obtain the evolutional dispersion patterns.

We provide a comprehensive algorithm for simulating the dispersion process:

Algorithm 2 Algorithm for simulating the dispersion process

1: Set up the implementation grid of $m \times n$ cells, where $m = n = 50$.
2: Set the initial pattern for a small number of affected cells on the grid.
3: **for** macroscale time step $t = 1, 2, \ldots$ **do**
4: Simulate multiple realizations of birth-death process and compute the average number of individuals at the microscale within each cell for microscale steps to reach the macroscale time step.
5: Compute the ratio of the average number of individuals and the cell capacity at the microscale and use the ratio as the cell density for dispersion process at the macroscale.
6: Simulate the dispersion process at the macroscale.
7: Convert the density to population size by multiplying by capacity to produce an initial population for the microscale simulation in each cell.
8: Alternate the microscale and macroscale simulations for sufficient time steps to obtain the ending cell states.
9: During the simulation, compute the statistics that are used to describe the evolution.
10: **end for**

4.2 Analysing Dispersion Patterns

We use two statistics to quantify the properties of dispersion patterns. The velocity of dispersion is used to quantify how quickly populations disperse under certain scenarios, and the complexity of dispersion patterns describes the roughness of the pattern boundaries and provides information about the shape and size of dispersion patterns. The two statistic results presented in Chapter 5 is computed by averaging of 500 realizations to reduce the effects of stochasticity.

4.2.1 Velocity of Dispersion

Since the dispersion patterns usually present irregular shapes, it is difficult to determine the extent to which populations move away from a starting location to a destination location. We use the change in the area of affected cells to measure the velocity of dispersion. For nondirectional movements, we assume the populations move in six directions with equal probabilities. We draw a circle that has the same area as the interaction neighbourhood, and use the change in the radius as a means to measure the velocity of dispersion.

The affected cells in the same interaction neighbourhood have individual densities. We compute the centre of mass for each interaction neighbourhood to use as centroid of the circle. The position of centre of mass is calculated as the weighted position of all cells in the same interaction neighbourhood. The radius of the circle is computed in a way that the area of the circle equals the total area of all affected cells in the interaction neighbourhood.

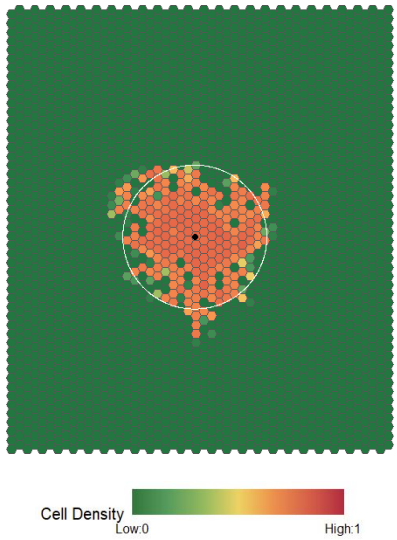


Figure 4.1: An example of an interaction neighbourhood. The colour of the cells represent the density of the cells. The black point in the middle of the interaction neighbourhood is the centre of mass. The radius of the circle is used to measure the area of the interaction neighbourhood.

4.2.2 Complexity of Dispersion Patterns

The boundary of interaction neighbourhood forms the outline of the interaction neighbourhood, and its complexity is a way to quantify the behaviour. We define the measure of complexity as the number of boundary edges divided by the number of affected cells in the interaction neighbourhood. The complexity increases if there is complex geometry in the boundary normalized by the number of cells in the interaction neighbourhood.

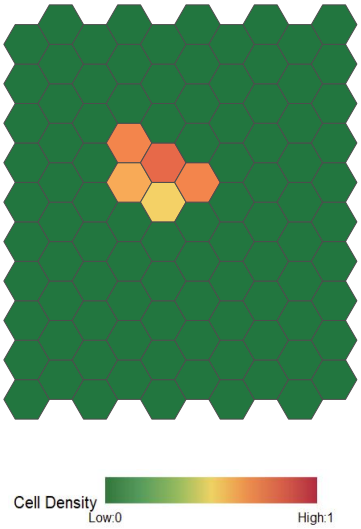


Figure 4.2: An example of an interaction neighbourhood with five affected cells.

There are five affected cells in Figure 4.2, and the total number of boundary edges is 16, according to the definition, the complexity of the pattern is thus $16/5 = 3.2$. The complexity can be used to assess the roughness of the pattern boundary. With the same number of affected cells, the complexity is smaller for the pattern in which the affected cells are more connected with each other. When the interaction neighbourhood becomes extremely large, it is not feasible to use the definition to compute the complexity. We provide an alternative formula for computing the complexity. For each affected cell in the interaction neighbourhood, we count the number of its adjacent affected neighbours n_a, where $a = \{1, \ldots, N_a\}$, and N_a is the total number of affected cells in the pattern. The formula for computing complexity is

$$\text{complexity} = \left(6N_a - \sum_{a=1}^{N_a} n_a\right) / N_a \tag{4.2.1}$$

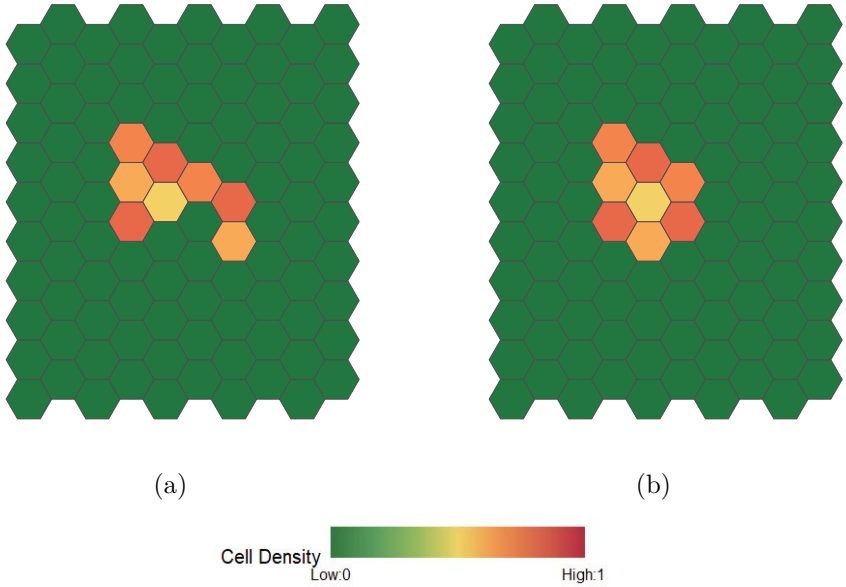


Figure 4.3: Two distinct interaction neighbourhood with identical number of affected cells.

There are identical number of affected cells in the interaction neighbourhood in Figure 4.3, but they display different shapes. The total number of boundary edges is 28 for pattern (a) and 21 for pattern (b). Thus, the complexity for two patterns is 3.5 and 2.6, respectively. In general, the complexity is smaller for patterns with a "smoother" boundary.

4.2.3 Distribution of Quantitative Statistics

We create two sets of experiments to study the distribution of quantitative statistics. In each set of experiments, we generate 1000 samples and 50 realizations for each combination

of parameter values for 30 time steps, then we take average of the 50 realizations to obtain an average of those quantitative statistics. We choose the parameter values in the experiments to obtain evolution patterns that grow without limit at a fixed speed. The experiment data obtained is also used in Chapter 6 for inversion purpose.

In the first set of experiment, we assume the percentage of redistributed population γ_1 and the percentage of survival population γ_2 are both normally distributed with mean being 0.8 and standard deviation 0.4/6, i.e. γ_1, γ_2 follows $N(0.8, 0.4/6)$. In the second set experiment, we let parameter ρ in the survival probability to follow a Beta distribution with a modified domain, i.e. ρ follows $Beta(2, 4)$ with $\rho \in [1/6, 5/6]$. We set the parameter ϕ to be $1/12$ and ξ to be 1 to reduce the impact of cell densities on the survival probability.

4.3 Parallel Computing

The simulations in this project were coded in R programming language and executed on the Compute Canada's high-performance computation system. We simulate 100 realizations for these experiments in Phase I to compute the average complexity of the patterns. Since each realization is independent of the other, and no effect is needed to communicate between the realizations, we use parallel computing to carry out the simulations simultaneously. When we investigate the evolutional patterns of two interaction neighbourhoods, we use apply functions in R to simulate the dispersion process for two groups at the same time.

Chapter 5

Numerical Investigations of Model Behaviour

In this chapter, we investigate the properties of the SCA model through numerical simulations. The SCA model can exhibit complex behaviour, even though the component parts are simple. The complexity arises from local interactions between the components and governed by simple rules which can lead to complex global behaviour and interesting geometric dispersion patterns. The model is nonlinear and stochastic. This complicates theoretical analysis, since techniques that apply to linear Markov chain models are not easily applicable. We therefore conduct a numerical investigation to provide insight into model behaviour and reveal some qualitative features of the model.

The simulations in this chapter aim to explore the behaviour of the model under various scenarios and parameter values. We vary aspects of the model in a systematic fashion. We present details of the simulations, including the experimental setup, parameter settings and simulation results.

We begin with an example of a general analysis that reveals some common characteristics of the model and the tools we use to investigate the behaviour. To give a sense of the dynamic behaviour, we present plots at specified times, see Figure 5.1. The figure shows the solution at six time steps, giving a sense of the evolution of the interaction neighbourhoods. As described, we also compute quantitative statistics of complexity of patterns and velocity of dispersion to describe the evolution patterns, illustrated in Figure 5.2. Detailed information for calculating the two statistics can be found in Chapter 4.

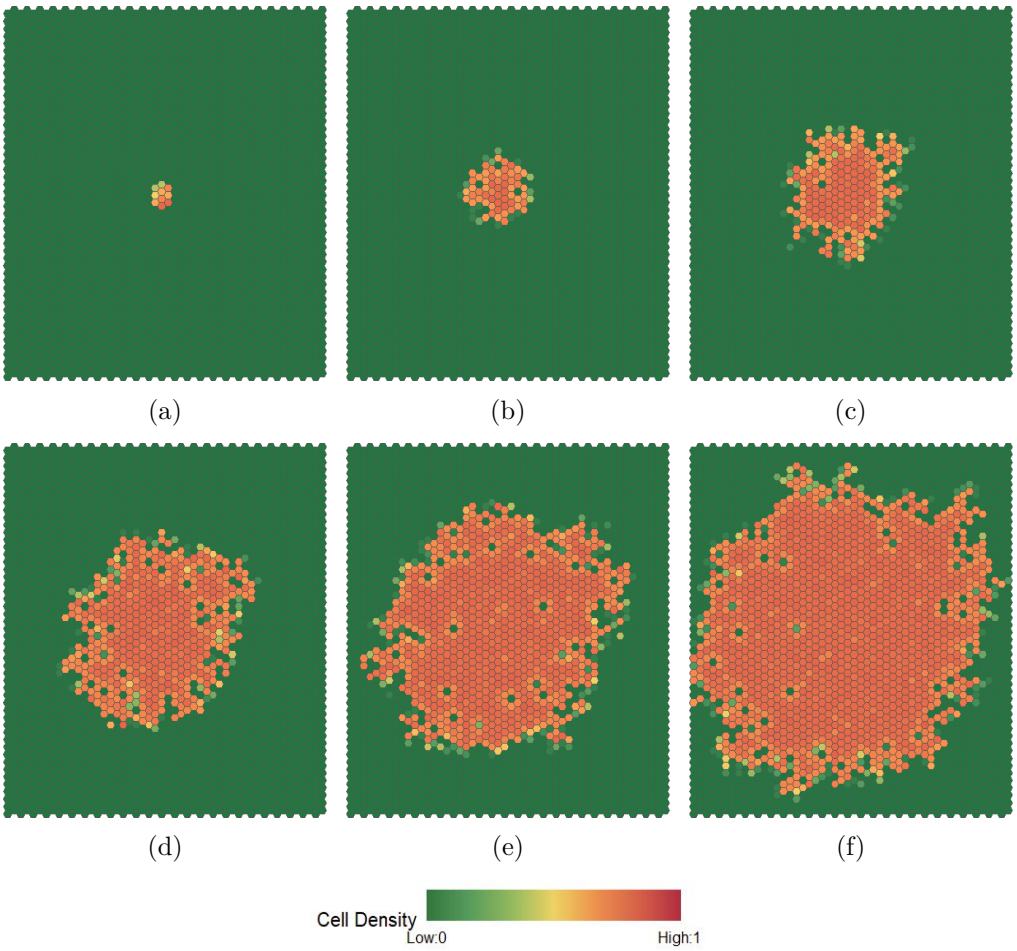
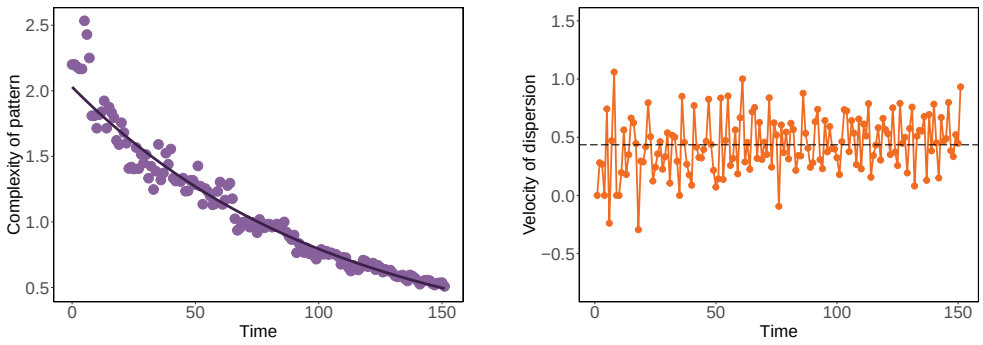
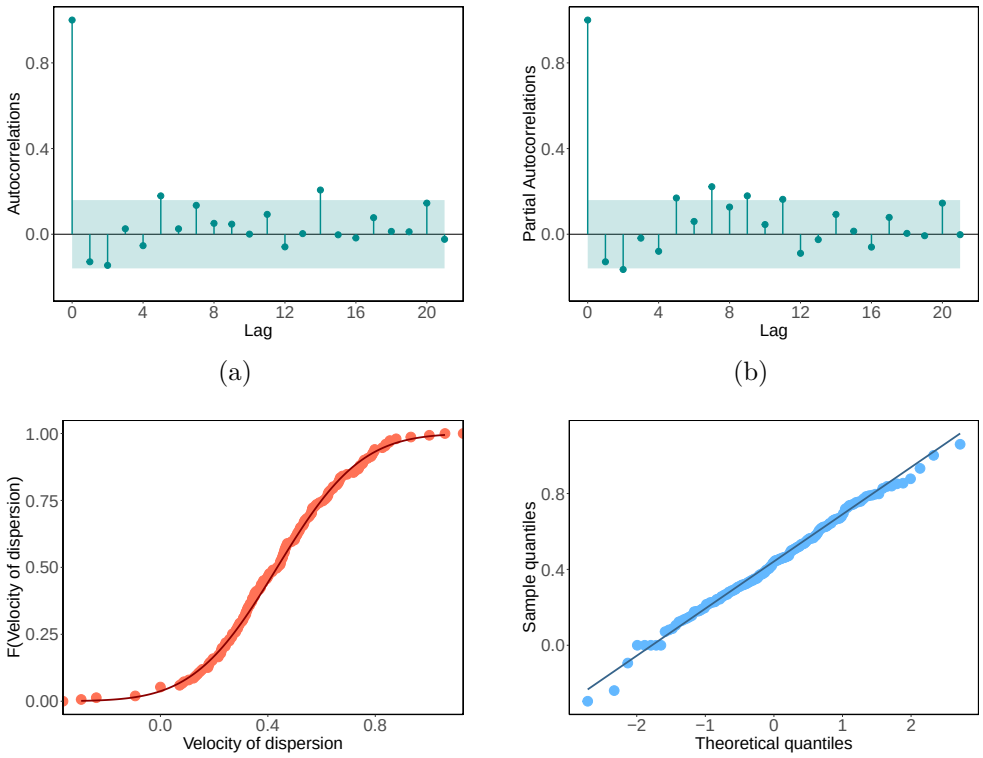


Figure 5.1: Evolution of patterns at specific time $t = 0, 30, 60, 90, 120, 150$. (a) $t = 0$. (b) $t = 30$. (c) $t = 60$. (d) $t = 90$. (e) $t = 120$. (f) $t = 150$.



(a) (b)

Figure 5.2: Plots of the complexity of the patterns at each time with fitted curve and velocity of dispersion of the model over time for one realization. (a) Complexity of patterns at each time with a fitted curve. The points represent complexity of patterns at each time, the curve represents the fitted model of the complexity over time with ($\ln(\text{complexity}) = 0.71 - 0.01 \times \text{time}, R^2 = .97, F(1, 150) = 5211, p < .001$). (b) Velocity of dispersion. The black dashed line represents the average dispersion velocity.



(a) (b)

(c) (d)

Figure 5.3: Plots for velocity of dispersion for one realization. (a) Autocorrelation plot. (b) Partial autocorrelation plot. (c) Empirical cumulative distribution function plot. (d) Normal QQ plot.

Generally, the complexity of patterns decreases exponentially over time, as the boundary becomes "smoother" and the interaction neighbourhood grows, see Figure 5.2. The velocity of dispersion oscillates around a straight line, which suggests that an interaction neighbourhood size grows at a predictable rate. According to the autocorrelation plot (ACF) and partial autocorrelation plot (PACF) in Figure 5.3, there is no apparent structure in the residuals, and the lag order of the correlations that lie outside the bands are not multiples of each other, so the ACF and PACF plots resemble white noise. As a complement to the visual examination, a Box-Ljung test ($\chi^2 = 29.30, df = 20, p = .08$) also suggests that the velocity of dispersion resembles white noise. The empirical cumulative distribution function (eCDF) plot and the normal QQ plot both indicate that the velocity of dispersion follows a Gaussian distribution and the Kolmogorov–Smirnov test for normality ($D = 0.11, p = .33$) verifies that the velocity of dispersion resembles Gaussian white noise.

Due to the stochastic nature of the model, the evolution patterns do not repeat even with identical initial patterns and parameter settings. We show evolution patterns from three realizations using the same initial pattern and parameter values in Figure 5.4. Despite the stochastic variation, we see in Figure 5.5 that the evolution patterns have roughly the same complexity of pattern, so the coefficients for three fitted models are fairly close. The velocity of dispersion for three realizations also behave similarly. Therefore, we only present the analysis results for realization 3 in Figure 5.6. There is no apparent structure in the ACF and PACF plots and the Box-Ljung test ($\chi^2 = 24.55, df = 20, p = .22$) indicates that the velocity of dispersion resemble white noise. The eCDF, normal QQ plot and the Kolmogorov–Smirnov test for normality ($D = 0.07, p = .95$) suggest that the velocity of dispersion follows a Gaussian distribution. Thus, we conclude that the velocity of dispersion for realization 3 resembles Gaussian white noise. The results are characteristic of the general trend in other realizations. These observations suggest that while a given initial pattern and parameter settings leads to different pattern sequences, the sequences have similar complexity and dispersion statistics.

We show evolution patterns for three different initial patterns with identical parameter settings in Figure 5.7. The complexity of patterns and velocity of dispersion indicate that different initial patterns lead to similar final patterns in the long run. We conclude that the model has a kind of long time stability property.

These simulation results all exhibit non-linear dynamics expected in the dispersion process. Though there are relatively few parameters, the model produces a wide range of behaviours, which suggests that it can be useful for modelling complex phenomena.

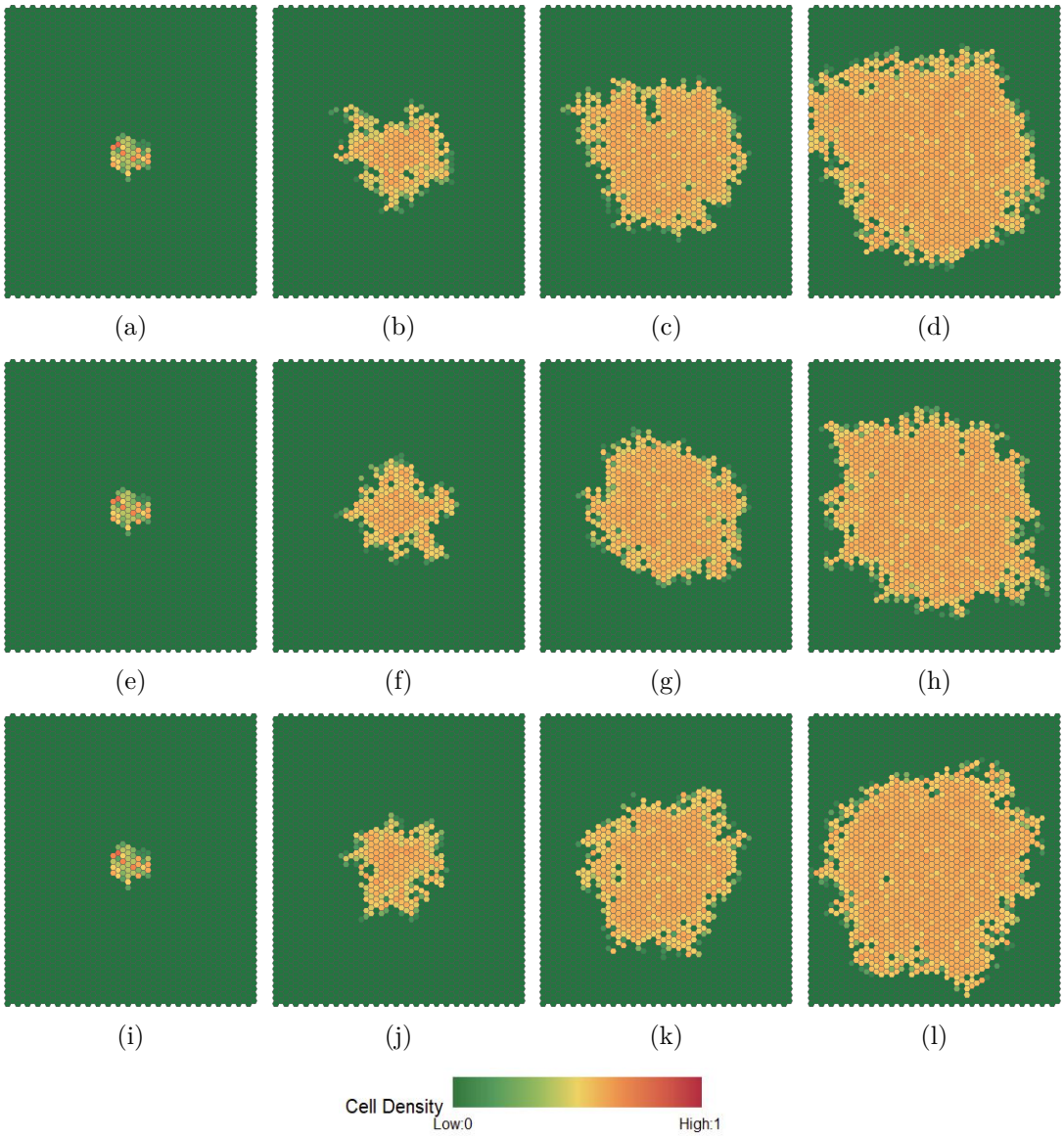
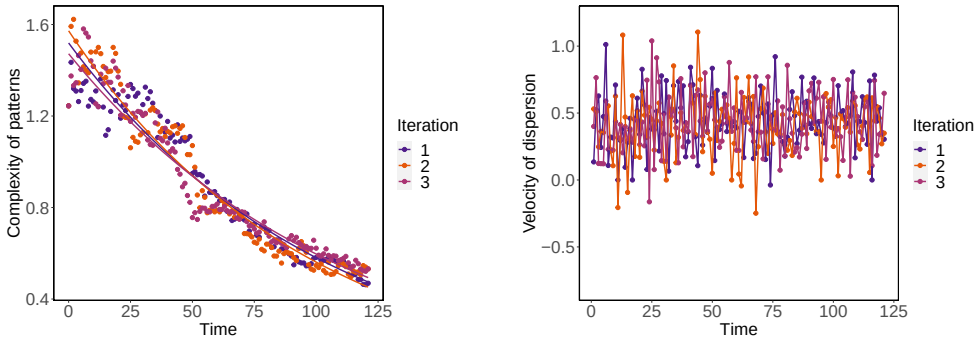
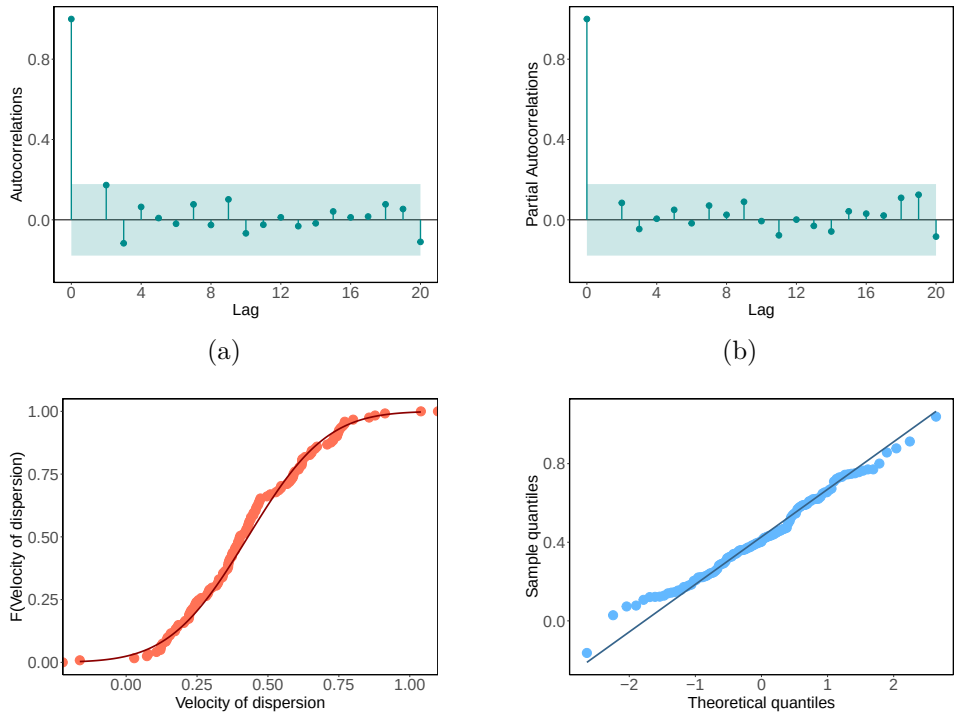


Figure 5.4: Evolution patterns at specific time $t = 0, 40, 80, 120$ for three realizations with the same initial pattern and same parameter settings. Each row presents the evolution patterns for one realization. (a) $t = 0$. (b) $t = 40$. (c) $t = 80$. (d) $t = 120$. (e) $t = 0$. (f) $t = 40$. (g) $t = 80$. (h) $t = 120$. (i) $t = 0$. (j) $t = 40$. (k) $t = 80$. (l) $t = 120$.



(a) (b)

Figure 5.5: Plots of complexity of the patterns at each time with fitted curve and velocity of dispersion with identical initial patterns and parameter settings for three realizations. (a) Complexity of patterns at each time with fitted curve. The points represent complexity of patterns at each time, and the curve represents the fitted model of the complexity over time, with ($\ln(\text{complexity}) = 0.42 - 0.01 \times \text{time}, R^2 = .95, F(1, 120) = 2253, p < .001; \ln(\text{complexity}) = 0.45 - 0.01 \times \text{time}, R^2 = .95, F(1, 120) = 2223, p < .001, \ln(\text{complexity}) = 0.39 - 0.01 \times \text{time}, R^2 = .95, F(1, 120) = 2198, p < .001$) for realization 1, 2, and 3, respectively. (b) Velocity of dispersion.



(a) (b)

(c) (d)

Figure 5.6: Plots for velocity of dispersion with identical initial patterns and parameter settings for one realization. (a) Autocorrelation plot. (b) Partial autocorrelation plot. (c) Empirical cumulative distribution function plot. (d) Normal QQ plot.

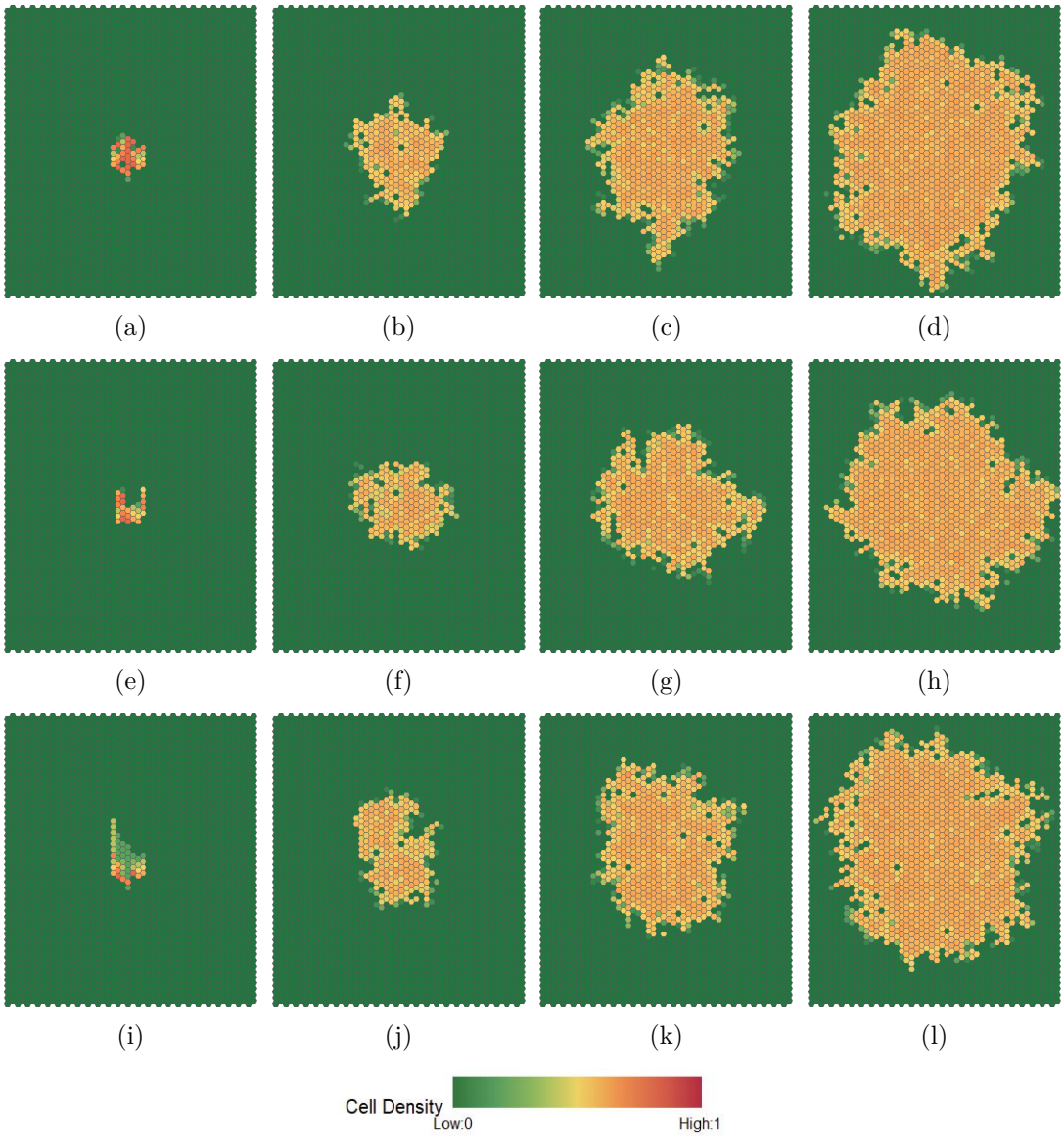
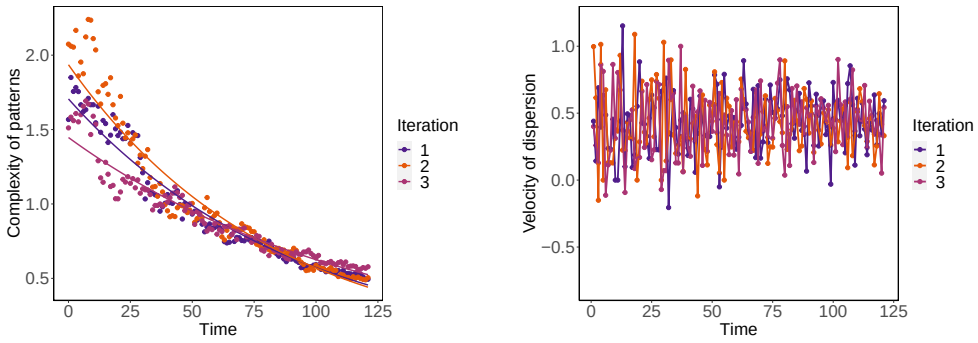


Figure 5.7: Evolution patterns at specific time $t = 0, 40, 80, 120$ for three realizations with different initial patterns and the same parameter settings. Each row presents the evolution patterns for one realization. (a) $t = 0$. (b) $t = 40$. (c) $t = 80$. (d) $t = 120$. (e) $t = 0$. (f) $t = 40$. (g) $t = 80$. (h) $t = 120$. (i) $t = 0$. (j) $t = 40$. (k) $t = 80$. (l) $t = 120$.



(a) (b)

Figure 5.8: Plots of complexity of the patterns at each time with fitted curve and velocity of dispersion with the same parameter settings for three different initial patterns. (a) Complexity of patterns at each time with fitted curve. The points represent complexity of patterns at each time, the curve represents the fitted model of the complexity over time, with $(\ln(\text{complexity}) = 0.53 - 0.01 \times \text{time}, R^2 = .97, F(1, 120) = 3390, p < .001; \ln(\text{complexity}) = 0.66 - 0.01 \times \text{time}, R^2 = .95, F(1, 120) = 2458, p < .001, \ln(\text{complexity}) = 0.37 - 0.01 \times \text{time}, R^2 = .95, F(1, 120) = 1992, p < .001)$ for realization 1, 2, and 3, respectively. (b) Velocity of dispersion.

5.1 Goals for the Simulation Study

The goals for the simulation study includes:

1. Develop an understanding of the dispersion process, including hidden behaviours.
2. Explore behaviour of the model for different scenarios or conditions.
3. Perform sensitivity analysis of the model to identify the critical parameters that influence the behaviour of the model.
4. Develop statistics that can quantify important behaviour of the model and develop an understanding of what kinds of model behaviour can be used to calibrate the model using experimental data.

5.2 Simulation Methodology

We focus on three properties: the population available to disperse, the probability that population disperses and the probability that population successfully immigrates to a new region. These properties jointly determine the patterns generated by the stochastic process model. The microscale model provides the population source for dispersion, while the change of parameter settings in the microscale model affects the population size during dispersion.

Extensive simulation studies suggest that the behaviour approaches three limiting situations, illustrated in Figure 5.9, which are

1. Fade away completely;
2. Settle into a fixed finite size;
3. Grow without limit at a fixed speed.

The experiments suggest that even though the underlying rules for the model are simple, the model can generate complex behaviour characteristic of dispersion processes found in nature.

5.2.1 Impact of Boundary Conditions

The model requires boundary conditions to give values at the boundaries of the domain. Boundary conditions are determined by the application of the model. In the previous discussion, we focus on the scenarios where the simulation runs are performed assuming there is no boundary effect involved in the dispersion process, i.e., the border of the grid does not affect the model behaviour. Periodic boundary conditions that connect the border cells are closest to simulating an infinite grid, however, they are computationally expensive to use. Besides, periodic boundary conditions are usually not realistic for real-world applications.

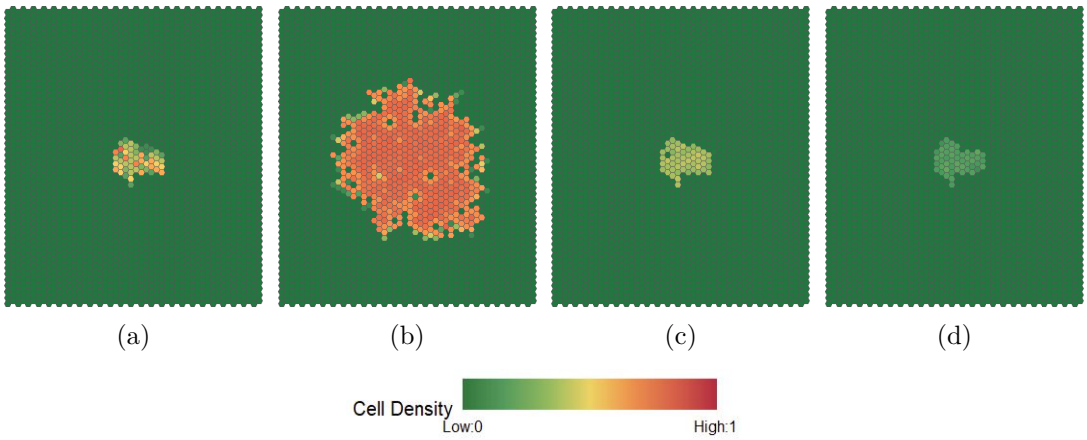


Figure 5.9: Initial pattern and classes of final patterns generated by the evolution of the model from the same initial configuration when the parameter settings vary. (a) Initial pattern. (b) Final pattern: growth without limit. (c) Final pattern: settle into a fixed finite size. (d) Final pattern: fade away.

We choose "absorbing" boundary conditions in the model for practical purposes. We create auxiliary "external" cells that have the same updating rule as cells in the grid, those auxiliary cells can receive a proportion of population from the cells on the boundary but they cannot send the population to other cells in the interaction neighbourhood. Computationally, absorbing boundary conditions are advantageous in the sense that they can simulate real-world applications with a minimum amount of computation and storage.

In this section, we aim to explore the boundary effects on model behaviours. In Figure 5.10, we observe that if there is sufficient density, the density tends to fills the entire space after a sufficiently long time, except for a few vacant cells on the grid, regardless of the grid size. Moreover, as time proceeds it becomes increasingly unlikely to find cells that are vacant.

The simulations suggest that cell densities vary continuously with respect to the parameter settings in the transition rule. Most cells that are on the boundary of the grid have a lower density than the cells that are in the centre of the grid, which is because those cells lose a small proportion of density from absorbing boundary conditions.

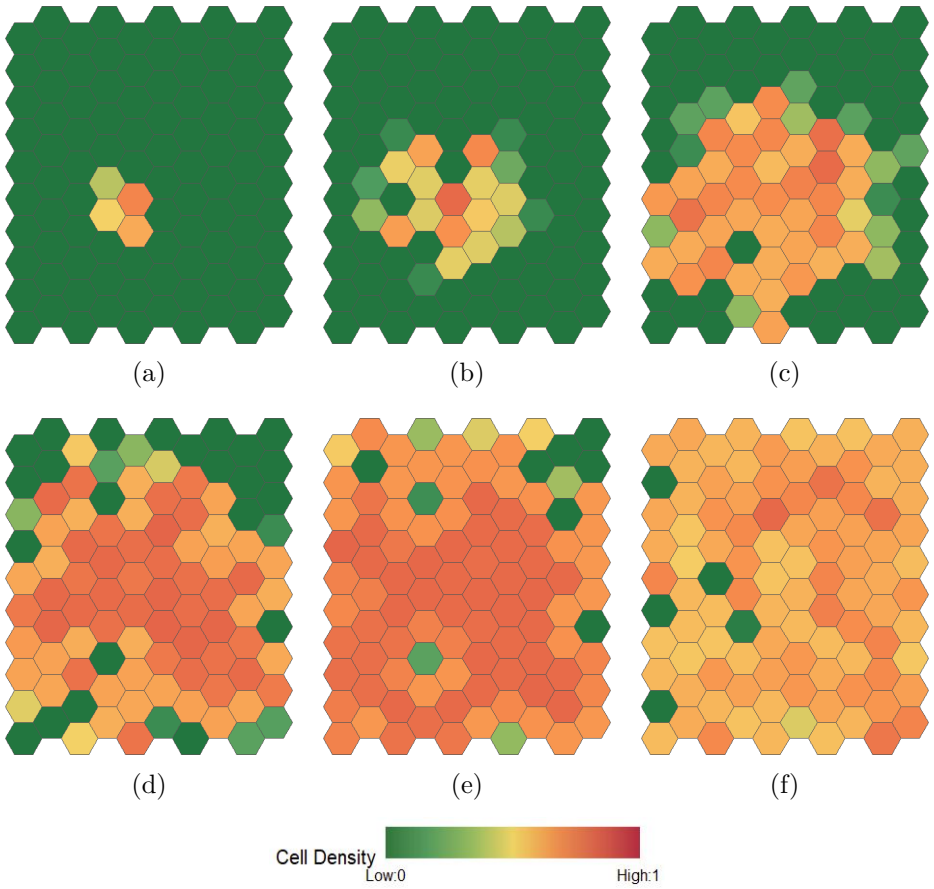


Figure 5.10: Evolution patterns at specific time $t = 0, 10, 20, 30, 40, 50$ on a domain size of 20×20 to reflect the impact of absorbing boundary conditions. (a) $t = 0$. (b) $t = 10$. (c) $t = 20$. (d) $t = 30$. (e) $t = 40$. (f) $t = 50$.

5.2.2 Experiment Setup

In the simulation studies, we evaluate the evolution patterns to verify that the model is capable of capturing expected behaviours common to the dispersion phenomena. In our application, we focus on scenarios when the model produces final patterns that grow without limit. We also use the complexity of patterns and dispersion velocity to provide additional information when necessary.

Since we conduct 1000 experiments and 50 realizations in each of those experiments to investigate the distribution of quantitative statistics including the average cell density, average number of affected cells, and average complexity of the interaction neighbourhood, it is not feasible to use a large grid size. Besides, in our previous discussions, the density tends to fill the entire grid regardless of the parameter settings when the model is run long enough and the quantitative statistics become indistinguishable. In this case, we want to conduct the experiments with the minimum grid size and run time that is capable of producing perceptible patterns that are sensitive to different parameter settings. Based on our exploration in model behaviour with different grid sizes and run time, we choose to simulate the model on a grid of 30×30 cells with 30 time steps to investigate the distributions of the quantitative measurements.

In the simulation studies to investigate model behaviours, we set up four phases to investigate the behaviour of the model. In each phase, we start with a specified interaction neighbourhood and then observe the impact of different parameter values on evolution patterns. In phase I, we conduct sensitivity analysis for each parameter in the model, investigating the impact of one parameter or one set of parameters on behaviours of the model. In phase II, we investigate the interactions between different parameter regimes of the model. We also investigate the change point for settling into a stable configuration. In phase III, we investigate the impact of geographical properties of the domain. We change the attributes of certain cells to make them suitable for various physical conditions, such as a lake or a region for which individuals can pass through. We also consider the evolution patterns affected by outer forces such as air flow. In phase IV, we investigate the evolution patterns of two different interaction neighbourhoods, which represent two distinct population groups with varying physical properties and we explore the evolution patterns of the two interaction neighbourhoods. All of the experiments require some stability in model behaviour, therefore we simulate the process on a grid of 50×50 cells that defines a domain big enough to observe the evolution patterns with up to 120 time steps before the boundary of the domain affects the dispersion patterns.

5.3 Simulation Parameters

There are seven parameters at the macroscale that determine the dispersion pattern. A detailed explanation is provided in Chapter 3. We focus on two probability parameters

to characterize the stochastic behaviour of the model, the inclusion probability and the survival probability. Those two probability rules are connected with different stages of the dispersion process. The inclusion probability is linked to the emigration stage in which the potential cells for dispersion are selected. The inclusion probability at a specific time point t is defined as 3.3.3, where $w_{ij}^{(t)}$ and $d_{ij}^{(t)}$ represent the inclusion weight and the maximum of relative density difference between neighbouring cells for cell (i, j), respectively. For a given cell, the two parameters α and β are used to decide the likelihood of being involved in the dispersion process.

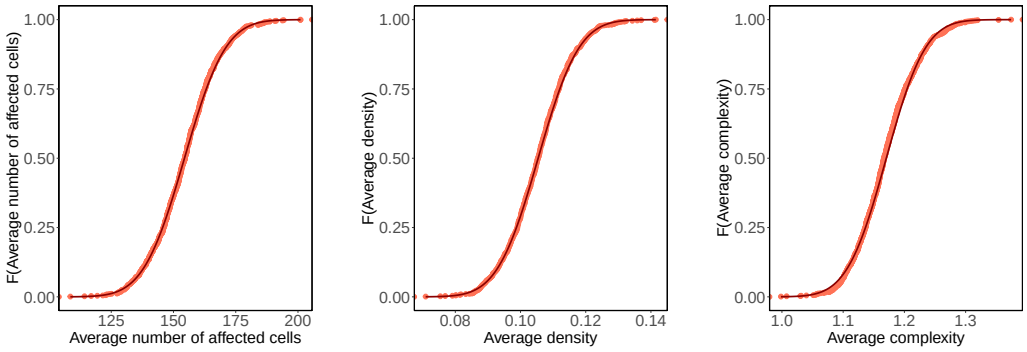
The survival probability is linked to the immigration stage in which the individuals establish population in new cells. It is defined as 3.3.4 for cell (i, j) at time t. There are three parameters ϕ, ρ and ξ in the survival probability that are used to determine the likelihood that the population successfully disperses to the new cells. In the dispersion process, the amount of individuals provides a source of cell densities that determines the pattern. The parameters γ_1 and γ_2 are used to regulate the cell density movements in each interaction neighbourhood. In addition, we investigate the impact of cell density. The cell density is regulated by the spread percentage γ_1 and survival percentage γ_2 at the macroscale, and the birth-death process at the microscale.

There are five parameters at the microscale including birth rate, death rate, number of microscale time steps in each macroscale time step, population capacity and number of replications of the model in each cell. However, the microscale process mainly contributes to the cell density, therefore, we only investigate the joint effect of the five parameters on dispersion patterns.

5.4 Distributions of Quantitative Measurements

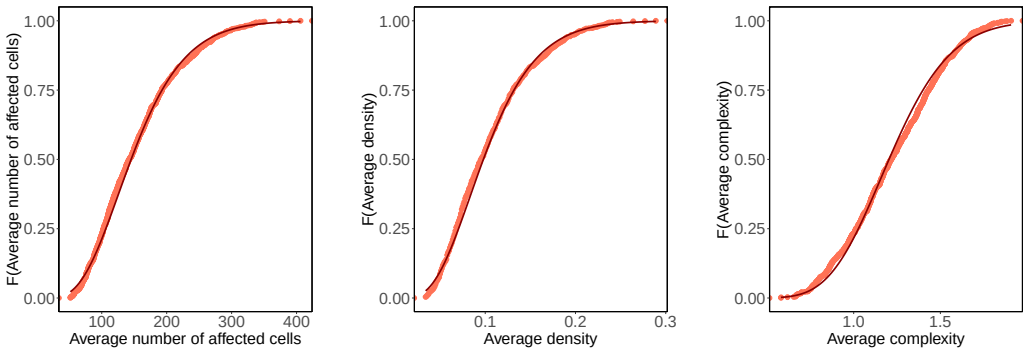
We aim to investigate the distributions of the three quantitative measurements including average number of affected cells, average density and average complexity of the evolution patterns. We set up two experiments with different priors on the parameters in the model. In the first experiment, we assign a normal prior $N(0.8, 0.4/6)$ for both spreading rate γ_1 and survival rate γ_2. In the second experiment, we assign a normal prior $N(0.8, 0.4/6)$ for γ_1, and a Beta prior beta(2, 4) for ϕ in the survival probability with $\phi \in [1/6, 5/6]$, we set $\rho = 1/12$ to make the quantitative measurements less dependent on the cell density. According to the eCDF plots in Figure 5.11 and the Kolmogorov-Smirnov test results, average number of affected cells ($D = 0.11, p = .23$), average density ($D = 0.53, p = .27$), average complexity ($D = 0.06, p = .84$), the estimated normal probability distribution is a good match to the empirical distribution of the three quantitative measurements in the first experiment. According to the eCDF plots in Figure 5.12 and the Kolmogorov-Smirnov test results, the average number of affected cells ($D = 0.05, p = .53$), average density ($D = 0.45, p = .46$), average complexity ($D = 0.09, p = .26$), and the estimated gamma

probability distribution appears to be a good match to the empirical distribution of the three quantitative measurements in the second experiment.



(a) (b) (c)

Figure 5.11: Empirical distribution functions along with the distribution functions for the normal probability with the corresponding estimated mean and variance for each quantitative measurement. (a) Average number of affected cells. (b) Average density. (c) Average complexity.



(a) (b) (c)

Figure 5.12: Empirical distribution functions along with the distribution functions for the gamma probability with the corresponding estimated shape parameters for each quantitative measurement. (a) Average number of affected cells. (b) Average density. (c) Average complexity.

5.5 Experiments

5.5.1 Phase I: Sensitivity Analysis for the Parameters

Generally, sensitivity analysis is the study of how uncertainty in the output of a model can be apportioned to different sources of uncertainty in the model input. Conducting sensitivity analysis has the potential to improve parameter calibration by making the process more systematic. The motivations of the sensitivity analysis below includes:

1. Determine which parameters have the most effect on model output.
2. Check if the model can be simplified by eliminating parameters that have a minor effect on the model output.
3. Identify critical regions in the space parameters, and identify parameters whose combined effect leads to complex behaviour.
4. Determine scenarios in which the input parameters have greatest effect on model output.

Sensitivity analysis can be performed at the local or global scale. A local analysis examines the sensitivity of one single parameter while a global analysis examines sensitivity with respect to the entire parameter domain. Due to the higher computational complexity, we prefer the local analysis and focus on the impact of a single parameter while keeping the other inputs fixed.

However, due to the large number of parameters and the complex relationships among the massively interacting parameters in the model, analytical exploration of the model behaviour is very challenging. Computational expense limits the number of simulations that can be performed, consequently, it is not feasible to evaluate the impact of every single parameter and every parameter combination on the model behaviour. Therefore, we perform sensitivity analysis for a few selected parameters of the model elements that govern the behaviour of the model.

5.5.1.1 Impact of Cell Density on Dispersion Patterns

There are two possible ways to vary the cell density in the model. One way is to vary the birth-death process at the microscale to vary the number of individuals produced and therefore to vary the input cell density at the macroscale. The other way is to vary the spread percentage and survival percentage at the macroscale.

5.5.1.1.1 Impact of the Birth-Death Process

We vary the death rate in the birth-death process with all the other parameters fixed to obtain different levels of population at the microscale. Therefore, we obtain different

levels of density at the macroscale. We choose high, medium, and low death rate at the microscale which corresponds to low, medium and high cell density at the macroscale, respectively. As we observe in Figure 5.13, the death rate that determines the population size at the microscale affects the complexity and the dispersion velocity of the interaction neighbourhood.

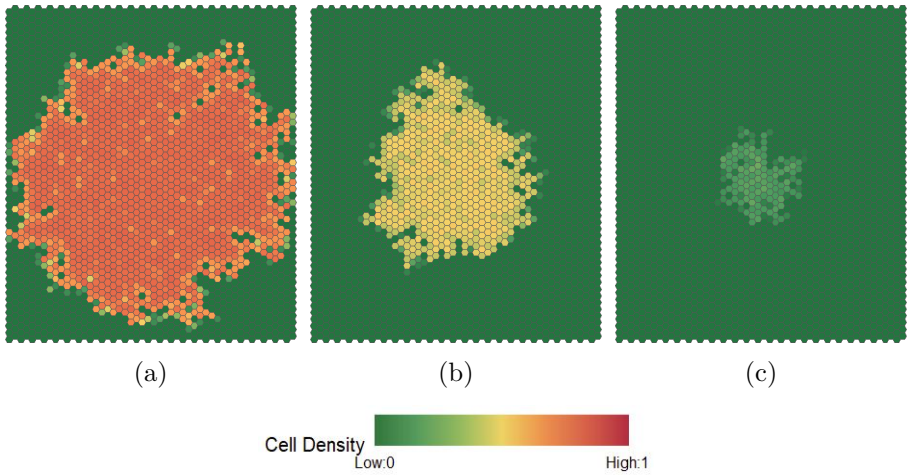


Figure 5.13: Final patterns at time $t = 100$ generated by the model from the same initial configuration with different death rates in the birth-death process at the microscale. (a) Low death rate ($d = 0.2$). (b) Medium death rate ($d = 0.5$). (c) High death rate ($d = 0.8$).

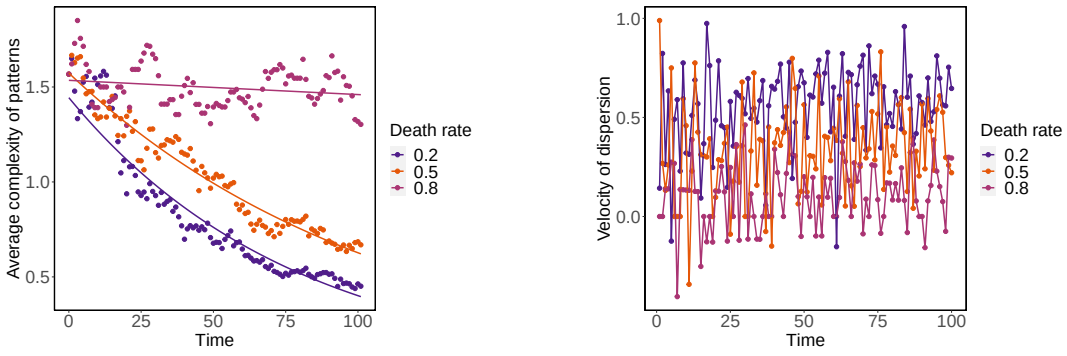


Figure 5.14: Average complexity of the patterns for 100 realizations with different death rates and velocity of dispersion of the model over the course of time for three realizations with different death rates at the microscale of the model. (a) Average complexity of patterns. The points represent the average complexity of patterns from 100 realizations at each time, and the curve represents the fitted model of the complexity over time, with$(\text{complexity} = 1.67 - 0.002 \times \text{time}, R^2 = .88, F(1, 99) = 697, p < .001)$, $(\ln(\text{complexity}) = 0.42 - 0.01 \times \text{time}, R^2 = .98, F(1, 99) = 6697, p < .001; \ln(\text{complexity}) = 0.49 - 0.01 \times \text{time}, R^2 = .99, F(1, 99) = 10980, p < .001)$ for death rate = 0.8, 0.5, 0.2. (b) Velocity of dispersion for one realization.

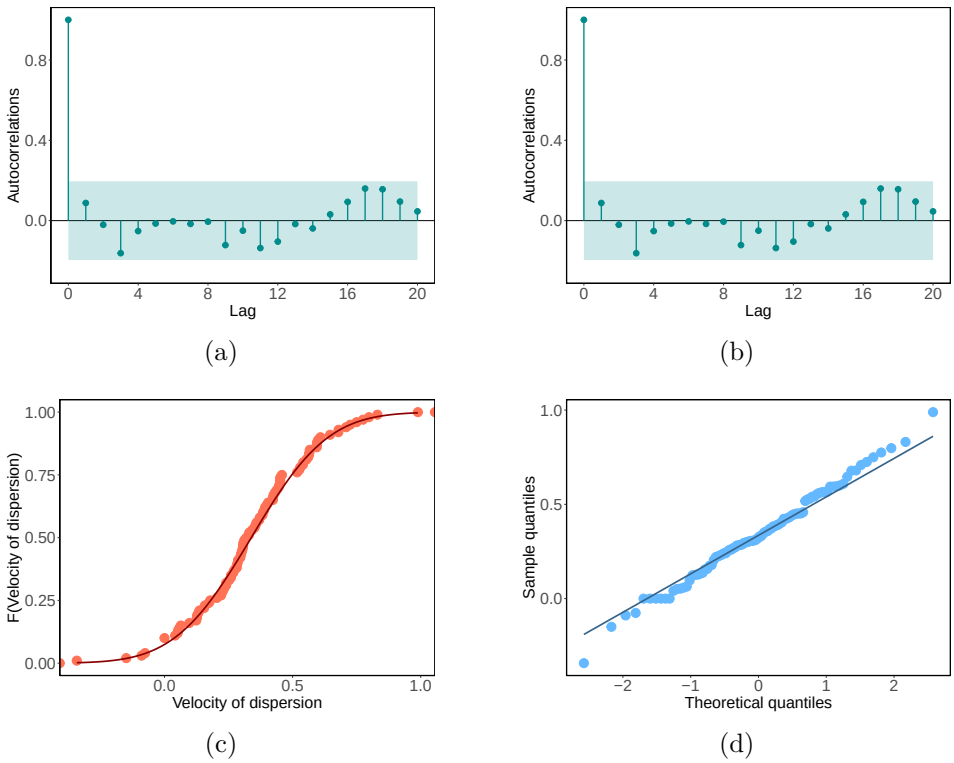
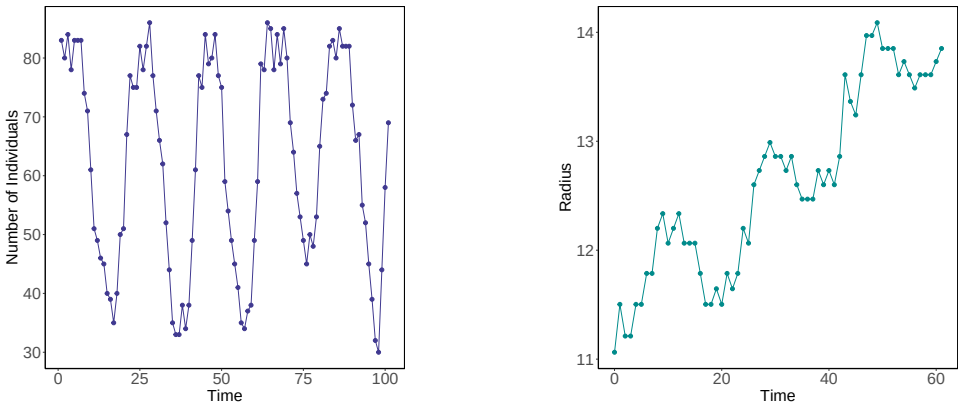


Figure 5.15: Plots for velocity of dispersion for the realization when death rate is 0.5. (a) Autocorrelation plot. (b) Partial autocorrelation plot. (c) Empirical cumulative distribution function plot. (d) Normal QQ plot.

In Figure 5.14, the average complexity decreases linearly when the death rate is at a very high level. This is because the average complexity generally decreases as interaction neighbourhood grows, but a high death rate in the birth-death process slows the growth of the interaction neighbourhood. While when the death rate is at level of 0.2 and 0.5, the average complexity decreases exponentially as the interaction neighbourhood grows at a much higher rate. When the death rate is at a lower level, the interaction neighbourhood expands at a much higher rate, as shown by the plot of velocity of dispersion oscillating around different levels of straight lines in Figure 5.14. This suggests that the magnitude of dispersion is dependent on population size. For the purpose of simplicity, we only present the analysis plots of the velocity of dispersion when the death rate is at the level of 0.5 in Figure 5.15, similar to our previous findings, the velocity of dispersion resemble Gaussian white noise. The velocity of dispersion have similar structure for the other two death rate levels.

We let the birth-death process show cyclic behaviour that occur in a 20 time step cycle, which represents the natural cyclic characteristics of the mountain pine beetle biology. The cell density at the macroscale also presents a cyclic behaviour, which lead to the same cyclic behaviour in dispersion patterns. We use the radius of the circle that equals the total area

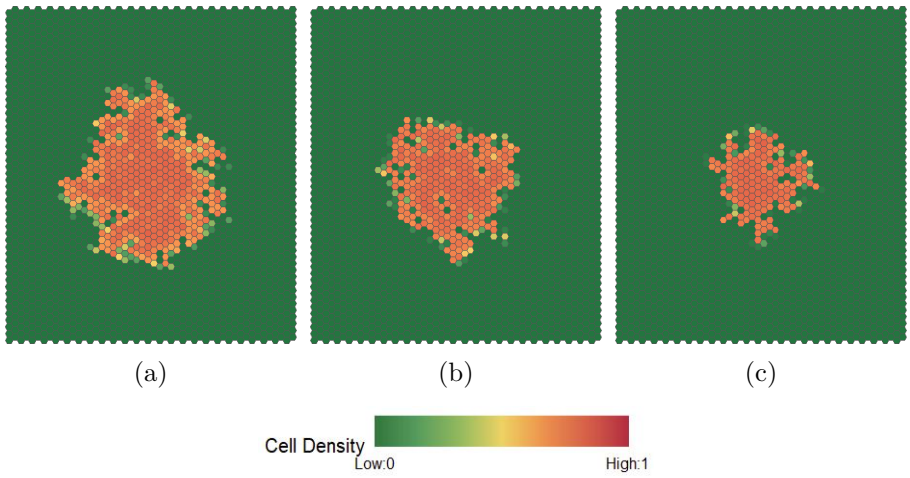
of all affected cells in the interaction neighbourhood to measure the neighbourhood growth over time. Apparently, the radius presents a cyclic behaviour in Figure 5.16. When there is enough individuals to pass into the macroscale model, the interaction neighbourhood grows, and when there is not enough individuals to pass into the macroscale model, the interaction neighbourhood stops growing and even shrinks in size.



(a) (b)

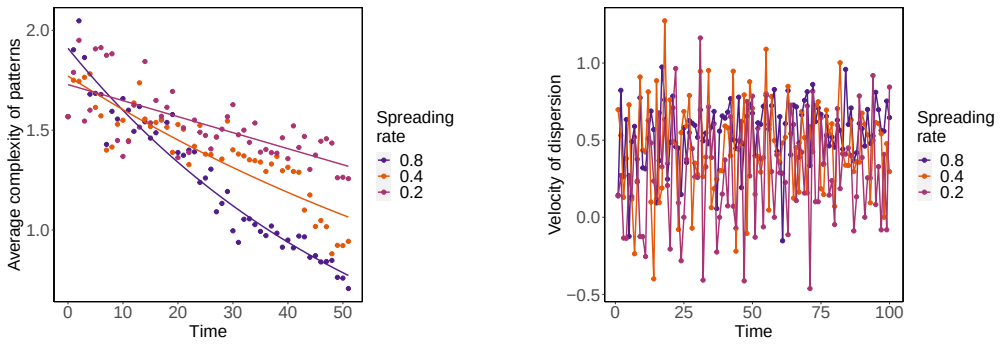
Figure 5.16: Cyclic behaviour of the birth-death process. (a) Number of individuals generated from the birth-death process over time given the initial population size is 90. (b) Radius of the circle that equals the total area of all affected cells in the interaction neighbourhood..

5.5.1.1.2 Impact of the Spreading Rate γ_1



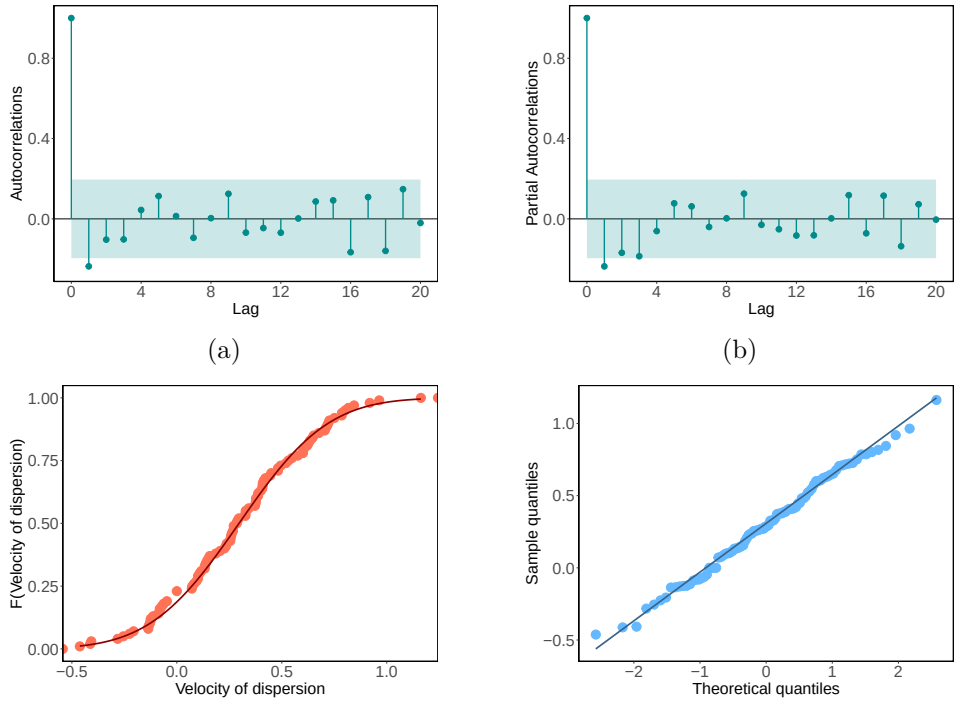
(a) (b) (c)

Figure 5.17: Final patterns at time $t = 50$ generated by the model from the same initial configuration with different spreading rates γ_1 at the macroscale. (a) High spreading rate ($\gamma_1 = 0.8$). (b) Medium spreading rate ($\gamma_1 = 0.4$). (c) Low spreading rate ($\gamma_1 = 0.2$).



(a) (b)

Figure 5.18: Average complexity of the patterns and velocity of dispersion of the model over the course of time for three realizations with different spreading rate at the macroscale of the model. (a) Average complexity of patterns. The points represent average complexity of patterns from 100 realizations at each time, the curve represents the fitted model of the complexity over time, with $\ln(\text{complexity}) = 0.42 - 0.01 \times \text{time}, R^2 = .98, F(1, 99) = 6697, p < .001$; $\ln(\text{complexity}) = 0.53 - 0.01 \times \text{time}, R^2 = .99, F(1, 99) = 25180, p < .001$, and $\text{complexity} = 1.69 - 0.01 \times \text{time}, R^2 = .99, F(1, 99) = 16720, p < .001$) for spread rate 0.8, 0.5, and 0.2, respectively. (b) Velocity of dispersion for one realization.



(a) (b)

(c) (d)

Figure 5.19: Plots for velocity of dispersion for one realization. (a) Autocorrelation plot. (b) Partial autocorrelation plot. (c) Empirical cumulative distribution function plot. (d) Normal QQ plot.

An alternative way to vary the cell density is to vary the spreading rate γ_1 at the macroscale of the model. The spreading rate regulates the percentage of populations that emigrate to new cells, therefore, different spreading rates lead to various model behaviours. As we observe in Figure 5.17, the size of the interaction neighbourhood is much larger when the spreading rate γ_1 is at a higher level, which is verified in the velocity of dispersion plot. This suggests the fact that the dispersion process is accelerated by the population size.

5.5.1.1.3 Impact of the Survival Rate γ_2

Another option to vary the cell density is to vary the survival rate γ_2 at the macroscale of the model. The survival rate determines the percentage of population that successfully immigrates into the new cells. We can see from the dispersion patterns in Figure 5.20, a larger survival rate leads to a faster increase in the size of the interaction neighbourhood, which once again indicates the positive influence of population size on dispersion process.

The average complexity of patterns decreases exponentially for the three survival rate levels, and the velocity of dispersion also resembles Gaussian white noise.

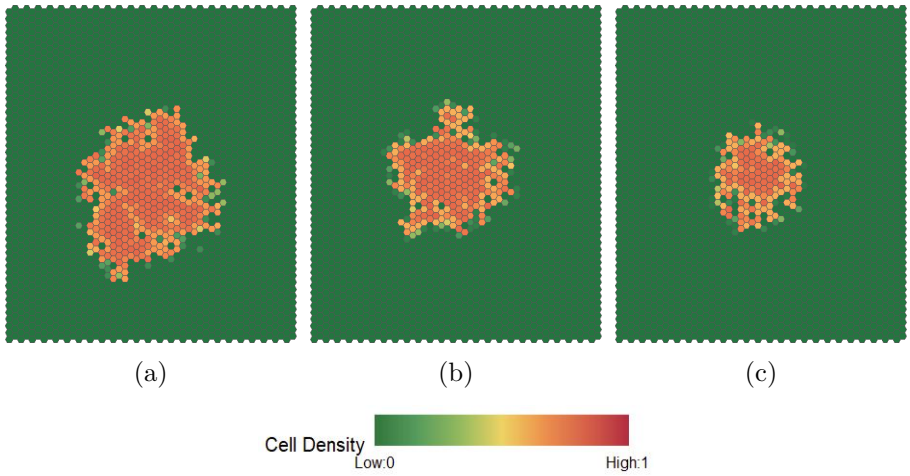
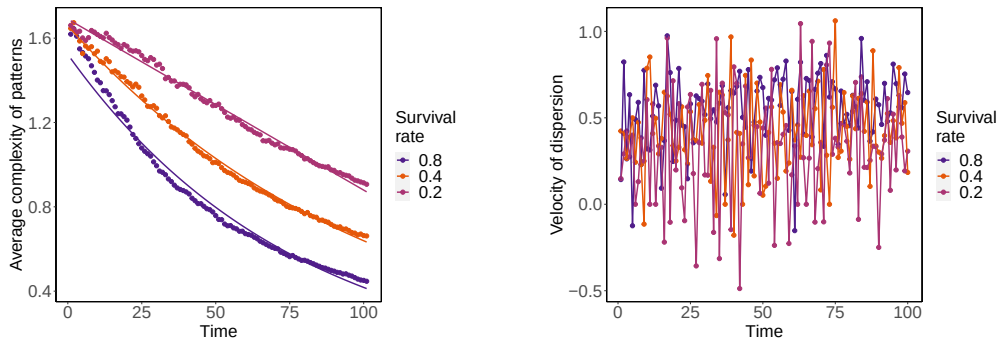


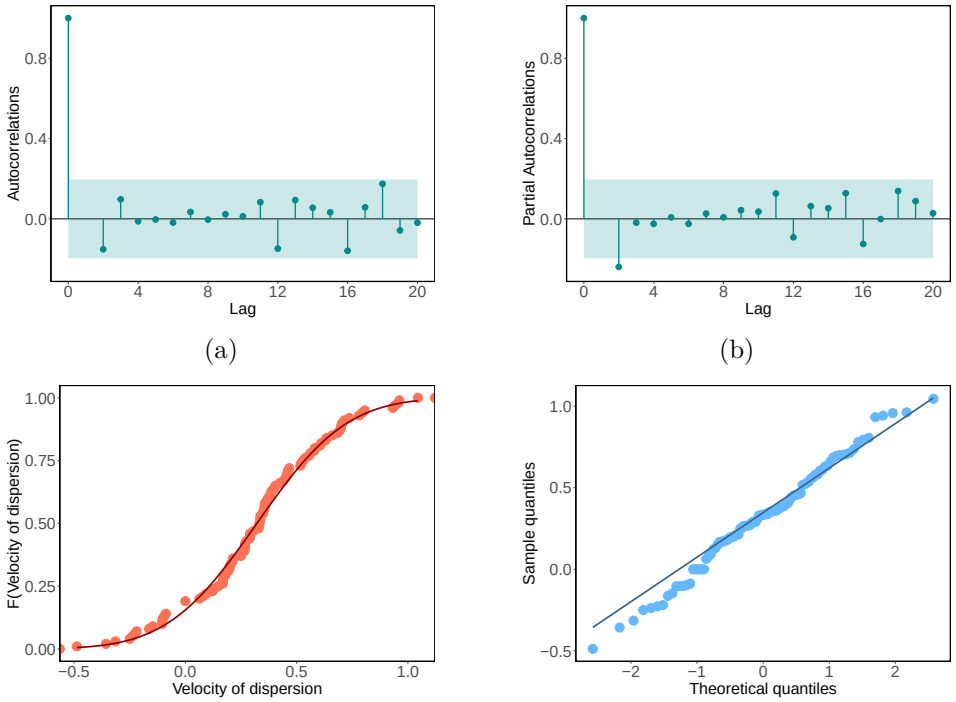
Figure 5.20: Final patterns at time $t = 50$ generated by the model from the same initial configuration with different survival rates γ_2 at the macroscale. (a) High survival rate ($\gamma_2 = 0.8$). (b) Medium survival rate ($\gamma_2 = 0.4$). (c) Low survival rate ($\gamma_2 = 0.2$).

Comparing the ending patterns from Figure 5.13, Figure 5.17, and Figure 5.20, we conclude that varying the cell densities in the model using three different options all influence the evolution patterns generated by the model. The spreading rate γ_1 and the survival rate γ_2 seem to have similar effects on the model behaviours while the birth-death process on the microscale appears to have a large effect on the model behaviour. Based on our observations of the trend of average complexity over time, the average complexity tends to decrease exponentially when there is enough density to disperse, and it tend to decrease



(a) (b)

Figure 5.21: Average complexity of the patterns and velocity of dispersion of the model over the course of time for three realizations with different survival rate levels at the macroscale of the model. (a) Average complexity of patterns. The points represent average complexity of patterns from 100 realizations at each time, and the curve represents the fitted model of the complexity over time, with ($\ln(\text{complexity}) = 0.42 - 0.01 \times \text{time}, R^2 = .98, F(1, 99) = 6697, p < .001; \ln(\text{complexity}) = 0.5 - 0.01 \times \text{time}, R^2 = .99, F(1, 99) = 22210, p < .001, \ln(\text{complexity}) = 0.55 - 0.01 \times \text{time}, R^2 = .99, F(1, 99) = 11380, p < .001$) for survival rate 0.8, 0.4, and 0.2, respectively. (b) Velocity of dispersion for one realization.



(a) (b)

(c) (d)

Figure 5.22: Plots for velocity of dispersion for one realization. (a) Autocorrelation plot. (b) Partial autocorrelation plot. (c) Empirical cumulative distribution function plot. (d) Normal QQ plot.

faster when the density is at a higher level. However, it tends to decrease linearly when the density available to disperse is at a very low level.

5.5.1.2 Impact of the Inclusion Probability

The inclusion probability is determined as 3.3.3, where $w_{ij}^{(t)}$ and $d_{ij}^{(t)}$ are defined as inclusion weights and difference in densities between adjacent cells that are described in previous chapter, and the parameters $\alpha_{ij}^{(t)}$ and $\beta_{ij}^{(t)}$ are used to distinguish the inclusion probability between high density cells and low density cells. The parameters can be varied for each cell (i, j) according to the local physical properties at each time step t. For simplicity, we assume the cells are uniform in the simulation runs, therefore all cells have the same α and β values. We vary the parameters α and β to investigate the impact of inclusion probability on dispersion patterns.

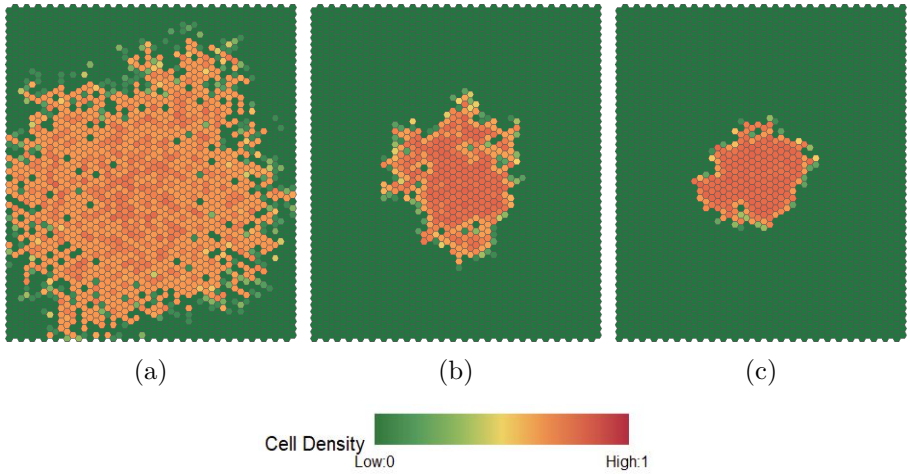
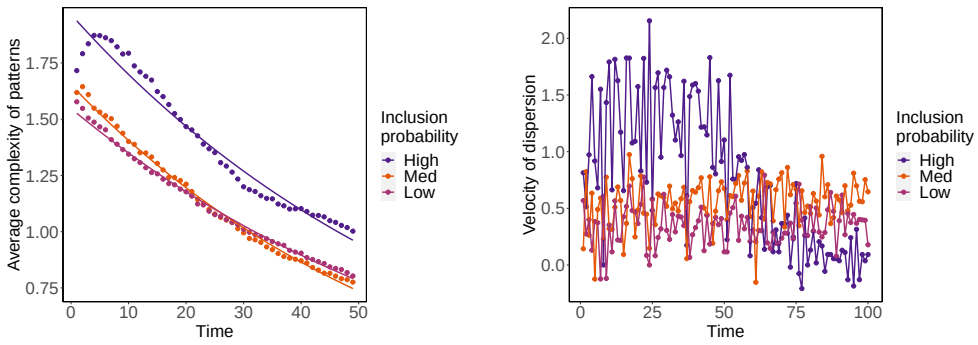


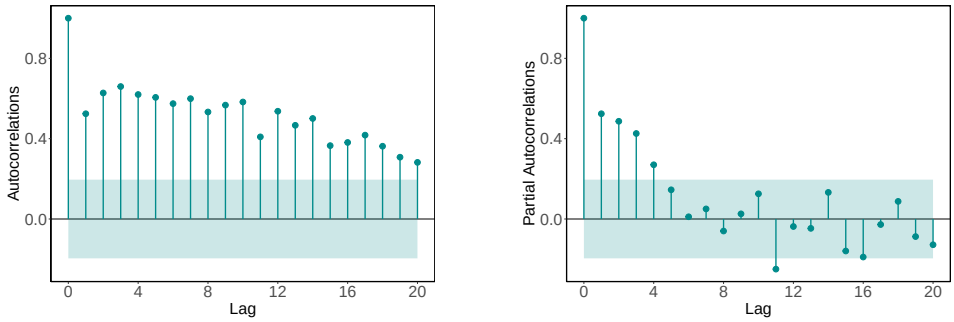
Figure 5.23: Final patterns at time $t = 50$ generated by the model from the same initial configuration with different inclusion probabilities at the macroscale. (a) High inclusion probability ($\alpha = 1, \beta = 0.2$). (b) Medium inclusion probability ($\alpha = 1, \beta = 1$). (c) Low inclusion probability ($\alpha = 1, \beta = 2$).

Given fixed values for ω and d, varying α and β jointly leads to varying inclusion probabilities. According to Figure 5.23, when the inclusion probability is large, the neighbouring non-affected cells of the affected cells are more likely to be included in the dispersion process, and the density is more likely to disperse to those non-affected cells. Therefore, we observe the size of the interaction neighbourhood grows faster with a larger inclusion probability. Furthermore, the interaction neighbourhood grows with a "rougher" boundary and there are lots of vacant cells inside the interaction neighbourhood. This suggests that the neighbouring non-affected cells are more likely to be affected during the dispersion process. On the contrary, when the inclusion probability is small, non-affected neighbouring cells are



(a) (b)

Figure 5.24: Average complexity of the patterns and velocity of dispersion of the model over the course of time for three realizations with different inclusion probabilities at the macroscale of the model. (a) Average complexity of patterns. The points represent average complexity of patterns from 100 realizations at each time, and the curve represents the fitted model of the complexity over time, with ($\ln(\text{complexity}) = 0.67 - 0.01 \times \text{time}, R^2 = .97, F(1, 47) = 1496, p < .001; \ln(\text{complexity}) = 0.5 - 0.02 \times \text{time}, R^2 = .99, F(1, 47) = 9088, p < .001; \ln(\text{complexity}) = 0.44 - 0.01 \times \text{time}, R^2 = .99, F(1, 47) = 10900, p < .001$) for large, medium and small inclusion probability, respectively. (b) Velocity of dispersion for one realization.



(a) (b)

Figure 5.25: Plots for velocity of dispersion for one realization of large inclusion probability. (a) Autocorrelation plot. (b) Partial autocorrelation plot.

less likely to be included in the dispersion process, the interaction neighbourhood grows at a much slower rate, and has a "smoother" boundary.

We observe some moderate deviation from the fitted model for average complexity of large inclusion probability, and this is due to the unpredictable and complex pattern boundaries resulted by the large inclusion probability. In addition, this unpredictability is also reflected in the velocity of dispersion as the dispersion velocity does not resemble white noise.

5.5.1.3 Impact of the Survival Probability

The survival probability is the probability of successful population dispersion to low density cells. It characterizes establishment of population in new areas. The survival probability determined as 3.3.4, depends on the population size that is received by the non-contributing cells. Similar to the scheme of inclusion probability parameters, $\phi_{ij}^{(t)}$, $\rho_{ij}^{(t)}$ and $\xi_{ij}^{(t)}$ are used to adjust the survival probability for each non-contributing cell at each time step t. We assume all cells take the same parameter values in the simulation runs. We vary those parameters jointly to study the impact of survival probability on dispersion patterns.

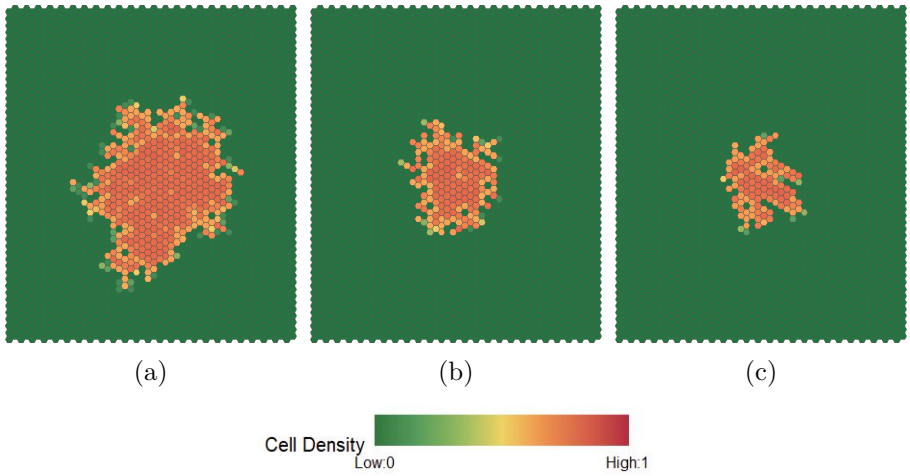
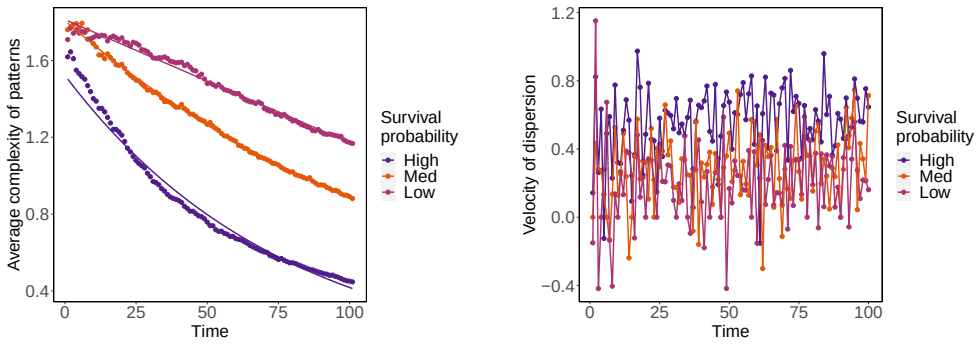


Figure 5.26: Final patterns at time $t = 50$ generated by the model from the same initial configuration with different survival probabilities at the macroscale. (a) High survival probability ($\phi = 1/3, \rho = 1/3$). (b) Medium survival probability ($\phi = 1/6, \rho = 1/6$). (c) Low survival probability ($\phi = 1/9, \rho = 1/9$).

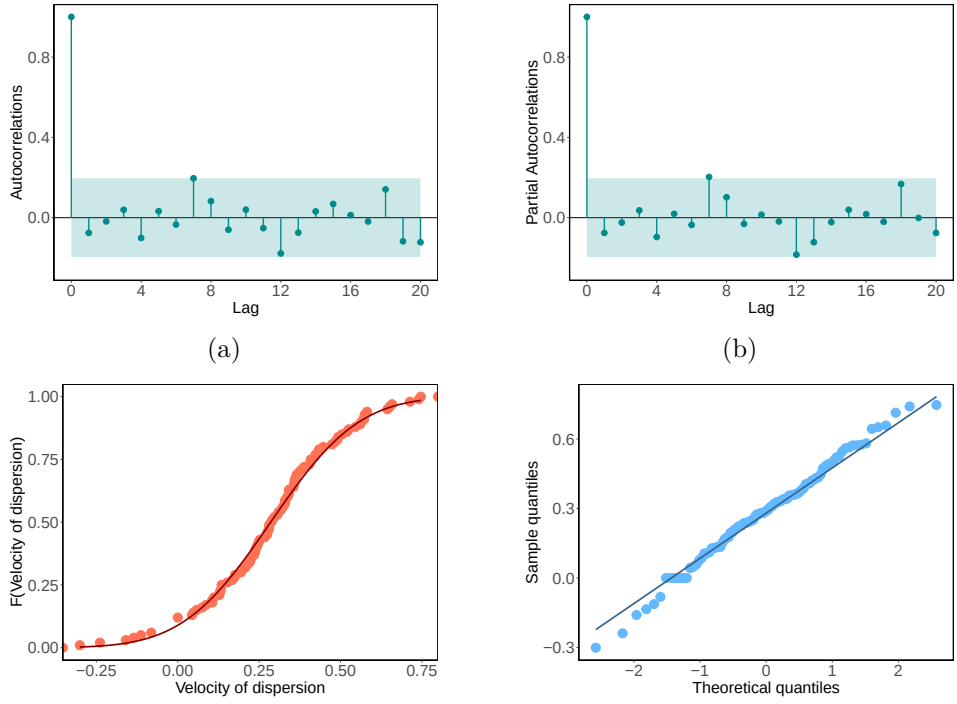
As we observe in Figure 5.26, the survival probability has a similar effect to the inclusion probability in terms of interaction neighbourhood size. The interaction neighbourhood grows faster and there are fewer vacant cells inside the interaction neighbourhood when the survival probability is large. This is because a large survival probability means that the population are more likely to establish in the new cells, thus, the non-affected cells are more likely to be affected during the dispersion process. The average complexity decreases linearly when the survival probability is at a very low level, this result is similar to the impact of birth-death-process at the micro scale. The velocity resemble Gaussian white noise.

Compared to Figure 5.23, while a large inclusion probability and survival probability both tend to accelerate the dispersion process, they lead to different evolution patterns. A large inclusion probability leads to a much uncertain and complex pattern.



(a) (b)

Figure 5.27: Average complexity of the patterns and velocity of dispersion of the model over the course of time for three realizations with different survival probabilities at the macroscale of the model. (a) Average complexity of patterns. The points represent average complexity of patterns from 100 realizations at each time, and the curve represents the fitted model of the complexity over time, with ($\ln(\text{complexity}) = 0.42 - 0.01 \times \text{time}, R^2 = .98, F(1, 99) = 6697, p < .001; \ln(\text{complexity}) = 0.6 - 0.01 \times \text{time}, R^2 = .99, F(1, 99) = 49100, p < .001; \text{complexity}) = 1.8 - 0.006 \times \text{time}, R^2 = .98, F(1, 99) = 7267, p < .001$) for large, medium, and small survival rates, respectively. (b) Velocity of dispersion for one realization.



(a) (b)

(c) (d)

Figure 5.28: Plots for velocity of dispersion for one realization of small survival probability. (a) Autocorrelation plot. (b) Partial autocorrelation plot. (c) Empirical cumulative distribution function plot. (d) Normal QQ plot.

5.5.2 Phase II: Interactions between the Model Regimes

Interactions due to varying parameters simultaneously also impacts the evolution patterns of the model. Varying cell density, the inclusion probability and the survival probability simultaneously leads to complex model behaviours. We aim to explore this effect. There are three ways to vary the cell densities in the simulation runs, i.e., the birth-death process at the microscale, the spreading rate γ_1 and survival rate γ_2 at the macroscale. But they behave similarly so we only vary the spreading rate γ_1 to vary the cell densities in phase II.

5.5.2.1 Varying Cell Density and Inclusion Probability Simultaneously

We aim to investigate the effect of varying cell density and inclusion probability simultaneously. In Figure 5.29, we see that when the cell population is large enough to support the dispersion to new areas and the inclusion probability is sufficiently large, the interaction neighbourhood grows with only a few vacant cells in the interior of the neighbourhood. However, when the inclusion probability is large but the cell density is small, the interaction neighbourhood grows but with a much rougher boundary and a higher number of vacant cells in the interior. This is because when the cell density is small, it is very unlikely to survive in the new cells due to the survival probability, leading to many vacant cells on the boundary and interior. When the cell density and inclusion probability are both small, the interaction neighbourhood grows at a very slow rate and there are no vacant cells in the interior.

5.5.2.2 Varying Cell Density and Survival Probability Simultaneously

Varying the cell density and inclusion probability simultaneously yields a more uniform pattern, as shown in Figure 5.30. When both are large, the interaction neighbourhood grows at a faster rate and when both small, the neighbourhood grows at a very slow rate. Compared to each plot in Figure 5.29, the boundary of the neighbourhood is much "smoother" with fewer vacant cells in the interior of the neighbourhood.

5.5.2.3 Varying Inclusion Probability and Survival Probability Simultaneously

Varying the inclusion probability and survival probability simultaneously yields very complex patterns. The inclusion probability determines how likely the cells are included in the dispersion process, and the survival probability determines how likely the population disperses to new cells. In Figure 5.31, when both of them are large, the interaction neighbourhood grows at a much faster rate with a very rough boundary, but there are a lot of vacant cells inside the neighbourhood; and when both of them are small, the interaction neighbourhood barely grows with a much smoother boundary, and there are almost no vacant cells inside the neighbourhood.

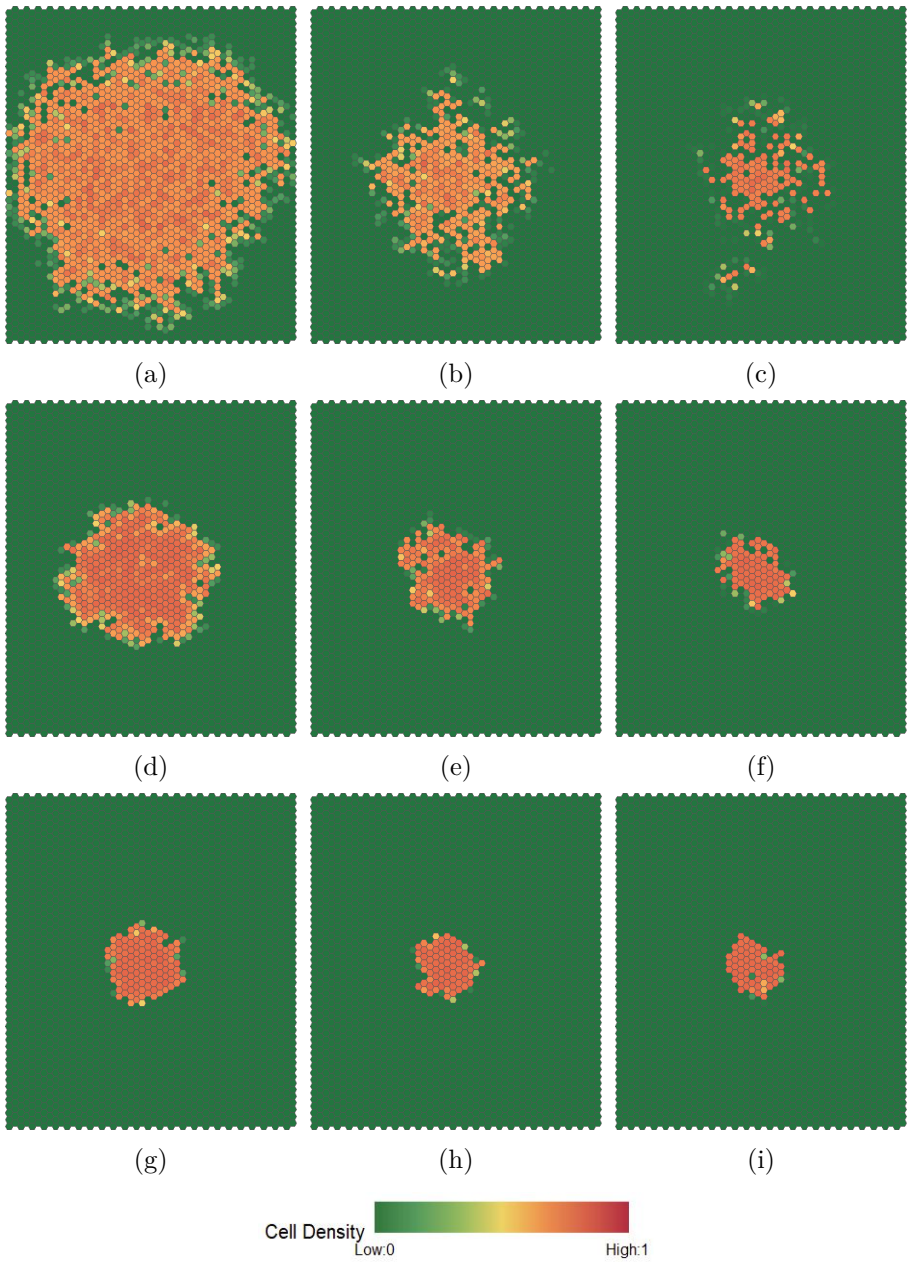


Figure 5.29: Model patterns at time $t = 30$ while varying cell density and inclusion probability simultaneously. (a) High cell density ($\gamma_1 = 0.8$) and large inclusion probability ($\alpha = 1, \beta = 0.2$). (b) Medium cell density ($\gamma_1 = 0.4$) and large inclusion probability ($\alpha = 1, \beta = 0.2$). (c) Low cell density ($\gamma_1 = 0.1$) and large inclusion probability ($\alpha = 1, \beta = 0.2$). (d) Medium cell density ($\gamma_1 = 0.4$) and large inclusion probability ($\alpha = 1, \beta = 0.2$).(e) Medium cell density ($\gamma_1 = 0.4$) and medium inclusion probability ($\alpha = 1, \beta = 1$).(f) Medium cell density ($\gamma_1 = 0.4$) and medium inclusion probability ($\alpha = 1, \beta = 1$). (g) Large cell density ($\gamma_1 = 0.8$) and small inclusion probability ($\alpha = 1, \beta = 3$). (h) Medium cell density ($\gamma_1 = 0.4$) and small inclusion probability ($\alpha = 1, \beta = 3$). (i) Small cell density ($\gamma_1 = 0.1$) and small inclusion probability ($\alpha = 1, \beta = 3$).

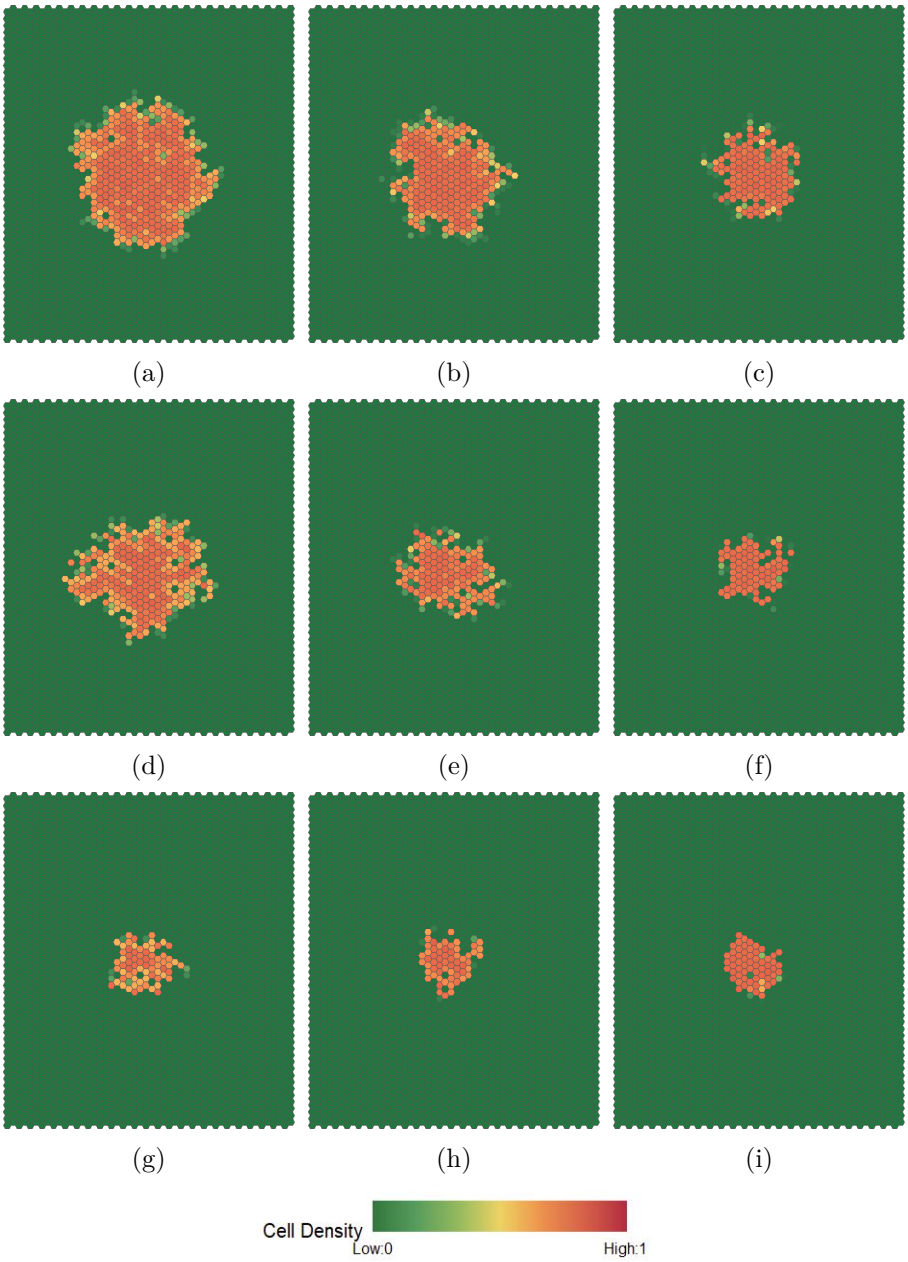


Figure 5.30: Model patterns at time $t = 30$ while varying cell density and survival probability simultaneously. (a) High cell density ($\gamma_1 = 0.8$) and large survival probability ($\phi = 1/2, \rho = 1/2$). (b) Medium cell density ($\gamma_1 = 0.4$) and large survival probability ($\phi = 1/2, \rho = 1/2$). (c) Low cell density ($\gamma_1 = 0.1$) and large survival probability ($\phi = 1/2, \rho = 1/2$). (d) Medium cell density ($\gamma_1 = 0.4$) and medium survival probability ($\phi = 1/3, \rho = 1/3$).(e) Medium cell density ($\gamma_1 = 0.4$) and medium survival probability ($\phi = 1/3, \rho = 1/3$).(f) Small cell density ($\gamma_1 = 0.4$) and medium survival probability ($\phi = 1/3, \rho = 1/3$). (g) Large cell density ($\gamma_1 = 0.8$) and small survival probability ($\phi = 1/9, \rho = 1/9$). (h) Medium cell density ($\gamma_1 = 0.4$) and small survival probability ($\phi = 1/9, \rho = 1/9$). (i) Small cell density ($\gamma_1 = 0.1$) and small survival probability ($\phi = 1/9, \rho = 1/9$).

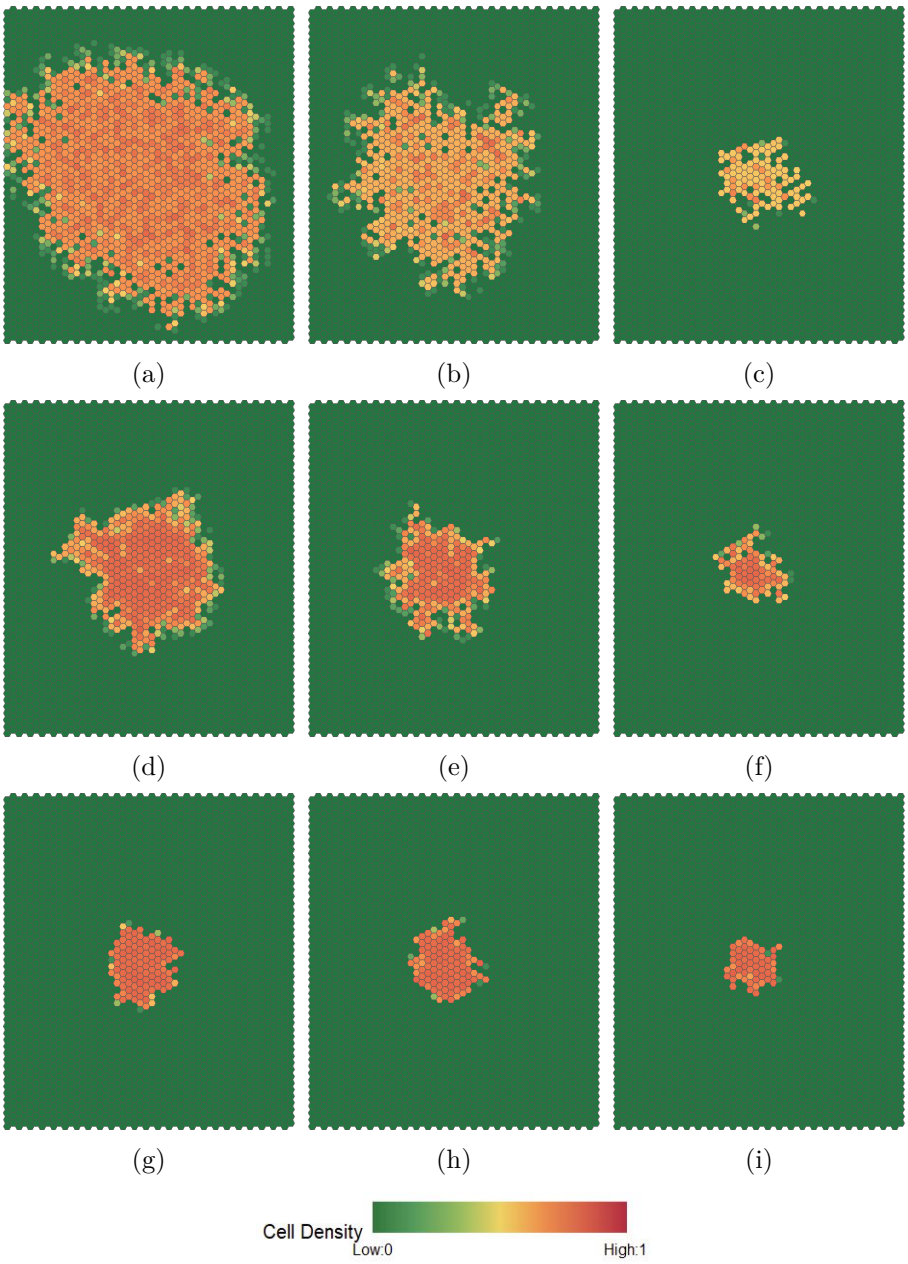


Figure 5.31: Model patterns at time $t = 30$ while varying inclusion probability and survival probability simultaneously. (a) Large inclusion probability ($\alpha = 1, \beta = 0.2$) and large survival probability ($\phi = 1/2, \rho = 1/2$). (b) Large inclusion probability ($\alpha = 1, \beta = 0.2$) and medium survival probability ($\phi = 1/3, \rho = 1/3$). (c) Large inclusion probability ($\alpha = 1, \beta = 0.2$) and small survival probability ($\phi = 1/9, \rho = 1/9$). (d) Medium inclusion probability ($\alpha = 1, \beta = 1$) and large survival probability ($\phi = 1/2, \rho = 1/2$).(e)Medium inclusion probability ($\alpha = 1, \beta = 1$) and medium survival probability ($\phi = 1/3, \rho = 1/3$).(f) Medium inclusion probability ($\alpha = 1, \beta = 1$) and small survival probability ($\phi = 1/9, \rho = 1/9$). (g) Small inclusion probability ($\alpha = 1, \beta = 3$) and large survival probability ($\phi = 1/2, \rho = 1/2$). (h) Small inclusion probability ($\alpha = 1, \beta = 3$) and medium survival probability ($\phi = 1/3, \rho = 1/3$). (i) Small inclusion probability ($\alpha = 1, \beta = 3$) and small survival probability ($\phi = 1/9, \rho = 1/9$).

5.5.3 Phase III: Impact of Inhomogeneous Cell Space

Physical environments may have inhomogeneous properties, e.g., population dispersion is often disturbed by heterogeneity influence such as the geographical situations or the outer forces. Thus, the conditions for dispersion vary cell to cell. For example, in mountain pine beetle infestation study, there are deforested areas such as prairie, grassland, rocky hills, large lakes or human constructions and dispersion is influenced by weather conditions such as wind. We modify specific cell properties in the grid or vary the transition rule to mimic this environmental inhomogeneity.

Figure 5.32: Mountain pine beetle infestation in a local inhomogeneous region of the Greater Yellowstone Ecosystem. This aerial photograph was taken by Jane Pargiter during the summer of 2007. Photo is from [123]. Licensed under a CC BY-NC license.

5.5.3.1 Geographical Properties of the Cell Space

We set up two experiments to investigate situations in which the population is only allowed to disperse through some cells in the domain. In this way, we aim to simulate the dispersion patterns to reflect the heterogeneous environment when there is a barrier for the population to disperse. We see in Figure 5.33 and Figure 5.34, the population only immigrates to specific cells as expected, and the dispersion patterns do not display the kind of boundaries seen in other simulations. The dispersion also takes much longer to spread.

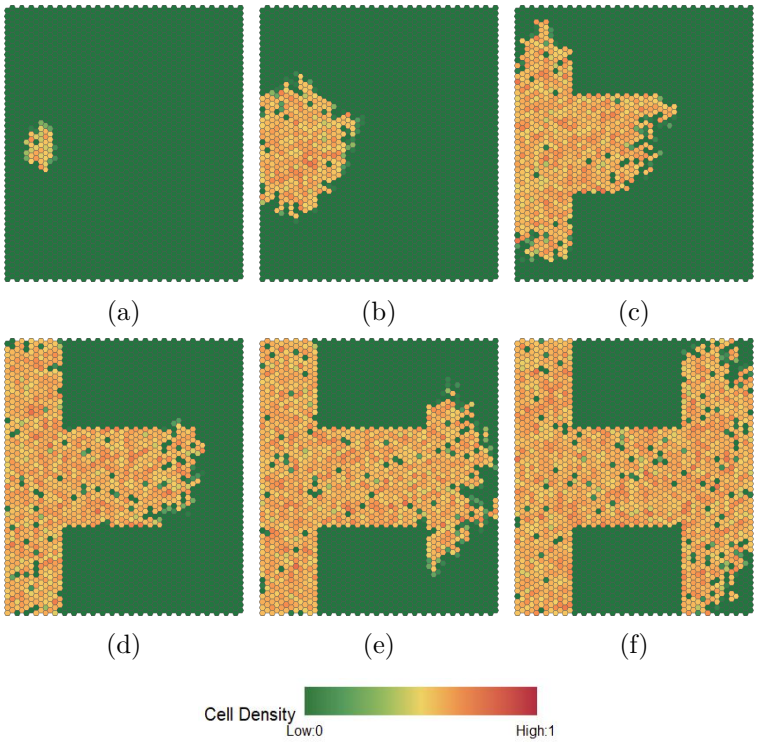


Figure 5.33: Evolution patterns when there is a region for which individuals can pass through in the domain. (a) $t = 10$. (b) $t = 100$. (c) $t = 200$. (d) $t = 300$. (e) $t = 400$. (f) $t = 500$.

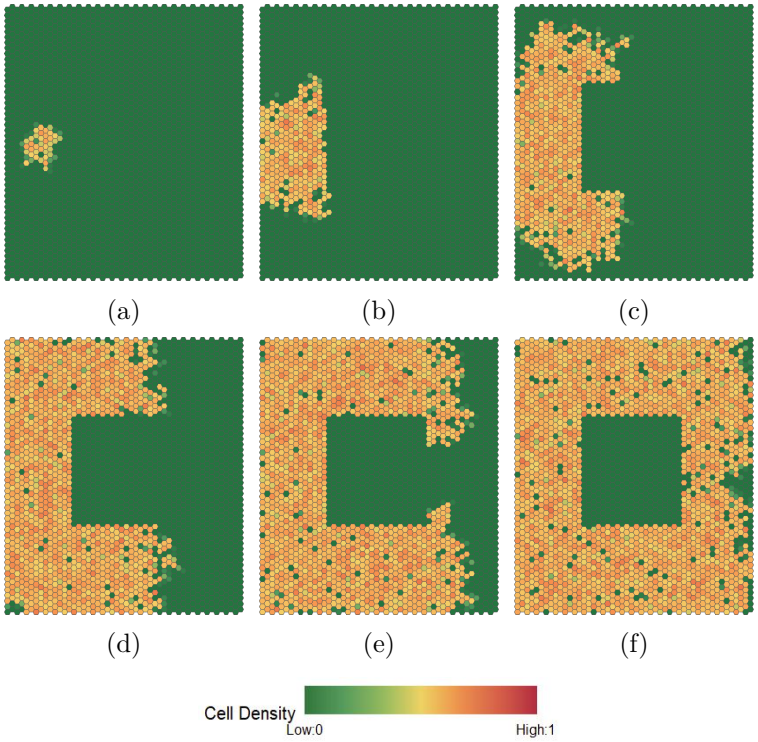


Figure 5.34: Evolution patterns when there is a lake in the domain. (a) $t = 10$. (b) $t = 100$. (c) $t = 200$. (d) $t = 300$. (e) $t = 400$. (f) $t = 500$.

5.5.3.2 Outer Forces

The dispersion process may be affected by environmental forces. For example, mountain pine beetles may be strongly affected by wind, i.e., they can fly to a larger region with greater spatial variability.

We model an effect like wind by including a probability that is assigned to cells on different locations of the grid and we encourage the directional movement of populations by varying the likelihood of including specific cells in the grid according to the wind speed and direction. For example, suppose the wind is blowing from West towards East. We assign a probability to increase the chance of including high density cells on the West and low density cells on the East while lowering the probability of including low density cells on the West and high density cells on the East. In this way, the population in high density cells in the West is more likely to disperse to low density cells in the East. By modifying the transition rule, we are able to treat the atmosphere as a variable instead of a simple "constant".

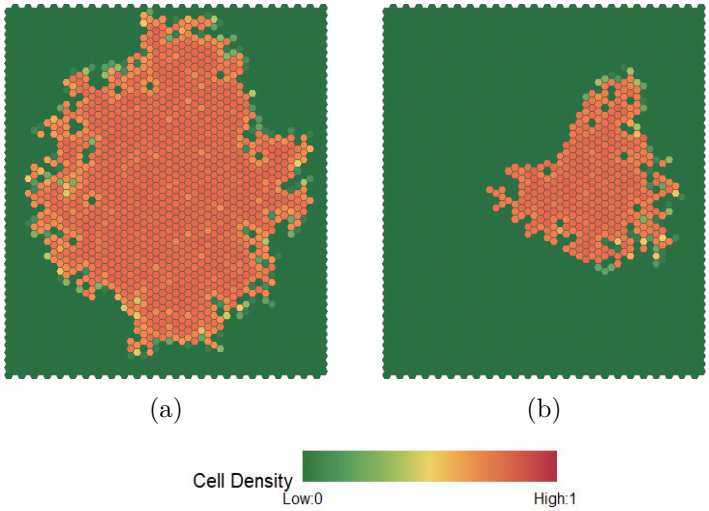


Figure 5.35: Evolution patterns at time $t = 120$ for a non-directional and directional movement to the East with the same initial configurations and parameter settings. (a) Non-directional movement. (b) Directional movement.

We see in Figure 5.35, compared to the non-directional movement (a), the population tend to disperse to the East, and there are fewer affected cells at the same time steps. The experiments in Phase III suggest that the model can provide flexibility for adapting to different situations by simulating the heterogeneous environment in practice.

5.5.4 Phase V: Evolutional Patterns of Two Groups of Cells with Different Properties

The previous simulation runs are all based on one interaction neighbourhood. In practice, there may be more than one interaction neighbourhood which have different physical properties, and we would like to simulate the dynamics of two distinct interaction neighbourhoods with different initial patterns and physical properties to observe the evolution when there are more than one interaction neighbourhood.

We assume there are two distinct interaction neighbourhoods in the grid that are far away from each other at the start of the dispersion process. The two interaction neighbourhoods have distinct properties including different cell densities, inclusion probability and survival probability. We assume that both are expanding in the grid. But the lower small interaction neighbourhood disperses at a much slower rate, and the upper large interaction neighbourhood disperses at a much faster rate. We see in Figure 5.36, the upper interaction neighbourhood quickly grows and approaches the lower small interaction neighbourhood. After certain time steps, the two interaction neighbourhoods merge with each other, and they behave like one interaction neighbourhood.

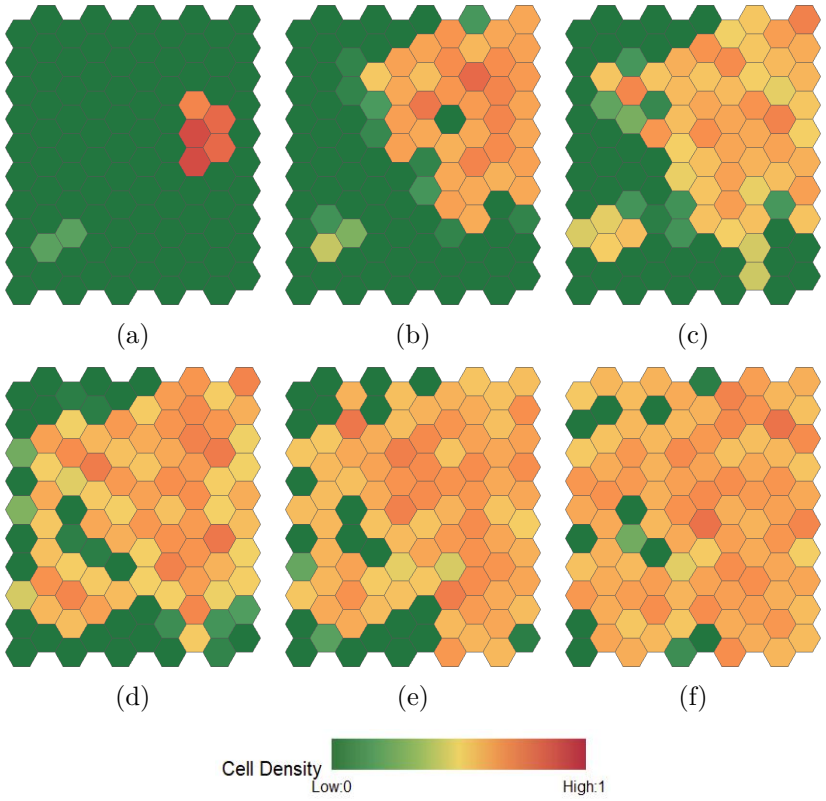


Figure 5.36: Evolution patterns of interactions between two interaction neighbourhoods when they both grow without limit. (a) $t = 0$. (b) $t = 20$. (c) $t = 40$. (d) $t = 60$. (e) $t = 80$. (f) $t = 100$.

5.6 Summary and Discussion

We investigate the various aspects of the multiscale SCA model behaviour. The numerical simulation results indicate that the SCA model capable of generating various dispersion patterns. In general, the model behaviour approaches three limiting situations: fade away completely, settle into a fixed finite size, and grow without limit at a fixed speed. The dispersion patterns depend on the a couple of factors such as the choice of the microscale model, the model parameters and the interactions between the model parameters. Some parameters have similar effect and variations among those parameters tend to generate patterns with almost identical physical characteristics. The microscale model provides the input of the macroscale SCA model, which mainly determines the number of affected cells and cell density at a specific time. The spreading rate and survival rate have similar effects on model behaviour as the microscale model. The inclusion probabilistic rule and the survival probabilistic rule mainly determine the smoothness of the boundary of the dispersion patterns, as well as the number of vacant cells inside the interaction neighbourhood.

Chapter 6

Model Calibration

An important property of a model is the degree to which the parameters in the model can be "calibrated" or determined from observed data on the model behaviour. Calibration problem is an essential component in using a model to forecast behaviour of the system being modelled. In this chapter, we conduct a numerical investigation of the calibration problem for the Stochastic Cellular Automaton (SCA) model.

The calibration problem is a form of inverse problem. In general, the map that implicitly relates the model parameters to the model output is not 1-1, so assumptions are required in order that the calibration problem admits a well defined solution. In this chapter, we consider two Bayesian approaches to the calibration problem, namely the Stochastic Inverse Problem (SIP) and Bayesian Inverse Problem (BIP) [16]. The SIP requires fewer assumptions than the BIP but requires sufficient data on the output to build a nonparametric approximate distribution, so we conduct the investigation of the SIP using simulated data. To consider a calibration using real data, we study the BIP approach for data consisting of graphical representations that indicate the progression of tree damage in a pine beetle infestation.

The main goal of this chapter is to numerically demonstrate that the SCA can be calibrated from observed data and to investigate the sensitivity of calibration for individual parameters in the SCA. A complication with calibration for the SCA is the stochastic nature of the output, which means that repeated simulations with the same parameters may yield different results. To handle this, we calibrate the model results for the expectation of model outputs. That is, we calibrate with respect to the output of the model obtained by averaging over a large number of simulations.

6.1 Introduction

The parameters in a physical process model of a complex system represent important physical characteristics of the system being modelled. Scientific inference using a model of a complex system, e.g., forecasting behaviour, involves evaluating the model for parameter values that are relevant for the current physical state of the system. However, in a great

number of cases, it is difficult or impossible to directly observe the parameters in a physical process model.

This raises the problem of model calibration. A form of inverse problem, in the stochastic setting, the model calibration problem is to determine a probability distribution on the model parameters from (noisy) observational data on behaviour of the solutions of the model. Model calibration involves a number of technical complications and there are many fundamental challenges. First, the models that are used to characterize the processes often have limitations on the output values, for example, the models may be imperfect or imprecise in terms of characterizing the system behaviour. Also, in reality, the data is usually uncertain, i.e., the data is affected by observational noise. Furthermore, the models may be computationally expensive to evaluate. Most importantly, typically the model is not 1-1 and there are multiple parameter values that produce the same data.

There are a variety of approaches to tackle the problems of calibration of model parameters from observed data and forecasting for time-dependent systems. These include data assimilation [178] [102], [10], [9], [119], Bayesian calibration [23] [90], [100], [113], and filtering [107], which all use a dynamics approach focused on small perturbations of a particular solution of the model. Some of these approaches are related to stochastic processes and time series analysis.

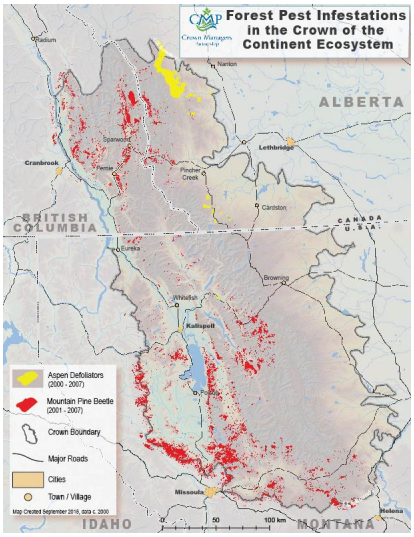
In this chapter, we use two different approaches to the model calibration problem, the Stochastic Inverse Problem (SIP) [16], [30], [21], [31] and Bayesian Inverse Problem (BIP) [129], [113]. We carry out calibration using both simulated data and real data in the form of graphical representations of tree damage from pine beetles.

We collect data on certain quantities that are considered to be quantities of interest related to the model solution, i.e., the dispersion patterns. We infer the parameter information from data based on these quantities. Therefore, the SCA model we build induces a measurable nonlinear map from the domain of physical parameters to the range of some observable quantities of interest that can be computed as some measurable features of the dispersion patterns.

6.2 Data for Mountain Pine Beetle Infestations

It is difficult to obtain data suitable for calibration of a model for pine beetle infestations. One possible data is obtained from aerial surveys by observers in aircraft that provide overall forest health status in broad-scale and gather information about detect MPB infestations. The aerial surveys identify the location of the insects and its infestation severity. Usually, multiple maps are combined to form digital infestations, and the information gathered from the air are combined with results obtained from ground surveys to estimate the severity of beetle infestation since it is impossible to determine the exact number of infested trees from the air.

An important issue in model calibration using graphical data is the choice of physical domain. This involves both the scale over which simulations can be carried out in reasonable time and the validity of representation of the model. We perform calibration using the graphical data from a region that was used to monitor the long-term ecological health of the Crown of the Continent Ecosystem. The map keeps record of dead (red) pine trees which indicate a high likelihood of mountain pine beetle attacks from 2001 to 2007. The area of dead pine trees is about 2865 km^2 and occupies about 4.4 % of the entire Crown of the Continent Ecosystem region.



(a)

Figure 6.1: An aerial survey map for mountain pine beetle infestation in the Crown of the Continent Ecosystem. Image of map is from the US Geological Survey database [58].

6.3 The Model Calibration Problem

To formulate the calibration problem, we assume the vector of parameters for the SCA model M is λ and the corresponding parameter domain that describes the physical characteristics of dispersion process is Λ, and the solution of the model is Y with $M(Y, \lambda) = 0$, where λ takes values in the domain Λ. The solution of the model can be written as $Y = Y(\lambda)$, which is an implicit function of the parameters. We collect data on a "Quantity of Interest", which is a function of the model solution that constitute observable quantities. Suppose the quantities of interest are computed by a function q defined on the model solution Y, then we summarize the model as $Q(\lambda) = q(Y(\lambda))$, where Q is the map from the parameter domain Λ to the "observable space" The map Q is rarely $1-1$, i.e., there are multiple parameter values that produce the same output value. For example, many combinations of parameters in the SCA model lead to identical dispersion pattern. We let $\mathcal{D}$ denote the range of Q on

Λ defined as $\mathcal{D} = \{Q(\lambda) : \lambda \in \Lambda\}$. Therefore, the inverse Q^{-1} of Q is the map from $\mathcal{D}$ to subsets of Λ,

$$Q^{-1}(y) = \{\lambda \in \Lambda : Q(\lambda) = y\}.$$

6.3.1 The Stochastic Forward Problem

The stochastic inverse problem is the inverse of the classic stochastic forward problem. We assume a stochastic model for the vector of model parameters λ, where Λ is the sample space, $\mathcal{B}_\Lambda$ is the Borel σ algebra, and $(\Lambda, \mathcal{B}_\Lambda, \tilde{P}_\Lambda)$ is the probability of possible sets of parameters, called the *data generating distribution*. This reflects uncertainty in the parameter values. In the inverse problem, we treat the range $\mathcal{D} = Q(\Lambda)$ as the domain for the model parameters, and our goal is to choose a *sample value* $\lambda \in \Lambda$ corresponding to an observed output $d = Q(\lambda) \in \mathcal{D}$. We choose the Borel σ algebra $\mathcal{B}_\mathcal{D}$ on $\mathcal{D}$ to form the measurable space $(\mathcal{D}, \mathcal{B}_\mathcal{D})$.

The *Stochastic Forward Problem (SFP)* is to compute the probability distribution $P_\mathcal{D}$ on $(\mathcal{D}, \mathcal{B}_\mathcal{D})$ that is induced by Q from the probability measure $\tilde{P}_\Lambda$ via

$$P_\mathcal{D}(B) = \tilde{P}_\Lambda(Q^{-1}(B)) \text{ for } B \in \mathcal{B}_\mathcal{D}.$$

The inverse map Q^{-1} of Q is the map from $\mathcal{D}$ to subsets of Λ defined by,

$$Q^{-1}(y) = \{\lambda \in \Lambda : Q(\lambda) = y\} \text{ for } y \in \mathcal{D}. \tag{6.3.1}$$

$P_\mathcal{D}$ is sometimes called the push-forward measure, and the SFP is sometimes described as propagating uncertainty through the model. The SFP provides a probability distribution on the potential experimental results under the varying experimental conditions. The SFP plays a central role in model forecasting and inference, and is a cornerstone for scientific and engineering inference. The SFP focuses on the behaviour of a model over a wide range of possible parameters, so the model can include a wide variety of behaviours in possible solutions.

The *ensemble approximation* for solving the SFP is computed by conducting experiments using the Monte Carlo method. According to the data generating distribution $\tilde{P}$, we collect samples $\{\lambda_i\}_{i=1}^n$ and compute the corresponding output values $\{q_i = Q(\lambda_i)\}_{i=1}^n$. We then use those output values to construct an approximation to $P_\mathcal{D}$.

6.3.2 The Stochastic Inverse Problem

The SIP is to compute a probability distribution P_Λ on $(\Lambda, \mathcal{B}_\Lambda)$ that induces a distribution $P_\mathcal{D}$ on $(\mathcal{D}, \mathcal{B}_\mathcal{D})$ when the data generating distribution $\tilde{P}$ is unknown and the collection of output values $\{q_i = Q(\lambda_i)\}_{i=1}^n$ is a sample from $P_\mathcal{D}$. The full theory for the existence, prop-

erties, and approximate computation of solutions of the SIP under very general conditions are presented in [16, 21, 30, 31, 37, 48–50, 198, 199]. The SCA model satisfies the conditions and needed for formulating and solving the SIP.

In general, there is an infinite number of solutions P of the SIP. We adopt a Bayesian approach to deal with this issue. We choose a prior distribution that specifies a unique solution of the SIP, which is a Bayesian posterior. In this case, the SIP is well-posed in the sense of Hadamard, and the posterior inverse probability distribution depends continuously a.e. on initial conditions and parameters. The solution of the SIP is approximated by a form of importance sampling. We choose an initial random sample in Λ according to some fixed distribution (typically uniform), and use the initial sample to produce an updated weighted sample with probabilities corresponding to the solution of the SIP. We use the public domain Python code BET [89] that provides a variety of tools for solving the SIP for simple and complex models.

In an experimental situation, choosing a good parameter domain Λ is part of the problem. It is important to choose a reasonable parameter domain Λ because Λ determines the range of the parameters for which model behaviour is being studied. If Λ is too small or too big or does not include the range of meaningful parameter values, then it may not be possible to solve the calibration problem. Furthermore, it is also crucial to choose a prior family of distributions. In the absence of additional information, we specify a non-informative prior with maximum entropy since we have no idea which parameter values can generate the output values and we assume any parameter value is equally likely. This prior is called the *uniform Ansatz* [16]. That is used in all computations.

6.3.3 The Bayesian Inverse Problem

In the BIP, we assume there is a true value $\tilde{\lambda}$ for the vector of model parameters λ, but we do not know the exact value. However, we can observe noisy values of the model output value $\tilde{y}$ which can be written as $\tilde{y} = Q(\tilde{\lambda})$. We model the uncertainty in the physical process by choosing a small sample space that contains the true value $\tilde{\lambda}$ and we assume a probability distribution on the domain to describe the uncertainty. Since the output values are stochastic, we model the output as $\tilde{y} = Q(\tilde{\lambda}) + \epsilon$, where ϵ is normal noise with mean zero and a fixed covariance matrix Σ. Given a collection of noisy model output values $\{\hat{y}_i\}_{i=1}^M$, the goal for BIP is to compute a posterior distribution on Λ from which we obtain an estimate of $\tilde{\lambda}$.

We choose a sample space $\bar{\Lambda}$ for the model parameter λ such that $\tilde{\lambda} \in \bar{\Lambda}$. Since the true parameter value $\tilde{\lambda}$ is a fixed constant and it is not affected by the random "noise", we assume that the noisy model output values, i.e., $\mathcal{Y} = \{\hat{y}_i\}_{i=1}^M$, are independent draws from $N(Q(\tilde{\lambda}), \Sigma)$. Given $Q(\lambda) \sim N(Q(\tilde{\lambda}), \Sigma)$, we define the conditional probability density as $p(\hat{y}|\lambda) \sim N(Q(\lambda), \Sigma)$ for $\hat{y} \in \mathcal{Y}$ and $\lambda \in \bar{\Lambda}$.

6.4 Model Parameters and Quantities of Interest

According to the simulation results in Chapter 5, the parameters in the SCA model affect the model output in various ways. Some parameters have similar effect and variations among those parameters tend to generate patterns with almost identical physical characteristics. It is critical to identify the appropriate parameters for the calibration problem because some parameters in the model reflect the domain knowledge of the dispersion process and some parameters represent the specific stages of the process can be validated experimentally. The microscale of the model is used to characterize dynamics and interactions between individual mountain pine beetles in one or a few neighbouring trees, this is a separate process and the model parameters should be validated in a separate exercise.

Table 6.1: List of parameters for the SCA model

Symbol	Definition
α	Shape parameter in the inclusion probability
β	Shape parameter in the inclusion probability
γ_1	Spreading rate
γ_2	Survival rate
ϕ	Shape parameter in the survival probability
ρ	Shape parameter in the survival probability
ξ	Shape parameter in the survival probability

The calibration problem is solved at the macroscale. We introduce two probabilistic rules in the model to characterize the stochastic feature of the dispersion process. The inclusion probability represents the host selection process, in which case beetles usually prefer large-diameter trees with a thicker phloem and low infestations to provide a good habitat. α and β are set to reflect these physical properties of the interaction neighbourhoods, such as temperature, wind, light and health conditions of the trees in the neighbourhood. This also applies to ρ and ξ in the survival probability. Therefore, α, β, ρ and ξ that are relevant to physical properties of the interaction neighbourhoods and environmental conditions can be validated using the domain knowledge. ϕ in the survival probability reflects the general success of host colonization in the new areas, this process is a complex scheme facilitated by pheromones and kairomones, which is more likely to be determined by the tree infestation maps. γ_1 and γ_2 jointly decide the severity of infestations, γ_1 characterizing the mass-attack strategy of beetles which is not apparent to human, and it should be determined by the infestation maps. γ_2 also characterizes the host colonization process, but it can be validated experimentally in the field recapture-release studies. For this reason, we identify γ_1 and ϕ as model parameters for the calibration problem, then $\lambda = (\gamma_1, \phi)$.

For the experiments with simulated data, the setup of calibration problem for the SCA model is based on using statistics of dispersion patterns, therefore, we choose the domain

of the model parameters according to model behaviours that yield interesting patterns. For the calibration using simulated data, the domain for γ_1 and ϕ are $[0.6, 1]$ and $\left[\frac{1}{6}, \frac{5}{6}\right]$, respectively. We set a domain for the parameter ϕ to ensure that the survival probability is large enough to generate a growing dispersion pattern. Thus, $\Lambda = [0.6, 1] \times \left[\frac{1}{6}, \frac{5}{6}\right] \subset \mathbb{R}^2$. We set the parameter $\rho = \frac{1}{12}$.

We use four measurable quantities of interest in solving the calibration problem. The quantities of interest include the number of affected cells, the average cell density, the complexity of dispersion pattern and the derivative of fitted exponential curve of the complexity. The derivative of fitted curve of complexity is only feasible for the first two experiments due to limitations in data generation. These quantities are observable and can be easily computed from a dispersion pattern. Furthermore, they provide measurable information about the dispersion patterns.

6.5 Experiments

We provide four experiments using the SIP and BIP approaches to solve the calibration problem. The first three experiments are based on the simulated data and the fourth experiment is based on the real data in Figure 6.1.

We generate 1000 samples with 50 realizations for each combination of parameter values for use in averaging. Since the dispersion patterns tend to behave similarly for large times, we consider the patterns at 30 time steps to obtain patterns that have significant variation. The data generation takes around 30 hours on the Compute Canada's high-performance computation system depending on the computation resources.

As for the quantities of interest, we compute the average of the number of affected cells, the average cell density and the complexity of dispersion patterns. We fit the exponential curve the average complexity to obtain the derivative.

6.5.1 Experiment 1

6.5.1.1 Data Generation

We adopt the BIP approach and assume the true values for γ_1 and ϕ are 0.7 and $\frac{1}{3}$, respectively. To generate data, we choose a data generating distribution $\tilde{P}$, draw a sample from the distribution, solve multiple realizations of the model for each parameter sample, then compute the indicated statistics. This yields the data for calibration. We then "forget" the data generating distribution and solve the BIP. Note that we do not expect to recover the data generating distribution in general. In this experiment, we add noise $N(0, 0.05^2)$ to the true values. We generate 1000 samples with 50 realizations for each combination of γ_1 and ϕ. We compute the four quantities of interest for the dispersion patterns as the

experimental data y. Therefore, Q maps a four dimensional domain into a two dimensional domain.

6.5.1.2 Results

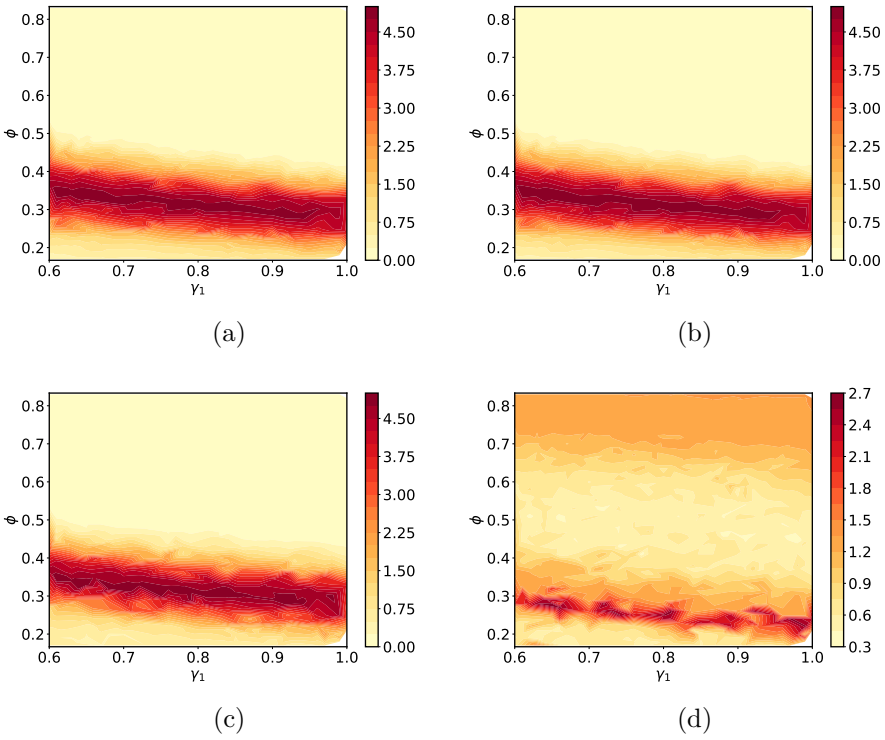
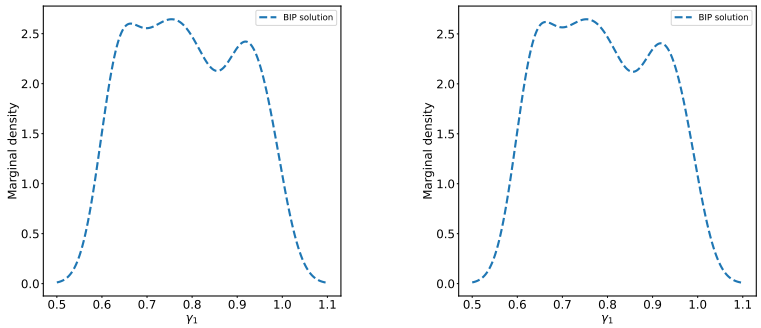


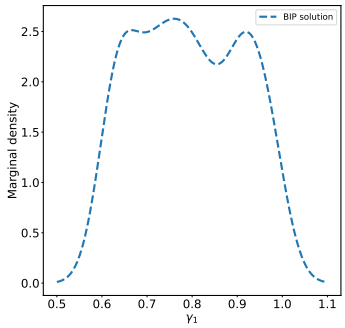
Figure 6.2: Bayesian posterior density contour plot for γ_1 and ϕ generated from four quantitative measurements. (a) Average number of affected cells. (b) Average cell density. (c) Average complexity of dispersion pattern. (d) Derivative of fitted curve of complexity.

The density contour plots of the inverse solution are shown in Figure 6.2. The plots show a density that has the form of a "ridge" that is parallel to the γ_1 axis, which indicates that the distribution of γ_1 is nearly uniform over the domain and many values of γ_1 lead to the same output data. This reflects the uniform Ansatz prior. This indicates that we chose a reasonable range for γ_1 in Λ. If we had the wrong interval for γ_1, the numerical method would indicate a singular distribution. As for ϕ, the contour lines in the first three plots are spaced very close to each other, then the values change rapidly, which indicate that the values for ϕ are concentrated within a small interval and the posterior density of ϕ is very peaked around a single value, which is close to 0.3. However, in the fourth contour plot that uses derivative of fitted curve for complexity as the quantity of interest, there are two concentric shapes, this means the posterior density of ϕ has a bi-modal shape. The peaks

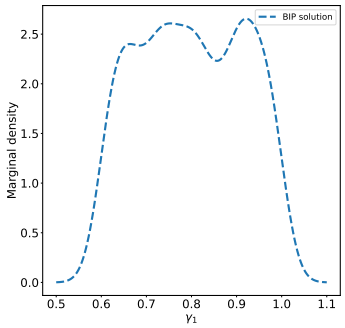


(a)

(b)



(c)



(d)

Figure 6.3: Marginal distribution for γ_1 generated from four quantitative measurements. (a) Average number of affected cells. (b) Average cell density. (c) Average complexity of dispersion pattern. (d) Derivative of fitted curve of complexity.

of the posterior are around 0.26 and 0.74. The results are also confirmed in the marginal distributions for γ_1 and ϕ in Figure 6.3 and Figure 6.4, respectively.

According to the plots, the model is more sensitive to the changes in ϕ than γ_1. It is not surprising that we can infer more information from the quantities of interest for the parameter ϕ because the characteristics of dispersion patterns, especially, the geometric shape of the boundaries, are mainly determined by the survival probability which determines how successful the population can build in new areas. However, different quantities of interest may provide distinct information about the parameter value.

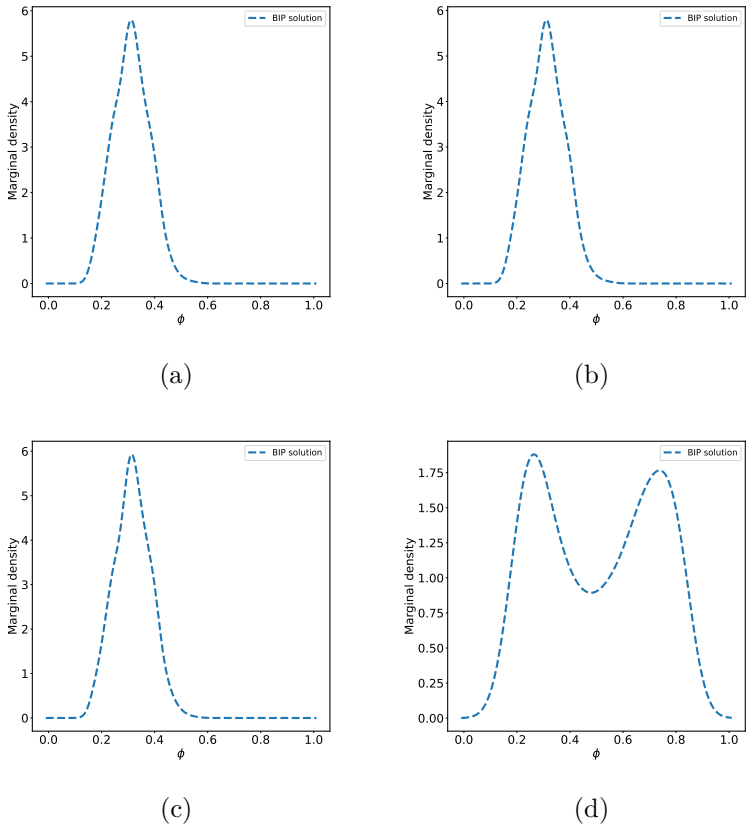


Figure 6.4: Marginal distribution for ϕ generated from four quantitative measurements. (a) Average number of affected cells. (b) Average cell density. (c) Average complexity of dispersion pattern. (d) Derivative of fitted curve of complexity.

6.5.2 Experiment 2

6.5.2.1 Data Generation

In this experiment, we adopt the SIP approach. For a data generating distribution, we assume a normal prior $N(0.8, (0.4/6)^2)$ for γ_1 and a beta prior Beta(2, 4) for ϕ with $\phi \in \left[\frac{1}{6}, \frac{5}{6}\right]$, treating them as independent.

6.5.2.2 Results

The density contour plots also show similar behaviours for γ_1 in Figure 6.5. Even though the contour lines show some irregular shapes in the fourth plot, the contour lines in general are almost parallel to the γ_1 axis, which indicate that the same output data can be matched with many values of γ_1 and the density for γ_1 is nearly uniform over the domain. This reflects the uniform Ansatz prior. This conclusion is confirmed in the marginal density plots in Figure 6.6.

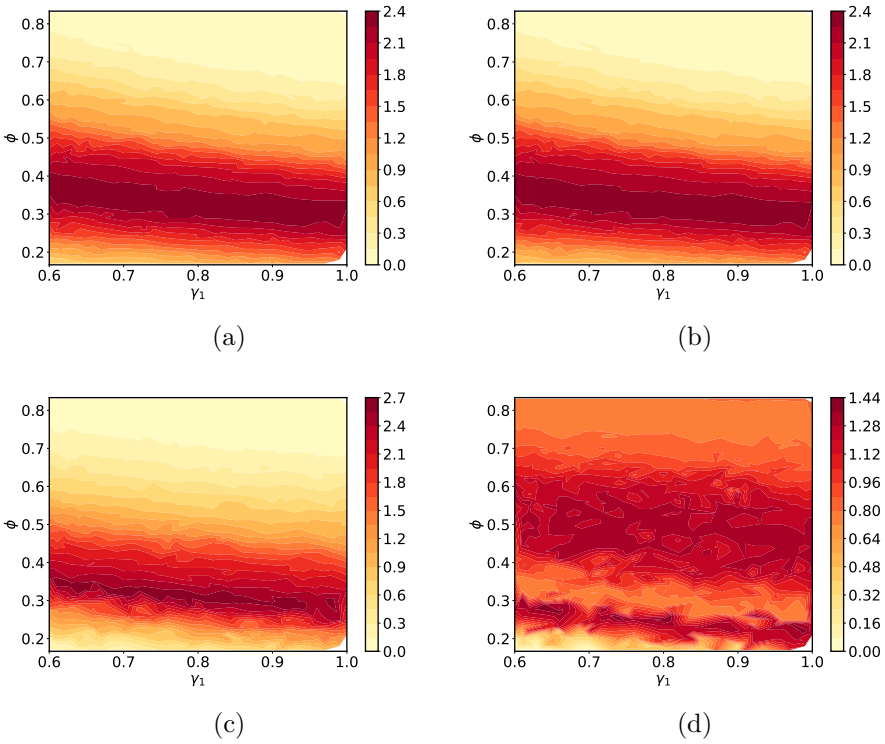


Figure 6.5: Bayesian posterior density contour plot for γ_1 and ϕ generated from four quantitative measurements. (a) Average number of affected cells. (b) Average cell density. (c) Average complexity of dispersion pattern. (d) Derivative of fitted curve of complexity.

For the first three quantities of interest, the contour lines are relatively close to each other, which indicates that the values of ϕ are concentrated within a small interval and the posterior density of ϕ is concentrated within a small interval. For the fourth quantity of

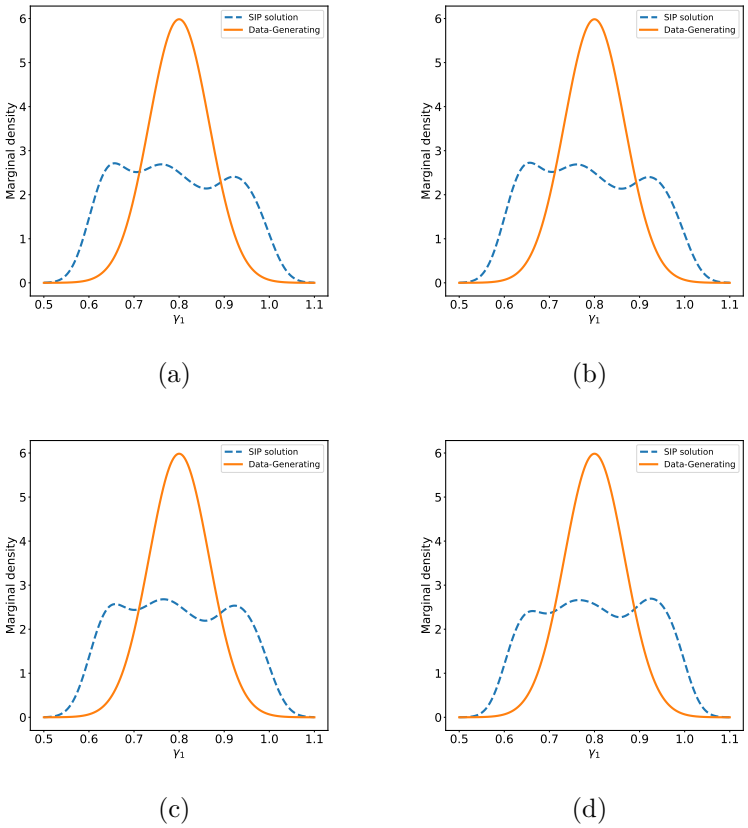


Figure 6.6: Marginal distribution for γ_1 generated from four quantitative measurements. (a) Average number of affected cells. (b) Average cell density. (c) Average complexity of dispersion pattern. (d) Derivative of fitted curve of complexity.

interest, the plot has fairly irregular contour lines and there are multiple concentric shapes in the plot, which points to a multi-modal posterior density for ϕ. This is also confirmed in Figure 6.7.

The marginal density plots for γ_1 in Figure 6.6 indicate the posterior generated from SIP approach has a much larger variance than the prior distribution, but the SIP approach can locate the values for γ_1 in the chosen domain. While the marginal plots for ϕ in Figure 6.7 display a high level of agreement between the posterior and prior densities except the one using derivative of fitted curve of complexity as the quantity of interest. This conclude that choosing the right quantity of interest can help to infer information for the parameter values.

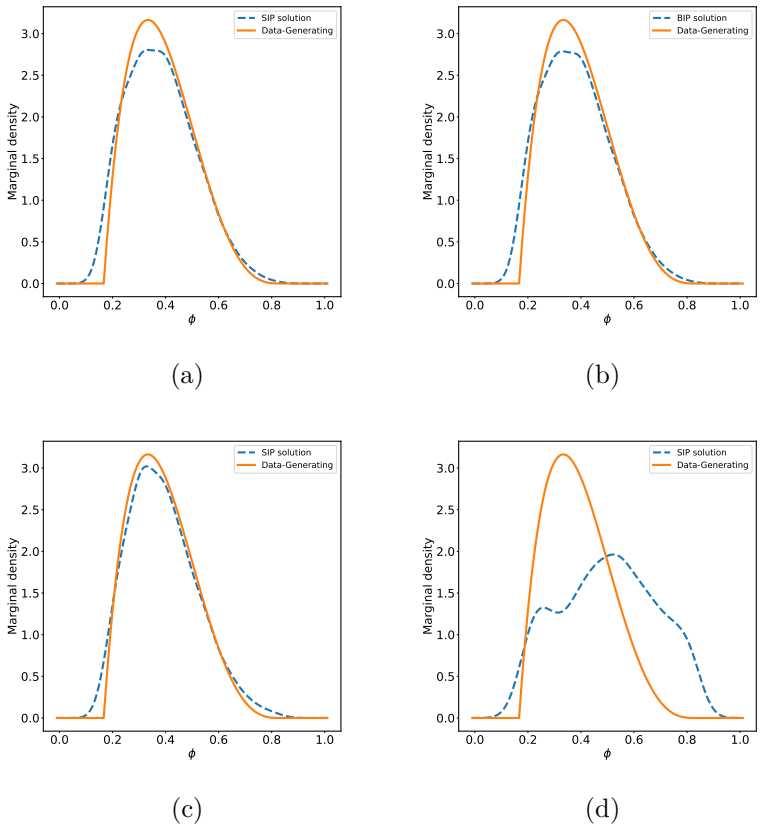


Figure 6.7: Marginal distribution for ϕ generated from four quantitative measurements. (a) Average number of affected cells. (b) Average cell density. (c) Average complexity of dispersion pattern. (d) Derivative of fitted curve of complexity.

6.5.3 Experiment 3

6.5.3.1 Data Generation

We conduct two experiments in this section. We assume the true value for γ_1 is 0.7 in both experiments, and we assume there is one true value for ϕ which is $\frac{2}{3}$ in the first experiment and there are two true values for ϕ in the second experiment, which are $\frac{1}{3}$ and $\frac{2}{3}$. We run the SCA model with the "true" parameter values for 30 time steps to obtain a dispersion pattern. In order to create the noisy observations, we add "noise" $N(0, 0.2^2)$ to the cells with non-zero density, and $N(0, 0.6^2)$ to the zero density cells that are adjacent to at least one non-zero density cell in the dispersion. If a cell has density that is above 1, then we set its density to 1; and if a cell has density that is below 0, we set its density to 0. We generate 1000 observations in both experiments with equal number of samples in the second experiment.

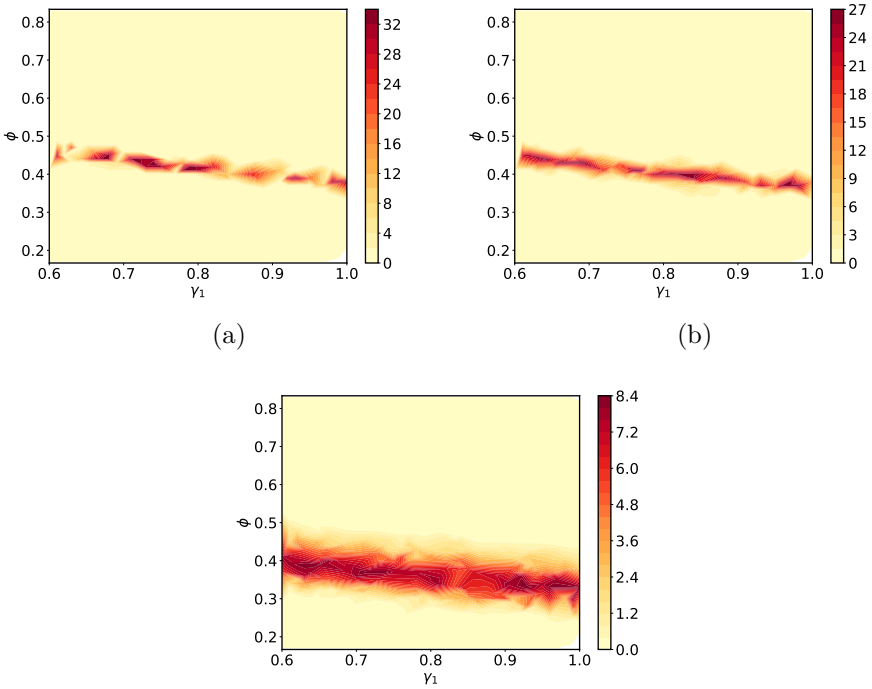
6.5.3.2 Results

The density contour plots in Figure 6.8 show distinct behaviours for γ_1. There is roughly one concentric shape in the first two plots while the contour lines are almost parallel to the γ_1 axis in the third plot. This concludes that the posterior density has a ridged shape using the average number of affected cells and average cell density as the quantities of interest, and the density has a nearly uniform behaviour when using the average complexity as the quantity of interest.

The density contour plots provide highly consistent indications for ϕ, there is one concentric shape, the posterior density for ϕ are concentrated within a very narrow region with a single mode.

The marginal densities for γ_1 in Figure 6.9 conclude that the experiment can locate most of the γ_1 values within the chosen domain. In Figure 6.10, the marginal densities for ϕ are all unimodal, but they peaked at slightly different values, 0.42, 0.4 and 0.38 for the three quantity of interest, respectively. They do not recover the "true" value of $\phi = \frac{2}{3}$, but they have very small variances.

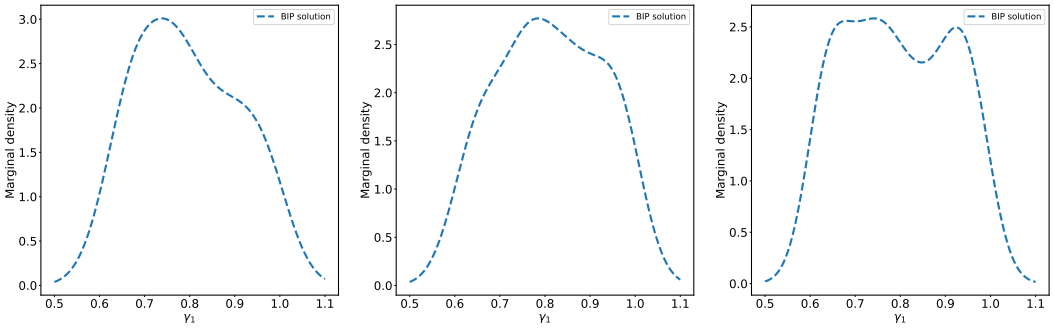
In the density contour plots in Figure 6.11, the contour lines are almost parallel to the γ_1 axis which indicates a nearly uniform posterior for γ_1, which is confirmed by the marginal density plot in Figure 6.12. There are two clear concentric areas for ϕ, this conclude a bi-modal posterior for ϕ, and the marginal densities in Figure 6.13 are all bi-modal with comparable variances. The marginal densities obtained from average number of affected cells and average cell density are peaked at the same values 0.41 and 0.64. However, the marginal density obtained from average complexity of dispersion pattern is peaked at slightly different values, which are 0.36 and 0.57. The marginal densities do not recover the "true" values for $\phi = \frac{1}{3}$ and $\phi = \frac{2}{3}$. Compared to the marginal densities in Experiment 3, the marginal



(a) (b)

(c)

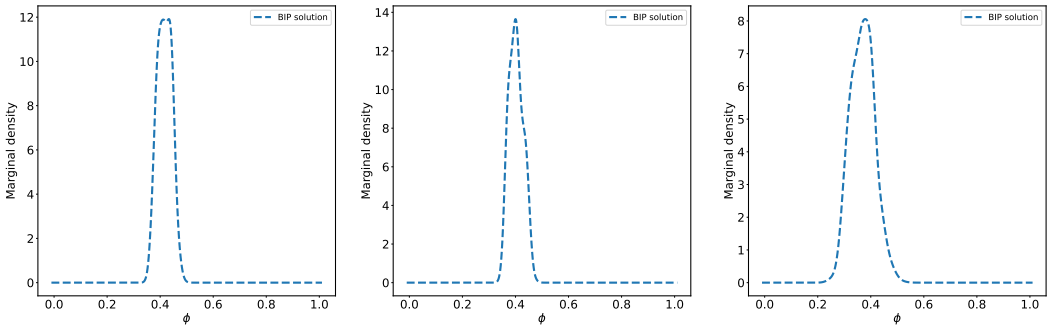
Figure 6.8: Bayesian posterior density contour plot for γ_1 and ϕ generated from three quantitative measurements. (a) Average number of affected cells. (b) Average cell density. (c) Average complexity of dispersion pattern.



(a) (b) (c)

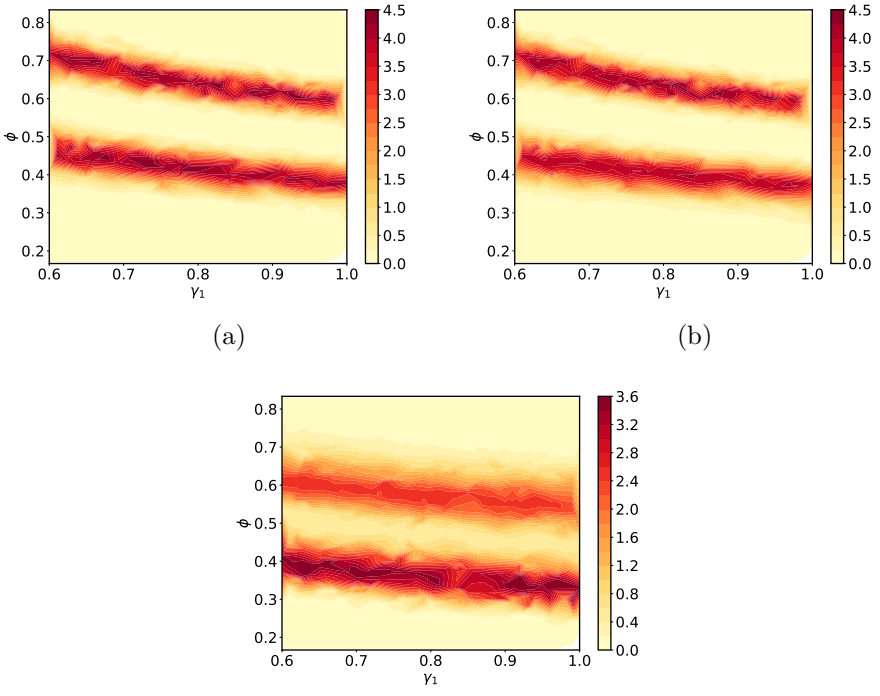
Figure 6.9: Marginal distribution for γ_1 generated from three quantitative measurements. (a) Average number of affected cells. (b) Average cell density. (c) Average complexity of dispersion pattern.

densities in Experiment 1 are more accurate, this is likely because the BIP solutions are affected by the "noise" added to the infestation map.



(a) (b) (c)

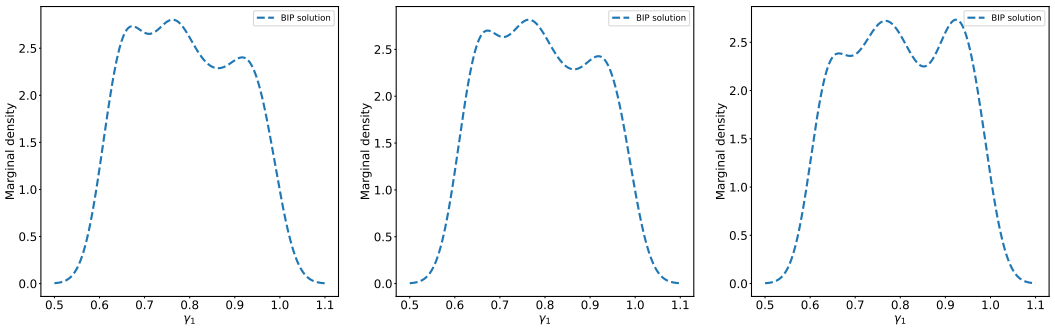
Figure 6.10: Marginal distribution for ϕ generated from three quantitative measurements. (a) Average number of affected cells. (b) Average cell density. (c) Average complexity of dispersion pattern.



(a) (b)

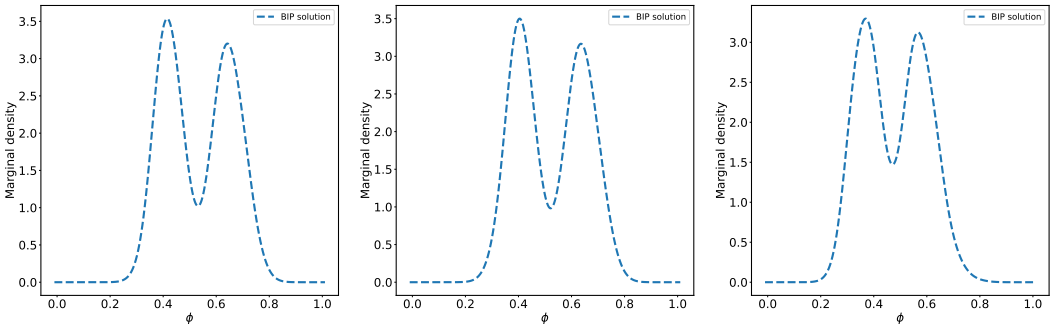
(c)

Figure 6.11: Bayesian posterior density contour plot for γ_1 and ϕ generated from three quantitative measurements. (a) Average number of affected cells. (b) Average cell density. (c) Average complexity of dispersion pattern.



(a) (b) (c)

Figure 6.12: Marginal distribution for γ_1 generated from three quantitative measurements. (a) Average number of affected cells. (b) Average cell density. (c) Average complexity of dispersion pattern.



(a) (b) (c)

Figure 6.13: Marginal distribution for ϕ generated from three quantitative measurements. (a) Average number of affected cells. (b) Average cell density. (c) Average complexity of dispersion pattern.

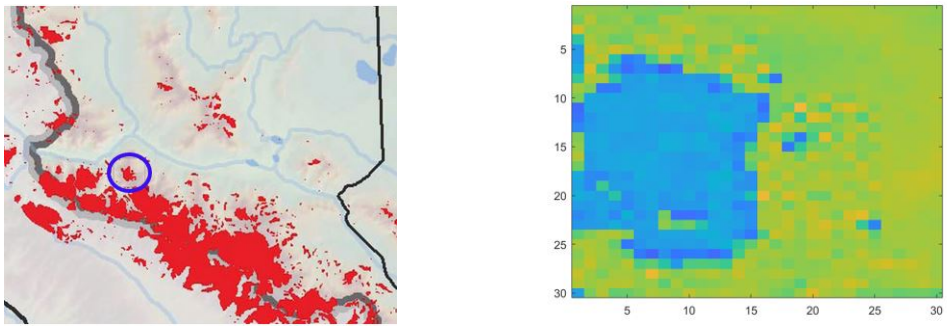
6.5.4 Experiment 4

6.5.4.1 Data Generation

The original aerial survey map in Figure 6.1 is in JPG format and it has 1291×1455 pixels translating into $2500km \times 3800km$. We use Photoshop to crop the image to a region of 32x32 pixels, representing a region of 58km x 58 km. At this scale, it is possible to identify well defined forested areas and there are many localized regions that hold a single mass of pine beetle infestation in a roughly ellipsoidal shape in the centre of a square. Given this is several years into an infestation, we believe that we should consider an infestation shape that is roughly ellipsoidal. We do not consider regions where the local infestation has a sparsely heterogeneous pattern at that scale as that indicates a very early stage of infestation locally and we do not know the initial conditions.

Algorithm 3 Algorithm for map conversion

1: Convert the colour image into grey scale with 256 shades.
2: Use imread in Matlab to read in the image and convert to a 32x32 matrix whose values are numbered from 1 to 256, divide the entries by 256 to get a grey density in each entry.
3: Remove the background colour greys by setting matrix values below a threshold (.8) to zero to form the mean vector for the noisy data.
4: Add iid zero mean normal noise to each nonzero value in the matrix, add $N(0, 0.2^2)$ to the cells with non-zero density, and $N(0, 0.6^2)$ to the zero density cells that are adjacent to at least one non-zero density cell. If a cell has density that is above 1, then we set its density to 1; and if a cell has density that is below 0, we set its density to 0.
5: Generate 100 observations.



(a) (b)

Figure 6.14: Cropped Image from aerial survey map for the MPB infestation in a local region. (a) Cropped image before transformation. (b) Cropped image after transformation.

6.5.4.2 Results

The density contour plots in Figure 6.15 indicate somewhat similar behaviours for γ_1. The contour lines have some irregular shapes in the first two plots but they are close to being parallel to the γ_1 axis, while the contour lines in the third plot do not show specific shapes and are almost parallel to the γ_1 axis. Therefore, the marginal density is closest to being uniform in the third plot in Figure 6.16.

Nevertheless, the density contour plots show consistent behaviours for ϕ, the values are concentrated within a very narrow region, which indicates a peaked shape around a single value, therefore, we calibrate ϕ to a small value interval. However, the observed peaked values are not quite the same, which are 0.42, 0.34 and 0.48, respectively, in Figure 6.17. The marginal densities all have very small variances.

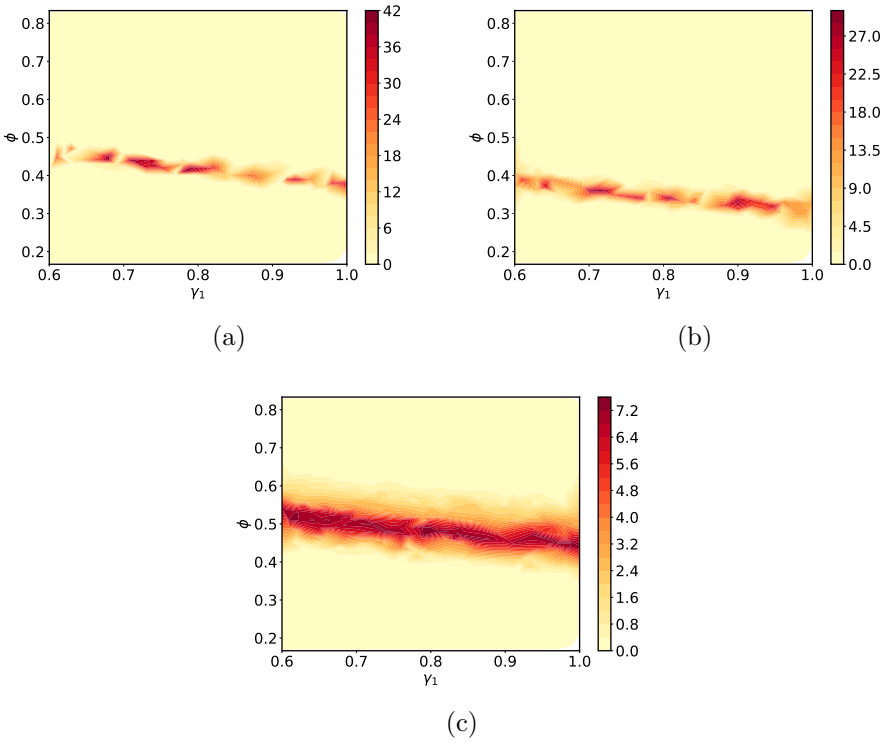
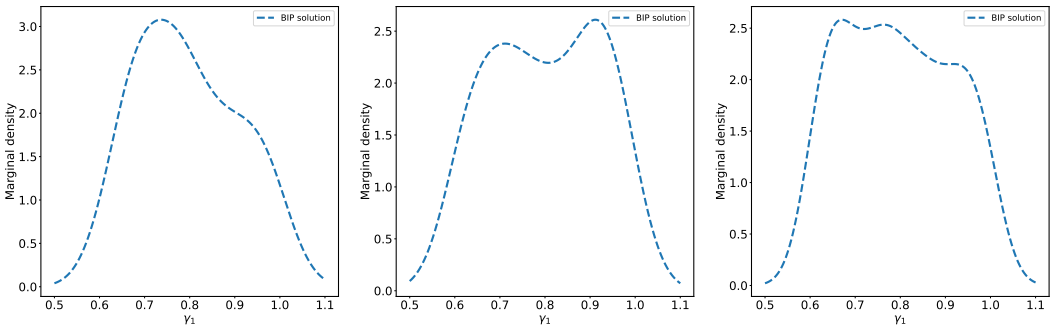
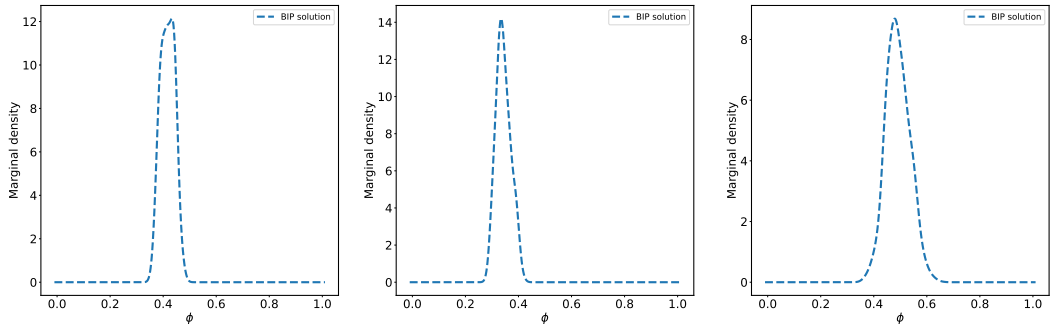


Figure 6.15: Bayesian posterior density contour plot for γ_1 and ϕ generated from three quantitative measurements. (a) Average number of affected cells. (b) Average cell density. (c) Average complexity of dispersion pattern.



(a) (b) (c)

Figure 6.16: Marginal distribution for γ_1 generated from three quantitative measurements. (a) Average number of affected cells. (b) Average cell density. (c) Average complexity of dispersion pattern.



(a) (b) (c)

Figure 6.17: Marginal distribution for ϕ generated from three quantitative measurements. (a) Average number of affected cells. (b) Average cell density. (c) Average complexity of dispersion pattern.

Chapter 7

Conclusion

7.1 Summary of Contributions

In this thesis, we construct a multiscale stochastic cellular automaton (SCA) model to characterize the macroscale dispersion process of a population considering collective behaviour of individuals at the microscale. The macroscale model is inspired by John Conway's Game-of-Life, but we use probabilistic rules to govern the cell states to describe stochasticity effects in the dispersion process. The model includes a bridge between components of the model at different scales, in which the simulation schemes at the macroscale employ the detailed information obtained from the microscale and the macroscale information initiates the microscale model. We apply the SCA model to the mountain pine beetle (MPB) infestation. The numerical simulation results suggest that the model is capable of producing patterns that resemble the physical characteristics of the MPB infestations, and the calibration process indicates that the model parameters can be determined from observed data.

A motivation for the construction of the SCA model is to use simple rules to generate complex behaviour. The modelling framework is of general interest and can be applied in other scientific practices to study a variety of phenomena. The transition rule of the model is simple, yet the model produces a wide range of behaviour using only a few parameters. In some cases at least, these parameters have physical interpretations. The structure of the model is flexible, the model can be integrated with various kinds of microscale models and can include multiple scales when there is an appropriate approach to link these scales.

Numerical simulations indicate that the SCA model is capable of generating various dispersion patterns. In general, the model behaviour approaches three limiting situations: fade away completely, settle into a fixed finite size, and grow without limit at a fixed speed. The dispersion patterns depend on the a couple of factors such as the choice of the microscale model, the model parameters and the interactions between the model parameters. Some parameters have similar effect and variations among those parameters tend to generate patterns with almost identical physical characteristics. The microscale model provides

the input of the macroscale SCA model, which mainly determines the number of affected cells and cell density at a specific time. The spreading rate and survival rate have similar effects on model behaviour as the microscale model. The inclusion probabilistic rule and the survival probabilistic rule determine the smoothness of the dispersion patterns, and the boundary conditions also play a critical role in determining the model output. In addition to visually evaluate dispersion patterns, we develop several statistics to quantify the dispersion patterns, which are also used in the model calibration process.

The numerical investigations of the calibration problem demonstrate that the SCA model can be calibrated from observed data. We use two approaches, the Stochastic Inverse Problem (SIP) and Bayesian Inverse Problem (BIP), to solve the calibration problem. For simulated data, we generate model output from a large number of realizations due to the stochastic nature of the SCA model, and we calibrate the model output by averaging over those realizations. For real data, we add "noise" to visual plot of a MPB infestation to obtain model output. We also investigate the sensitivity of the SCA model output to individual parameters in solving the calibration problem.

7.2 Future Investigations

We list a few possible subjects of further research.

7.2.1 Theoretical Analysis

In the thesis, we develop a two-dimensional SCA model to characterize the dispersion process, and we perform numerical simulations to study the two-dimensional SCA model behaviour because the SCA model is highly non-linear and stochastic which complicates theoretical analysis, and techniques that apply to linear Markov chain models are not applicable. One possible future research subject is to develop a formal theoretical analysis for the SCA model. A first step might be to derive a one-dimensional version.

7.2.2 Integration with Other Microscale Models

We use the birth-death process in the thesis to describe individual behaviour for the MPB infestation, and we convert the population size at the microscale to the cell density used as the model input at the macroscale. The model can integrate with other ecological models when there is an appropriate way to upscale the model output to the SCA model, i.e., to compute the microscale model output as the cell density for the SCA model.

7.2.3 Application to Other Fields

The multiscale SCA modelling framework is of general interest and can be applied in other scientific practices to study a variety of phenomena. Two potential applications of im-

mediate interest are the nuclear radiation damage and infectious disease transmission. The multiscale SCA model requires an appropriate method to describe the individual behaviour and to upscale the microscale model output.

7.2.4 Construction of Cascaded Stochastic Cellular Automaton Model

The SCA model in the thesis includes two scales, but the model can be expanded to handle a wider range of scales. One possible subject of further research is to utilize the SCA model flexibility to include more than two scales in the SCA model by constructing a "cascaded" SCA model. For example, in Figure 7.1, we assume individual behaviour occurs within each cell and the cells of the same colour belongs to the same interaction neighbourhood with highly similar physical properties. We call cells of the same colour the "child" unit, and the entire region the "parent" unit. We implement the SCA model within each "child" unit, in which the "child" unit for colouring areas do not intersect. We treat each "child" unit as cells in the "parent" unit and use the output of the "child" unit as the input for the SCA model on the "parent" unit. This cascaded SCA model can be performed efficiently in parallel. This interest in this approach is that it can be applied to a much larger region.

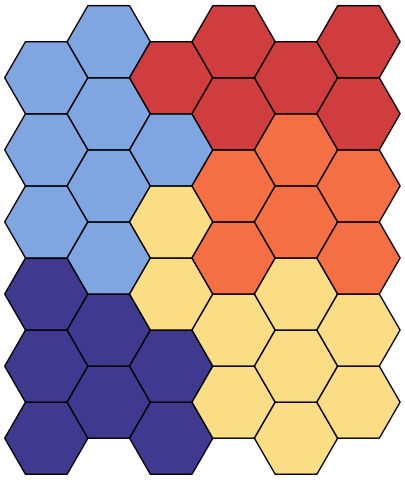


Figure 7.1: A plot of cascade stochastic cellular automaton model. The hexagons represent cells in the "child" unit, and the hexagons of the same colour form the interaction neighbourhoods of the "child" unit. The interaction neighbourhoods of the "child" unit form the cells of the "parent" unit.

Bibliography

[1] Mountain pine beetle.

[2] Mountain pine beetle, January 2024.

[3] Brant Abbott, Brad Stennes, and G. Cornelis van Kooten. Mountain pine beetle, global markets, and the british columbia forest economy. *Canadian Journal of Forest Research*, 39(7):1313–1321, 2009.

[4] Alexandru Agapie, Anca Andreica, and Marius Giuclea. Probabilistic cellular automata. *Journal of computational biology : a journal of computational molecular cell biology*, 21, 07 2014.

[5] Karim Ahmed and Anter El-Azab. Phase-field modeling of microstructure evolution in nuclear materials. *Handbook of Materials Modeling: Applications: Current and Emerging Materials*, pages 2313–2334, 2020.

[6] Mark Alber, Adrian Buganza Tepole, William R. Cannon, Suvranu De, Salvador Dura-Bernal, Krishna Garikipati, George E. Karniadakis, William W. Lytton, Ellen Kuhl, and Linda Petzold. Integrating machine learning and multiscale modeling—perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences. *npj Digital Medicine*, 2(1), 11 2019.

[7] M.U. Altaf, T. Butler, C. Dawson, I. Hoteit, X. Luo, and T. Mayo. Improving short-range ensemble kalman storm surge forecasting using robust adaptive inflation. *Monthly Weather Review*, 141:2705–2720, 2013.

[8] M.U. Altaf, T. Butler, T. Mayo, X. Luo, C. Dawson, A.W. Heemink, and I. Hoteit. A comparison of ensemble kalman filters for storm surge assimilation. *Monthly Weather Review*, 142:2899–2914, 2014.

[9] J.L. Anderson. An ensemble adjustment kalman filter for data assimilation. *Monthly weather review*, 129(12):2884–2903, 2001.

[10] JD Annan, JC Hargreaves, NR Edwards, and R. Marsh. Parameter estimation in an intermediate complexity earth system model using an ensemble kalman filter. *Ocean Modelling*, 8(1):135–154, 2005.

[11] Senthil Athithan, Vidya Shukla, and Sangappa Biradar. Dynamic cellular automata based epidemic spread model for population in patches with movement. *Journal of Computational Environmental Sciences*, 2014:1–8, 02 2014.

[12] Jackson P. Audley, Christopher J. Fettig, A. Steven Munson, Justin B. Runyon, Leif A. Mortenson, Brytten E. Steed, Kenneth E. Gibson, Carl L. Jørgensen, Stephen R. McKelvey, Joel D. McMillin, and Jose F. Negrón. Impacts of mountain pine beetle outbreaks on lodgepole pine forests in the intermountain west, u.s., 2004–2019. *Forest Ecology and Management*, 475:118403, 2020.

[13] Barbara Bentz. Mountain pine beetle, dendroctonus ponderosae (coleoptera: Curculionidae, scolytinae). In John L. Capinera, editor, *Encyclopedia of Entomology*, volume 4, pages 2494–2497. Springer, Dordrecht, Heidelberg, Germany, 2nd edition edition, 2008.

[14] A. Berryman and M. Ashraf. Effects of abies grandis resin on the attack behavior and brood survival of scolytus ventralis (coleoptera: Scolytidae). *Canadian Entomologist*, 102:1229–1236, 1970.

[15] Alan A. Berryman, Kenneth F. Raffa, Jeffrey A. Millstein, and Nils Chr. Stenseth. Interaction dynamics of bark beetle aggregation and conifer defense rates. *Oikos*, 56(2):256–263, 1989.

[16] Derek Bingham, Troy Butler, and Don Estep. Inverse problems for physics-based process models. *Annual Review of Statistics and Its Application*, 11(1), 2024.

[17] V.I. Bogachev. *Measure Theory (Volume 2)*. Springer-Verlag Berlin Heidelberg, 2007.

[18] Christopher Bone, Suzana Dragicevic, and Arthur Roberts. A fuzzy-constrained cellular automata model of forest insect infestations. *Ecological Modelling*, 192(1):107–125, 2006.

[19] J. H Borden, L. C Ryker, L. J Chong, H. D Pierce Jr, B. D Johnston, and A. C Oehlschlager. Response of the mountain pine beetle, dendroctonusponderosae hopkins (coleoptera: Scolytidae), to five semiochemicals in british columbia lodgepole pine forests. *Canadian journal of forest research*, 17(2):118–128, 1987.

[20] G. Box and N. Draper. *Empirical Model-Building and Response Surfaces*. John Wiley & Sons, 1987.

[21] J. Breidt, T. Butler, and D. Estep. A measure-theoretic computational method for inverse sensitivity problems I: Method and analysis. *SIAM Journal on Numr. Anal.*, 49(5):1836–1859, 2011.

[22] Clive M Brown, Pekka J Nuorti, Robert F Breiman, A Leroy Hathcock, Barry S Fields, Harvey B Lipman, Gerald C Llewellyn, Jo Hofmann, and Martin Cetron. A community outbreak of legionnaires' disease linked to hospital cooling towers: an epidemiological method to calculate dose of exposure. *International Journal of epidemiology*, 28(2):353–359, 1999.

[23] J. Brynjarsdóttir and A. O'Hagan. Learning about physical parameters: the importance of model discrepancy. *Inverse Problems*, 30(11):114007, oct 2014.

[24] D. G. Burnell. A dispersal-aggregation model for mountain pine beetle in lodgepole pine stands. *Population Ecology*, 19(1):99, 1977.

[25] T. Butler, M.U. Altaf, C. Dawson, I. Hoteit, X. Luo, and T. Mayo. Data assimilation within the advanced circulation (adcirc) modeling framework for hurricane storm surge forecasting. *Monthly Weather Review*, 140:2215–2231, 2012.

[26] T. Butler, P. Constantine, and T. Wildey. A posteriori error analysis of parameterized linear systems using spectral methods. *SIAM Journal on Matrix Analysis and Applications*, 33:195–209, 2012.

[27] T. Butler, C. Dawson, and T. Wildey. A posteriori error analysis of stochastic differential equations using polynomial chaos expansions. *SIAM Journal on Scientific Computing*, 33:1267–1291, 2011.

[28] T. Butler and D. Estep. A numerical method for solving a stochastic inverse problem for parameters. *Ann. Nucl. Energy*, 52(0):86–94, 2013.

[29] T. Butler, D. Estep, and P. Nishant. *A Ramble Through Probability: How I Learned to Stop Worrying and Love Measure Theory.* SIAM, Philadelphia, 2022. under contract.

[30] T. Butler, D. Estep, and J. Sandelin. A computational measure theoretic approach to inverse sensitivity problems II: A posteriori error analysis. *SIAM Journal on Numr. Anal.*, 50(1):22–45, 2012.

[31] T. Butler, D. Estep, S. Tavener, C. Dawson, and J.J. Westerink. A measure-theoretic computational method for inverse sensitivity problems III: Multiple quantities of interest. *SIAM/ASA Journal on Uncertainty Quantification*, 2(1):174–202, 2014. doi:10.1137/130930406.

[32] T. Butler, D. Estep, S. Tavener, T. Wildey, C. Dawson, and L. Graham. Solving stochastic inverse problems using sigma-algebras on contour maps. preprint available at arXiv:1407.3851, 2014.

[33] T. Butler, L. Graham, D. Estep, C. Dawson, and J.J. Westerink. Definition and solution of a stochastic inverse problem for the Manning's n parameter field in hydrodynamic models. *Adv. in Water Resour.*, 78(0):60 – 79, 2015.

[34] T. Butler, L. Graham, S. Mattis, and S. Walsh. A measure-theoretic interpretation of sample based numerical integration with applications to inverse and prediction problems under uncertainty. *SIAM Journal on Scientific Computing.* accepted, preprint available at digital.auraria.edu/IR00000054/00001.

[35] T. Butler, A. Huhtala, and M. Juntunen. Quantifying uncertainty in material damage from vibrational data. *Journal of Computational Physics*, 283:414 – 435, 2015.

[36] T. Butler, J. Jakeman, M. Pilosov, S. Walsh, and T. Wildey. A New Approach to Optimal Experimental Design Using Singular Values of Jacobians for Nonlinear Observable Maps. submitted, 2018.

[37] T. Butler, J. Jakeman, and T. Wildey. Combining push-forward measures and Bayes' rule to construct consistent solutions to stochastic inverse problems. *SIAM Journal on Scientific Computing*, 40(2):A984–A1011, 2018.

[38] T. Butler, J. Jakeman, and T. Wildey. Convergence of probability densities using approximate models for forward and inverse problems in uncertainty quantification. *SIAM Journal on Scientific Computing*, 40(5):A3523–A3548, 2018.

[39] T. Butler and S. Mattis. Enhanced surrogate modeling with adjoints for goal-oriented measure-theoretic inversion. preprint available at http://digital.auraria.edu/IR00000052/00001, 2016.

[40] T. Butler and T. Wildey. Utilizing adjoint-based error estimates for surrogate models to accurately predict probabilities of events. *International Journal for Uncertainty Quantification*, 8(2):143–159, 2018.

[41] Troy Butler, Clint Dawson, and Tim Wildey. Propagation of uncertainties using improved surrogate models. *SIAM/ASA Journal on Uncertainty Quantification*, 1:164–191, 2013.

[42] Troy Butler, T Wildey, and Tian Yu Yen. Data-consistent inversion for stochastic input-to-output maps. *Inverse Problems*, 36(8):085015, aug 2020.

[43] K. Buxton. Lodgepole pine trees killed by mountain pine beetle near bonaparte lake, bc.

[44] M ELKIN C and L REID M. Low energy reserves and energy allocation decisions affect reproduction by mountain pine beetles, dendroctonus ponderosae. *Functional Ecology*, 19(1):102–109, 2005.

[45] D. Cacuci. *Sensitivity and Uncertainty Analysis: Theory*, volume I. Chapman & Hall/CRC, 1997.

[46] Ignatius Cahyanto, Lori Pennington-Gray, and Jeffrey Wehrung. The mountain pine beetle: A study of tourism businesses' perceptions of the risk of ecological disturbances. *Journal of Park and Recreation Administration*, 36:24–40, 01 2018.

[47] Muffy Calder, Claire Craig, Dave Culley, Richard de Cani, Christl A. Donnelly, Rowan Douglas, Bruce Edmonds, Jonathon Gascoigne, Nigel Gilbert, Caroline Hargrove, Derwen Hinds, David C. Lane, Dervilla Mitchell, Giles Pavey, David Robertson, Bridget Rosewell, Spencer Sherwin, Mark Walport, and Alan Wilson. Computational modelling for decision-making: where, why, what, who and how. *Royal Society Open Science*, 5(6):172096, June 2018.

[48] J. Chi, D. Estep, and H. Wang. A sliced inverse Bayesian approach for stochastic inverse problems: The continuum case, 2022. in preparation.

[49] J. Chi, D. Estep, and H. Wang. A sliced inverse Bayesian approach for stochastic inverse problems: The discrete case, 2022. in preparation.

[50] Jiarui Chi. Sliced inverse approach and domain recovery for stochastic inverse problems, 2021. Ph.D. Thesis, Department of Statistics.

[51] Alex M. Chubaty, Bernard D. Roitberg, and Chao Li. A dynamic host selection model for mountain pine beetle, dendroctonus ponderosae hopkins. *Ecological Modelling*, 220(9):1241–1250, 2009.

[52] I. Colomina and P. Molina. Unmanned aerial systems for photogrammetry and remote sensing: A review. 92:79–97, 2014.

[53] L. J. Corbett, P. Withey, V. A. Lantz, and T. O. Ochuodho. The economic impact of the mountain pine beetle infestation in British Columbia: provincial estimates from a CGE analysis. *Forestry: An International Journal of Forest Research*, 89(1):100–105, 11 2015.

[54] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, Sep 1995.

[55] S. Cotter, M. Dashti, and A. Stuart. Approximation of Bayesian inverse problems. *SIAM Journal of Numerical Analysis*, 48:322–345, 2010.

[56] Koby Crammer and Yoram Singer. On the algorithmic implementation of multiclass kernel-based vector machines. *J. Mach. Learn. Res.*, 2:265–292, March 2002.

[57] Katherine F. Crowley, Gary M. Lovett, Mary A. Arthur, and Kathleen C. Weathers. Long-term effects of pest-induced tree species change on carbon and nitrogen cycling in northeastern u.s. forests: A modeling analysis. *Forest Ecology and Management*, 372:269–290, 2016.

[58] Science Coordinator Crown Managers Partnership. Mountain pine beetle in the crown of the continent ecosystem, 2010. The aerial survey map of mountain pine beetle infestation in the Crown of the Continent Ecosystem is from US Geological Survey database, https://www.sciencebase.gov/catalog/item/560312d0e4b03bc34f5449aa.

[59] D. D'Ambrosio, S. Di Gregorio, and G. Iovine. Simulating debris flows through a hexagonal cellular automata model: Sciddica $s_{3^{\smile}hex}$. *Natural Hazards and Earth System Sciences*, 3(6):545–559, 2003.

[60] D. D'Ambrosio, Salvatore Di Gregorio, and Giulio Iovine. Simulating debris flows through a hexagonal cellular automata model: Sciddica s3-hex. *Natural Hazards and Earth System Sciences - NAT HAZARDS EARTH SYST SCI*, 3:545–559, 11 2003.

[61] Renato Cordeiro de Amorim and Christian Hennig. Recovering the number of clusters in data sets with noise features using feature rescaling factors. *Information Sciences*, 324:126 – 145, 2015.

[62] Renato Cordeiro de Amorim and Boris Mirkin. Minkowski metric, feature weighting and anomalous cluster initializing in k-means clustering. *Pattern Recognition*, 45(3):1061 – 1075, 2012.

[63] Alberta Sustainable Resource Development. Mountain pine beetle : management strategy, 12 2007.

[64] Amalesh Dhar and Chris DB Hawkins. Regeneration and growth following mountain pine beetle attack: A synthesis of knowledge. *BC journal of ecosystems and management*, 12(2), 2011.

[65] Adam Duarte and Ivana Mali. Modeling dispersal processes for ecological systems. In Brian Fath, editor, *Encyclopedia of Ecology (Second Edition)*, pages 154–163. Elsevier, Oxford, second edition edition, 2019.

[66] Jacob P. Duncan, J. Powell, Luis F. Gordillo, and J. Eason. A model for mountain pine beetle outbreaks in an age-structured forest: Predicting severity and outbreak-recovery cycle period. *Bulletin of Mathematical Biology*, 77:1256–1284, 2015.

[67] D. Elfverson, D. Estep, F. Hellman, and A. Malqvist. Uncertainty quantification for approximate p-quantiles for physical models with stochastic inputs. *SIAM ASA Journal on Uncertainty Quantification*, 2:826–850, 2014.

[68] BC Environment and British Columbia. Ministry of Forests. *Bark Beetle Management Guidebook*. Forest practices code of British Columbia. Forest Service, British Columbia, 1995.

[69] D. Estep, A. Malqvist, and S. Tavener. Nonparametric density estimation for randomly perturbed elliptic problems I: Computational methods, a posteriori analysis, and adaptive error control. *SIAM Journal on Scientific Computing*, 31:2935–2959, 2009.

[70] M. L. Evenden, C. M. Whitehouse, and J. Sykes. Factors Influencing Flight Capacity of the Mountain Pine Beetle (Coleoptera: Curculionidae: Scolytinae). *Environmental Entomology*, 43(1):187–196, 02 2014.

[71] Geir Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans (1978–2012)*, 99(C5):10143–10162, 1994.

[72] Geir Evensen. The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dynamics*, 53:343–367, 2003.

[73] William Feller. *An introduction to probability theory and its applications*. Second edition. John Wiley & Sons Inc., Wiley New York, 1971.

[74] Christopher J Fettig, Christopher Asaro, John T Nowak, Kevin J Dodds, Kamal J K Gandhi, Jason E Moan, and Jeanne Robert. Trends in Bark Beetle Impacts in North America During a Period (2000–2020) of Rapid Environmental Change. *Journal of Forestry*, 2022.

[75] B. Fitzpatrick. Bayesian analysis in inverse problems. *Inverse Problems*, 7(5):675, 1991.

[76] Andreas Flache and Rainer Hegselmann. Understanding complex social dynamics: a plea for cellular automata based modelling. *Journal of Artificial Societies and Social Simulation*, 1(3), 1998.

[77] Ashlee N. Ford Versypt. Multiscale modeling in disease. *Current Opinion in Systems Biology*, 27:100340, 2021.

[78] M. M. Furniss, L. N. Kline, R. F. Schmitz, and J. A. Rudinsky. Tests of Three Pheromones to Induce or Disrupt Aggregation of Douglas-Fir Beetles (Coleoptera: Scolytidae) on Live Trees. *Annals of the Entomological Society of America*, 65(5):1227–1232, 09 1972.

[79] Joslin G., D. Bingham, J. Holloway, M. Grosskopf, C. Kuranz, and E. Rutter. Prediction and computer model calibration using outputs from multifidelity simulators. *Technometrics*, 55(4):501–512, 2013.

[80] D. Galbally, K. Fidkowski, K. Willcox, and O. Ghattas. Nonlinear model reduction for uncertainty quantification in large-scale inverse problems. *International Journal For Numerical Methods in Engineering*, 81:1581–1608, 2010.

[81] R. I. Gara, D. R. Geiszler, and W. R. Littke. Primary Attraction of the Mountain Pine Beetle to Lodgepole Pine in Oregon. *Annals of the Entomological Society of America*, 77(4):333–334, 07 1984.

[82] Martin Gardner. The fantastic combinations of jhon conway's new solitaire game "life". *Scientific American*, 223:20–123, 1970.

[83] Martin Gardner. Mathematical games: on cellular automata, self-reproduction, the garden of eden and the game "life". *Scientific American*, 224:112–117, 1971.

[84] D. R. Geiszler, V. F. Gallucci, and R. I. Gara. Modeling the dynamics of mountain pine beetle aggregation in a lodgepole pine stand. *Oecologia*, 46(2):244–253, 1980.

[85] Kenneth E. Gibson. Management guide for mountain pine beetle dendroctonus ponderosae hopkins. USDA Forest Service, 2010.

[86] W.R. Gilks, S. Richardson, and D.J. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman % Hall, 1996.

[87] R. W. Gorte. Mountain pine beetles and forest destruction: Effects, responses, and relationship to climate change. 2009.

[88] L. Graham, T. Butler, S. Walsh, C. Dawson, and J. Westerink. A measure-theoretic algorithm for estimating bottom friction in a coastal inlet: Case study of Bay St. Louis during Hurricane Gustav (2008). *Monthly Weather Review*, 145(3):929–954, 2017.

[89] L. Graham, S. Mattis, S. Walsh, T. Butler, M. Pilosov, and D. McDougall. BET: Butler, Estep, Tavener Method v3.0.0, 2020. DOI: 10.5281/zenodo.3936258, Documentation: http://ut-chg.github.io/BET/.

[90] M. Grosskopf, D. Bingham, M. Adams, W.D. Hawkins, and D. Perez-Nunez. Generalized computer model calibration for radiation transport simulation. *Technometrics*, 00, 2020.

[91] E. Matthew Hansen. Forest Development and Carbon Dynamics after Mountain Pine Beetle Outbreaks. *Forest Science*, 60(3):476–488, 09 2013.

[92] Sarah J. Hart, Thomas T. Veblen, Karen S. Eisenhart, Daniel Jarvis, and Dominik Kulakowski. Drought induces spruce beetle (dendroctonus rufipennis) outbreaks across northwestern colorado. *Ecology*, 95(4):930–939, 2014.

[93] Brian J. Harvey, Daniel C. Donato, and Monica G. Turner. Recent mountain pine beetle outbreaks, wildfire severity, and postfire tree regeneration in the us northern rockies. *Proceedings of the National Academy of Sciences*, 111(42):15120–15125, 2014.

[94] Justin Heavilin and James Powell. A novel method of fitting spatio-temporal models to data, with applications to the dynamics of mountain pine beetles. *Natural Resource Modeling*, 21(4):489 – 524, 2008.

[95] D. Henderson, R. Boys, and D. Wilkinson. Bayesian calibration of a stochastic kinetic computer model using multiple data sources. *Biometrics*, 66:249–256, 2010.

[96] L. Hernández Encinas, S. Hoya White, A. Martín del Rey, and G. Rodríguez Sánchez. Modelling forest fire spread using hexagonal cellular automata. *Applied Mathematical Modelling*, 31(6):1213–1227, 2007.

[97] Molly Heuss, A. D'Amato, and K. Dodds. Northward expansion of southern pine beetle generates significant alterations to forest structure and composition of globally rare pinus rigida forests. *Forest Ecology and Management*, 434:119–130, 2019.

[98] Jeffrey A. Hicke, Bingbing Xu, Arjan J.H. Meddens, and Joel M. Egan. Characterizing recent bark beetle-caused tree mortality in the western united states from aerial surveys. *Forest Ecology and Management*, 475:118402, 2020.

[99] D. Higdon, J. Gattiker, B. Williams, and M. Rightley. Computer model calibration using high-dimensional output. *Journal of the American Statistical Association*, 103(482):570–583, 2008.

[100] D. Higdon, M. Kennedy, J. Cavendish, J. Cafeo, and R. Ryne. Combining field data and computer simulations for calibration and prediction. *SIAM J. Sci. Comput.*, 26:448–466, 2004.

[101] Umeaa (Sweden). Inst. foer Skogsteknik) Holmgren, P. (SLU and T Thuresson. Satellite remote sensing for forestry planning - a review. *Scandinavian journal of forest research*, 13(1):90–110, 1998.

[102] I. Hoteit, B. Cornuelle, A. Kohl, and D. Stammer. Treating strong adjoint sensitivities in tropical eddy-permitting variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 131:3659–3682, 2005.

[103] J. Idier, editor. *Bayesian Approach to Inverse Problems*. Wiley, New York, 2010.

[104] Andrew Ilachinski. *Cellular Automata*. WORLD SCIENTIFIC, 2001.

[105] Michael J. Jenkins, Justin B. Runyon, Christopher J. Fettig, Wesley G. Page, and Barbara J. Bentz. Interactions among the Mountain Pine Beetle, Fires, and Fuels. *Forest Science*, 60(3):489–501, 08 2013.

[106] Kelsey L. Jones, Victor A. Shegelski, Nathan G. Marculis, Asha N. Wijerathna, and Maya L. Evenden. Factors influencing dispersal by flight in bark beetles (coleoptera: Curculionidae: Scolytinae): from genes to landscapes. *The TRIA-Net Project: collaborative ecological genomics for forest health*, 01(01):1024–1041, 2019.

[107] Rudolph Emil Kalman et al. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45, 1960.

[108] Ioannis Karafyllidis and Adonios Thanailakis. A model for predicting forest fire spreading using cellular automata. *Ecological Modelling*, 99(1):87–97, 1997.

[109] Jarkko Kari. Theory of cellular automata: A survey. *Theoretical Computer Science*, 334(1):3–33, 2005.

[110] Merrill Kaufmann, Gregory Aplet, Michael Babler, William Baker, Barbara Bentz, Michael Harrington, Brad Hawkes, Laurie Huckaby, Michael Jenkins, Dan Kashian, Robert Keane, Dominik Kulakowski, Ward McCaughey, Charles McHugh, Jose Negron, John Popp, William Romme, Tania Schoennagel, Wayne Shepperd, and Thomas Veblen. *The Status of Our Scientific Understanding of Lodgepole Pine and Mountain Pine Beetles – A Focus on Forest Ecology and Fire Behavior*. Jan 2008.

[111] Markus Kautz, Peter Anthoni, Arjan J. H. Meddens, Thomas A. M. Pugh, and Almut Arneth. Simulating the recent impacts of multiple biotic disturbances on forest carbon cycling across the united states. *Global change biology*, 24(5):2079–2092, 2018.

[112] Lori J. Kayes and Daniel B. Tinker. Forest structure and regeneration following a mountain pine beetle epidemic in southeastern wyoming. *Forest Ecology and Management*, 263:57–66, 2012.

[113] M. Kennedy and A. O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, 2001.

[114] A. Ross Kiester and Kevin Sahr. Planar and spherical hierarchical, multi-resolution cellular automata. *Computers, Environment and Urban Systems*, 32(3):204–213, 2008. Discrete Global Grids.

[115] Vlastimil Křivan, Mark Lewis, Barbara J. Bentz, Sharon Bewick, Suzanne M. Lenhart, and Andrew Liebhold. A dynamical model for bark beetle outbreaks. *Journal of Theoretical Biology*, 407:25–37, 2016.

[116] Dean Koch, Mark A. Lewis, and Subhash Lele. The signature of endemic populations in the spread of mountain pine beetle outbreaks. *Bulletin of Mathematical Biology*, 83(6):65, May 2021.

[117] Peter Kumberger, Felix Frey, Ulrich S. Schwarz, and Frederik Graw. Multiscale modeling of virus replication and spread. *FEBS Letters*, 590(13):1972–1986, 2016.

[118] Mélodie Kunegel-Lion and Mark Lewis. Factors governing outbreak dynamics in a forest intensively managed for mountain pine beetle. *Scientific Reports*, 10:7601, 05 2020.

[119] Evan Kwiatkowski and Jan Mandel. Convergence of the square root ensemble kalman filter in the large ensemble limit. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):1–17, 2015.

[120] O Langmead and Charles Sheppard. Coral reef community dynamics and disturbance: A simulation model. *Ecological Modelling - ECOL MODEL*, 175:271–290, 07 2004.

[121] Joseph R Lentino, Mari Ann Rosenkranz, Jacqueline A Michaels, Viswanath P Kurup, Harold D Rose, and Michael W Rytel. Nosocomial aspergillosis: a retrospective review of airborne disease secondary to road construction and contaminated air conditioners. *American journal of Epidemiology*, 116(3):430–437, 1982.

[122] J. Loeppky, D. Bingham, and W. Welch. Computer model calibration or tuning in practice. *University of British Columbia, Vancouver, BC, Canada, Report*, (221), 2006.

[123] JA Logan, WW Macfarlane, and L Willcox. Effective monitoring as a basis for adaptive management: a case history of mountain pine beetle in Greater Yellowstone Ecosystem whitebark pine. *iForest - Biogeosciences and Forestry*, (1):19–22, 2009. This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

[124] Jesse A Logan, Peter White, Barbara J Bentz, and James A Powell. Model analysis of spatial patterns in mountain pine beetle outbreaks. *Theoretical Population Biology*, 53(3):236–255, 1998.

[125] E. Louise Loudermilk and Wendell P. Cropper. Multiscale modeling of longleaf pine (pinus palustris). *Canadian Journal of Forest Research*, 37(11):2080–2089, 2007.

[126] Pierre-Yves Louis and Francesca Nardi. *Probabilistic Cellular Automata : Theory, Applications and Future Perspectives*. 01 2018.

[127] Jean Mairesse and Irène Marcovici. Around probabilistic cellular automata. *Theoretical Computer Science*, 559:42–72, 2014. Non-uniform Cellular Automata.

[128] Sharon Martinson, Tiina Ylioja, Brian Sullivan, Ronald Billings, and Matthew Ayres. Alternate attractors in the population dynamics of a tree-killing bark beetle. *Population Ecology*, 55:95–106, 01 2012.

[129] Youssef M. Marzouk, Habib N. Najm, and Larry A. Rahn. Stochastic spectral methods for efficient bayesian solution of inverse problems. *Journal of computational physics*, 224(2):560–586, 2007.

[130] Erik Matthysen. Multicausality of dispersal: a review. *Dispersal ecology and evolution*, 27:3–18, 2012.

[131] S. Mattis, T. Butler, C. Dawson, D. Estep, and V. Vessilinov. Parameter estimation and prediction for groundwater contamination based on measure theory. *Water Resources Research*, 51:7608–7629, 2015.

[132] S.A. Mattis, K.R. Steffen, T. Butler, C.N. Dawson, and D. Estep. Learning quantities of interest from dynamical systems for observation-consistent inversion. *Computer Methods in Applied Mechanics and Engineering*, 388:114230, 2022.

[133] B A Maurer. Scale. In Sven Erik Jorgensen and Brian Fath, editors, *Encyclopedia of Ecology*, pages 3154 – 3159. Elsevier Science, 2008.

[134] Arjan J. H. Meddens, Jeffrey A. Hicke, and Charles A. Ferguson. Spatiotemporal patterns of observed bark beetle-caused tree mortality in british columbia and the western united states. *Ecological Applications*, 22(7):1876–1891, 2012.

[135] Garrett W. Meigs, Robert E. Kennedy, Andrew N. Gray, and Matthew J. Gregory. Spatiotemporal dynamics of recent mountain pine beetle and western spruce budworm outbreaks across the pacific northwest region, usa. *Forest Ecology and Management*, 339:71–86, 2015.

[136] Eli K. Melaas, Mark A. Friedl, and Andrew D. Richardson. Multiscale modeling of spring phenology across deciduous forests in the eastern united states. *Global Change Biology*, 22(2):792–805, 2016.

[137] Melanie Mitchell. *Computation in Cellular Automata: A Selected Review*, chapter 4, pages 95–140. John Wiley & Sons, Ltd, 1998.

[138] Henry A. Moeck and Clarence S. Simmons. Primary attraction of mountain pine beetle, dendroctonus ponderosae hopk. (coleoptera: Scolytidae), to bolts of lodgepole pine. *The Canadian Entomologist*, 123(2):299–304, 1991.

[139] Ned Morgan. Bark beetles & new terrain: How sun peaks rose to the challenge, November 2014.

[140] Rebecca E Morrison, Todd A Oliver, and Robert D Moser. Representing model inadequacy: A stochastic operator approach. 2017.

[141] Klaus Mosegaard and Albert Tarantola. Monte Carlo sampling of solutions to inverse problems. *Journal of Geophysical Research*, 100(B7):12431–12447, 1995.

[142] Ran Nathan. The challenges of studying dispersal. *Trends in Ecology & Evolution*, 16(9):481–483, 2001.

[143] H. Owhadi, C. Scovel, and T. Sullivan. On the brittleness of Bayesian inference. *SIAM Review*, 57(4):566–582, 2015.

[144] J. Panchal, S. Kalidindi, and D. McDowell. Key computational modeling issues in integrated computational materials engineering. *Computer-Aided Design*, 45(1):4 – 25, 2013. Computer-aided multi-scale materials and product design.

[145] Shayn M. Peirce-Cottler and Alison Marsden. Multiscale computational modeling of biomedical systems: current approaches and payoffs. *Current Opinion in Biomedical Engineering*, 11:A1–A3, 2019.

[146] Colin A. Penn, Lindsay A. Bearup, Reed M. Maxwell, and David W. Clow. Numerical experiments to explain multiscale hydrological responses to mountain pine beetle tree mortality in a headwater watershed. *Water Resources Research*, 52(4):3143–3161, 2016.

[147] M. Plumlee. Bayesian calibration of inexact computer models. *Journal of the American Statistical Association*, 112(519):1274–1285, 2017.

[148] James A. Powell, Martha J. Garlick, Barbara J. Bentz, and Nicholas Friedenberg. Differential dispersal and the allee effect create power-law behaviour: Distribution of spot infestations during mountain pine beetle outbreaks. *Journal of Animal Ecology*, 87(1):73–86, 2018.

[149] James A. Powell, Jesse A. Logan, and Barbara J. Bentz. Local projections for a global model of mountain pine beetle attacks. *Journal of Theoretical Biology*, 179(3):243–260, 1996.

[150] James Price, Daniel McCollum, and Robert Berrens. Insect infestation and residential property values: A hedonic analysis of the mountain pine beetle epidemic. *Forest Policy and Economics*, 12:415–422, 07 2010.

[151] Robert A. Progar, Nancy Gillette, Christopher J. Fettig, and Kathryn Hrinkevich. Applied chemical ecology of the mountain pine beetle. *Forest Science*, 60(3):414–433, 09 2013.

[152] Ravi Radhakrishnan. A survey of multiscale modeling: Foundations, historical milestones, current status, and future prospects. *Aiche Journal. American Institute of Chemical Engineers*, 67, 2021.

[153] Kenneth F Raffa, Briah H Aukema, Nadir Erbilgin, Kier D Klepzig, and Kimberly F Wallin. Interactions among conifer terpenoids and bark beetles across multiple levels of scale: an attempt to understand links between population patterns and physiological processes. *Recent Advances in Phytochemistry 39: 79-118*, 2005.

[154] Kenneth F. Raffa, Jean-Claude Grégoire, and B. Staffan Lindgren. Chapter 1 - natural history and ecology of bark beetles. In *Bark Beetles*. Elsevier Inc, 2015.

[155] A.P. Ramakrishnan. Dispersal–migration. In Sven Erik Jørgensen and Brian D. Fath, editors, *Encyclopedia of Ecology*, pages 930–938. Academic Press, Oxford, 2008.

[156] Todd Redding, Rita Winkler, Pat Teti, D. Spittlehouse, Sarah Boon, John Rex, and Steve Chatwin. Mountain pine beetle and watershed hydrology. *Ecosystems*, 9:33–50, 01 2009.

[157] R. W. Reid. Biology of the mountain pine beetle, dendroctonus monticolae hopkins, in the east kootenay region of british columbia i. life cycle, brood development, and flight periods. *The Canadian Entomologist*, 94(5):531–538, 1962.

[158] R. W. Reid. Biology of the mountain pine beetle, dendroctonus monticolae hopkins, in the east kootenay region of british columbia.: Iii. interaction between the beetle and its host, with emphasis on brood mortality and survival. *The Canadian Entomologist*, 95(3):225–238, 1963.

[159] Srujan Rokkam, Anter El-Azab, Paul Millett, and Dieter Wolf. Phase field modeling of void nucleation and growth in irradiated metals. *Modelling and Simulation in Materials Science and Engineering*, 17(6):064002, aug 2009.

[160] William H. Romme, Jon Clement, Jeffrey A. Hicke, Dominik Kulakowski, Lee H. Macdonald, Tania Schoennagel, and Thomas T. Veblen. Recent forest insect outbreaks and fire risk in colorado forests: A brief synthesis of relevant research. 2006.

[161] Victoria A. Saab, Quresh S. Latif, Mary M. Rowland, Tracey N. Johnson, Anna D. Chalfoun, Steven W. Buskirk, Joslin E. Heyward, and Matthew A. Dresser. Ecological Consequences of Mountain Pine Beetle Outbreaks for Wildlife in Western North American Forests. *Forest Science*, 60(3):539–559, 10 2013.

[162] J. Sacks, W. Welch, T. Mitchell, and H. Wynn. Design and analysis of computer experiments. *Statistical Science*, pages 409–435, 1989.

[163] L. Safranyik, Hugh Barclay, Alan Thomson, and W. Riel. A population dynamics model for the mountain pine beetle, dendroctonus ponderosae hopk. (coleoptera : Scolytidae). 01 1999.

[164] L. Safranyik and A.L. Carroll. The biology and epidemiology of the mountain pine beetle in lodgepole pine forests. In L. Safranyik and W.R. Wilson, editors, *The Mountain Pine Beetle: A Synthesis of Biology, Management, and Impacts on Lodgepole Pine*, pages 3–66. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, British Columbia, 2006.

[165] L. Safranyik, D. A. Linton, R. Silversides, and L. H. McMullen. Dispersal of released mountain pine beetles under the canopy of a mature lodgepole pine stand. *Journal of Applied Entomology*, 113(1-5):441–450, 1992.

[166] Ugur Sahin, Selman Uguz, Hasan Akın, and Irfan Siap. Three-state von neumann cellular automata and pattern generation. *Applied Mathematical Modelling*, 39(7):2003–2024, 2015.

[167] F. Sahli Costabal, J.S. Choy, K.L. Sack, J.M. Guccione, G.S. Kassab, and E. Kuhl. Multiscale characterization of heart failure. *Acta Biomaterialia*, 86:66–76, 2019.

[168] Palash Sarkar. A brief history of cellular automata. *ACM Comput. Surv.*, 32(1):80–107, March 2000.

[169] Satu Schaeffer, Manuel Jimenez-Lizarraga, Sara Rodriguez-Sanchez, Gerardo Cuéllar-Rodríguez, Oscar Aguirre-Calderón, Angel Reyna-González, and Alan Escobar. Detection of bark beetle infestation in drone imagery via thresholding cellular automata. *Journal of Applied Remote Sensing*, 15, 03 2021.

[170] Santiago Schnell, Ramon Grima, and Philip K. Maini. Multiscale modeling in biology: New insights into cancer illustrate how mathematical tools are enhancing the understanding of life from the smallest scale to the grandest. *American Scientist*, 95(2):134–142, 2007.

[171] Cornelius Senf, Rupert Seidl, and Patrick Hostert. Remote sensing of forest insect disturbances: Current state and future directions. *International Journal of Applied Earth Observation and Geoinformation*, 60:49–60, 2017.

[172] Irfan Siap, Hasan Akin, and Selman Uğuz. Structure and reversibility of 2d hexagonal cellular automata. *Computers and Mathematics with Applications*, 62(11):4161–4169, 2011.

[173] Jonathan Silvertown, Senino Holtier, Jeff Johnson, and Pam Dale. Cellular automaton models of interspecific competition for space–the effect of pattern on process. *Journal of Ecology*, 80(3):527–533, 1992.

[174] G.Ch Sirakoulis, I Karafyllidis, and A Thanailakis. A cellular automaton model for the effects of population movement and vaccination on epidemic propagation. *Ecological Modelling*, 133(3):209–223, 2000.

[175] Jen Skerritt. Millions of beetles are wiping out forests all across the world, August 2020.

[176] Raymond Edward Smallman. *Modern physical metallurgy*. Elsevier, 2016.

[177] R. V. Solé, S. A. Levin, Mercedes Pascual, Manojit Roy, Frédéric Guichard, and Glenn Flierl. Cluster size distributions: signatures of self–organization in spatial ecologies. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 357(1421):657–666, 2002.

[178] J. V. T. Sorensen, H. Madsen, and H. Madsen. Parameter sensitivity of three Kalman filter schemes for assimilation of water levels in shelf sea models. *Ocean Modelling*, 11:441 – 463, 2006.

[179] A. M. Stuart. Inverse problems: A Bayesian perspective. *Acta Numerica*, 19:451–559, 2010.

[180] WY Tan and S Piantadosi. On stochastic growth processes with application to stochastic logistic growth. *Statistica Sinica*, pages 527–540, 1991.

[181] Lina Tang and Guofan Shao. Drone remote sensing for forestry research and practices. *Journal of forestry research*, 26(4):791–797, 2015.

[182] A. Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, Philadelphia, 2005.

[183] Caz Taylor. Effects of natal dispersal and density-dependence on connectivity patterns and population dynamics in a migratory network. *Frontiers in Ecology and Evolution*, 7:354, 09 2019.

[184] Tommaso Toffoli and Norman Margolus. *Cellular Automata Machines: A New Environment for Modeling*. MIT Press, Cambridge, MA, USA, 1987.

[185] R. Tuo and J. Wu. A theoretical framework for calibration in computer models: parametrization, estimation and convergence properties. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):767–795, 2016.

[186] Michael. Unser, Julien. Fageot, and John Paul. Ward. Splines are universal solutions of linear inverse problems with generalized tv regularization. *SIAM Review*, 59(4):769–793, 2017.

[187] A. W. van der Vaart. *Asymptotic statistics / A. W. van der Vaart.* Cambridge series on statistical and probabilistic mathematics. Cambridge University Press, Cambridge, UK ; New York, NY, USA, 1998.

[188] Henry van den Bedem and James S. Fraser. Integrative, dynamic structural biology at atomic resolution—it's about time. *Nature Methods*, 12:307–318, 2015.

[189] J. P. Vité and G. B. Pitman. Bark beetle aggregation : Effects of feeding on the release of pheromones in dendroctonus and ips. *Nature*, 218(5137):169–170, Apr 1968.

[190] Charles T. Vollmer. *Statistical upscaling of stochastic forcing in multiscale, multiphysics modeling*. PhD thesis, Colorado State University, 2019.

[191] Joseph Walpole, Jason A. Papin, and Shayn M. Peirce. Multiscale computational models of complex biological systems. *Annual Review of Biomedical Engineering*, 15(1):137–154, 2013. PMID: 23642247.

[192] Aaron Weed, Matthew Ayres, Andrew Liebhold, and Ronald Billings. Spatio-temporal dynamics of a tree-killing beetle and its predator. *Ecography*, 40:221–234, 02 2016.

[193] Stephen Wolfram. Cellular automata as models of complexity. *Nature*, 311:419–424, 1984.

[194] Stephen Wolfram. *A New Kind of Science*. Wolfram Media, 2002.

[195] Stephen Wolfram. *Cellular Automata and Complexity*. 03 2018.

[196] Thomas Worsch. Simulation of cellular automata. *Future Generation Computer Systems*, 16(2):157–170, 1999.

[197] Michael A Wulder, Caren C Dymond, Joanne C White, Bob Erickson, L Safranyik, and B Wilson. Detection, mapping, and monitoring of the mountain pine beetle. *The Mountain Pine Beetle: A Synthesis of Biology, Management, and Impacts on Lodgepole Pine Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, BC*, pages 123–154, 2006.

[198] L. Yang, T. Butler, and D. Estep. Solution of the stochastic inverse problem through disintegration by the range and importance sampling, 2022. in preparation.

[199] L. Yang and D. Estep. Solution of the stochastic inverse problem in infinite dimensions through regularization of the solution space, 2022. in preparation.

[200] Lei Yang. Infinite dimensional stochastic inverse problems, 2018. Ph.D. Thesis, Department of Statistics.

[201] Su Ye, John Rogan, Zhe Zhu, Todd J. Hawbaker, Sarah J. Hart, Robert A. Andrus, Arjan J.H. Meddens, Jeffrey A. Hicke, J. Ronald Eastman, and Dominik Kulakowski. Detecting subtle change from dense landsat time series: Case studies of mountain pine beetle and spruce beetle disturbance. *Remote Sensing of Environment*, 263:112560, 2021.

[202] James F. Ziegler, M.D. Ziegler, and J.P. Biersack. Srim – the stopping and range of ions in matter (2010). *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 268(11):1818–1823, 2010. 19th International Conference on Ion Beam Analysis.

[203] Ilya E. Zlobin. Linking the growth patterns of coniferous species with their performance under climate aridization. *Science of The Total Environment*, 831:154971, 2022.