AGENT-BASED MODELLING OF WEST NILE VIRUS PROPAGATION IN SOUTHERN MANITOBA, CANADA

by

Hamid Reza Nasrinpour

A Thesis submitted to the Faculty of Graduate Studies of

the University of Manitoba

in partial fulfillment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

Department of Electrical and Computer Engineering

University of Manitoba

Winnipeg, Manitoba, Canada

ABSTRACT

This research addresses the design and development of a data-driven Agent-Based Modelling (ABM) framework to simulate transmission and spread of West Nile Virus (WNV) among heterogeneous mobile humans, various bird species, and *Culex* genus mosquitoes over a geographic region at a province-scale. A diverse variety of topics and techniques regarding the data collection phase is presented, as modelling WNV has many disparate attributes. WNV is a mosquito-borne disease influenced by avian species as their amplifying hosts. A significant amount of data, such as home range, flight speed, WNV competence index, etc., regarding over 150 bird species along with their population estimates and locations in Manitoba, Canada are estimated and assembled. The primary contribution of this thesis is the development and validation of a data-driven Cellular Difference Equation (CDiffE) scheme for adoption in WNV-ABMs or other mosquito-borne disease ABMs. The migration patterns of different bird species, nocturnal biting activities of *Culex* mosquitoes, daily temperature and rainfall, and land cover impact are incorporated into the CDiffE model. The CDiffE model at its core employs difference equations, which are computationally faster than commonly used differential equation models. The proposed CDiffE model is cellular to capture heterogeneity of various geographical areas. The CDiffE has been rigorously verified and validated. While the whole system is designed from an ABM perspective at a cellular level, it exhibits biologically compatible behaviour at the macro-level scale. The proposed CDiffE demonstrates high accuracy in predicting real-world mosquito population trends and geographical distributions, evidenced by the mosquito trap data from Manitoba. The proposed CDiffE model updates on an hourly step to act as an environment for a comprehensive ABM of WNV spread among peripatetic humans. Such a hybrid ABM is successfully built on top of the proposed CDiffE scheme to study the impact of human

movements on the prevalence of the virus. The human movement component is modeled on data available from cell phone trajectories as well as census and demographic datasets. Simulation results clearly illustrate the importance of human movement patterns and demonstrate the need for real-world data. Yet human mobility is often disregarded within current WNV modeling efforts.

ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere gratitude to my truly kind advisors, Drs. Robert McLeod and Marcia Friesen, for providing me with guidance, support, and knowledge throughout the years. You both have been incredibly encouraging to me. I thank you for your patient and giving me the opportunity to develop this work. I am fortunate to have been a member of your lab. Bob, you are the friendliest supervisor I have ever had.

I would like to thank my research committee, Drs. Ken Ferens and Pourang Irani, for their valuable feedback and support during my studies at the University of Manitoba. I especially acknowledge and thank Dr. Abba Gumel who dedicated his time and expertise to guide me and critique my work during my visit to Arizona State University. I am deeply appreciative for the time Dr. Nathaniel Osgood spent on reviewing this thesis, and for all the helpful comments he provided.

Special thanks to Ryan Neighbour, a dear friend, for constructive suggestions and the hours of brainstorming and discussions. Thanks are also extended to Alex Reimer for assistance in writing this dissertation, and preparing the datasets of this project; the Manitoba Health Dept. for input on the trap data in Manitoba; Dr. Ahmed Abdelrazec for consultation on designing the mathematical model. I would also like to acknowledge the NSERC and the University of Manitoba for financial support over the years.

Finally, I would like to express my sincere appreciation towards my dearest Yasamin for her encouragement and continuous support in the completion of this thesis. Yasamin, thank you for all you have done.

DEDICATION

To my best friend and lovely wife, Yasamin, you are my source of inspiration and motivation. I could not have done this without you, my love. Your love and faith in me kept me going when nothing else would. I did this with you by my side, and I would not want it any other way.

To my wonderful mom and dad, who always supported my dreams.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABM	Agent-based model / Agent-based modelling
BAM	Boreal Avian Modeling
BCR	Bird Conservation Region
CDC	Centers for Disease Control
CDiffE	Cellular Difference Equation
CNS	Central Nervous System
CO2	Carbon Dioxide
Cx.	Culex
DBF	dBase format / Database File
DD	Degree-Days
DE	Differential Equation
DiffE	Difference Equation
DLL	Dynamic-link library
EDR	Effective Detection Radius
EIP	Extrinsic Incubation Period
FID	Feature ID
GIS	Geographical Information System
GLMM	Generalized Linear Mixed Model
GPS	Global Positioning System
KMZ	Keyhole Markup language Zipped
LST	Land Surface Temperature
MBBA	MBBA

mm	millimetre
MS	Microsoft
MTS	Manitoba Telecom Services
NDV	Normalized Difference Vegetation
ODE	Ordinary Differential Equation
OSM	Open Street Map
PBF	Protocol-buffer Binary Format
PIF	Partners In Flight
SIR	Susceptible-Infected-Recovered / Susceptible-Infectious-Removed
SQL	Structured Query Language
USGS	U.S. Geological Survey
UTM	Universal Transverse Mercator
WGS	World Geodetic System
WNV	West Nile Virus

Chapter 1: INTRODUCTION

1.1 Introduction to Agent-Based Modelling and West Nile Virus

This thesis research is to develop and validate an agent-based model (ABM) of West Nile Virus (WNV) transmission, using the Java-based Anylogic simulation software, in particular for southern Manitoba, Canada. While most agent-based models for health applications integrate one type of agent (e.g., humans acting and interacting to spread a contact-based infection such as influenza), the proposed research is significantly more complex in that it considers three distinct types of agents: humans, mosquitoes, and birds, each of which are subsequently broken down into a near-realistic diverse array of agent profiles in terms of behaviours and interactions. The proposed ABM uses weather and GIS information to spatially model the spread of WNV under given climate conditions at hourly time-steps.

1.1.1 Agent-Based Modelling

Agent-based simulation or Agent-based modelling (ABM) [1]–[3] is a natural and intuitive way of simulating systems where individual agents (e.g., people or mosquitoes) play significant roles. In this bottom-up approach, the system contains a set of autonomous individuals, i.e., agents, interacting based on a set of rules within an environment. From the micro-level interactions between agents of the same or different types, the macro-level patterns of the system emerge. In its simplest form, a cellular automata (e.g., game of life [4]) can be considered an ABM where the agents are the cells governed by simple rules based of their neighbouring cell states (i.e., the environment). The conceptual depth of ABMs is due to capturing complex phenomena which are not explicitly programmed by the modeller, and cannot be explained by

the individual-level rules [5]. Often the outcome is greater than sum of all the individual components in an ABM.

A heterogeneous population of agents is inherently suitable to an agent-based model where e.g., each person can have their own profile of movement and interaction. ABMs are generally well suited to model social networks when either the agents or the topology of the interactions is heterogeneous or complex. ABMs are naturally designed to model interaction between heterogeneous agents, and in WNV epidemiology, the main means of spread of the virus is the interaction cycles within different agent types. This makes ABMs an ideal tool to investigate the epidemiology of WNV.

The value of ABMs is being established through a growing literature base in various fields. However, the modelling approach itself is still considered a relative newcomer in relation to mathematical models and more established simulation models of social and human agent systems. Although ABM has its roots in the 1960s [6] or even earlier, it has gained significant attention in last decade as a powerful method for modelling various systems with advances in computing capabilities and in the amount of human-centric data available. It is increasingly applied across many disciplines, ranging from natural, social, and physical sciences, engineered systems and beyond to business, and operations management [3], [7], [8].

ABMs are recognized as computational models that permit a distinctive approach to complementary empirical research [9], [10]. In ABMs, dynamic social systems are modeled as a collection of highly stochastic agents (in many cases, primarily humans), their individual profiles or characteristics, their individual behaviours, and interactions between agents and between agents and the environment. Agents are purposeful and autonomous entities able to assess their situation, make decisions, and compete or co-operate with one another on the basis of a set of

rules. As noted, the conceptual depth of an ABM is derived from its ability to model behaviour that may be counterintuitive and/or to discern a complex behavioural whole that is greater than the sum of its parts. ABM provides a natural description of a system that can be calibrated and validated by representative expert agents and is flexible enough to be tuned to high degrees of sensitivity in agent behaviours and interactions. ABMs are particularly well suited to dissipative irreversible system modelling in which agent behaviour is complex, non-linear, stochastic, and may exhibit memory or path-dependence [2], [11]. Such systems support quantized or individuated characterization, and particularly exhibit heterogeneous aspects.

Yet, the full potential of ABM remains unrealized as the methods to exploit massive amounts of data are still emerging. The key challenges of an ABM approach applied to socio-ecological systems are improving agent decision and adaption models, improving validation and verification, and improving spatial representation and levels of abstraction [12]. Further, the key challenges for the ABM approach to advance and realize its potential include validating agent behaviours and emergent phenomena, better agent behavioural models, improved simulation analytics, and improving hybrid and large scale ABMs [13]–[15].

1.1.2 West Nile Virus Background

West Nile virus (WNV) is an arbovirus (arthropod-born virus) which was first isolated from a feverish woman in 1937 [16]. Its name comes from the district of West Nile in Uganda. In Egypt, at the beginning of the 1950s, the ecology of the virus was learned and its symptoms were discovered. The virus was detected in humans, birds and mosquitoes. Since then, outbreaks have been found in various European countries [17]. In 1996, there was a major human epidemic in Bucharest, Romania, where WNV became a concern for public health. The arrival of WNV in the American continent occurred in 1999, and more specifically in New York [18]. In Canada,

WNV reached southern Ontario in 2001, while the first human cases were detected in August 2002. In Canada, the mosquito infection rate and mortality of birds have been used as a proxy for WNV transmission risk. The first appearance of infected birds in Manitoba was in 2002 [19]. The highest number of human cases associated with WNV for Manitoba was reported in 2007 and 2003, with 587 and 143 cases, respectively [20]. This figure includes asymptomatic, neurological, non-neurological syndrome, and unclassified cases with positive test results [20]. In the most recent update on WNV information in Manitoba, a total of 24 human cases reported as of December 9, 2016 [20].

A majority of people who become infected with WNV do not show any noticeable symptoms. As these people remain asymptomatic, these cases are not usually reported in statistics. Others who become ill with WNV mostly show mild flu-like symptoms within two weeks after infection [21]. This period is called the incubation period. Mild symptoms include headache, fever, fatigue and body aches. However, a few people who are infected will develop serious symptoms that could affect the central nervous system (CNS) and cause neurologic illnesses such as encephalitis/meningitis (acute inflammation of the brain/surrounding membranes). People of older ages or with weakened immune system are generally at a higher risk for severe illness. The severe symptoms includes serious headache, high fever, stiff neck, vomiting, blurred vision, confusion, coma and paralysis [6, 7]. Symptoms may last for months or years without treatment. Unfortunately, there is no vaccine for humans yet, and the infection can result in death.

1.1.3 Transmission of West Nile Virus

Mosquitoes of mainly the *Culex* genus are the vectors of WNV infection [23], i.e., they carry and transmit the virus to other animals including humans. Under certain environmental

conditions, adult female mosquitoes lay their eggs. As such, they need to take a blood meal from their hosts to obtain sufficient protein and iron for egg-laying. During the next stages of the mosquito life-cycle, eggs hatch into larvae, and then begin molting their skins until they change into pupae that evolve into adult mosquitoes [24]. The newly emerged adult may inherit the virus but with low probability [25], [26]. This is called vertical transmission within the mosquitoes, which has a negligible transmission rate [27].

It is believed the main means of transmission and spread of WNV is through birds [28], [29]. An infectious mosquito can infect a healthy bird by feeding on it. An infectious bird can in turn infect a healthy mosquito that bites the bird. In this transmission cycle, birds act as so-called amplifying hosts since the virus is amplified in their bloodstream. Amplifying hosts keep a high enough level of pathogen that it can be transmitted to the next feeding mosquitoes. On the other hand, there are incidental or dead-end hosts in the WNV epidemiology that cannot pass the pathogen to another host or feeding mosquitoes. Dead-end hosts of WNV are various kinds of mammals including dogs, cats, sheep, goats, deer, horses and humans. Horse seroprevalence in some countries has been used as a proxy for WNV transmission risk [30]. Horses provide an ideal surveillance tool in that they are highly affected by the virus (with paralysis and death also being possible), and thus infection can be more readily detected. The whole transmission cycle of WNV is shown in Figure 1-1.

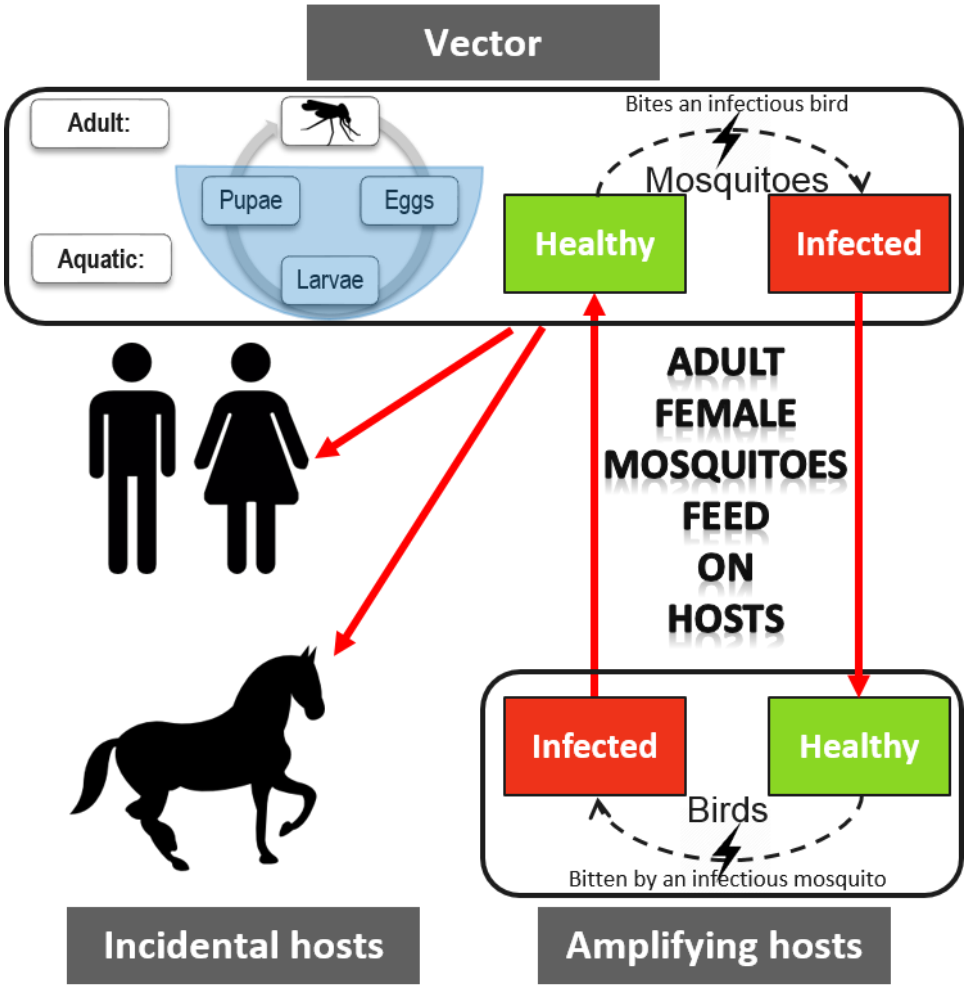


Figure 1-1 Mosquito life cycle and West Nile Virus transmission cycle diagram

1.2 Overview of Key Factors in a WNV-ABM

The primary objective of this thesis is to model the dynamics of WNV spread in southern Manitoba. Before an attempt to make any model, a number of key factors, each of which has a particular role in WNV transmission, need to be recognized. In this section, main factors affecting WNV transmission cycle are reviewed. These factors become inputs into an ABM and govern the behaviours and interactions of the three types of agents: humans, birds, and mosquitoes.

1.2.1 Weather

The duration of each stage of the mosquito life cycle as well as their biting rate and rate of virus maturation are a function of weather conditions. Birds' migratory behaviour is influenced by climate. Human outdoor activities which expose them to mosquitoes are weather-dependant. Among the three agents, the impact of weather on mosquito conditions has the most significance influence on the emergence of WNV. This impact can be indexed and measured in different ways. All these indices are associated with temperature and/or precipitation.

1.2.1.1 Precipitation

There is no doubt that humidity is required for mosquitoes to lay their eggs. However, the desired level and frequency of rainfall for mosquitoes is controversial [30]–[32]. Any influence of rainfall on mosquito dynamics affects the WNV transmission. Further, various mosquito species considered can have different preferences with regard to water in their habitat. For instance, while *Culex* species put their eggs directly on the water surface, some other species produce eggs on damp soil [24]. An in-depth discussion on the impacts of precipitation needs to address the difference in the mosquito species across the different studies.

Heavy rainfall obviously increases the water surface which may favor some mosquitoes, but it can also flush their habitat [33]. It may be safe to generally assume that wet weather conditions amplify occurrence of mosquito-borne diseases [34]. However, drought can unfortunately lead to similar situations as well. More precisely, drought, similar to heavy rainfall, has a two-sided ambiguous effect on mosquito habitat. At first glance, drought reduces the necessary water for mosquito breeding, but further investigation provides facts to reinforce a drought hypothesis in favor of mosquito habitat [34]–[36]. The basic idea of the hypothesis is that a mild winter can help mosquito larvae survive. If this winter is followed by a dry spring and summer, it drives

both birds and mosquitoes to come closer to the remaining shrinking water, which in turn increases rates of contact between mosquitoes and birds. Moreover, most mosquito predators (e.g., dragonflies and frogs) cannot survive a drought condition, plus concentrated stagnant water has an ideal high level of organic nutrients for mosquitoes [37].

1.2.1.2 Temperature

Temperature generally has a positive correlation with WNV activities. Temperature affects both mosquito population dynamics and the virus amplification. In the *Culex* genus, higher temperatures expedite the development of mosquitoes, increase the reproduction rate and the number of blood meals, boost vector competency and chances of the virus transmission by reducing the extrinsic incubation period (EIP)[24], [30], [31], [33], [38]–[43]. Decreasing the life-cycle of mosquitoes sometimes also decreases the longevity of mosquitoes at more extreme temperatures over 30° C. In other words, higher temperatures intensify the transmission of the virus, yet cool temperatures (in Summer around 15-20° C) could increase the life-time of mosquitoes from weeks to months [44]. Both effects are not desired from a public health perspective.

The fluctuation of temperature generally does not favour the spread of WNV [43]. The development of mosquitoes and the virus both need warm weather over a period of time. A commonly used index to describe variations in mosquito population and their infection rates are cumulative degree-days [45]. Degree-days index for a given day shows the number of degrees Celsius the mean temperature of the day is below or above a pre-defined base threshold. For *Culex tarsalis* (*Diptera*: *Culicidae*) mosquitoes, the base number is estimated to be 14.3° C, and the activation of virus (EIP) requires 109 Degree-Days (DD) [41].

Other indexes used to measure temperature conditions in the WNV literature include Land Surface Temperature (LST) [46], [47], which may be computed from remote sensing imagery, and the elevation [48], [49], which is negativity linked to temperature [30].

1.2.2 Landscape Features

Habitat quality for mosquitoes and birds depends on landscape features. In the WNV literature, landscape information is mainly used to determine habitat quality for mosquitoes. For mosquitoes, vegetation type and water bodies are the two main elements that need to be described by landscape variables or indexes. A common approach to describe landscape features is to classify different regions of land into various categories based on how the land is being covered or used by nature/humans. As such, it is sometimes called land cover or land use classification. These categories may include urban area, rural area, grassland, agricultural cropland, various forest types, sands, rocks, roads and water. Any existing land cover dataset categorizes the land into different number/type of classes as their intended usage can be different. For example, classification of land for agricultural purposes is different from that of nature conservation groups. When the classes are defined, remote sensing satellite images are usually used to classify each region. Then, for each class, mosquito habitat quality parameters can be assigned, usually according to an expert's opinion.

An alternative method in WNV modelling is to directly calculate mosquito-related indexes for the land. As an example, normalized difference vegetation index (NDV), which assess the land cover from a vegetation perspective with a numerical value between -1 and 1, has been identified and used as a risk factor for mosquito abundance models [34], [46], [50]. Other landscape parameters used in the WNV literature to describe mosquito habitat include road

density, stream density, slope percent, soil type, sumps along the roads and distance to the nearest wetlands/bogs [34], [46], [51], [52].

1.2.3 Birds and Mosquitoes Species as Distinct Agent Types in a WNV-ABM

1.2.3.1 Birds

In North America, at least 59 different species of mosquitoes [53] [54] and over 225 species of birds [23], [55] have been found infected with WNV. However, not all of the mosquito species act as bridge vectors by feeding on both birds and mammals, including humans. There are several studies with data on different mosquito species and bridge vectors [56]–[58]. Only some mosquito genera have been reported in various regions of the world, and their population count heavily depends on weather conditions and landscape features [59]. Among birds, some species such as *corvidae* (corvids), commonly referred to as the crow family, are known to be competent hosts. The crows become severely ill after the infection and often die, whereas many other species do not become ill and develop WNV antibodies to resist against infection [30]. Host competence is typically determined by the duration and magnitude of viremia. The contribution of different bird species to WNV transmission to mosquitoes is determined by the combination of competence, abundance, and mosquito feeding patterns. Several studies have quantitatively determined which host species are more important in different parts of North America [60], [61]. Many other studies offers insights on mosquito feeding patterns and the role of different bird species (and their community structure) in WNV amplification [62]–[65].

Avian hosts (birds) are believed to be responsible for large-scale spread of WNV mainly due to their food seeking activities and somewhat to their migration [29], [52]. Yet no model exists to explain the distribution of wild birds primarily because of longer flight spans compared to mosquitoes, coupled with the cost of bird monitoring and data collection and various rates of

migration within a species [30]. Traits and behaviour (e.g., roosting) of each species of birds also play an important role in the prevalence of WNV [66]–[68]. As an example, American crows are susceptible to become infected via means other than mosquito bites, such as vertical transmission and eating infected dead or nestling birds [69]. Studies on the relative role of birds in WNV transmission have also reported a positive correlation between seroprevalence and bird weight, which may be due to higher CO_2 production and the longer expected lifespan of larger birds [70]. Nestling behaviour of birds might be another factor to play a role in WNV transmission. Nestlings are more exposed to mosquito bites as they are too young to move, have fewer feathers and immature immune systems [30]. Competent avian hosts also include the house finch, house sparrow, mourning dove, gray catbird, European starling, northern cardinal and corvids such as the American robin, American crow, western scrub-jay, yellow-billed magpie and blue jay [69], [71]–[74]. Among these species, some are preferred over the others by each mosquito species. However, in general, mosquito blood feeding patterns depend mainly on accessibility of hosts in a certain area [75].

1.2.3.2 Mosquitoes

Mosquito species differ in their preferred breeding habitats, biting behaviour, flight range, life-cycle period, host preference, etc. It is believed that differences in mosquito habitat preferences are the reason why both extreme (wet and dry) weather conditions could result in the outbreaks [34]. A recent study on weather drivers of WNV across North America can be found in the work of Paull et al. [32]. A summary of these differences for a few mosquito species are reported in Table 1-1 [28], [34], [76].

Table 1-1 Differences in mosquito species properties [28], [34], [76]

Mosquito Species	**Habitat Preference**	**Flight Range**	**Host Preference**	**Activity Time**	**Life Cycle (days)**
Aedes aegypti	Shaded artificial containers, Tree holes	200 m	Mammals	Crepuscular/day	10 - 20
Ochlerotatus sollicitans	Salt marshes, Freshwater	2.5 km	Large mammals	Crepuscular/day	7 - 10
Culex pipiens	Urban area, Water with high organic content	2 km	Birds	Crepuscular/night	10 - 14
Culex tarsalis	Every fresh water source except tree holes	>6 km	Opportunistic	Crepuscular/night	10 - 14
Psorophora columbiae	Rice fields, Temporary pools and ditches	>16 km	Opportunistic	Day/Night	4 - 10

Culex (*Cx*.) *pipiens* is the main competent WNV vector in the eastern United States and Canada, including southern Ontario and Quebec [54], [55]. *Cx. pipiens* is one of the most important vectors capable of amplifying WNV in the bird population [28]. These primarily bird biting mosquitoes may switch their hosts to mammals and humans during the late summer and early fall [54]. They are mostly found in urban areas and have habitat preferences for water with high organic content [34], [77], [78]. The specie can also lay eggs in artificial containers such as cans, tires or stagnant water in any trash container. Catch basins or storm drains in cities can be an ideal place for them to reproduce [79].

Culex restuans mosquito is a competent primarily bird biting vector for WNV, which feeds infrequently on mammals [28], [55]. Its biting activity occurs mainly at night, and it has a similar breeding habitat to *Cx pipiens*.

Culex tarsalis is extremely efficient in preserving and intensifying WNV and is considered to be the main vector in the Prairie provinces of Alberta, Manitoba and Saskatchewan [20], [43], [55], [80], [81]. This species, similar to other *Cx*. species, can transmit WNV to their offspring [25]. *Culex tarsalis* feeds multiple times mostly on birds and sometimes mammals [55], [82]. The larvae of this principal bridge vector in Manitoba can tolerate a wide range of water conditions [83]. The species can be found in almost any fresh water source except tree holes, including temporary water bodies (e.g., bird baths and used tires), alkaline lakes, and salty wetlands [55], [79], [83]. Larval habitat can be shared with other species such as *Cx. pipiens* and several species of *Aedes* and *Anopheles* [83]. A female one can produce eggs multiple times during a season [55], [79]. The eggs on the surface of water are attached together, forming rafts of around 200 eggs [24], [83]. A proportion of up to 95% of larvae can be lost by predation [83]. The life cycle duration depends on temperature. It takes approximately 14 days at 21° C and only 10 days at 26° C [24] to go through their life cycle. In spring, the female adults feed mostly on birds at dusk, and during late summer they switch hosts to feed on mammals [54], [83]. They are reported to be persistent biters even during the day and dawn [55], [84], [85]. Their foraging flight is typically less than one kilometer, but it can be extended well beyond six kilometers up to 27 km [28], [83]. The combination of traits that characterizes *Cx. mosquitoes* are essential knowledge when developing transmission models.

1.3 Overview of Related Work

1.3.1 West Nile Virus non-ABM Literature

According to [30], different approaches to model WNV transmission risk or spread can be categorized into three sometimes overlapping classes of "risk factor analysis," "landscape epidemiology," and "transmission dynamic modelling."

1.3.1.1 Risk Factor Analysis

Risk factor analyses try to identify various variables associated with WNV occurrence through observing the natural environment and making a connection between conditions and WNV prevalence. It bears noting that risk factor models are normally associational than causal. In these models, mostly birds (in North America) and horses (in Europe and Middle East) have been used as infection markers to measure WNV prevalence [30]. Investigated variables such as climate, host competences and landscape characteristics are employed in statistical models to describe variations in WNV occurrence. These studies may be spatial and include landscape characteristics that could move it under the class of "landscape epidemiology."

1.3.1.2 Landscape Epidemiology

Landscape epidemiology, which is also known as spatial epidemiology, can be used to produce risk maps over various landscapes. These approaches mostly use Geographical Information System (GIS) software to represent geographic distribution of disease as a result of several elements. Spatial statistical tools may be used to model data based on covariates [30]. These models can implement three types of layers: mosquitoes, birds and incidental hosts. Each layer can independently measure the degree of WNV severity based on landscape characteristics.

The layers can be then combined together based on a weighted linear combination to produce risk indices [86].

1.3.1.3 Transmission Dynamic Modelling

The third approach of transmission dynamic modelling includes integrative studies where spatio-temporal co-occurrence of vectors and hosts can be modeled [30]. There are studies to introduce theoretical formulation for mosquito host feeding patterns [87]. Others have used t-test or Generalized Linear Mixed Models (GLMMs) to model competency of avian hosts [30]. Most notably, difference or differential equation (DE) models have been utilized to model WNV transmission dynamics. Thomas and Urena formulated a difference equation for WNV evolution in a mosquito–bird–human community with a focus on mitigation via pesticide [88]. Their model has no geographic element, and the incubation period is disregarded to simplify the model [88]. Wonham et al. developed a single-season classic susceptible-infectious-removed (SIR) DE model for WNV transmission in a bird-mosquito population [89]. Their focus is on estimating the disease reproduction number and the chance of an outbreak [89]. Bowman et al. propose a single-season DE model of WNV transmission dynamics in a mosquito–bird–human population to assess personal protection and mosquito reduction strategies [27]. Mosquito and bird compartments of their model have no exposed state. Also the effect of bird-to-bird vertical transmission and migration of birds are not studied in their model [27]. Cruz-Pacheco et al. [90] formulate and analyse a DE model of WNV transmission with a focus on hosts and estimating the competence of bird species. Simpson et al. [73] developed a DE model for WNV transmission with a focus on host feeding preferences using one vector species (*Cx. pipiens*) and two categories of preferred and non-preferred bird hosts

In this context, ABMs can be designed to combine features and strengths of all the DE models mentioned above into a single system. ABMs, intrinsically, can incorporate biodiversity of birds and mosquitoes, their heterogeneous contacts, and their interaction with mobile humans, without making any homogenous assumptions about the population or hard-coding any statistical rules/correlations at the system-level. Whereas exploring any additional factor in a DE model requires hard-coding of some new parameters or even re-designing the whole model at a population-level. Additionally, human behaviour and their community structure has an indisputable effect on the spread of an infectious virus (e.g., flu) [91], [92]. For instance, human behaviour can directly change variables in the abundance of mosquitoes and birds through larviciding, pesticiding, bird hunting, etc. Yet, these factors are generally neglected in WNV models as humans are dead-end WNV hosts. Weather and GIS information can easily be injected into ABMs to track and potentially predict the spatio-temporal dynamics of WNV in a given landscape under customized weather conditions. Depending on the level of details implemented in an ABM, the role of each characteristic/trait of various mosquitoes and birds in the WNV epidemiological system may be investigated. Also, to capture even more dynamics and phenomena, the model alterations can be conducted at an individual-level, if need be. More importantly, the impact of some uncertain intervention strategies and control scenarios can be effectively assessed. For example, one may wonder if building a school near a pond or a small stagnant lake (i.e., a potential hot-spot for mosquitoes) would make any difference to WNV prevalence in an area. In a WNV-ABM, it is very straightforward to discover how agents (e.g., people/mosquitoes) would respond to a change in the environmental conditions. Other capacities of ABMs can be summarized as: (1) representing geography at a continuous level (rather than a quantized level, which are typically vastly cruder representations in aggregate models); (2)

representing individual decision-making (e.g., with respect to personal protective behaviour, possible vaccination, municipal mosquito spraying, etc.) in a situated way that draws on local perception; (3) keeping track of individual longitudinal information about a person's history. This allows reporting statistics and comparing/calibrating against individual longitudinal empirical data.

1.3.2 Related Agent-Based Models

ABMs have been extensively applied in health care applications [93]–[96] as well as in somewhat limited mosquito-related studies. A wind and odor driven ABM for host-seeking behaviour of mosquitoes is proposed in [97]. They explicitly formulate their ABM based on the *Culex* mosquito feeding behaviour on roosting birds. The model is 2D (i.e., at a fixed height from the ground), and generates movement trajectories for each mosquito as an individual agent. In [98], using the Repast toolkit, an ABM is proposed to simulate population of *Aedes* mosquitoes, which are the primary vector of Dengue and Chikungunya. Their agents include mosquitoes, humans, dogs, and cats. The mosquito agents can fly randomly, look for resting places, look for blood meal, bite hosts, etc. An ABM for simulation of dengue transmission in Thailand with a focus on vaccination practices is proposed in [99]. This ABM includes aggregated mosquito (per building) and human agents who can be either at home or work. Mosquito agents tend to stay at the same location. However, individual infectious mosquitoes may migrate to other locations with some low probability. In simulating multi-year epidemics, the human populations are assumed to become older and more susceptible to dengue. In [100], an ABM for Malaria transmission is proposed where different intervention strategies are explored. A number of different *Anopheles* mosquitoes and humans are among the agents. In this ABM, a probabilistic decision-tree defines the life-cycle and behaviour of mosquito agents. In [101], a temperature-

driven ABM of *Anopheles gambiae* mosquitoes with an emphasize on its life-cycle is proposed. These mosquitoes are Malaria vectors. In [102], an ABM for the population dynamics of *Aedes aegypti* mosquitoes is proposed where two control strategies for Zika virus are simulated. The geographical scope of their ABM consists of three zones of human locations, vegetation, and breeding sites. The mosquito behaviour is driven by the present zone and monthly average temperature.

There are other ABMs which focus on the impact of human behaviours in spread of these diseases, where an aggregation of mosquitoes is an agent. Typically, the dynamics of an aggregated mosquito agent update is based on a DE model. For example, in [103] an ABM is used to simulate spread of Chikungunya where mosquitoes have a network-patch model. In a network-patch model [104], mosquitoes are divided into patches of high, medium, and low densities. Each patch can contain a number of location nodes (i.e., buildings in a town). Humans can move from one node to another node within a virtual network. A similar approach is used in [105], where each network node is associated with one patch. A framework for modelling of mosquito-borne pathogen (e.g., Malaria and Dengue) transmission is proposed in [106], where humans spend various amounts of time at various virtual nodes. Another ABM for dengue transmission with houses arranged in 16 different mosquito patches is proposed in [107], where humans can visit these houses.

It is notable that humans in the above-mentioned mosquito-borne diseases act as amplifying hosts, in contrast to WNV. In addition, generally these models do not need to explicitly account for a third agent type, i.e., birds. Indeed, often those models do not include any hosts within the virus cycle other than humans. The presence of such additional hosts, in general, makes WNV a more complex disease to model and more appropriate for an ABM approach.

1.3.3 West Nile Virus Agent-Based Models

1.3.3.1 The WNV-ABM by Li et al. (2005)

The usage of ABMs in the WNV literature is rather scant. In [52], using the Repast toolkit, a WNV-ABM is proposed. The scope of their model is an area of around 165 km^2 in Cook county, Illinois, US. This area is modelled as a raster map where each cell of the array represents one acre (about 4046 m^2). Landscape is classified into 63 classes of different land-uses, such as commercial, wetlands, agricultural, etc., each of which are associated with a habitat quality index of one to three. Each cell of the raster map contains various land-use classes. A weighted mean of the fixed values of land-use parameters are associated with the parameters of the cell.

The ABM models birds and humans as individual agents being capable of flying or moving around the map as part of their daily activities. However, the population of mosquitoes and the transmission of the disease are controlled by a set of proposed differential equations. Their equations are controlled by weather-dependant parameters such as soil surface moisture. The weather data (i.e., temperature, precipitation and humidity) are updated hourly.

Land-use parameters of each cell determine the initial distribution/density of birds, mosquitoes and human, habitat quality (food abundance) for birds and the likelihood of human outdoor activities. Chances of human outdoor activities are also affected by their age, which is grouped into different classes of ‘infants,’ ‘children and teenagers,’ ‘young and middle-aged,’ and ‘seniors.’ Basically, every day birds fly into neighbour cells of higher habitat quality based on the parameters of cells and return to their home cell at night. If humans are outdoor, and infectious mosquitoes are present at the same cell at the same time, the transmission of WNV may occur. The mathematical equation of transmission is based on the prevalence of infection

rate for humans, the number of infectious mosquitoes, and the total number of mosquitoes in the cell.

There are heavily simplified assumptions throughout their paper. It is claimed that the three bird species of black-capped chickadee, blue jay, and American crow are modeled within the ABM. However, it is not known how they differentiate between these birds in the model. The equations for transmission of WNV between birds and mosquitoes are not reported. The mosquito model was too general to include any specific species or to distinguish different stages of its life-cycle. Finally, there is no report on any kind of model validation or results.

1.3.3.2 The WNV-ABM by Bouden et al. (2008)

An ABM with no human component is proposed by Bouden et al. [51]. The scope of their model is southern Quebec, Canada, with a daily time step and weekly assessments. GIS information is used in calculation of the locations and initial population of birds and mosquitoes. The map is divided into municipalities as the reference areas e.g., for visualization and mosquito habitat. Weather data (temperature and precipitation) are used in mathematical equations to compute the dynamics of the mosquito population at each time step. A DLL (Dynamic-link library) of the BIOSIM software [108] is used to interpolate values for temperature and precipitations at certain locations on the map based on four neighboring weather stations and elevation data.

Two bird-biting species of *Cx. pipiens* and *Cx. restuans* are considered for the mosquitoes. The mosquitoes are modeled as intelligent density maps which are defined by a set of parameters assigned to municipalities in the map. Two main stages of mosquito-life cycle, ‘adult mosquitoes’ and ‘larvae’ are separated in the equations. The dynamics of the population of the two stages and birds, as well as the transmission of disease are controlled by a DE model based

on Wonham et al.'s mathematical model [89]. In a new proposed DE, different bird species are distinguished. Climate effects are included in the new proposed DE model based on Madder and colleagues' work [109]. Unfortunately, the new proposed DE model is not explained in their paper.

Sumps along the roads are considered to be the main reservoirs of mosquitoes and larvae. Total length of roads for each municipality is computed from GIS information. On average, for each 30 linear meters of road, a sump is considered, and 20% of sumps are assumed to contain larvae. Heavy rainfall and larvicide spraying are supposed to flush sumps, killing a large proportion of larvae. The parameters such as emerging number of adult mosquitoes from each sump are set according to expert opinion.

The main bird species studied is the American crow. The birds are divided into two classes of American crow and generic birds (i.e., other bird species known to carry WNV). The North American breeding bird survey data [110], ornithologists' opinions and another database are employed to adjust the population of birds. Changes in the population of crows are calibrated based on data consisting of reports of counts of dead birds. Roost agents are used to represent a group of birds belonging to a roost. Location of roosts and the average number of crows per square kilometer for each region of the map are extracted from the data. Moving behaviour of a roost agent (a group of birds) is modeled with a particle system proposed by Reeves [111], wherein birds' flight speed and home range are crucial model parameters. Each roost has different velocity and direction for their movement, and they can fly away during a day up to a maximum radius. The assumption is that generic birds belong to the municipality and do not leave it while searching food during the day, whereas crows may leave their home municipality. All birds return home to spend night in their roosts.

The WNV propagation is simulated from July 1 until October 1. The user can specify a day of heavy rainfall or larviciding in a given municipality. Many of parameters of the model can also be modified by the user. Due to lack of data, simulation results only for some municipalities are calibrated. For the calibration, the trend of average number of reported *Cx.* mosquitoes, reported infected mosquitoes and reported infected (dead) crows over 2003-2006 during the months of simulations are used to make similar trends in the output of the system. Their most complete data was for Laval municipality in 2003, where the number of *Cx.* mosquitoes in the simulation and field data had a similar trend. They conclude that an important limit of the system is the lack of field data.

The lack of a human component is the main shortcoming of their ABM. This limitation can be improved by a complementary ABM where strategies to minimize the exposure between humans and mosquitoes, e.g., applying insect repellent or wearing long-sleeved shirts, could be modeled and evaluated. From a public health point of view, defining WNV risk index based on the infection rate among humans would be better matched to reality than that of birds and mosquitoes. This is especially true for the Province of Manitoba where the bird surveillance system as an early indicator of WNV has not been in effect since 2006. In addition, allocating sumps along the roads as the only reservoir for mosquito larvae may be true for the urban mosquito *Cx. pipiens*, but it is not true in general. For example, in Manitoba, Canada the main recorded vector is *Cx. tarsalis*.

In this research thesis, a more complete hybrid ABM is developed and validated to address some of these deficiencies.

1.4 Organization of the Work

This section outlines how this thesis is organized. Chapter 2 explains the details and procedures of the data collection and preprocessing phase, in particular for landscape and bird species data. The obtained data drive the ABM, proposed in Chapter 3 and Chapter 4. Most of these data can be found in Appendix A. Chapter 3 presents and discusses the core of the proposed ABM which is indeed a Cellular Difference Equation (CDiffE) scheme. Further extraction and mining of WNV data of various bird species, and the analysis and validation of the ABM (i.e., CDiffE) output are also reported in Chapter 3. Chapter 4 elucidates how to develop a hybrid ABM on top of the proposed CDiffE scheme. It expands on the human component of the CDiffE scheme. The integration of mobile human agents into the CDiffE scheme is essential to complement the proposed hybrid ABM. Human population and movement data prepressing is explained in Chapter 4 as well. Chapter 5 provides concluding remarks, contribution to knowledge in the research area and a summary of the thesis. Appendix A gives some Java code snippets in addition to the data discussed in Chapter 1. Appendix B provides supplementary figures regarding the ABM simulations. These figures include weekly distribution of mosquitoes across the province, and average daily temperature and rainfall values observed in the simulations for a number of years. Appendix C presents an agent-based model of meme propagation in the Facebook online social network in order to compare its dynamics and trends against the spread of an infection disease such as WNV.

Chapter 2: DATA PREPARATION

2.1 Introduction

This chapter examines the data inputs into an agent-based model and simulation of WNV using the Anylogic software [112], with specific focus on data collection and compatibility, and preparation or processing techniques.

Although Anylogic is a powerful multi-paradigm modelling framework, there are few user group resources or forums available for its users. To the best of author's knowledge, there is only one active user community in LinkedIn for the Anylogic modelling software. In addition to more traditional simulation, Anylogic (v. 7.0.3) also has relatively recent support for GIS simulation and modelling. Within GIS, Esri shapefiles are the most commonly used [113]–[117]. The shapefile format includes vector data representing location, shape, and attributes of geographic features such as lakes, mountains, buildings, and roads.

However, it can be quite difficult to format shapefiles in a way that a modeller could easily apply or use them within the Anylogic framework for GIS-based simulations. This chapter explains (in a tutorial-based style) the procedures used to prepare the data required to develop an ABM of WNV spread in southern Manitoba, Canada. The region of interest is an area of approximately 148,812 km^2 partially covered by grasslands (Canadian Prairies), where the primary WNV vector is *Cx. tarsalis*.

As noted, WNV is carried and transmitted by mosquito vectors. Birds and humans are among the hosts for the infection. A WNV model requires, at a minimum, data on these three agent types. The mosquito-related data include (but are not limited to) weather for population dynamics, landscape features for habitat preferences, and twilight times and daylight duration for setting peak periods of mosquito agents' biting activities. A conceptual ABM may model the

area as a grid, in which each cell has different properties regarding mosquito population dynamics. Such data would be used to tune or modulate the mosquito parameters of each cell according to weather, landscape, and daylight conditions, ultimately governing the mosquito behaviours and interactions. The bird-related data that are collected include nesting/roosting locations, population estimates of each species, home range areas, breeding season months, communal or solitary living habits, and typical flight speeds. A conceptual ABM may distribute and initialize the bird agents of different species based on the population estimates and roosting locations. The movement patterns of birds may be determined based on the home range, flight speed, and living habits of each species. For instance, in each time-step, birds could pick a random flight speed in a certain range and fly up to a certain maximum distance. The algorithms used for movement simulation may be different depending on whether the species are solitary, and whether they are mating at a particular time of the year. The human-related data necessary to incorporate realistic human movement patterns include census counts, street networks, and coordinates of cellular telephone towers providing service for a number of anonymized mobile users, where mobile phones act as proxies for their users. A conceptual ABM may initially distribute human agents over the map based on the census data. Human agents may then move inside the street network according to cellular phone tower or trajectories provided by the data, with cellphones serving as proxies for individuals. The high-level architecture of such an ABM is illustrated in Figure 2-1.

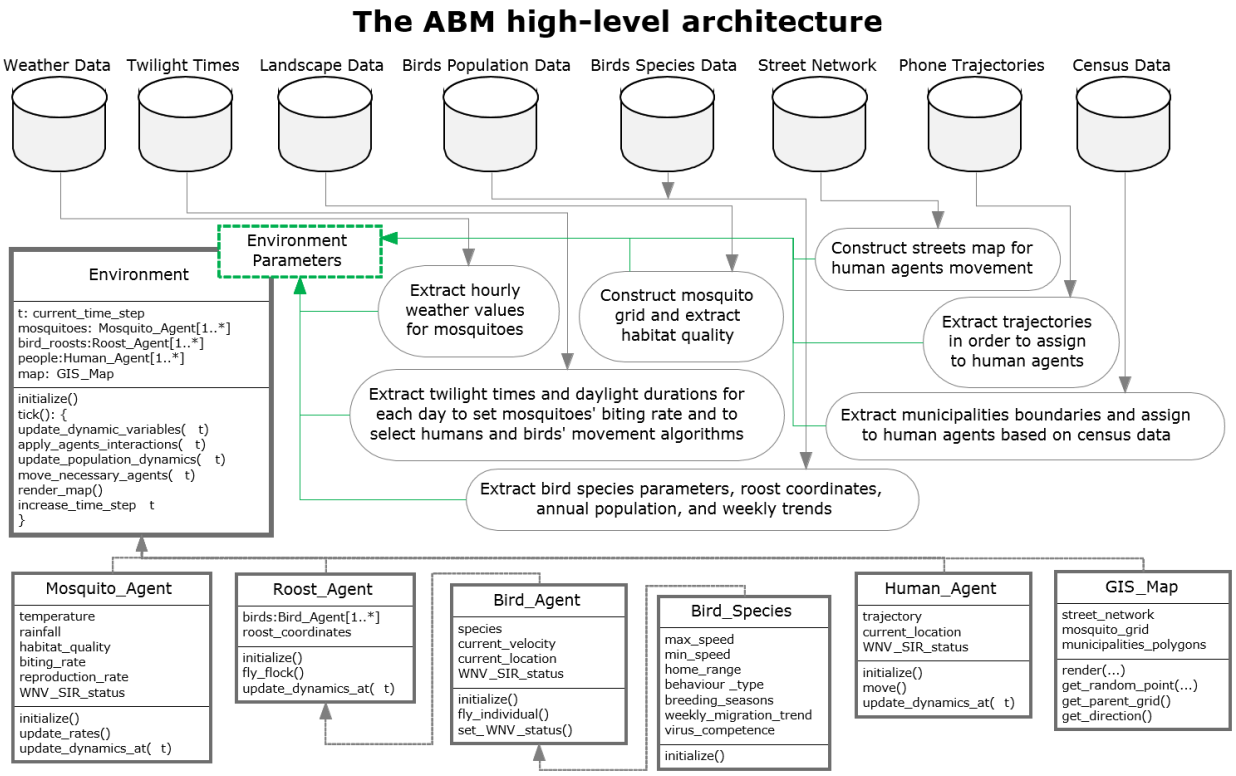


Figure 2-1 High-level architecture of an arbitrary WNV-ABM

The remainder of this chapter describes and presents the methodology and results for collecting, assembling and reformatting some of these data for each agent type of a conceptual ABM of WNV propagation.

2.2 Procedures to Extract Agent Data

The validity and relevance of any ABM relies on incorporating as much real and meaningful data as possible to characterize the environment and the agents. This section examines the collection and processing techniques of data most relevant for an ABM associated with WNV. These include agent data related to mosquitoes, birds, and humans. Similar data processing would be required for other mosquito transmitted diseases as well for whatever ABM framework being utilized elsewhere.

The applied techniques are described in some detail, in order to assist one using Anylogic and Esri ArcGIS, which combine to provide a powerful toolbox to modellers, particularly those working on geo-simulations. The details provided here would significantly reduce frustration for other modellers who are beginning to utilize the software, as there is a high-level of subtleties in the mechanism/interface of both suites of software. The techniques also illustrate a primarily challenge in ABM, that of combining different and often disparate datasets.

2.2.1 Mosquito Data

2.2.1.1 Weather

Preparing weather data is a simple process, and does not need to be discussed in detail. This is primarily due to the familiarity everyone has with weather, as well as readily available data. For this work, one weather station per (urban and rural) municipality was considered. The weather data (including precipitation and temperature) for the years 2002-2014 were downloaded from Canada National Climate Data and Information Archive and BioSim databases [108] whenever possible. For the municipalities where there was no weather station, and for filling in any missing data in the real data, BioSim was used to provide simulated weather data. As an input point for the BioSim built-in simulator [108], whenever possible, the coordinates of an existing weather station were used. Otherwise, a location provided by Wikipedia and confirmed on Google maps was used as the coordinates of a municipality. These data were combined with the gathered hourly and daily data. Wherever simulated data were used (in 114 out of 118 municipalities included), the data were flagged for future reference.

2.2.1.2 Landscape Features

Recall that a common approach to describe landscape features is to classify different regions of land into various categories based on ground cover, ground features, as well as land use. Remote sensing satellite images, from the Manitoba Remote Sensing Centre in our case, are used to classify each region.

This and other geospatial data are typically available in the shapefile format. The GIS library and components available in Anylogic have the ability to work with shapefiles. As such, one way to add land-based habitat characteristics for mosquitoes in a WNV model in any GIS-integrated software (such as Anylogic) is through shapefiles. A square grid shapefile of southern Manitoba was chosen, where each cell represents a 5 km × 5 km mosquito site. Within each cell, the area of each land cover class can be calculated and recorded in the shapefile database file. Then the shapefile can be loaded in Anylogic. So, for any given coordinate in the map, one could retrieve the covering mosquito cell and its associated information in the shapefile database. Here, the procedure to create such a shapefile using the Esri ArcMap (part of the ArcGIS software package) is explained in detail. It is noted that the changes in the land cover over the span of simulation years are negligible.

First, the land cover data for each region of Manitoba was downloaded from the Manitoba Initiative Data Warehouse [118] in the shapefile format. The class of land cover for each feature (polygon) in the data was identified with a *GridCode* number in the shapefile database file. The coordinate system of these shapefiles is the UTM (Universal Transverse Mercator). This is a projected coordinate system that enables the ArcMap to calculate geometric properties of a polygon feature such as area or perimeter. For this reason, the coordinate system was kept unchanged at this time.

All shapefiles were then added together to make a single general map of land cover using the *append* command in the geo-processing toolbox of ArcMap. Next, the geometry of the new map was repaired using the features toolset under the data management toolbox which is part of the geo-processing toolbox. Repairing the geometry was necessary to fix some common geometry problems (such as empty parts or duplicate vertices). The outcome was a map (or shapefile) with standard geometric specifications as shown in Figure 2-2, where blue colors represent water body or wetlands; green represents different forest types, orange represents agricultural or forage cropland.

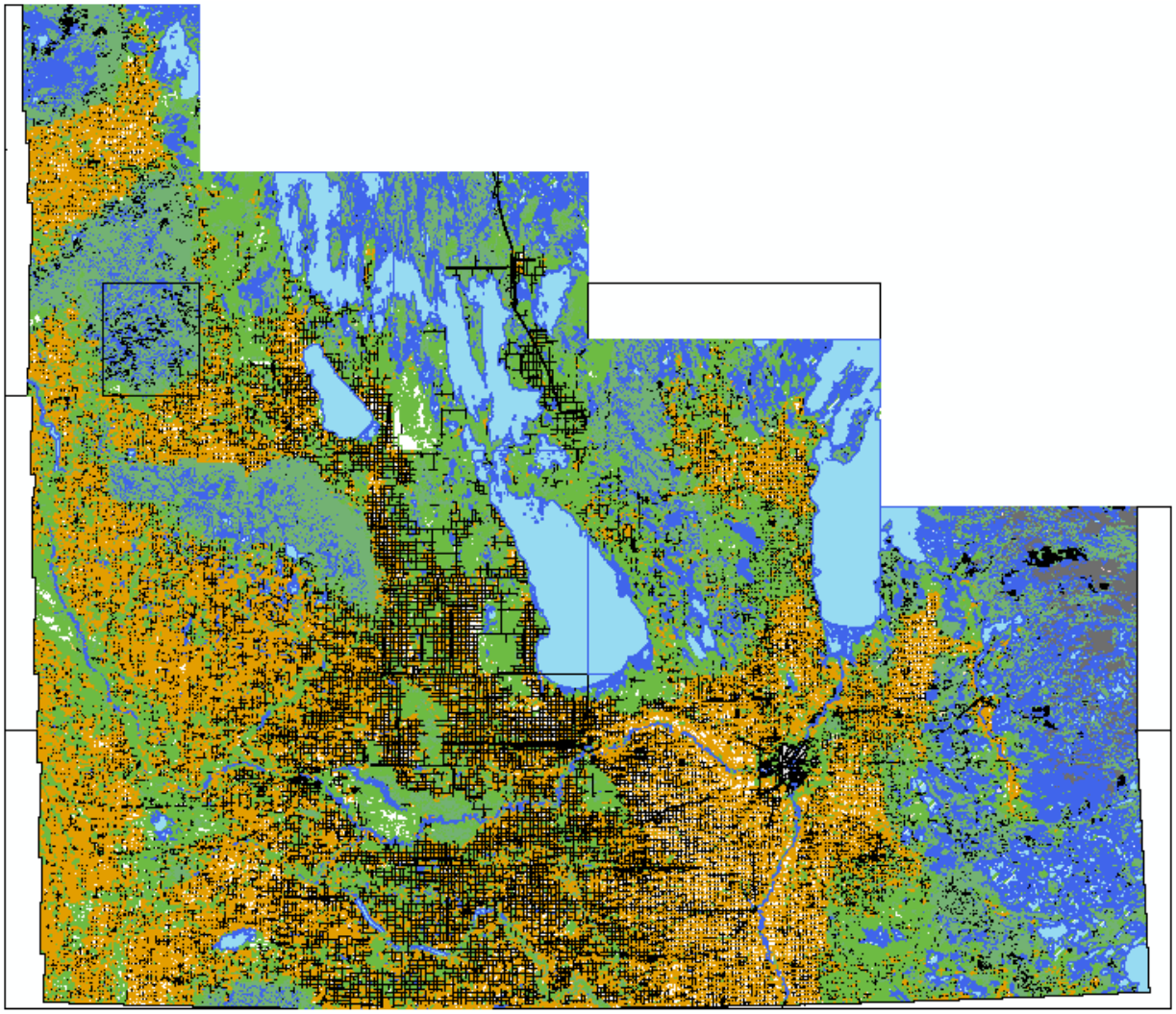


Figure 2-2 Manitoba Land Cover map (shapefile) by combining data from all regions

At this stage, the 5 km × 5 km mosquito grid had to be built to incorporate the land cover data from the previously prepared shapefile. An overlay grid of southern Manitoba (i.e., region of interest) with an accuracy of 5 km by 5 km using the *Fishnet* command in the geo-processing toolbox of ArcMap was created. The output of this procedure was a rectangle shaped map containing many square cells as shown in Figure 2-3, where the (urban and rural) municipalities are also shown (in color) for a better visual clarity of where the mosquito grids are located. It is noted that the coordinate system of this grid was the UTM, the same as that of the data source and data frame of all the layers in the ArcMap project.

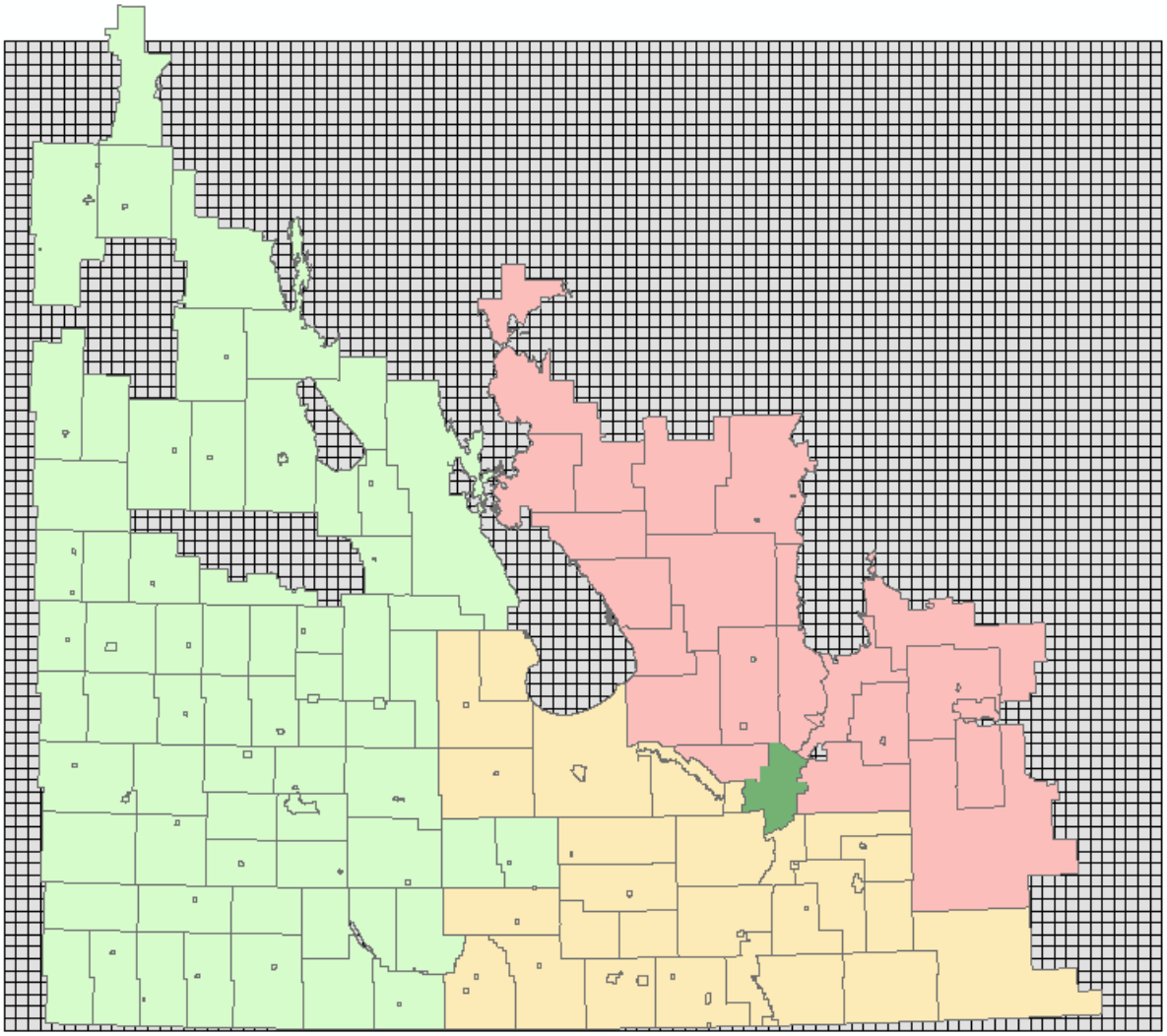


Figure 2-3 Mosquito grid beneath the map of municipal boundaries

Subsequently, the land cover shapefile from the previous step had to be attached to the mosquito grid. To do so, the *identity* command in the geo-processing toolbox was applied to find the geometric intersection of the grid and land cover map by setting the grid as the *identity feature*, and the land cover map as the *input feature*. As a result, the land cover map or portions thereof that overlap the grid obtain the attributes of the grid, which are basically the square cell ids. This means for every single pair of a square cell and a land cover polygon with some overlapping area, a new polygon feature is created in the new map (or shapefile). Figure 2-4 illustrates an example of this operation. The new map is called the land cover grid. It is notable that polygons resulting from the intersection of the square (mosquito) grid and the land cover map are not necessarily squares (in UTM). However as most of these polygons are still shaped as squares, the map is called a grid. For all the entries (polygon features) with a known square cell id in the land cover grid, the land cover *GridCode* is also known. For the next step, the geometric area of each of these entries is required. Therefore, a new field called *Area_Sq_M* was added to the shapefile database using the *attribute table* in the ArcMap. Then the area of each feature in square meters was calculated and stored in this field using the *calculate geometry* feature in the *attribute table*.

There are ways to find the proportional area of each land cover class for each square cell using ArcMap. Once these metadata are calculated, they can be stored in the shapefile database file. The extension of these files is DBF, they and can be accessed in a classical FoxPro database query. However, there are two techniques that a developer should consider. First, the filename plus its extension has to be less than eight characters. Second, there is only one table per database file. So, the table name in a Structured Query Language (SQL) *select* command is the same as the filename, and the database address in the connection string only includes the location

without the filename. Sample Java code is provided in Appendix A, illustrating how to connect to a shapefile database (DBF) file using a Microsoft (MS) Access dBase driver connection string.

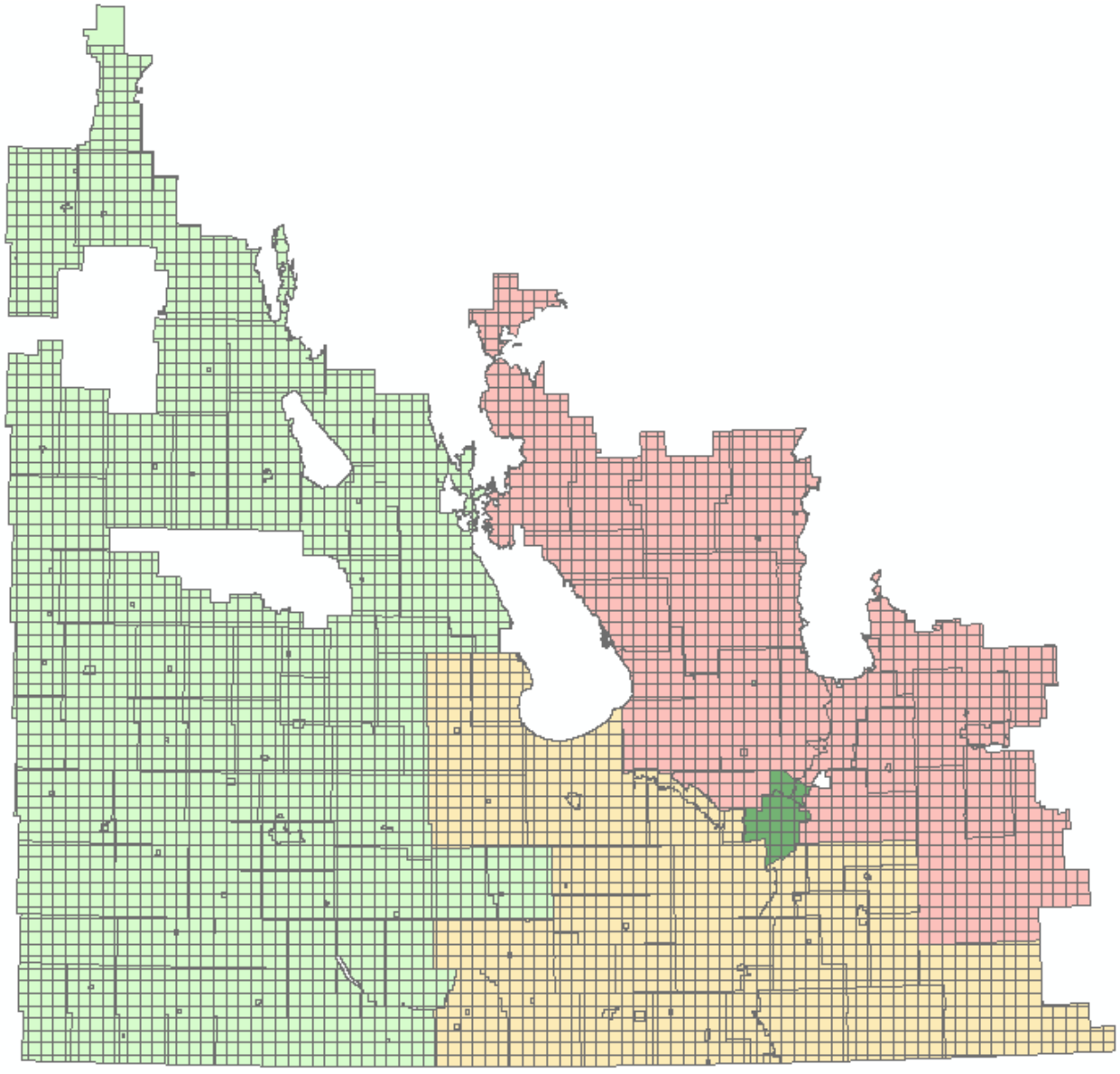


Figure 2-4 Result of the *identity* command by setting mosquito grid as the *identity feature* and municipal boundaries as the *input feature*

Given the limitation of shapefile databases, a decision was made to clone the shapefile database into a MS access database, and apply necessary queries and adjustments there. As such, the land grid shapefile database was exported into a text file. The text file was imported into a MS Access dBase. Then, using a small C# application, the proportional area of each land cover class within each square cell was calculated and stored in a MS Access table where the primary key was the square cell id. This means for any given mosquito cell, if the cell id is known, a simple SQL query could reveal the exact information regarding the land cover within the cell. The Anylogic GIS library is helpful in these instances. Once a shapefile is loaded into a GIS map component of Anylogic, for any given pair of latitude and longitude, the id (or any other attribute) of the shapefile feature at the same point is accessible using the *findPoliticalArea* function. So, in our case, Anylogic can be set to return the square cell id for the polygon feature over a given coordinate. The cell id can, in turn, be used to query land cover information of the area.

The last step is to prepare the mosquito grid shapefile to be loaded into Anylogic. For this, the mosquito grid should be clipped to reduce the number of unnecessary cells where no information about the land cover is present. Mostly regions near the boundaries or outside of the province have no land cover information. The only possible impact of removing these regions is the case where some healthy birds fly out of the province and come back home infectious because of the infectious mosquitoes present within these disregarded regions. This is a systematic error and is simply not considered in the model. So, first, a mask of all the cells with useful information is created to clip out the remaining cells. The mask creation procedure is as follows.

1. The *dissolve* command from the geo-processing tool is applied by setting the the land cover grid (including the grid and land cover data) as the input.
2. The “Create multipart feature” option is unchecked; no field is added to the *dissolve* or *statistics* fields.

This would give a single polygon for the whole map of the land cover grid, which could be used as the boundary of the region of interest. At this stage, a hole was noticed in the single boundary (mask) polygon which was due to missing data in the land cover map. As such, a filling donut holes procedure was necessary as follows.

1. The *editor toolbar* was added to the ArcMap toolbar. Then an edit session was started by selecting the *editor toolbar*.
2. The mask in the *create features* bar was selected as the active layer. Then a template as the *construction tool* from the box below it was selected.
3. A rectangle or a polygon over the donut hole was drawn. Then edits were saved and the edit session ended.
4. Once again, the shapefile was dissolved to merge all polygons in this shapefile together. At this stage, the mask without any holes was ready, as shown in Figure 2-5. The area in this figure indicates locations where the land cover data are known.

To exclude the unnecessary mosquito cells, either the *clip* or *intersection* command must be applied on the mask and the mosquito grid. It is noted that while the *intersect* method saves a copy of feature id (FID) fields of both shapefiles in the new shapefile, the *clip* method keeps no record of FIDs. Therefore, if one uses the *clip* method, FID of each cell of the grid (i.e., mosquito cell id) should be copied into a new field beforehand.

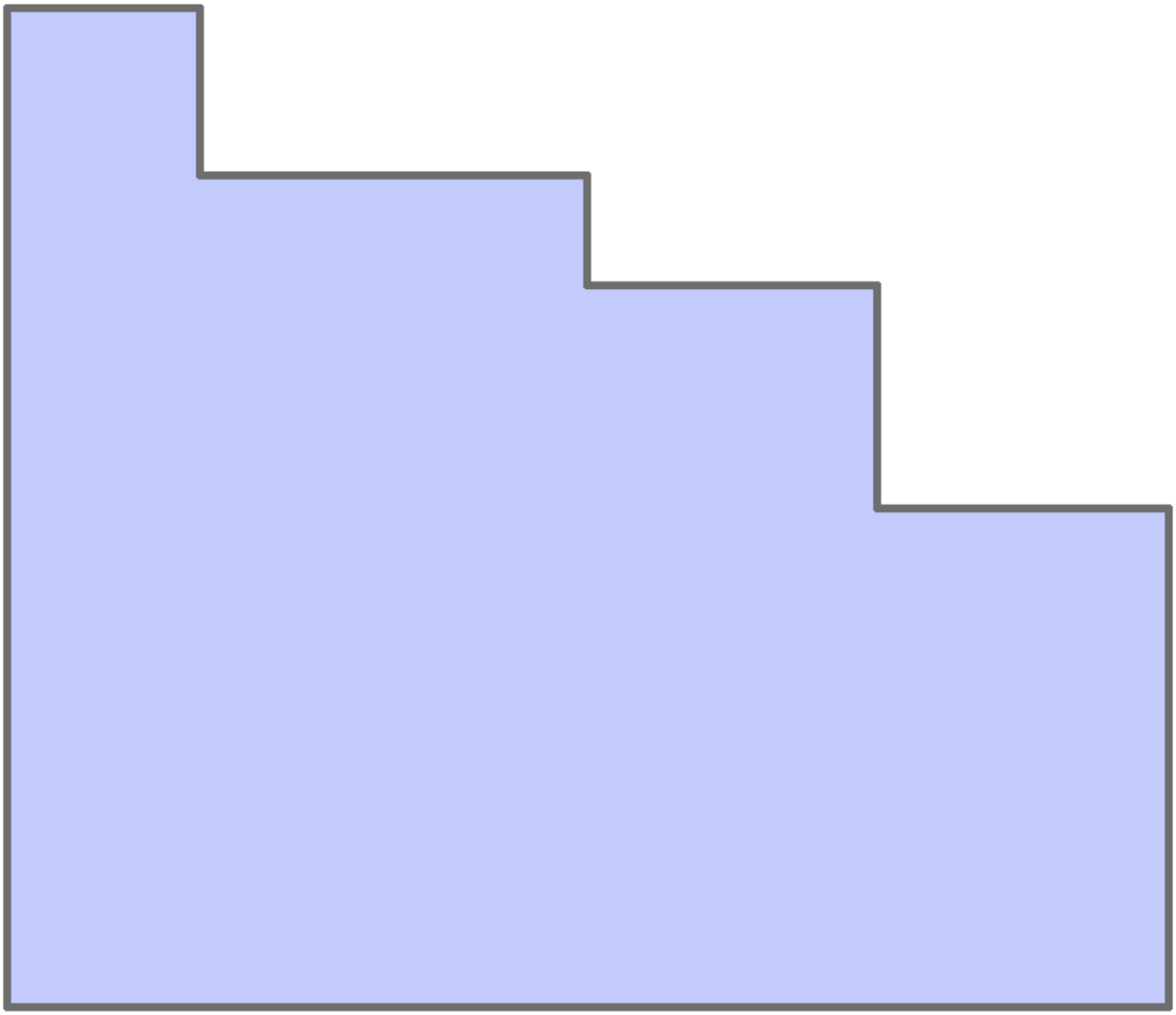


Figure 2-5 Boundary mask of land cover grid

Finally, the coordinate system has to be projected from UTM to World Geodetic System (WGS) 1984 so that it is consistent with Anylogic. One more vital technique is that the shapefile must have at least two fields (other than the default FID and Shape) so that it can be loaded within Anylogic, and in particular for use with the *findPoliticalArea* function.

2.2.2 Bird Data

2.2.2.1 Population Estimate

Many different data sources had to be combined to produce the bird population database. Since detailed population maps were not readily available for most species, a process had to be developed to estimate the population of individual species within relatively small areas. There would first be an emphasis on the geographic distribution, and then on temporal trends.

Two approaches were developed to create these estimates. The first was a "top-down" approach, which relied heavily on population estimates for large regions and relative abundance maps for local distributions. The second method could be considered a "bottom-up" approach. This approach used localized point count surveys and species-specific correction factors to estimate population. Through the use of these two separate approaches, it is possible to establish population estimates that would be suitable for the models considered in this thesis. As with many ABM approaches attempting to use as realistic and meaningful data as possible, best guess estimations were required. As more accurate data becomes available, the veracity of the estimates improves.

2.2.2.1.1 Partners In Flight (PIF) Approach

The first "top-down" approach used the U.S. Geological Survey (USGS) abundance maps, which were created by USGS from their 50 roadside stops breeding bird surveys conducted at peak breeding season (June for most species) [110]. The files were downloaded from their website in the form of shapefiles [110]. Each cell in these maps contained a relative abundance value representing the average number of birds observed by the survey in that area. The data from these surveys had been extrapolated and processed so that a map of the entire United States and southern Canada was available (see Figure 2-6).

These abundance maps were then combined with a 10 km by 10 km grid, provided by the Manitoba Breeding Bird Atlas (MBBA) [119]. This was done to make the data compatible with the breeding location data from the MBBA. The grid was downloaded as a Keyhole Markup Language Zipped (KMZ) file from the MBBA website and combined with the abundance map shapefiles using a Python script in the Esri ArcMap.

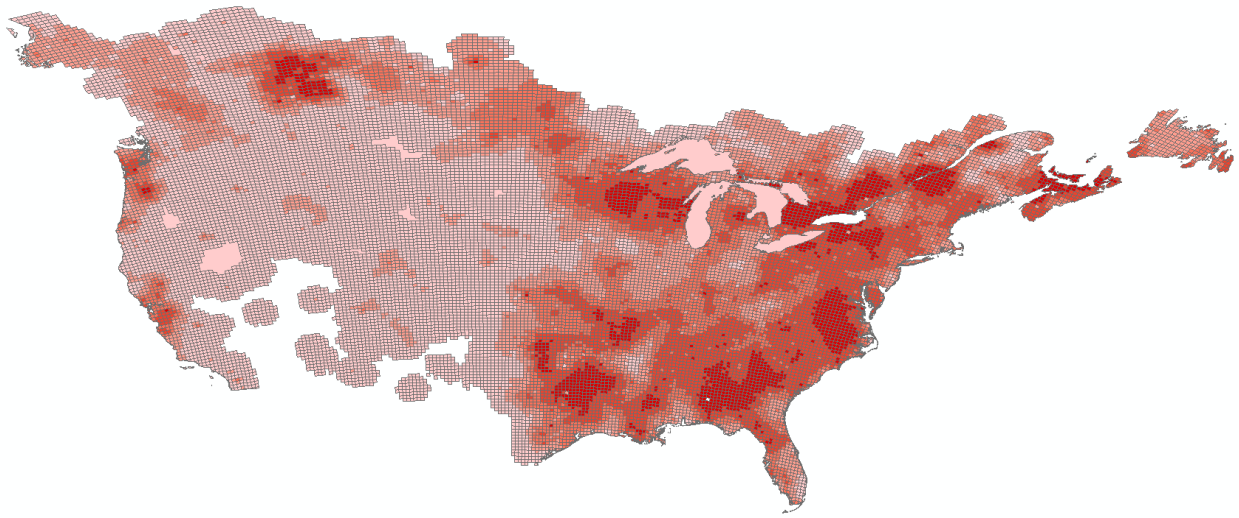


Figure 2-6 An example of the USGS abundance map, where the darker areas denote areas of high species abundance

Each 10 km by 10 km square received the relative abundance value of the abundance map cell that covered it. If a square was covered by parts of two or more cells, the relative abundance value was taken as the weighted mean between all the cells' values, with more weight being given to those cells that covered the majority of the 10 km by 10 km square. If the square was not completely covered by the abundance map cells, the parts that were not covered were considered to be covered by a cell with a relative abundance value of zero. This assumed that the species in question did not live beyond the edges of the abundance map. In many places, this will have been a valid assumption, but at the edges of the USGS study area, this may cause an

underestimation of the abundance. In this way, the relative abundance data was combined with the MBBA 10 km by 10 km squares as shown in Figure 2-7.

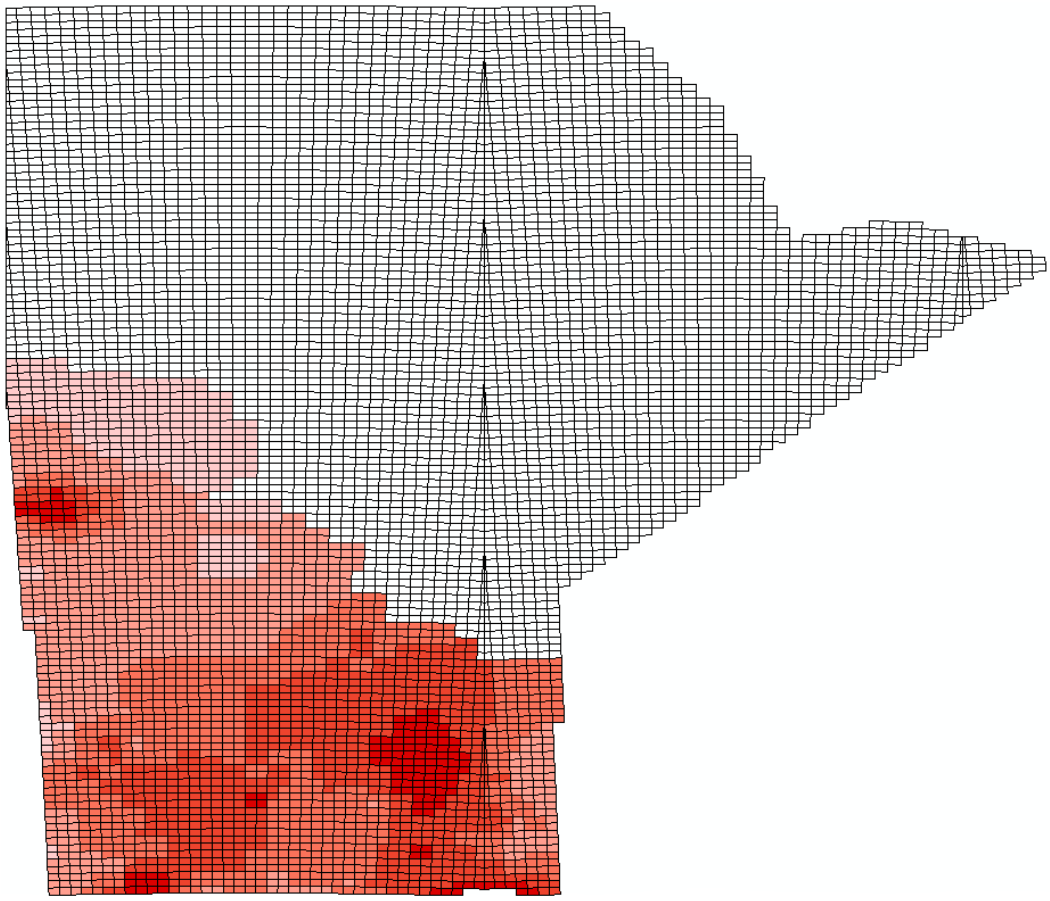


Figure 2-7 USGS abundance data for Manitoba quantized into a 10 km by 10 km grid

Next, the relative abundances were combined with the Partners in Flight (PIF) population estimates for regions in Manitoba [120]. Each PIF population estimate was an estimated population of a certain species for an individual Bird Conservation Region (BCR) in Manitoba. The study area contained three BCR regions, and the PIF gave aggregate population estimates for each of the three regions for all of the species (see Figure 2-8 [121]). The population was distributed between the squares in each region in order to create real population estimates for each square. A greater population was given to squares with a higher relative abundance. Also,

some squares were not exactly 10 km by 10 km, and so a greater population was also given to the larger squares. The equation for the estimated real population of each square is given as follows.

$$Population\ of\ BCR \times \left(\frac{Relative\ Abundance(RA) of\ the\ square}{\sum RA\ of\ squares\ of\ the\ same\ species\ in\ the\ region} + \frac{Area\ of\ the\ square}{\sum Area\ of\ squares\ in\ the\ region} \right) / 2$$

Next, in order to achieve an idea of how the population changed over the course of a year (mostly because of migration), the population estimates of each square were combined with weekly abundance estimates of bird species in Manitoba made available by the Manitoba Naturalist Society [122]. The population estimate was assumed to be the population at the time of maximum abundance in June when the USGS point counts were conducted. From here, the rest of the data was scaled accordingly. In this way, the single population estimate was extrapolated over the year.

Finally, the squares were filtered by whether or not the bird species bred in that area. Using the breeding status data provided by the MBBA, it was possible to remove each square that did not contain a nesting or roosting area within it. Considering the requirements of the WNV agent-based model, no differentiation was made between roosting and breeding areas. The idea was to associate migratory birds to their exact nesting/roosting location, needed by the ABM. The migratory birds at that time of year (i.e., summer) would have a nesting/roosting place within the model area as it is the breeding season. Thus, removing those areas would not remove our migratory bird species from the model.

An assumption was made that by filtering out squares after dividing up population estimates, the number of birds in each square was underestimated. However, efforts to divide up the population after filtering by breeding status led to negligibly different results (<0.5%), since most squares where a species was found contained breeding, and so this effect could be disregarded.

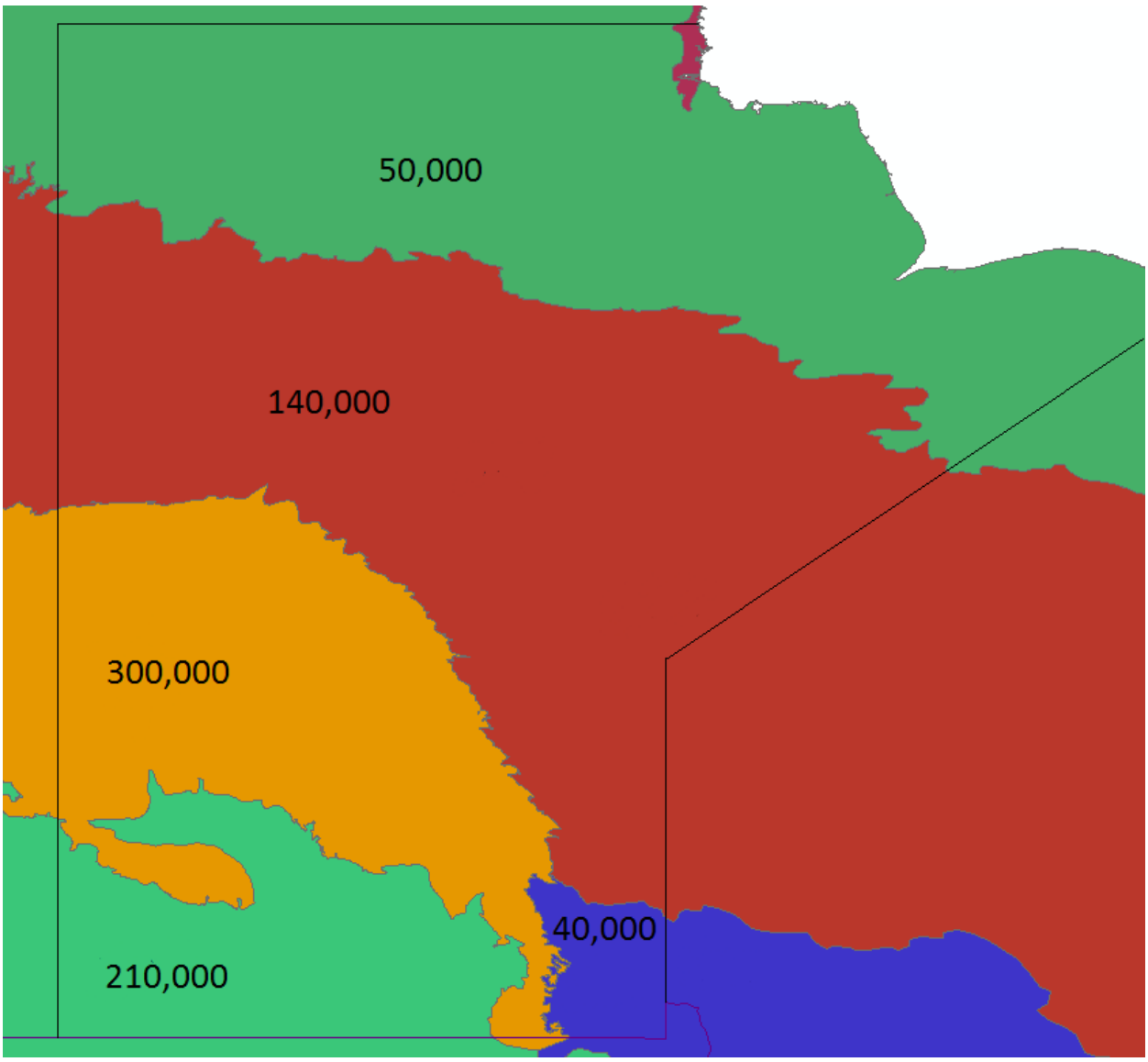


Figure 2-8 Bird Conservation Regions in Manitoba with example population estimates for each region [121] that did not include birds in the same BCR outside of Manitoba

2.2.2.1.2 Boreal Avian Modeling (BAM) Approach

The population was also estimated using a second process, the “bottom-up” approach. In this approach, we began with point count data that was provided by the MBBA. This data was only for squares where breeding was suspected, and gave a larger sample size than that offered by the USGS abundance maps within Manitoba.

The MBBA point counts were conducted by a participant standing in several pre-determined locations inside each 10 km by 10 km square and recording the number of birds that they observed or heard within 5 minutes. This system did not record all the birds in the area, but it did give a relative index into how many birds were in the area. In order to convert the point count data into a real population estimate for the square, correction factors needed to be applied.

Correction factors were obtained from the Boreal Avian Modeling (BAM) project [123]. Although BAM takes many different and complex correction factors into consideration as they create their own population estimates, only two are considered here: the effective detection radius and singing rate.

The effective detection radius (EDR) is defined as the distance from the point of observation where as many birds were detected beyond this radius as were undetected within this radius [123]. This factor took into consideration the fact that a bird further away would be less likely to be detected, and that certain species would be harder to detect at further distances. Thus, when the point count data were considered, it was reasonable to assume that the point count numbers correlated to the number of birds within the EDR.

The singing rate was given as the rate at which a bird sang out per minute, or similarly, the proportion of a bird population that sang at least once in one minute. Since most point counts depend on hearing the sound of a bird more than seeing it to identify the species, the singing rate

gives a useful approximation of how many birds remain quiet, and thus undetected, during the point count. By multiplying the singing rate by the number of minutes spent in observation, one can find the proportion of birds that sang out and had a chance of being identified in the point count. If the point count was long enough, this proportion would rise above 100%, as birds began to sing more than once. However, the observer for each point count was trained to count individual birds, not individual bird songs, and so multiple bird songs by the same bird could be discarded [124]. Therefore, the singing rate-time product was capped at 100%.

The population estimate for each square was calculated as follows.

$$Area\ of\ the\ square \times \left(\frac{Birds\ Observered}{Point\ Counts\ \times\ \pi(EDR)^2 \times (Singing\ rate\ \times Average\ time\ spent)}\right)$$

After calculating population estimates for each square, the population was again modulated to show annual changes. This was done in the same way as the first "top-down" approach using the annual abundance data. In this way, both approaches were used to create usable population estimates for the model.

2.2.2.2 Species Data

Other data, collected and inferred from a diverse variety of sources regarding each species, are reported in Appendix A. These data include home range area, breeding season months, communal or solitary living habits, and typical flight speeds.

The spring/summer home range of all birds was considered to be a circular area. As such, the average radius in meters was calculated and reported in Appendix A. The home range of a given species depends heavily on the habitat and food abundance conditions. The top priority was to collect the Manitoba data. Where possible, reported mean home ranges for the landscapes neighbouring the province were considered. When there were no data for similar landscapes, the mean of the home ranges (as reported in different sources and weighted by the sample sizes) was

used. It was also desirable to avoid underestimating the average home range as the conceptual agent-based model required the ceiling of an average home range to be set as the maximum flight distance for birds.

Breeding timing and living habit data were primarily collected from the Birds of North America online database [125]. The breeding season range goes up to, but does not include the end month. The living habits were categorized into three groups: solitary roosting behaviour, year-round communal roosting, and semi-annual communal roosting. Semi-annual communal roosting included those species that roost individually or in pairs during the breeding season, and then form flocks for migration in the fall. In general, the roosting behaviours of birds are often difficult to categorise. These designations represent an estimate for which type of model of each species' behaviour may best fit in the specific agent-based model, and are not intended to reflect any more universal definition of communal roosting.

The mean flight speed is reported in the format of m/s in Table A.I in Appendix A. For many species, the data on flight speed were either very sparse or non-existent. For such species, certain approximations had to be made, such as using the reported speed for a similar species of that genus or taking the average of reported speeds of the whole family. It was decided to find the typical flight speed at which birds fly/forage during a day. However, in most sources, it was not clear as to the type of speed that they measured. Values were generally either "minimum power speed" (V_{mp}) or "maximum range speed" (V_{mr}). If a certain species had both V_{mp} and V_{mr} available, the smaller V_{mp} would have been recorded as the flight speed of the species. Whenever there were only reports on the maximum flight speed, the minimum number in the given range of maximum flight speeds was used as the typical speed. It is notable that a bird's normal flight speed going from perch to perch is much less than the numbers reported in Table A.I in

Appendix A. These values were treated as the ceiling of typical speed in the birds' movement component of the conceptual agent-based model, and the birds' minimum flight speed was set at two m/s.

2.2.3 Human data

The Open Street Map (OSM) of the province was downloaded from GeoFabrik.de as a highly compressed Protocol-Buffer Binary Format (PBF) file. The street map was then extracted from this binary file using the routing features of Anylogic. A sample street map of southern Manitoba including only trunk, primary, and secondary roads can be seen in Figure 2-9. The A* pathfinding algorithm can be applied on this network for human agents' routing. Census data for each municipality of the province for initial location of human agents were downloaded from Statistics Canada. Municipal boundaries in Manitoba were downloaded as a shapefile from the National Resources Canada website. The population of human agents then had to be distributed within these boundaries according to the census data for each municipality. Anylogic has a *GISRegion* component where one can call the function *randomPointInside* to randomly choose a point inside the given region. Therefore, this function can be used to initialize the human agents' population inside each municipality. However, the shapefile polygon features, representing municipal boundaries, first had to be converted to Anylogic *GISRegion(s)*. Up to Anylogic version 7.2, the built-in converter was not fully functional to convert all the shapefile features to Anylogic *GISRegion(s)*. A sample Anylogic Java code for this non-intuitive conversion can be found in Appendix A. The code also gives hints to developers on how to extract all point coordinates of objects in a shapefile. It is notable that each shapefile has a different number of nested layers of objects. This may be the reason why the built-in converter does not work for all shapefiles.

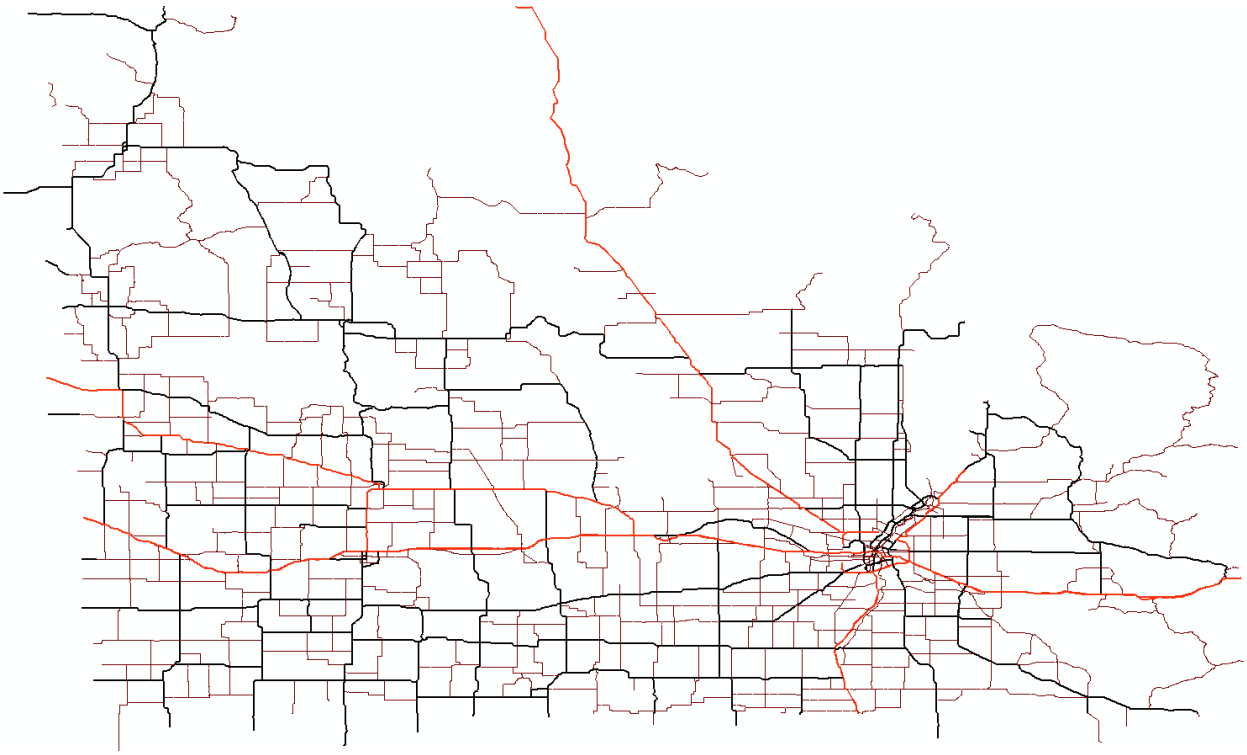


Figure 2-9 Southern Manitoba street network showing only trunk, primary, and secondary roads

For the purpose of this chapter and brevity, human movement patterns, data collection, and processing methods are not further discussed here, but are discussed in Chapter 4. In addition to readily available census data as mentioned above, other sources include those related to personal cellular devices as well as technologies being developed for intelligent transportation systems.

2.3 Results

2.3.1 Mosquito Data

The final weather database has daily and hourly values of weather temperature and rainfall for 118 rural and urban municipalities in Manitoba from 2002 to 2014.This dataset could be used for any weather-dependent studies in the southern Manitoba region. The final land cover database has information on 6067 square cells in southern Manitoba. For each cell, the

corresponding municipality and weather station, total area, coordinates, and area of different land cover classes present in the cell are known. The land cover classes used in this dataset include Agricultural Field, Deciduous Forest, Water Body, Range and Grassland, Mixed-wood Forest, Wetland – Marsh, Wetland – Treed Bog, Treed Rock, Coniferous Forest, Fire-Burnt Area, Open Deciduous Forest/Shrub, Agricultural – Forage Crop, Cultural Features, Forest Cut Block, Sand and Gravel, Roads and Rail Lines, Wetland – Fen, and Lichen heath. The land cover database can be used in geo-spatial studies in southern Manitoba. Such studies are of particularly high importance for agricultural purposes. One limitation of the land cover database and the extraction procedure is that the land cover was assumed to be static. Generally, more work on land cover and in particular on dynamic land cover changes is ongoing. For instance, Murray-Rust et al. proposed an open-frame agent-based model to capture changes in land cover [126]. As a future work, the land cover extraction procedure could be automated using Python scripts in ArcGIS. Such an automated procedure makes the system adjustable in response to changes in land cover data. The land cover database in conjunction with the weather database could benefit studies that focus on the forecasting of mosquito-borne diseases in southern Manitoba.

Future work on weather data may focus on improving algorithms used for simulating missing data, in particular, hourly rainfall values for a specific area if hourly estimates are necessary. In this thesis, the BioSim software was applied for this purpose. Depending on the fidelity of the application, one may need to include more weather stations for Manitoba by going through the same procedures explained earlier.

2.3.2 Bird Data

The final bird database contains information on 152 different bird species. For each species, there are population estimates for each of the 2056 square cells, which are roughly 10 km by 10

km areas located in southern Manitoba. Each of these population estimates was also used to create a weekly population estimate for each square to represent weekly impacts of migration. Only squares where some evidence of breeding had been found were included for each species, as these squares were assumed to also contain nocturnal roosts for the species [51]. This dataset could be used by other researchers working on topics such as modeling birds' movements, bird interactions in various agent-based models, and geo-simulations including birds. The assets developed here could also be of considerable value in light of the significant importance of understanding the effects of climate change on risk of zoonoses such as WNV, and the impact of climate change, and land use and habitat disruption on ecosystem health.

Due to fundamental differences between the two population estimation approaches, their assumptions, and availability of data, the two estimates show a degree of disparity in some areas. The chart in Figure 2-10 shows a comparison of population estimates of American crow species for various locations (squares) in Manitoba. The BAM population estimates are calculated using the methodology previously discussed. The EDR-USGS estimates in the figure are calculated in a similar manner, but use the USGS 50 stops data for point counts instead of the MBBA point counts. The PIF population estimates, as mentioned before, are calculated from the regional PIF estimates and abundance maps created from the USGS 50 stops data. The similarity between the EDR-USGS and PIF estimates in most of squares reveals the importance of availability of stops and point counts data. By contrast, the difference between the two estimates in the other squares confirms the differences between the two "top-down" and "bottom-up" approaches. Figure 2-11 shows the physical location of selected squares in Figure 2-10. In order to achieve an idea of how the two estimates compare against some historical data in Manitoba, an average density for a number of species was calculated from the average BAM and PIF estimates for the species

across the province. In Figure 2-12 and Figure 2-13, these densities are compared against the Manitoban bird densities calculated for 1966-94 by Downes and Collins [127] and reported in the work of Kirk et al. [128] for two groups of species of short-distance and long-distance migrants. If a species winters mostly within Canada or North America, it is categorized as a short-distance migrant. If they mainly spend winters in Central or South America, they are considered long-distance migrants [128]. In general, for the long-distance migrants, the BAM densities have less disparity than the birds densities provided by Downes and Collins [127], compared to the PIF. Whereas, for the short-distance migrants, the PIF densities are closer to the birds densities provided by Downes and Collins [127].

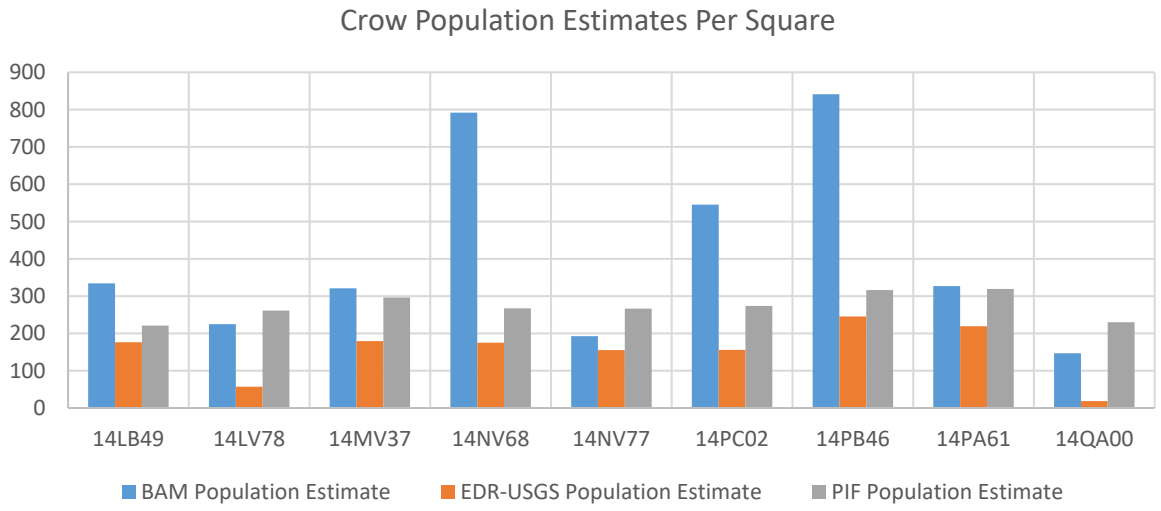


Figure 2-10 Comparison of crow population estimates between the BAM and PIF methods in some selected squares

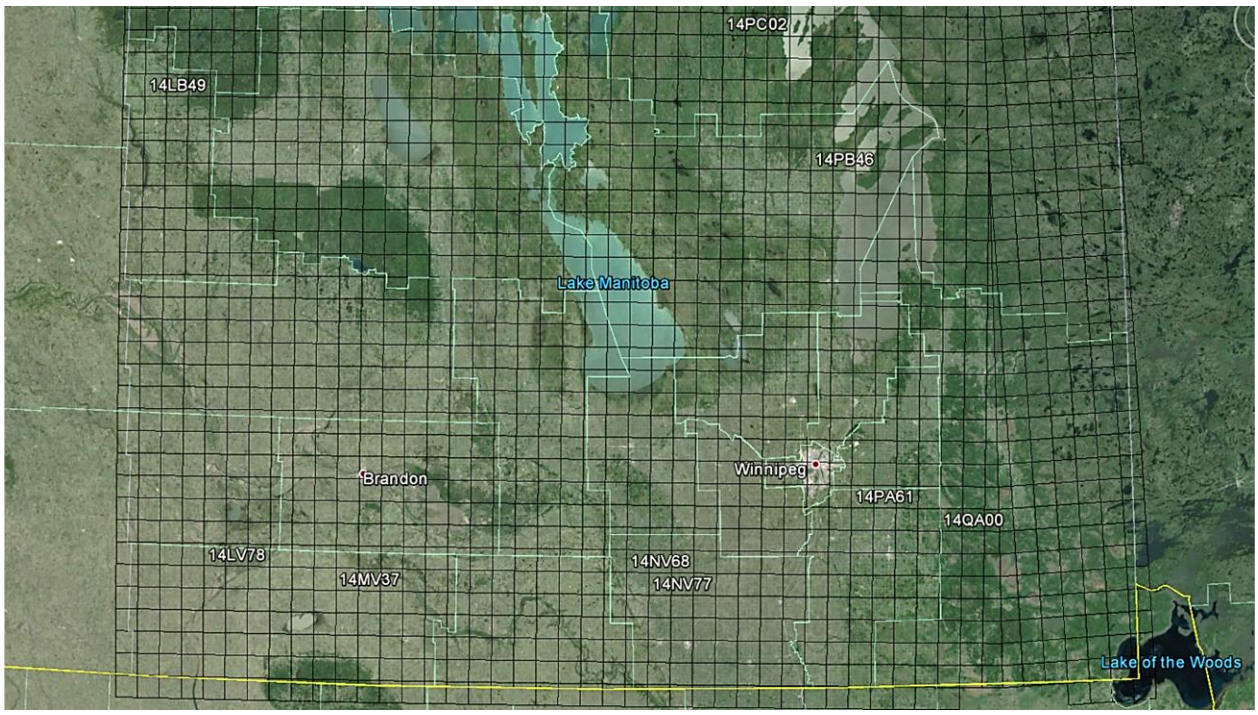


Figure 2-11 Physical location of some selected squares of bird roosts in Manitoba

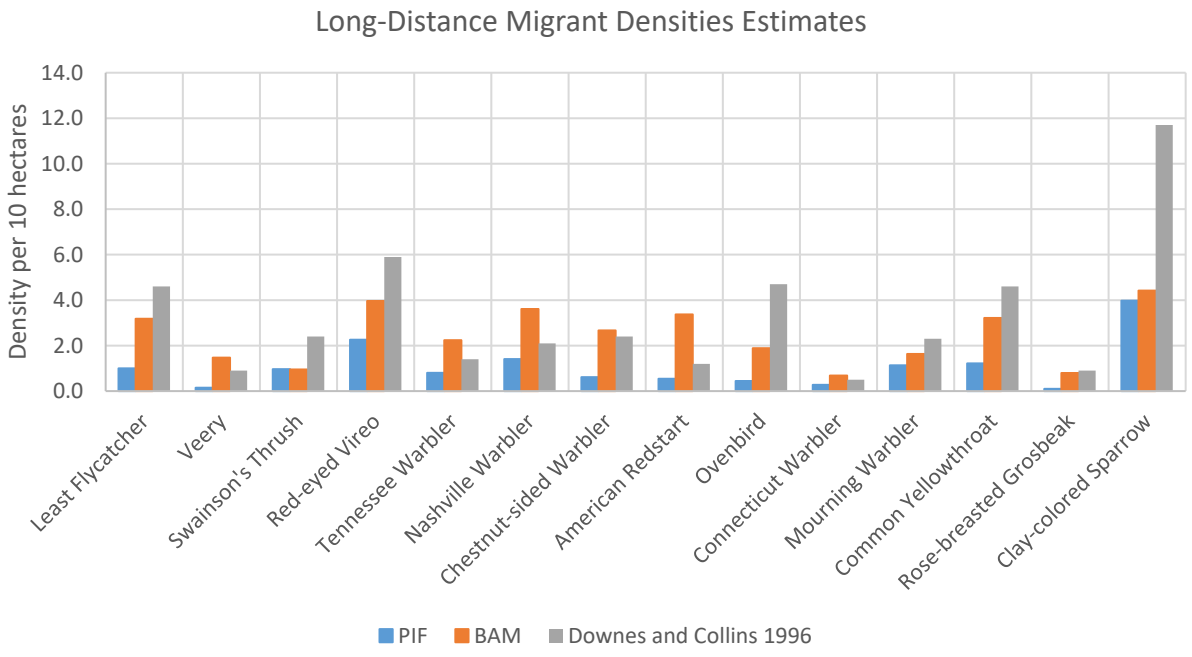


Figure 2-12 Comparison of the mean densities of BAM and PIF against the densities in [127] for a number of long-distance migrant species

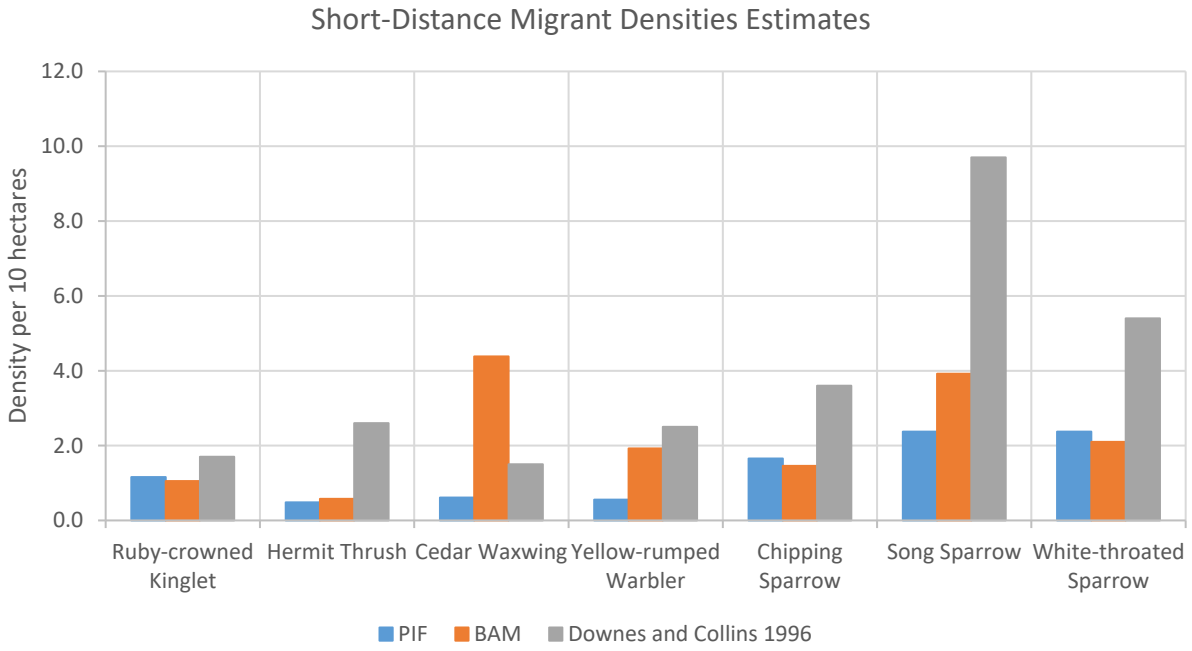


Figure 2-13 Comparison of the mean densities of BAM and PIF against the densities in [127] for a number of short-distance migrant species

Both bird population estimate approaches have several assumptions and shortcomings built in. The "top-down" approach used USGS abundance maps that were based on a very small sample size (< 2% of the birds' project area, or approximately 4,112 out of 205,600 km^2), and these points had only been surveyed approximately once a year (50 stops a year). Also, the PIF regional population estimates, although useful for large-scale conservation efforts and approximations [120], are not very accurate at a small scale. Thus, the first approach contains rather sparse data that has been heavily processed and extrapolated to pertain to a large detailed area. The "Bottom-up" approach sought to fix some of these problems. It was based on the MBBA point counts, which were available for all of the squares over several years. These squares were again not surveyed very many times over the year (15 times over 5 years) but much more of the study area was covered by these point counts [124]. Also, the correction factors used

to convert point counts into population estimates were in part designed by BAM to make up for the weaknesses in the PIF population estimates. However, the correction factors used here did not take every factor into account that could have influenced the point counts. The EDR and possibly the singing rate would have varied with different habitats and vegetation types, but the approach described here did not take this into account. The EDR has a large effect on the population estimate, and is consistently smaller than the radius used in the PIF estimates. As a result, the population estimates are likely to be higher than the PIF estimates. There is also likely a habitat bias against certain hard to reach habitats. The point counts were done by volunteers, and mostly beside roads, so the more remote locations were less likely to be surveyed. This may have caused an over- or under-representation, depending on the species and its preferred habitat. In addition to this, there is evidence that point counts done next to roads obtained biased results for some species. More work on correction factors and population estimates, in general, is ongoing, and future studies and data should be able to improve on these processes.

Bird species data, including home range area, breeding months, communal or solitary living habits, and typical flight speeds are presented in Table A.I in Appendix A. As mentioned in the bird species data section (2.22.2.2.2), as well as Appendix A, many of the reported species data are estimations of some kind. Therefore, many more field studies and work on bird species - although improving - are consistently required.

2.3.3 Human Data

The procedures, explained in this chapter, for preparing the human data, has two outputs. The first output details a database of human census in municipalities of Manitoba with their boundary coordinates. These coordinates are saved as Anylogic *GISRegion* objects on file. The second output involves the street network of the province in a compatible format with Anylogic

GIS components. Chapter 4 discusses issues regarding extracting the trajectories for human agents according to cellular phone tower data.

2.4 Summary of Chapter 2

2.4.1 Conclusions

Anylogic simulation software in combination with Esri ArcGIS provides a powerful toolbox for developers and modellers to simulate almost any GIS-based environment or process. In this chapter, the application of interest was WNV propagation in the province of Manitoba. The land cover data of Manitoba was rasterized in an optimum sized shapefile compatible with Anylogic. Some hints and techniques regarding working with shapefiles in Anylogic were reviewed. A database of over 150 different bird species vulnerable to WNV, including their nesting locations, population estimates, home range radii, roosting behaviour, and start and end of breeding season was collected. The street network for Manitoba, extracted from OpenStreetMap, was loaded into Anylogic to be used in its pathfinding library. The procedures for collecting, combining, and reformatting all these data are explained in details in a tutorial-based style to benefit other modellers working in similar areas.

Researchers are constantly exploiting new non-traditional sources of data for modelling different human diseases. For example, in a relatively recent study, google search data have been used for modelling transmission dynamics of the Zika virus [129]. On the contrary, in this chapter, more traditional data sources were gathered and prepared in a suitable way for agent-based modelling of WNV. Inevitably, modeling natural or environmental processes depends heavily on the availability of appropriate data. This is particularly true for verification and validation of models, as models and simulations would not gain significant attention unless they

are shown to closely resemble reality. This resemblance can only occur with meaningful data; therefore, the importance of data cannot be overstated. The pertinent procedures and an overview of resultant data for WNV geo-modelling are presented in this chapter.

2.4.2 Applications

There are some limitations in the presented data mining and assembling procedures, each of which was discussed in the corresponding section. Notwithstanding, similar mining methods could be adopted by other researchers to compile such dataset according to their own specific needs for other geographic areas e.g., estimating population and location of birds in other provinces/states in North America. The research of this chapter should be useful to others working on a variety of mosquito-borne diseases, such as Zika, Dengue, and Chikungunya, by providing the data relating to Manitoba and/or a systematic path to follow for producing and processing such crucial data. The final complied dataset here could also be used to model mosquito population dynamics, for instance in order to evaluate control strategies. Some recent studies in this area can be found in the work of Ewing and Cobbold [130] and Marini et al. [131].

Chapter 3: Cellular Difference Equation Scheme for West Nile Virus Modelling

3.1 Introduction

In this chapter, a data-driven Cellular Difference Equation (CDiffE) model, including human component and multiple species of birds, is proposed to be used in ABMs for WNV propagation. As one of the goals here was to develop a WNV simulation software for Manitoba, the implementation of the CDiffE is based on the data available from (or appropriate for) this region. However, the model design is more general and applicable to other areas. Out of the WNV models reviewed in section 1.3.3, the work of Bouden et al. [51] is conceptually most similar to the proposed work here. However, key differences to the work here are found in several key areas. For instance, the authors in [51] use differential equations applied at the level of the municipality, the two category of birds are treated as mobile roost agents, and time advances are implemented on a daily step. The work here applies differences equations, which are inherently faster than differential equations for numerical simulation. Also, the work here applies a difference equation to each mosquito cell, which is a much more fine-grained implementation. The CDiffE scheme here consists of many different bird species, yet their movement are considered at a less detailed scale. Finally, our work implements hourly time-steps, trading speed for more fine-grained outputs.

Compared to difference equation models, differential equation models have been applied more often in both WNV and ABM literature, and this chapter demonstrates that difference equation models offer a valuable approach to optimizing the trade-off between a model's computational complexity and its inherent ability to consider multiple agent types at very fine-

grained and individual profile levels. Difference equations are inherently faster for numerical computations, as they do not require a slow ordinary differential equation (ODE) solver to be triggered in each iteration of the algorithm. That is vital for ABMs as they generally suffer from computational complexities, as such, difference equations fit better within their realm. The proposed difference equation model is cellular, as it requires the spatial input domain to be rasterized. The process of creating an overlay grid of the input area of the model (i.e., southern Manitoba in this study) was explained in Chapter 2. Landscape feature impact on mosquito population is integrated in the proposed difference equation through the land cover combinations of each cell of the input grid. The impact of weather (i.e., temperature and rainfall) is added to mosquito breeding, development, biting, and death rates based on previously validated studies [132], [133]. Due to nocturnal activities of mosquitos of interest (i.e., *Cx. tarsalis* for southern Manitoba), the biting rate is further adjusted according to daylight conditions. The proposed equations are then altered to include weekly migration patterns of different species of birds. The distribution of birds is considered to be scattered across the input domain according to data specific to Manitoba as outlined in Chapter 2. Estimating birds-related WNV parameters for each species based on data from biological studies is discussed in this chapter. A simple but computationally fast method for modelling birds' movement component based on their home-range is introduced and implemented to be adopted in bird agents of an ABM. Finally, the model output is validated against carbon dioxide baited trap data from the Manitoban government in terms of Pearson and Spearman correlations.

3.2 Methods

In this section, the proposed cellular model is described. The region of study (i.e., Manitoba province) is modeled as a grid map where each cell represents a 5 km x 5 km area, called

“mosquito site.” Figure 3-1 shows the scope of the model and the mosquito grid with MapQuest tiles. Each mosquito site (or cell) has its own unique landscape features and is assigned to a rural municipality's weather station. There are around 6,000 mosquito sites covering the scope of the model in southern Manitoba, of which over 5,000 sites are active mosquito sites. If data regarding the bird populations within the neighbourhood of a mosquito site exists, the mosquito site is marked as active. Around 150 different species of birds, including local and migratory birds known to be infected with WNV, are present in Manitoba in summer. As noted, the data about the location and population count of these different bird species are collected (see Chapter 2). These locations are referred to as bird roosts. A roost site is where birds spend their night and sleep. Unlike the mosquito sites, sizes of different roosting sites are different. A roost site is not a single point, but rather an area (of at most 10 km x 10 km) around a center point. A roosting site may be shared by more than one bird species at a time. In the model, due to differences of various bird species (such as different virus competence indices), a bird roost is considered for every single species of bird that are located at the same roosting site. For example, if at a certain point (longitude and latitude) both bird species of American Crow and Raven are present, two different bird roosts with the same coordinate are added to the model. From an Agent-Based Modelling perspective, mosquito sites and bird roosts can be considered as agents in the model.

Using the BAM (Boreal Avian Modelling) estimation method, there are over 70,000 bird roosts in the model. During the breeding season, many birds (not all) tend to sleep at roosts near their nesting area within their daily home range. The home range is considered the area used by a bird for daily foraging activities. Each mosquito site is associated with a number of bird roosts depending on the size and coordinates of the roost, and the home range of the bird. On average, each active mosquito site is partially overlapped by (and consequently linked to) 30 bird roosts.

An arbitrary connection between some bird roosts and mosquito sites is displayed in Figure 3-2, where mosquito sites are drawn as squares and bird roosts are blue circles.

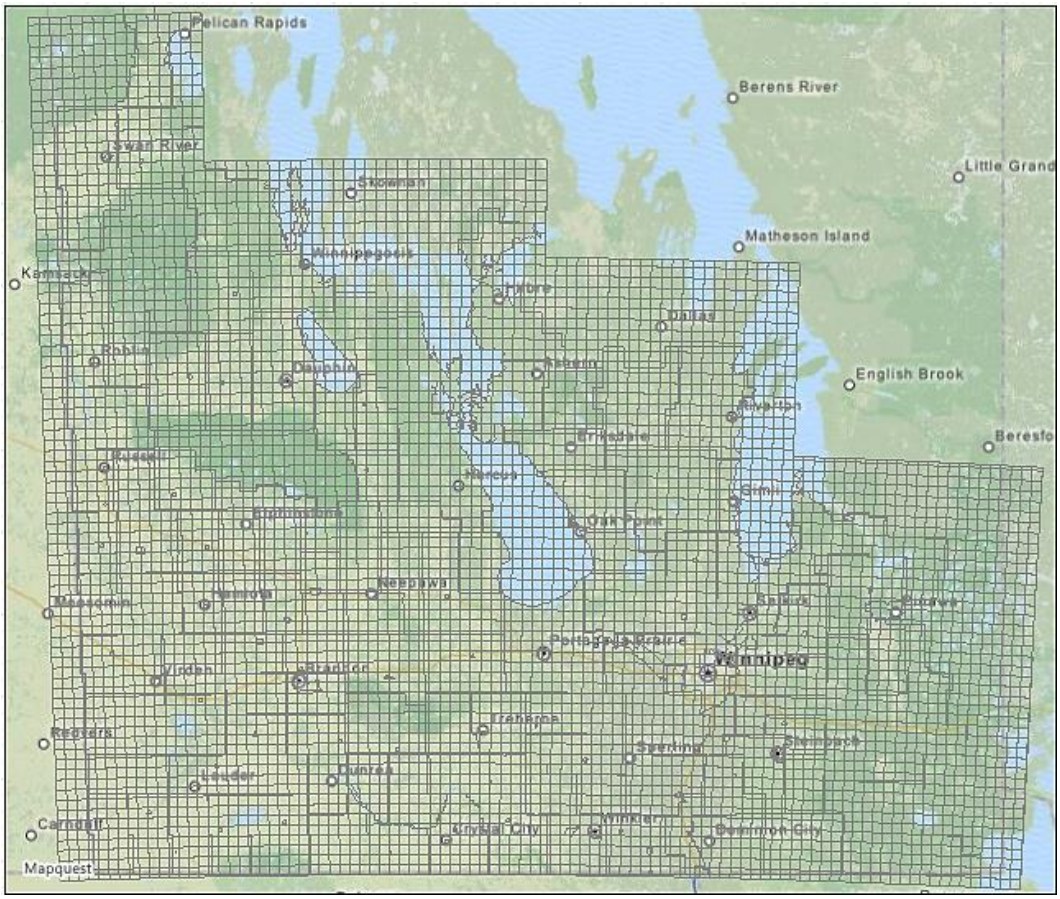


Figure 3-1 Southern Manitoba map showing the mosquito grid as an overlay

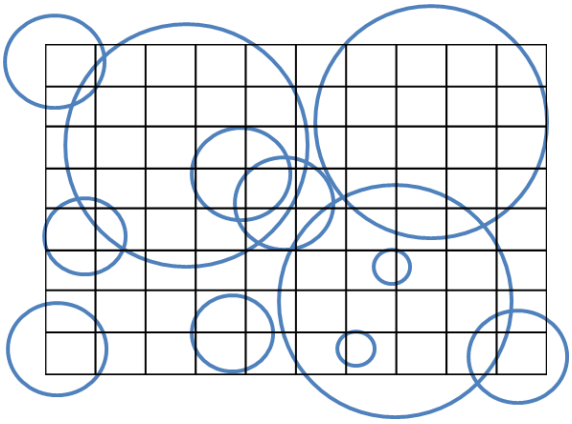


Figure 3-2 An arbitrary grid map displaying mosquito sites and bird roosts

3.2.1 Difference Equation Model Formulation

An instance of the proposed difference equation model is cloned in each mosquito cell (site). In every hourly time-step, the equations update for all the active mosquito sites. The dynamics of all agents (mosquito sites, bird roosts, and humans) update based on the proposed set of difference equations, which consider the interactions of parameters shown in Table 3-1. The key parameters are Daily mean Temperature, *T*, Daily Rainfall, *R*, and the Landscape descriptor, *L*. The proposed difference equation model is built upon the differential equation model of Wonham et al. [89] using the approach in Lewis et al. [134], with further extensions to include humans, various bird species, and impact of temperature, rainfall, and landscape features (so-called land use/cover). Initially, the natural birth and death rate of birds were also included in the model, but after adding the impacts of emigration and immigration of birds, those effects were no longer necessary to retain, as the population of birds were reset to their real-world population count on a weekly basis, no matter how many were born or died during the week. We begin by rewriting the corresponding set of difference equations of Lewis et al. [134] as follows with the parameters definition in Table 3-1.

$$M_a(t+1) = r\,(1-\mu_a)\big(M_s(t) + M_e(t) + M_i(t)\big) + (1-\mu_a)(1-\gamma)M_a(t) \tag{3.1}$$

$$M_s(t+1) = (1-\mu_m)(1-\beta_m)^{bB_i(t)/B_t(t)}\,M_s(t) + (1-\mu_m)\,\gamma\,M_a(t) \tag{3.2}$$

$$M_e(t+1) = (1-\mu_m)\left(1-(1-\beta_m)^{\frac{bB_i(t)}{B_t(t)}}\right)M_s(t) + (1-\mu_m)(1-k_m)M_i(t) \tag{3.3}$$

$$M_i(t+1) = (1-\mu_m)k_m M_e(t) + (1-\mu_m)M_i(t) \tag{3.4}$$

$$B_s(t+1) = (1-\beta_b)^{bM_i(t)/B_t(t)}\,B_s^{j}(t) \tag{3.5}$$

$$B_i(t+1) = (1-\delta)\big[1-(1-\beta_b)^{bM_{i(t)}/B_{t(t)}}\big]B_{s(t)} + (1-\delta)(1-\zeta_b)B_i(t) \tag{3.6}$$

$$B_r(t+1) = \zeta_b B_i(t) \tag{3.7}$$

Table 3-1 Parameters and variables of the CDiffE model

	Mosquito	Birds	Human
State variables			
Aquatic stage (including eggs, larvae, and pupae)	M_a		
Susceptible (adult)	M_s	B_s	H_s
Infectious (adult)	M_i	B_i	H_i
Recovered (adult)		B_r	H_r
Total adults		B_t	H_t
Core parameters per time-step (hourly)			
Reproduction number (combining egg laying and larval hatching)	r		
Maturation probability (i.e., developing into adult mosquitoes)	γ		
Natural death probability (for aquatic and adult mosquitoes)	μ_a , μ_m		
Probability of death due to infection (for birds)		δ , δ^j	
Probability of virus transmission to	β_m	β_b	β_h
Mosquito biting on host (no. of bites per mosquito per time-step)		b , b^j	b_h
Probability of recovery from virus		ζ_b	ζ_h
Virus incubation probability	k_m		
Bird migration (weekly) parameters			
Weekly count of population (given by the data)		B_w	
Number of (weekly) immigrant birds		B_m	
Proportion of (weekly) non-emigrant birds (stayed in the model)		η	
Dynamic parameters			
Mean daily positive Temperature (in Celsius)	T		
Daily Rainfall (in mm)	R		
Mean correlation between mosquito land cover and virus	L		
Daylight hours	D		
Other variables/parameters			
Species index (for birds)		j	
Normalized host competency index (for birds)		c^*	
Hourly average of b, b^j, or b^h (per a 24-hour period)		$\bar{b}$	$\bar{b}$
Hourly expected number of contacts with an infectious agent	e_m	e_b	e_h
Hourly time-step	t	t	t
Weekly time-step	t_w	t_w	t_w
Ratio of mosquito biting activities between nights and days	$\omega > 1$		
Ratio of biting on birds to biting on humans	$\lambda > 1$		
Importance of landscape feature for mosquito habitat	w_L		

* The definition of additional related symbols can be found in Table 3-3.

The core of the proposed (and implemented) difference equation model has some fundamental changes compared to Lewis et al.'s. First, in the proposed difference equation model, by disregarding the virus incubation period, the exposed state for mosquitoes (M_e) are removed to decrease the complexity of the model for computationally intensive applications such as those associated with ABM. This means equations (3.3) and (3.4) are merged together. Second, the exponents in the equations (3.2), (3.3), (3.5), and (3.6) must be modified to include multiple bird species, making the contact rate and the probability of transmission of virus from an infectious mosquito to a bird different for various bird species. This implies that the probability of transmission of virus from an infected bird to a mosquito (β_m) is the same among all the bird species. Removing this assumption is discussed in section 3.5.2. Some other important assumptions regarding the difference equations can be found in section 3.2.3. Lastly, other extensions are necessary at the core of the proposed difference equation model to incorporate the human components and the impact of changes in the population of birds due to migration. By applying these modifications, the core of the proposed difference equation model is as follows.

Mosquito Equations:

$$M_a(t+1) = r\,(1-\mu_a)\big(M_s(t) + M_i(t)\big) + (1-\mu_a)(1-\gamma)M_a(t) \tag{3.8}$$

$$M_s(t+1) = (1-\mu_m)(1-\beta_m)^{e_m(t)}\,M_s(t) + (1-\mu_m)\,\gamma\,M_a(t) \tag{3.9}$$

$$M_i(t+1) = (1-\mu_m)\big(1-(1-\beta_m)^{e_m(t)}\big)\,M_s(t) + (1-\mu_m)M_i(t) \tag{3.10}$$

where

$$e_m(t) = \frac{\sum_j b^j B_i^j(t)}{H_t(t) + \sum_j B_t^j(t)} \tag{3.11}$$

Bird Equations:

$$B_s^j(t+1) = \left[\left(1-\beta_b^j\right)^{e_b^j(t)} B_s^j(t)\right]\eta^j(t_w) + B_m^j(t_w) \tag{3.12}$$

$$B_i^j(t+1) = \left[\left(1-\delta^j\right)\left(1-\left(1-\beta_b^j\right)^{e_b^j(t)}\right)B_s^j(t) + \left(1-\delta^j\right)\left(1-\zeta_b^j\right)B_i^j(t)\right]\eta^j(t_w) \tag{3.13}$$

$$B_r^j(t+1) = \left[\,\zeta_b^j B_i^j(t) + B_r^j(t)\,\right]\eta^j(t_w) \tag{3.14}$$

where

$$e_b^j(t) = \frac{b^j M_i(t)}{H_t(t)+\sum_j B_t^j(t)} \tag{3.15}$$

$$B_m^j(t_w) = \begin{cases} B_w^j(t_w) - B_t^j(t_w) & ; B_w^j(t_w) > B_t^j(t_w) \\ 0 & ; otherwise \end{cases} \tag{3.16}$$

$$\eta^j(t_w) = \begin{cases} 1 & ; B_w^j(t_w) > B_t^j(t_w) \\ \frac{B_w^j(t_w)}{B_t^j(t_w)} & ; otherwise \end{cases} \tag{3.17}$$

Human Equations:

$$H_s(t+1) = (1-\beta_h)^{e_h(t)} H_s(t) \tag{3.18}$$

$$H_i(t+1) = \left(1-(1-\beta_h)^{e_h(t)}\right)H_s(t) + (1-\zeta_h)H_i(t) \tag{3.19}$$

$$H_r(t+1) = \zeta_h H_i(t) + H_r(t) \tag{3.20}$$

where

$$e_h(t) = \frac{b_h M_i(t)}{H_t(t)+\sum_j B_t^j(t)} \tag{3.21}$$

is the expected number of times a human is bitten by an infectious mosquito at the time-step *t*. Similarly, $e_b^j(t)$ is the expected number of times a bird of species *j* is bitten by an infectious mosquito, and e_m (*t*) is the expected number of times a mosquito bites an infectious bird. Also, the two parameters of $B_m^j(t_w)$ and $\eta^j(t_w)$ keep track of changes in the population of birds of species *j* in week t_w according to the real weekly population data (B_w). Therefore, at the beginning of each week, the total number of each bird species $B_t^j(t_w)$ will match the real-world

count of the bird species $B_w^j(t_w)$ that are present during the week (t_w). A value of less than one for η implies the real-world estimated population is less than the number of bird species within the simulation. As such, a proportion of birds from all compartments are assumed to leave the model scope (e.g., due to emigration); whereas when η is greater than one, (immigrant) birds are added to the susceptible compartment. As noted in Table 3-1, the variables t and t_w indicate time-steps; whereas the subscript t in B_t and H_t denotes the total value of its associated state variable (i.e., B for bird population and H for human population).

Similar to the model of Lewis et al. [134], the manner in which the difference equation model is formed implies the following ordering of events:

1) Maturation, infection, and transfer between the states occur at the beginning of the time-step.
2) Death either natural or due to the disease occurs at the end of the time-step.

This means the events listed in (1) are conditioned upon surviving mortality in the previous time-step.

3.2.2 Weather and Landscape Impacts

The core difference equation model is further modified to take into account the impacts of daily weather (i.e., T and R) variations and landscape features (L). As a result of these variations in temperature and rainfall, reproduction per time-unit, r, hourly maturation (or development) probability, γ, and hourly probability of natural death of aquatic and adult mosquitoes, μ_a and μ_m, in mosquito dynamics, equations (3.8 – 3.10) update on a daily basis according to the weather functions proposed in [132]. The reproduction, r, depends also on the different classes of land use/cover present in a mosquito site/cell. As such, the daily update equations for these core parameters are as follows.

$$r = r(T, R, L) = \left[\alpha_b e^{-a_b\,(T-T_b)^2}\right]\left[\frac{(1+s_b)\,e^{-r_b(R-R_b)^2}}{e^{-r_b(R-R_b)^2}+s_b}\right][1+w_L L]$$

$$\gamma = \gamma(T, R) = \left[\alpha_d e^{-a_d\,(T-T_d)^2}\right]\left[\frac{(1+s_d)\,e^{-r_d(R-R_d)^2}}{e^{-r_d(R-R_d)^2}+s_d}\right]$$

$$\mu_a = \mu_a(T, R) = [c_a(T-T_a)^2 + d_a]\left[1+\frac{e_a R}{1+R}\right]$$

$$\mu_m = \mu_m(T) = [c_m(T-T_m)^2 + d_m]$$

where T and R are daily temperature and rainfall at time-step t. Temperature is truncated to be always positive. The landscape descriptor, L, evaluates the habitat suitability of a mosquito cell based on the land cover types present in the cell. Its value is calculated by a special linear combination of land cover correlations of a mosquito site, which will be explained shortly; and the parameter w_L basically indicates the importance of land cover of a mosquito site. The rest of parameters in this set of functions controls how the temperature and rainfall-dependent functions react to the variations of weather conditions. A discussion of these weather functions parameters can be found in [132]. For the scope of this thesis, it is only important to tune these parameters according to the weather and mosquito surveillance data in southern Manitoba. This calibration procedure is explained in section 3.3.

In addition to the function modifications above, the mosquito biting rate (i.e., number per unit of time) is also a function of temperature and host preference. The host preference is added to the model by multiplying a coefficient between zero and one to the mosquitoes' biting rate, as discussed in subsection 3.2.3. The effect of temperature on (hourly) biting rate is assumed to be similar to that for Malaria transmission. As such, the quadratic biting rate function used in [133] for Malaria transmission is employed as follows.

$$b = b(T) = \frac{1}{24}\lfloor -0.00014\,T^2 + 0.027\,T - 0.322\rfloor^+$$

where

$$\lfloor f(x) \rfloor^{+} = \begin{cases} f(x) & ; if\ f(x) > 0 \\ 0 & ; otherwise \end{cases}$$

The reproduction parameter, *r*, also depends on the different classes of land use/cover present in a mosquito site/cell (i.e., the parameter *L*). As shown in Figure 3-3, for each mosquito site, the proportional area of each land cover class is known. According to the Manitoba Land Initiative [118], there are 18 known and one unknown class of land cover identified in our land cover/use dataset. On the other hand, Bowden et al. [135] report Pearson correlations between 14 types of land cover and West Nile Virus disease incidence in humans in the United States. The detailed description of their land cover types can be found in [136]. The Pearson correlation values and mapping from their types of land cover to our classes of land cover can be found in Table 3-2. These correlation values combined with the proportional areas of each type of land cover in our mosquito sites defines the parameter *L* as follows

$$L = \sum_{i} l_i \times p_i$$

where l_i is the proportional area of different land cover classes, and p_i is the Pearson correlation for each class of land cover. For example, the value of *L* for the mosquito site/cell in Figure 3-3 is as follows.

$$\begin{aligned} L_{\text{Figure 3-3}} &= 0.2\, p_{Residential} + 0.33\, p_{AgriculturalField} + 0.15\, p_{Rocks} + 0.32\, p_{WaterBody} \\ &= 20\% \times (-0.19) + 33\% \times 0.3 + 15\% \times (-0.05) + 32\% \times 0 = 0.0535 \end{aligned}$$

In other words, the landscape descriptor, *L*, is defined as the weighted average of correlations between a mosquito cell and human WNV cases in the United States. The impact of this parameter (*L*) on mosquito reproduction is controlled by another weight parameter, denoted w_L.

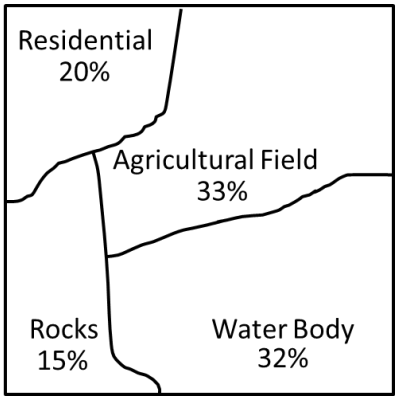


Figure 3-3 An example of land use/cover classes in a mosquito cell/site

Table 3-2 Adopted Pearson correlation values from [135] for land cover classes reported in [118] for Manitoba

Land cover class in [135]	**Manitoba's land cover class**	**Correlation**
Barren Land	Bare Rock/Gravel/Sand	-0.05
Deciduous forest	Deciduous forest	-0.51
Evergreen forest	Coniferous forest	-0.29
Mixed forest	Mixed-wood forest; Treed rock; Wildfire Areas; Forest cutovers	-0.32
Shrub (non-significant)	Open deciduous forest; Shrub	0
Grassland	Grassland/rangeland	0.46
Pasture	Forage crops	-0.28
Crops	Agricultural cropland	0.3
Woody wetland	Treed Bog	-0.19
Herbaceous wetland	Marsh; Fen	0.1
Ave. of all developed area	Cultural features; Roads/trails	-0.19
UNKNOWN	Water bodies; Lichen heath	0

3.2.3 Multiple Bird Species

Before describing how our implementation of the proposed CDiffE model differentiates between various bird species, a few facts and assumptions are reviewed:

1. For most species, susceptibility to infection is 100% [137], that is, we can assume that the transmission probability from an infected mosquito bite to a bird is 100%. To be clear, this virus transmission makes a bird infected, but not all of the infected birds become infectious.
2. There is no infected bird compartment in our proposed difference equation model, but only the infectious bird compartment (or state). This means a bird is marked as infectious only if the infected bird has the competency to transmit the virus to another mosquito vector. Otherwise, the infected bird remains a susceptible bird in the model.
3. Reservoir competence indices of an infected bird describe the relative proportion of vector mosquitoes that become infectious after feeding on such a bird [74], [137]. These indices help estimate the probability of transmission from bird to mosquito, given the contact rate. This is important in determining when an infected bird species become infectious to mosquitoes.
4. Given the way the difference equation model is currently formulated, it does not differentiate (i.e., per bird species) virus transmission probabilities from bird to mosquito. That is, β_m cannot be a vector or an array. This is discussed in 3.5.2.
5. For the estimation of the mosquito contact rate with each bird species, host-seeking mosquitoes' abundance data and dynamic WNV data are necessary. Unfortunately, we did not have access to these data from Manitoba. Additionally, host preference of *Cx.*

tarsalis mosquitoes depends on the species available in a region [138], i.e., *Cx. tarsalis* mosquitoes have opportunistic host-seeking habits [72], [139].

Considering items 1 – 4, the virus transmission probability from bird to mosquito (i.e., β_m) is the same for all bird species, but the virus transmission probability from mosquito to bird (i.e., β_b) is adjusted based on the reservoir competence index for each bird species. That is:

$$\begin{cases} \forall_j \, \beta_m^j = \beta_m \\ \beta_b^j = c^j \beta_b \end{cases}$$

where c^j is the host competence index for the bird species j. A host competence index is a metric to indicate relative number of mosquitoes which become infectious as a result of biting the infected host [74], [137]. Collecting these indices and other bird species' parameters is explained in subsection 3.2.3.1.

Given item 5, and the tendency to lower the computational complexity of the model, in our implementation, the biting rate (i.e., the number of bites made per mosquito per time step) is the same among different bird species; and mosquito's biting rate for humans is a proportion of birds' biting rate. That is:

$$\begin{cases} \forall_j \, b^j = b \\ b_h = \frac{1}{\lambda} b \end{cases}$$

where λ is the ratio of mosquito biting activities on birds to biting activities on humans.

The next two parameters to calculate for bird species are recovery, ζ_b, and death due to infection, δ, probabilities per unit of time. The procedure to estimate these two probabilities and associated rates are explained in subsection 3.2.3.2.

3.2.3.1 Bird Species' Competence Index

Host competence indices are calculated from species viremia level reported in [74], [137], [148], [149], [140]–[147]. Indices are then normalized to be between zero and one. In the difference equation model, these normalized indices are used as weights to adjust the virus transmission probability from mosquito to bird (i.e., β_b) for each species. The exact reasoning and formula to derive host competence indices can be found in [74]. Basically, if the host viremia is less than an infectious threshold in a day, its infectiousness is set at zero for the day. If not, the infectiousness is calculated. The average daily infectiousness of birds of the same species is then summed over the viremic period to produce the species index [74]. The infectiousness threshold used in these studies is for *Cx. pipiens* mosquitoes, and in general *Cx. tarsalis* mosquitoes are competent at a lower viremia level. However, the threshold is not as important as the fraction of infectious mosquitoes used in the calculation of species index [74]. Besides, these competence indices are normalized and used as relative weights. As such, the resulting effect on the difference equation model will be almost the same. For bird species that had not been tested, the average values from other species of the same family or order were used. Strictly speaking, for most of missing species, the order average was used except for a few Passeriformes where some data for the family was available. Normalized host competence indices for some of the bird species can be found in Table 3-3 on page 69.

Table 3-3 Birds' species estimated WNV parameters

Common name	**c***	p_d*	p_r*	n_d*	n_r*	$24\theta_d$*	δ*	$24\theta_r$*	ζ_b*
Blue Jay	0.7565	75%	25%	4.7	4	0.295	0.01221	0.072	0.00299
American Crow	0.5931	100%	0%	5.1	N/A	0.196	0.00814	0	0
Common Grackle	0.4460	33%	67%	4.5	4	0.089	0.00370	0.277	0.01148
House Finch	0.4057	100%	0%	7	N/A	0.143	0.00593	0	0
House Sparrow	0.3917	50%	50%	4.7	5	0.147	0.00613	0.139	0.00576
Black-billed Magpie	0.3563	100%	0%	6	N/A	0.167	0.00692	0	0
American Robin	0.3287	-	-	-	3	0.112	0.00466	0.333	0.01379
Song Sparrow	0.3084	-	-	-	5	0.112	0.00466	0.200	0.00830
American Kestrel	0.2477	-	-	-	3	0	0	0.333	0.01379
Brewer's Blackbird	0.2177	-	-	-	4	0.044	0.00185	0.250	0.01036
Great Horned Owl	0.2133	-	-	-	4	0	0	0.250	0.01036
Red-tailed Hawk	0.1623	-	-	-	5	0	0	0.200	0.00830
Red-winged Blackbird	0.1349	-	-	-	3	0.044	0.00185	0.333	0.01379
Northern Mockingbird	0.0934	-	-	-	2	0.112	0.00466	0.500	0.02062
European Starling	0.0541	-	-	-	4	0.112	0.00466	0.250	0.01036
Northern Flicker	0.0452	-	-	-	2	0	0	0.500	0.02062
Swainson's Thrush	0.0376	-	-	-	1	0.112	0.00466	1	0.04081
Mourning Dove	0.0285	-	-	-	3	0	0	0.333	0.01379
Gray Catbird	0.0248	-	-	-	1	0.112	0.00466	1	0.04081
Rock Pigeon	0.0013	-	-	-	1	0	0	1	0.04081
Brown-headed Cowbird	0	-	-	-	1	0	0	1	0.04081
Ring-necked Pheasant	0	-	-	-	1	0	0	1	0.04081

* *c: relative competence index*

p_d : percent of birds who died due to infection

p_r : percent of birds who recovered

n_d : mean number of days till death due to infection occured

n_r : number of infectiosness days

$24\theta_d$: daily death rate

δ : hourly death probability

$24\theta_r$: daily recovery rate

ζ_b : hourly recovery probability

3.2.3.2 Bird Species' Recovery and Death Rates and Probabilities

To begin with, assume the experimental studies are available for a species. In this case, for each species, the following information is obtained [74], [137]: the overall fraction of birds who died, (p_d), the overall fraction of birds who recovered, (p_r), the average number of days to death, (n_d), and the average number of days to recover from being infectious, (n_r). From a statistical perspective, the average probability of death is p_d for a period of n_d days, and the probability of recovery in an interval of n_r days is p_r on average.

Recall that the difference equation model is formed such that the events of death due to WNV infection and recovery from WNV infection are two independent events (see equation (3.13) or (3.6)). This means that infectious birds remain infectious if they do not die and at the same time do not recover. Therefore, in estimating each of these two hourly probabilities, the impact of the other is disregarded. As such, a Poisson model with rate parameter θ is fitted to the problem in order to determine the probability of the number of death events (or recovery events) that occurred per unit of time. Let k_h be the number of events in an interval of h hours (or time-steps). Thus, the probability of n events occurring in h hours is given by the probability mass function (PMF) of a Poisson distribution as follows.

$$P(k_h = n) = Poisson(k_h; \theta h) = \frac{(\theta h)^{k_h} e^{-\theta h}}{k_h!} \tag{3.22}$$

where θ is the hourly rate (i.e., expected number of events per time-step), and (θh) is the average number of events in h hours.

Consequently, the probability of observing at least one event (e.g., death or recovery) per unit of time (i.e., h = 1) is given by

$$P(k_1 > 0) = 1 - P(k_1 = 0) = 1 - e^{-\theta} \tag{3.23}$$

Now the information regarding each species from the literature can be applied to solve the equation (3.22) for θ, and the hourly probability is then given by equation (3.23). That is, if θ_d represents the hourly rate of dying due to infection, the probability of *not* dying due to infection in n_d days (i.e., $24n_d$ hours) is

$$P\left(k_{24n_d} = 0\right) = 1 - p_d = e^{-\theta_d(24n_d)}$$

where $0 \leq p_d < 1$ and $n_d > 0$ can be taken from the literature in [137]. This yields

$$\theta_d = -\frac{\ln(1-p_d)}{24n_d} \tag{3.24}$$

As such, the hourly probability of dying due to infection, δ, and leaving the infectious state can be derived from the corresponding rate as

$$\delta = 1 - e^{-\theta_d} \tag{3.25}$$

Similarly, if θ_r represents the hourly rate of recovery from infection, this rate will be calculated as

$$\theta_r = -\frac{\ln(1-p_r)}{24n_r} \tag{3.26}$$

where $0 \leq p_r < 1$ and $n_r > 0$ can be taken from the literature in [74], [137]; and consequently, the hourly probability of recovery, ζ_b, and leaving the infectious state, is derived from the corresponding rate as follows

$$\zeta_b = 1 - e^{-\theta_r} \tag{3.27}$$

For the corner case where $p_d = 1$ (and $p_r = 0$), i.e., all birds die due to infection (and no one recovers), the hourly recovery probability, ζ_b, is simply set to zero, but the hourly death probability, δ, is *not* set to one, thus avoiding changing the infectious birds' state in only one time-step as desired. Therefore, in this case, instead of using the equation (3.24), the hourly *rate* of dying due to infection, θ_d, in equation (3.25) is deliberately set to $1 / 24n_d$. That is

$$\delta = 1 - e^{-\frac{1}{24n_d}} \quad ; \zeta_b = 0 \tag{3.28}$$

Similarly, for the corner case where $p_r = 1$ (and $p_d = 0$), i.e., when all birds recover from infectiousness (and no one dies), the hourly death probability, δ, is simply set to zero, but the hourly recovery probability, ζ_b, is not set to one. Herein, the recovery probability, ζ_b, is given by equation (3.27) where the recovery *rate* of θ_r takes the value of $1 / 24n_r$ instead of taking its value from the equation (3.26). That is

$$\zeta_b = 1 - e^{-\frac{1}{24n_r}} \quad ; \delta = 0 \tag{3.29}$$

In summary, where the experimental studies are available for a species, death and recovery rates are estimated, and the corresponding probabilities are then calculated according to equations (3.24) - (3.27). Herein, the parameter n_r for species is set to the number of days that the species maintain a high level of viremia in their bloodstream to be infectious for mosquitoes. These values are taken from the data used in [74]. The other three parameters of p_d, p_r,and n_d are taken directly from [137] for a limited number of species. These values for a number of bird species can be found in Table 3-3. Where the experimental data are missing for a species, estimation of death and recovery rates for each species were made based on the following rules of thumbs and assumptions.

1. For bird species with no reported experimental studies, the average values from other species of the same family or order were used when there were some data available regarding at least one species in these orders/families.
2. Orders and families reported as incompetent in [150] were given a zero death rate, and one recovery rate; unless there was explicitly some data reported regarding at least one species in these orders/families. This assumption means incompetent birds do not die due to infection and will recover from infectiousness in one time-step.

3. When “no signs of clinical illness” were explicitly reported as in [137] for a certain order, a zero death rate was given to all their species, unless there were explicitly some data reported regarding at least one species in this order. This assumption means that if birds of a certain order do not show any symptoms of the disease, they will not die due to infection.
4. In all other cases where there was no general information available regarding a certain order, the recovery rate of a species was set to the average recovery rate of all birds with known data. The death rate of species of these orders was assumed to be zero. This means birds of high WNV mortality such as passerines (and especially corvids) are assumed to be already identified and studied by biologist. Therefore, birds with no experimental studies must have a negligible (or even zero) death rate. Even though these species may have been found dead while their bloodstream had an infection, yet one cannot confirm whether the death was a direct result of infection without proper experimental studies.

Derived hourly rates and probabilities of recovery from infectiousness and death due to infection for a number of bird species can be found in Table 3-3. It is notable that a natural death is not possible for birds as the model is seasonal (i.e., it runs only for one spring and summer per simulation year). However, the data-driven weekly variation in bird population is accounted for in the proposed CDiffE model.

3.2.4 Mosquito Nocturnal Activities

The majority of host-seeking activities of *Culex* mosquito occurs at night [151], [152]. Therefore, an average hourly biting probability should at least be split into high and low biting activity values for nights and days, respectively. Assuming that the parameter $\bar{b}$ indicates the average hourly probability of biting a certain species (e.g., humans), the following equations

derives the mosquito biting probabilities for night and day based on daylight hours, *D*, in a given day and ratio of biting activities between night and day, ω, such that the average hourly value holds true for a 24-hour day.

$$\begin{cases} \frac{b_{night}}{b_{day}} = \omega > 1 \\ \sum b_{night} + b_{day} = 24\bar{b} \end{cases} \quad \Rightarrow \quad D b_{day} + (24 - D)\omega b_{day} = 24\bar{b}$$

$$\Rightarrow \quad \begin{cases} b_{day} = \frac{24}{(24 - D)\omega + D}\bar{b} \\ b_{night} = \omega b_{day} \end{cases}$$

where b_{day} and b_{night} are mosquito biting probabilities during days and nights, respectively. This means these two parameters would replace the parameter b in the equations (3.11) and (3.15) (and the parameter b_h in equation (3.21)) based on the current hour of the day.

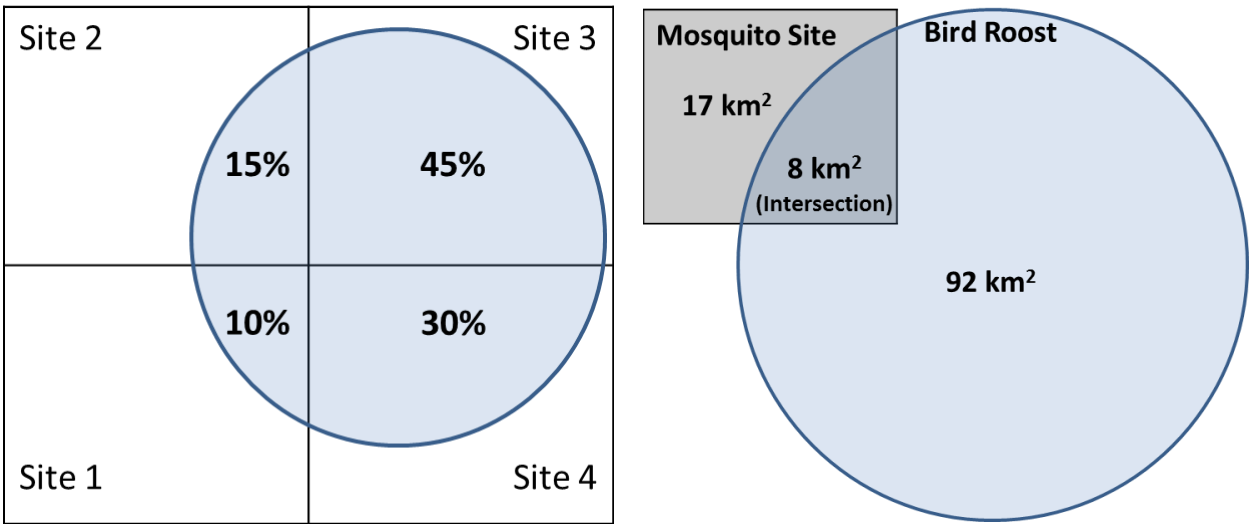
3.2.5 Avian Flow

In WNV epidemiology, the movement of birds is not as critical to the model as their roosting activities. The reason is that the majority of birds get bitten and infected while they roost. As such, spotting a bird roost is more influential than modelling bird (long-distance) movement patterns. Yet, short flights to surrounding areas in search of food are the primary way of spreading the infection. Consequently, the bird agents and avian flow of the system are designed based on this concept. Generally, the breeding seasons of the birds in the system are within the span of the simulation time (i.e., spring/summer). Given this, and the tendency to simplify the computational complexity of the avian flow, birds were assumed to remain close to their nesting locations at night, and fly around during the day within their home-range but still close to their nests.

A typical bird roost is covered by a number of mosquito site agents. After each time-step, some birds of a roost may become infectious as a result of contact with different mosquito (site) agents. Therefore, at the beginning of the next step, there is a probability that a bird that became infectious from the mosquito site X would spread the infection to the mosquito site Y, as long as both mosquito sites X and Y reside within the home range of the bird. Basically, at each time-step, the total population of birds of a roost is split proportionally among all the mosquito sites covered by the birds *at the time*, assuming a homogenous and well-mixed population of bird agents. For example, consider the case when a bird roost covers four mosquito sites called one, two, three, and four such that they are covering 10%, 15%, 45%, and 30% of the total area of the roost, respectively as shown in Figure 3-4a. So, if the total population of birds in this roost is 1000 and of which 100 are infectious, then 100 birds (including 10 infectious) will be present over the mosquito site one, 150 birds (including 15 infectious) will be present over the mosquito site two, 450 birds (including 45 infectious) will be present over the mosquito site three, and 300 birds (including 30 infectious) will be present in the mosquito site four. From the difference equation perspective, this assumption is inferred as multiplying all the bird state variables (B_s, B_i, B_r, and B_t) for a given mosquito site by a factor proportional to the intersection area of the roost and the mosquito site. Strictly speaking, this factor is defined as the proportion of intersection area to the roost area i.e., 8/100 for the roost/site pair in Figure 3-4b.

Moreover, unlike the classical assumption of difference and differential equation models where every vector (mosquito) could bite and infect every host (birds), it is considered here that only a certain proportion of mosquitoes in a site could bite birds of a certain roost provided that the roost is *not* completely covered by the mosquito site. For example, consider the case where a mosquito site with an area of 25 km^2 has an intersection area of 8 km^2 with a roost with an area

of 100 km^2 as shown in Figure 3-4b. Under classical DE models, all the mosquitoes scattered in the 25 km^2 region could bite the birds of the roost present in the 8 km^2 intersection region. However, in the proposed CDiffE model, only mosquitoes present in the 8 km^2 intersection region could bite the birds of the roost there. From the difference equation perspective, this assumption is implemented by multiplying the contact rate, b^j, in equations (11) and (15) by a factor proportional to the intersection area of the roost and the mosquito site. Strictly speaking, this factor is defined as the proportion of intersection area to the mosquito site area i.e., 8/25 for the roost/site pair in Figure 3-4b.



a) Blue circular roost covered by four mosquito sites; bird roost population are distributed proportionally among the mosquito sites

b) A Mosquito site / bird roost pair with an overlapping area; 8/25 of the mosquitoes may bite 8/100 of the birds.

Figure 3-4 Examples of mosquito sites and bird roosts relative position to each other

Each species of bird has different home ranges and may fly up to a certain maximum distance for their food seeking activities. While the average number of birds present in the simulation per mosquito cells can be calculated, the actual number of birds per each mosquito

cell is quite different at each hour. At the beginning of the simulation, the figure is over 11,000. It then goes up with a few fluctuations during the spring and summer until the last weeks of simulations where it gradually decreases to an approximate value of 8,000.

3.3 Parameter Calibration and Model Validation

The obtained Centers for Disease Control (CDC) carbon dioxide baited trap data in Manitoba [20] does not provide sufficient information regarding the number of infected mosquitoes for validation and tuning purposes, as the data only indicate if there is at least one infected mosquito in a given week in a community, and they do not provide any information about the number of infected pools. That is, the trap data are not reliable for assessing infection counts, but they are appropriate for total mosquito count. Therefore, the parameter calibration is achieved by comparing the total mosquito count in the model against the trap data. The trap data are the number of mosquitoes per collection night, which can be inferred as the average weekly count of total mosquitoes. These data were available from CDC week 22 to CDC week 36 (a total of 15 weeks) for approximately 30 communities in Manitoba from 2003 to 2014 [20]. As the data from 2003 is incomplete and not consistent with the other years, it was disregarded. The data from 2004 to 2013 are used as training data for calibration and tuning purposes, and the data from 2014 is then used for validation as the test data to assure the model is not over-fitted. Alternatively, one could have performed a k-fold cross-validation on the limited available dataset by partitioning the data into different training and test years and averaging the results over all the rounds to obtain an estimate of model generalization error. This is often carried out in statistical prediction models, particularly for selecting hyper-parameters when few data are available, and it would have been especially crucial if the intent was to report on an accuracy metric for a finalized software package for WNV prediction in Manitoba for all years. However,

the calibration process was very time-consuming in this work. As well, the goal here was to achieve acceptable behaviour to proceed with the CDiffE model for a higher-level ABM to study various scenarios. After performing the calibration procedure, either the test year or the average of all years was used for further studies of other phenomena of interest.

In the model, the count of total adult mosquitoes from all mosquito sites of each community is recorded only at mid-night in a single time-step. Although recording the mosquito counts during the whole night was possible, tallying each mosquito only once was desired to avoid overestimating the nightly population counts. This is conceptually similar to having mosquito traps run for an hour each night. This figure is then divided by the number of days in a week and the total area of mosquito sites in each community, so that the average nightly count/density per one km^2 is obtained for each community each week. For calibration purposes, the output for each simulated year is defined as a matrix of these average weekly mosquito counts. The matrix has around 30 rows (for communities) and 15 columns (for weeks). The correlation between this simulation output matrix and the trap data matrix are then calculated for each year. The average correlation between simulation output and trap data for years 2004 to 2013 is used as the fitness function of the optimization algorithm. While the optimization algorithm could be any metaheuristics search algorithm, the Grouped Bees Algorithm [153] in conjunction with the OptQuest optimization engine [154] is employed in this work.

The correlation between the simulation output and the trap data indicates how well the model predicts weekly trends of mosquito population counts in different locations/communities. It also captures the variations in mosquito distributions over different locations/communities in a certain week.

There are two general choices for correlation calculations: Pearson and Spearman rank order, each with their own benefits and drawbacks. The Pearson correlation compares the absolute numbers. It is more sensitive to outliers, such as wrong trap data in this application, yet it uses most of the information available from the data. It is notable that other than many missing data, it was confirmed that some traps may have failed for various reasons for any given week. This failure information was not, however, available in the obtained data. Alternatively, Spearman rank order correlation does not need such accurate data, as it only considers the ranks (orders) of numbers. For example, as long as the model's output for week two is higher than week one and less than week three, and this is also the case in the trap data, the correlation is 100%. The Spearman rank correlation is only good to achieve the relative trend, and not the amplitude of changes. Therefore, the data across different communities for every single week were summed up and treated as the total count for the province for each week. This figure is compared against the total density of mosquitoes in those communities in the model for each week to be used in the sensitive Pearson correlation; whereas the Spearman rank correlation was applied to compare every single pair of weeks and communities. The next step was to decide on how to combine these correlation metrics, as it was not known which correlation choice would have produced a better fit for the parameters. After a few weeks of preliminary calibration attempts, a linear combination of both metrics with equal weights was chosen as the final fitness function.

Unexpectedly, while the calibration/optimization process was running, it was consistently observed that the number of mosquitoes produced by the simulations is significantly lower than what data suggested, by several orders of magnitude. This is primarily due to large and complex input search space, lengthy simulation run-times, and some input parameters with high

sensitivity. Sensitivity analysis on the temperature and weather-related parameters can be found in [132]. This pre-mature convergence likely happens because the optimization process tends to fix some parameters near the boundaries of input intervals as constants while adjusting other parameters, and the fact that boundary values produce fewer mosquitoes than expected. Additionally, in many weeks, there should be few (i.e., almost zero) mosquitoes according to the trap data, so the optimization is likely trying to produce a low number of mosquitoes in general. Thus, the fitness function was altered to apply penalties to the solutions where the total density of mosquitoes generated by the simulation differs from the total number of mosquito trap counts by several powers of Euler's number, *e*, in each year. To apply this penalty, the trap data had to be scaled up to obtain a more accurate estimate of the actual number of mosquitoes in the field. However, the ratio of the number of captured mosquito to the actual population of mosquito varies by region and month [155]. The possible influence of weather conditions on trap effectiveness is also very likely to be strong. Based on the experiments in [155], it may be inferred that a captured *Cx. tarsalis* mosquito represents around 300 mosquitoes over one km^2 in August. Consequently, a simplifying assumption was made that the scaling factor of 300 remains the same in the entire simulation time.

Moreover, the optimization process produced some solutions with excellent performance in some years, but very poor matching trends in other years. Therefore, a minor penalty was applied to the solutions where the correlation was negative in a year. This penalty was imposed to achieve a general solution for producing a fair result in all years from 2004 to 2013. In the end, the calibration process resulted in acceptable trends of mosquito population dynamics for both test and training data. Detailed results of the calibration procedure can be found in following section.

3.4 Results and Discussion

3.4.1 Calibration Results

For the scope of this chapter, the human component is not yet added to the simulations implementations. This assumes H_t is zero in all corresponding equations. Depending on how the weights of correlation metrics are set and penalties in the fitness function assigned, the calibration procedures can end up with very different set of solutions. Here, the solutions for three general cases of correlation-oriented, population-oriented, and balanced-approach are presented as examples. The difference in these cases is the weight of population penalty used in their fitness function during the optimization process. Three evaluation metrics of Spearman's rank-order correlation, Pearson correlation and logarithmic-scale population ratio for each solution is given in Table 3-4, where the logarithmic-scale population ratio is defined as

$$\text{Pop. Ratio} = \frac{\ln(P_s)}{\ln(P_t \times 300)}$$

where P_t is the total number of weekly captured mosquitoes from the trap data for each year and P_s is the sum of average weekly density of mosquitoes per one km^2 in selected communities (i.e., corresponding to the trap data communities) generated by the simulation for each year. The values of equation parameters for each solution are given in Table 3-5 on page 84.

According to Table 3-4, the correlation-oriented solutions have higher correlations (both Spearman and Pearson) with an average of approximately 60% in all years (including both training and test years). The Population-oriented solutions have the lowest correlation values but the highest population ratio, with an average of approximately 73% in all years. The balanced-approach solutions fall in between, with an average of approximately 52% correlation and 68% population ratio.

Table 3-4 Calibration results for three different sets of solutions

	Correlation-oriented			**Population-oriented**			**Balanced-approach**		
	Pearson	Spearman	Pop. Ratio	Pearson	Spearman	Pop. Ratio	Pearson	Spearman	Pop. Ratio
Training years									
2004	62%	60%	54%	76%	65%	65%	73%	66%	61%
2005	90%	71%	54%	68%	65%	70%	86%	67%	61%
2006	61%	39%	64%	22%	22%	79%	38%	29%	72%
2007	44%	52%	53%	4%	15%	70%	24%	29%	62%
2008	86%	70%	53%	63%	69%	64%	78%	70%	60%
2009	18%	32%	64%	23%	28%	76%	23%	27%	70%
2010	77%	71%	63%	28%	56%	78%	55%	63%	69%
2011	79%	58%	63%	54%	55%	75%	66%	56%	70%
2012	73%	52%	62%	39%	42%	76%	56%	45%	69%
2013	35%	45%	66%	12%	50%	79%	28%	43%	72%
Average	**62%**	**55%**	**59%**	**39%**	**47%**	**73%**	**53%**	**50%**	**68%**
Test year									
2014	73%	64%	59%	51%	61%	74%	62%	60%	68%

Generally, the correlation for some years with a mild summer (i.e., 2009 and 2013) and the year with the highest number of mosquitoes in the traps (i.e., 2007) have the lowest values. A correlation value of 50% to 60% for weekly variations over various regions may not be considered a high value. However, in a WNV study on birds infection rates in Ontario, Canada [156], the Pearson correlation between observed and predicted values for monthly variations in Ontario (i.e., only a single location) fall between 60% to 63% for training and test data using an artificial neural network technique. It is notable that neural networks are theorized to be

universal approximators with acceptable accuracy. Although direct numerical comparison between the two models is not necessarily meaningful, there are conceptual links between the two. Generally, a high number of mosquitoes is correlated with high prevalence of infection in both mosquitoes and birds, assuming the presence of the virus. However, this should not be taken to imply that neural networks cannot achieve greater predictive power for WNV epidemiology. The value of the proposed CDiffE scheme, and in particular agent-based modelling, in this context is in construction of a virtual simulation framework in order to assess many different influential factors in WNV epidemiology before putting them in practice. Very specifically, it is important to emphasize that a mechanistic approach such as that explored here is able to look at the effect of counterfactual interventions in a way that the neural network approach is not able to. This reflects the fact the neural network is trained to the current data generating process, but counterfactual interventions alter that data generating process. ABMs also allow for improving the underlying theory over time, as learning is made concerning accuracy of observations in different areas (e.g., along different generative pathways) in ways that neural network (and statistical models in general) do not. Such simulators will help to elucidate the underlying mechanism of WNV propagation. Generally transparent simulation methods such as ABM provide insight into the subject of study in comparison with the black box approach offered by soft computing techniques such as neural networks. Further discussion of the table is presented in section 3.4.2.

Table 3-5 Numerical values of equation parameters for each of the three solutions of calibration process

	Correlation-oriented	**Population-oriented**	**Balanced-approach**
α_b	12	12	12
a_b	1E-5	1E-06	3.866E-6
T_b	22	25	24
s_b	2	1.799	1.819
r_b	0.006	0.008	0.009
r_d	0.004	0.006	0.006
T_d	23	26	25
R_b	20	20	17
a_d	0.04	0.04	0.029
α_d	1E-5	6.36E-05	1.671E-5
s_d	2.245	3.655	3.774
R_d	22	18	17
c_a	1E-10	1E-10	1.019E-10
d_a	1.36E-4	6.64E-05	1.688E-4
T_a	27	28	25
e_a	0.001	0.0001	0.001
c_m	1.474E-4	3.79E-05	4.587E-5
T_m	25	28	28
d_m	2.725E-4	0.001	9.919E-4
w_L	0.1	0.1	0.1

All the simulation runs were given an initial (small) number of 2000 × (1 + w_L L) mosquitoes in each mosquito site in the CDC week 21 (i.e., approximately the last week of April). It is notable that the parameter α_b and w_L are the same in all solutions as the model is highly sensitive to changes of these two parameters (see [132] for a discussion on the sensitivity of model to the reproduction rate). Accounting for the impact of w_L, an α_b value of 12 is associated to a maximum of 288 eggs, laid by each adult female mosquito during a day, which is in agreement with biological studies [157]. The second fixed parameter of w_L defines the impact of landscape features on mosquito habitat. While it may have been expected to have a high value of close to one for w_L, its value was set at 0.1 by the optimizer. The reason can be explained via Figure 3-5 where the parameter L (i.e., land use correlation with the number of WNV human cases) for southern Manitoba is depicted, and the dark blue color denotes a non-positive correlation.

As shown in Figure 3-5, land use in approximately the half of southern Manitoba (i.e., the area of interest of this study) has a negative correlation with WNV human cases in comparison to the whole North America. As such, high values of w_L could have significantly reduced the total number of mosquitoes generated in many grid cells, which is *not* desired and is penalized in the fitness function of the optimization process. So, a low positive value of 0.1 was selected (by the optimizer) as w_L to avoid underproduction of mosquitoes and to achieve a high Pearson correlation at the same time. However, if in a specific application where the population census is not as important as having a higher Pearson correlation, the value of w_L could be carefully increased. Once the value of w_L is updated, an optimizing search for tuning other parameters is, however, recommended. A report on the variations of w_L is given in subsection 3.4.4.

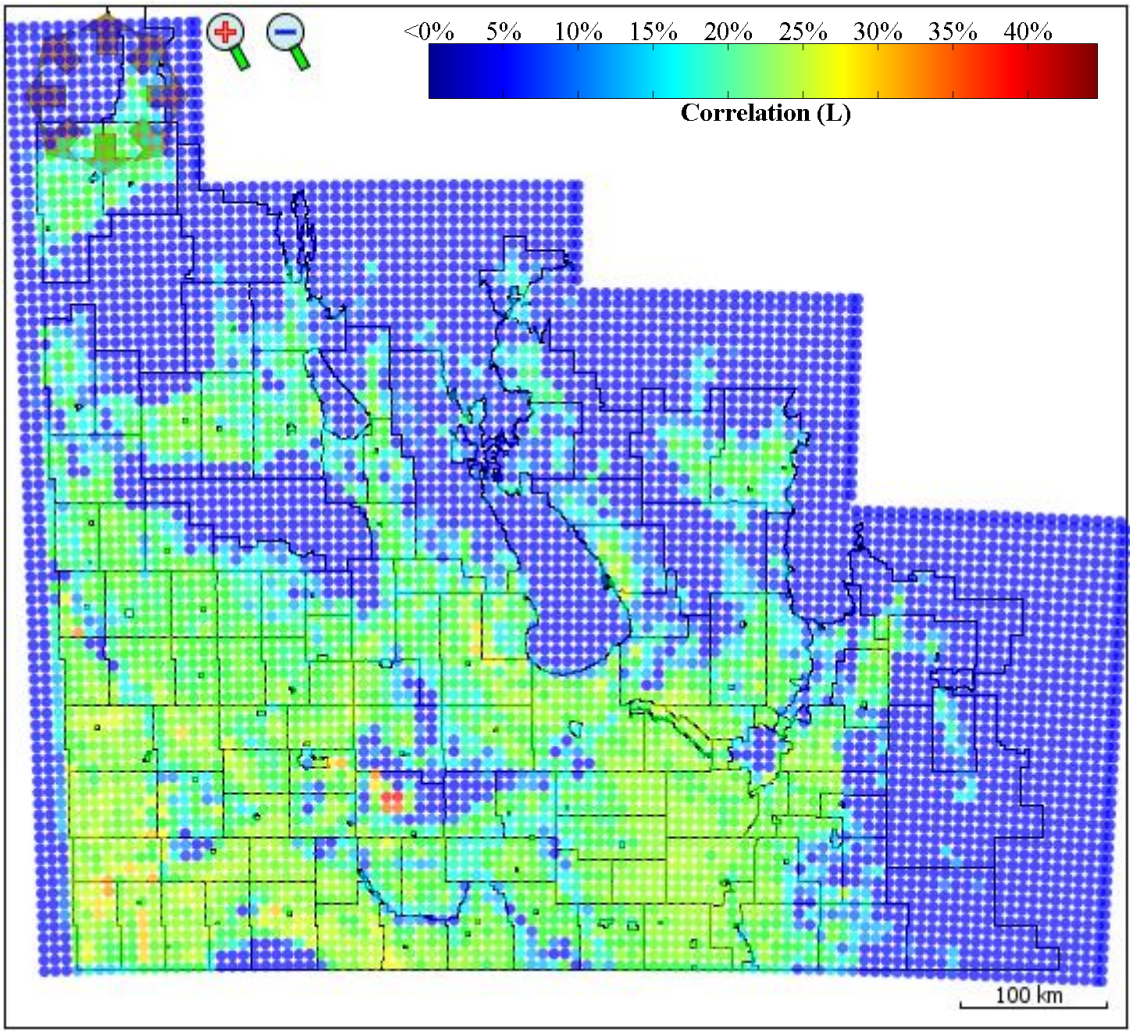


Figure 3-5 Weighted average land cover correlation with the number of WNV human cases, i.e., the parameter L

3.4.2 Model Validation

Overall, according to Table 3-4, the model has an acceptable predictive power of approximately 60% in terms of correlation between the trap data and simulation output for mosquito population dynamics. To provide a better visual insight into these quantities, two more assessing figures are presented. First, Figure 3-6 compares the scaled weekly trend of total trapped mosquitoes from the data against the weekly mosquito densities generated by simulation using the balanced-approach solution. The figure is depicted for all the years (i.e., training and

test) from 2004 to 2014. In most years, the trend of mosquito population dynamics in simulation output is close to the total number of mosquitoes collected weekly in Manitoba traps.

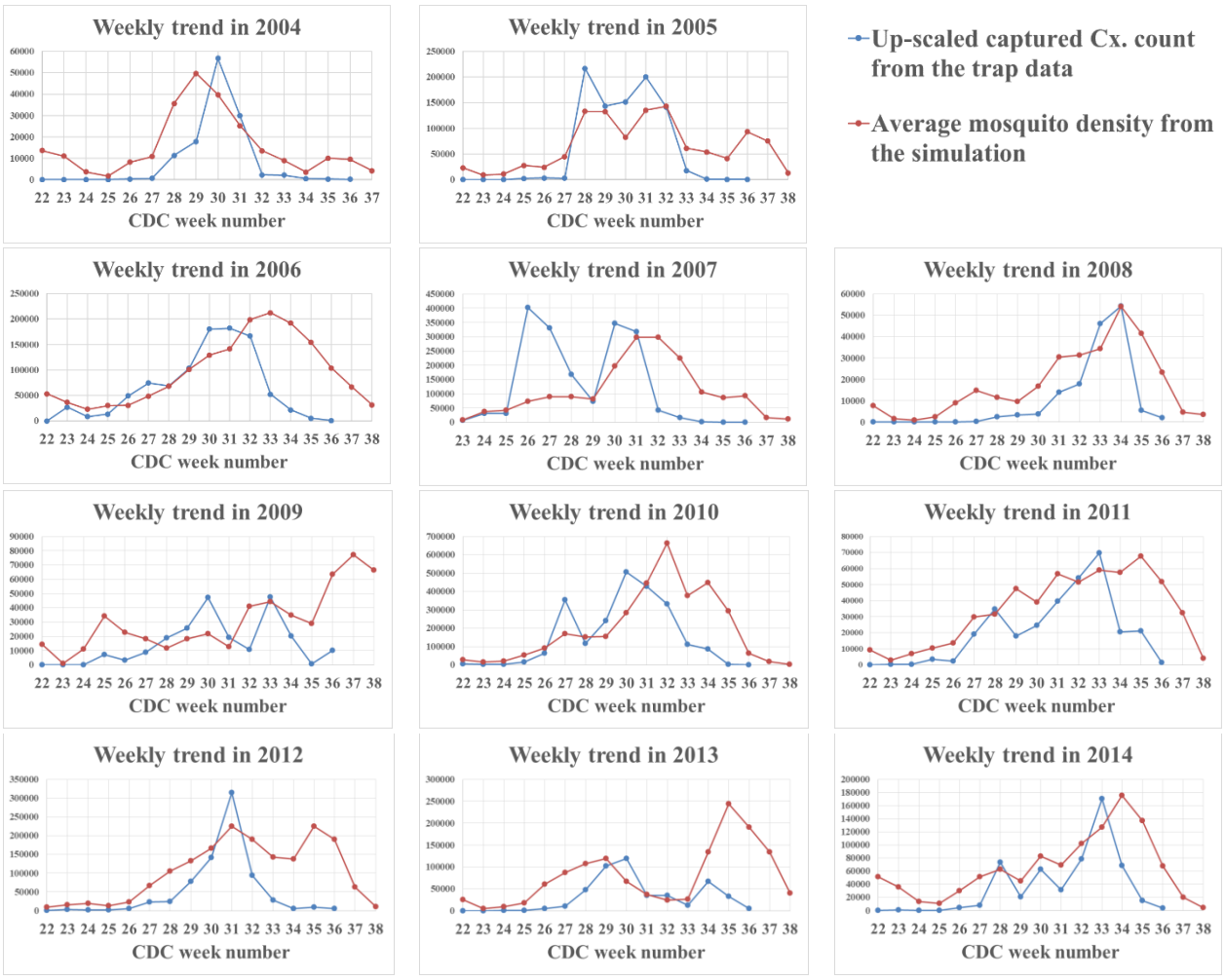


Figure 3-6 Weekly trends of sum of scaled trap data versus simulation's weekly mosquito densities for 2004 to 2014

The worst performance is for years 2009 and 2007. 2007 is when the highest number of WNV human cases and captured mosquito counts were observed in Manitoba. For some reasons, in 2007, a primary peak of mosquito (host-seeking) activity was observed in week 26 whereas the simulation does not capture any unusual increases in mosquito population. This might be due to unexpected changes in bird populations. In current simulations, bird populations are assumed

to have the same migration patterns and roosting locations in all of the simulated years. It is notable that trap data provides information on female host-seeking mosquitoes, not actually the count for the entire mosquito population. This means if there are not many hosts to feed on, there may not be many host-seeking mosquitoes, despite having a highly mosquito-populated area. On the other hand, if there are no real available hosts, then the traps, as they mimic a possible host by releasing CO2, may even capture more mosquitoes.

In the selected communities, years 2009 and 2013 with 2004 are among the years with the coolest/mildest summers. They have the lowest number of days with an average (per-community) daily temperature of greater than 17 °C. Also, this temperature value never reaches 24 °C in 2009 and 2004. Yet, as shown in Table 3-4, the model has an acceptable predictive power in 2004, which caused us to re-examine the previous suggestion that the model accuracy deteriorated when summers were mild. So, we further analysed the weather conditions in 2009 and 2013 where we initially believed the mild temperature made the model inaccurate. Upon further investigation of weather conditions in 2009 and 2013, we realized that there are days where both temperature and rainfall rise to values, respectively, above 20 °C and 10 mm at the same time. As a result, in 2009 and 2013, the model fails to fully take into account the impact of heavy rainfall in these days. That is, the model fails to kill a large proportion of larvae in these situations due to flushing, as is the case in reality (evidenced by the trap data). This suggests, in general, that the model has learned to be highly sensitive to temperature (in developing more mosquitoes) but not sensitive enough to rainfall to determine the exact threshold of rainfall values to flush a correct amount of larvae. This is partially due to the data we had, which biased the model to give more weight to temperature, as this was the safest approach for the model to

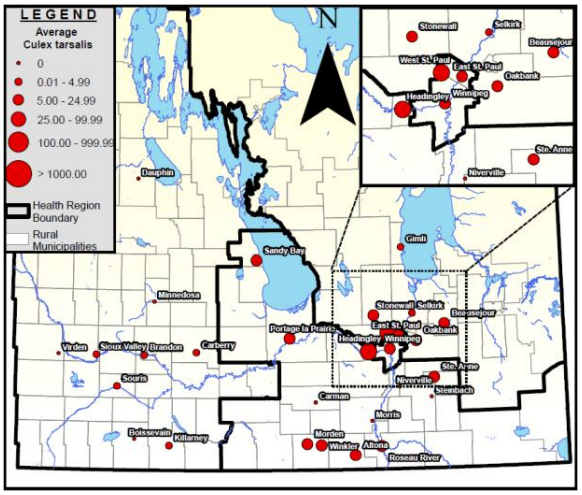
gain the maximum average fitness in all years. Arguably, there needs to be also a consideration of the impact of rainfall on the probability of capturing a given mosquito.

Surprisingly, it was observed that the same conditions occurred other times, and the model prediction suffered in those weeks as well. However, the impacts were strong enough in some years to significantly drop the correlation values. Generally, when rain of around 10 mm to 25 mm with temperature of occasionally 20°C (or slightly above it) occurs in the same week, the model fails to flush a sufficient number of larvae. Therefore, a visible increase (or maybe a jump depending on temperature) in the number of mosquitoes in the model is often observed when a drop or only a slight increase is expected. Examples of such co-incident events include week 27 in 2005, week 32 in 2006 and 2007, weeks 25 and 32 in 2009, week 32 in 2010, weeks 26 and 34 in 2013, and week 34 in 2014. Average (per-community) daily temperature and rainfall values for a number of years are shown in the figures in Appendix B.

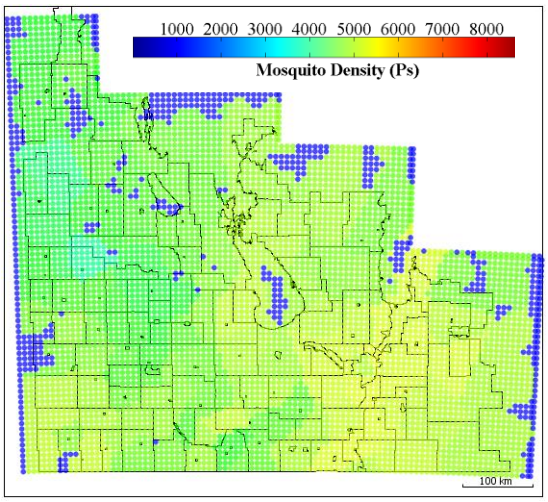
As can be seen, in general, in the last weeks of simulation, the system has more mosquitoes than the trap data suggests. This is most likely due to diapause, when mosquitoes become dormant towards the fall based on the changing daylight duration. Put differently, in those relatively cool weeks, there are still mosquitoes in the area, but they are not active enough to get captured at the traps. Incorporating the diapause process would be particularly important to obtain a better fit of mosquito population, if the simulation had to capture whole year dynamics. Moreover, always in week 22, the simulation has more mosquitoes than the trap data, which likely is because of the initial simulation conditions.

For further visual comparison, Figure 3-7 compares the mosquito geographic distribution in a number of weeks between the simulation and trap data in Manitoba. For each week, in Figure 3-7, from left to right, two images are shown: (1) trap data of the week on the left (image source: [20]), (2) mosquito density generated by the simulation during the same week on the right. It is immediately clear from the screenshots in Figure 3-7 that the model output correctly identifies the mosquito patterns at the province-scale. In addition, Figure 3-8 shows degree day accumulation base 14.3 °C [20], [158] for further visual comparison. The degree day function used in Figure 3-8 is derived from [159]. The simulation screenshots for this experiment were taken while the simulation was running for the test year of 2014. Screenshots and figures of other weeks are provided in Appendix B.

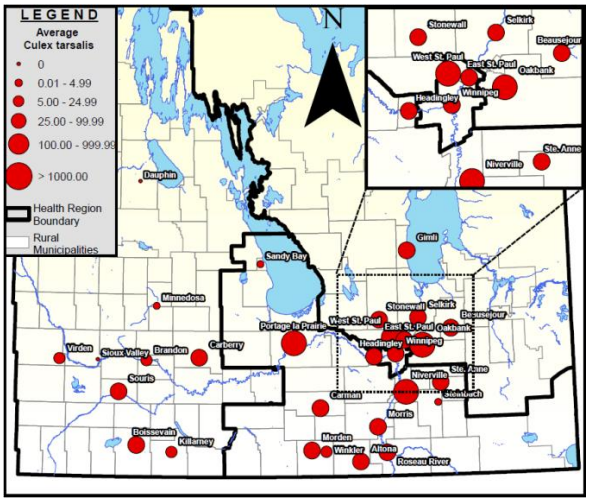
Cumulative degree days maps [158] are shown in Figure 3-8 as the Manitoba government as well as many modellers explicitly employ degree days as an indicator of WNV activities. Recall that the development of mosquitoes and the virus both need warm weather over a period of time. Reference [45] gives an example of degree days WNV models. Although degree days were not explicitly used in the proposed cellular model, the model output considers it to some extent. This confirms the importance of temperature on the mosquito life-cycle and WNV activities. The impact of temperature is explicitly assessed and reported in section 3.4.4. Figure 3-7 also suggests the Victoria beach area could be a potential hot spot for mosquito activities; however, no mosquito trap is installed there to judge this conjecture. This must be mainly due to the generally warm weather conditions of this area over a long period of time. In fact, in most years in our dataset, the vicinity of the Victoria beach was one of the hottest areas in the province from the middle of July until the first weeks of September.



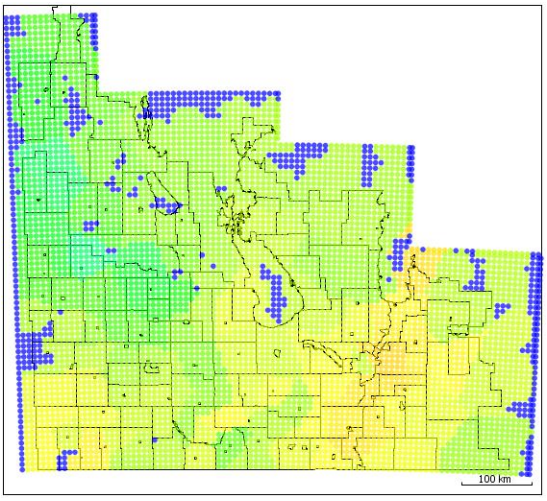
Trap data in week 27 in 2014



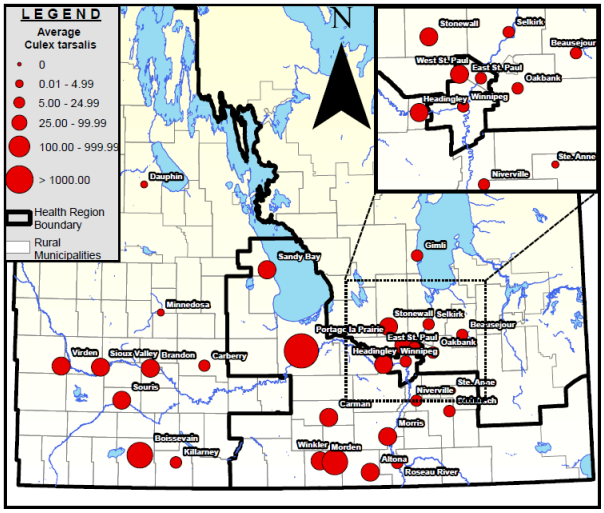
Simulation state in week 27 in 2014



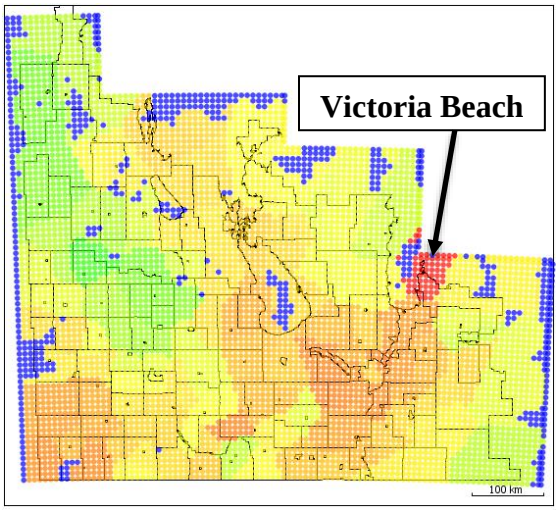
Trap data in week 30 in 2014



Simulation state in week 30 in 2014

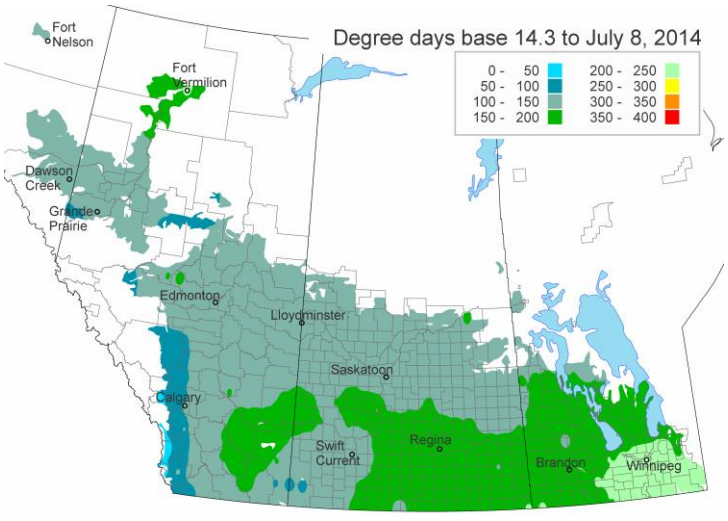


Trap data in week 34 in 2014

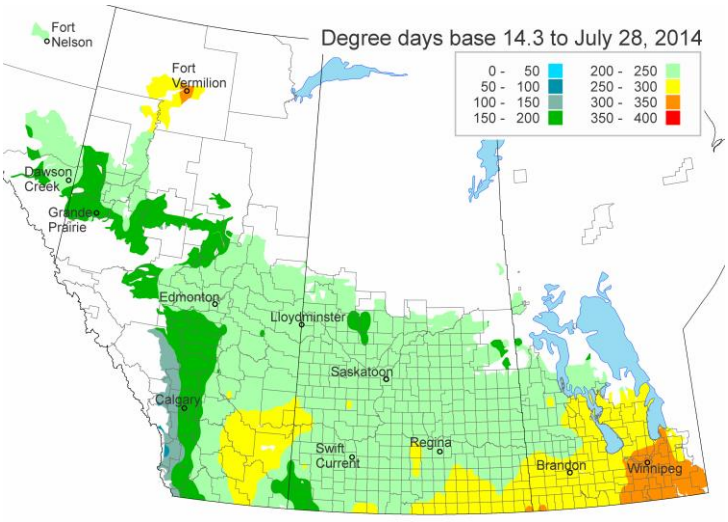


Simulation state in week 34 in 2014

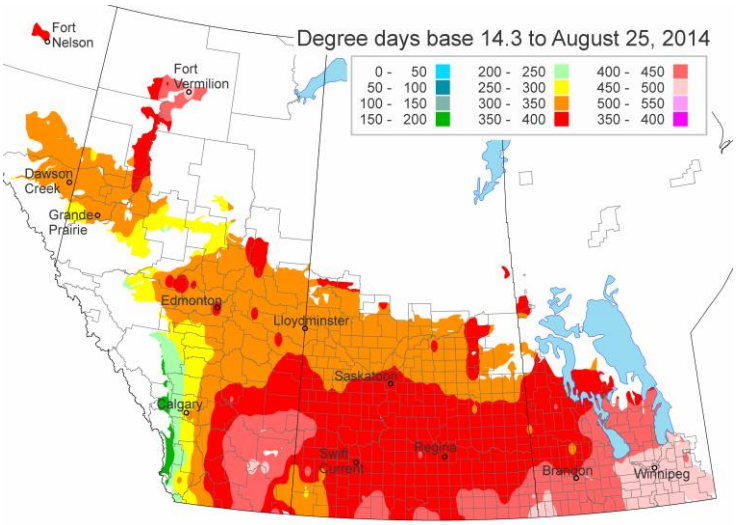
Figure 3-7 Weekly distribution of mosquitoes according to the trap data (source: [14]) and simulation



Week 27 in 2014



Week 30 in 2014



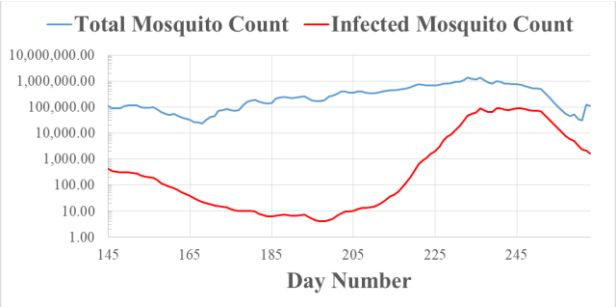
Week 34 in 2014

Figure 3-8 Map of cumulative degree-days (source: Agriculture and Agri-Food Canada [158])

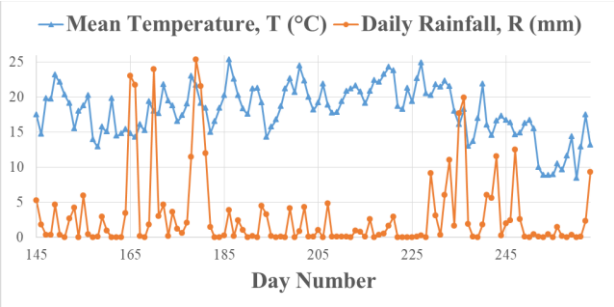
3.4.3 Infection Propagation

Simulations of infection propagation demonstrated that the birds are the main factors for spreading WNV to distant areas. The test year of 2014 with a small alteration in its temperature conditions was selected as the input for the simulation. The mean daily temperature was increased by 1 °C to capture more visible patterns of WNV spread. Alterations in temperature and rainfall values are discussed in subsection 4.4. A total of half a percent (0.005) of mosquito population was set to be infectious by infecting half of the population of one percent of mosquito cells/sites. The number of initially infected mosquitoes is high enough to initially infect a few of birds before those mosquitoes die out. Therefore, it is quite reasonable to assume some competent birds at those initially infected mosquito sites are infectious until they either recover or die. Figure 3-9 shows variation of daily rainfall, temperature, and the number of total mosquitoes and infected ones for this simulation in a logarithmic-scale. Figure 3-10 on page 95 illustrates the spread of the infection across the province over a number of weeks. Week 21 in Figure 3-10 shows the initial location of infected mosquito sites in a lime color. These mosquitoes infect some birds around them. By the middle of week 29 (i.e., day no. 194), the majority of the infected mosquitoes die out as the map is entirely blue, but there still exist some infectious birds. Beginning the week 32, the infection in mosquitoes has begun emerging, mostly in the areas close to the initially infected mosquito sites as these areas have the highest number of infectious birds, and they may still have infected mosquitoes. The infectious birds from these areas could fly over other mosquito sites and spread WNV to them. During the weeks 34-35, a rather sharp increase occurs in the population of mosquitoes in mosquito-favored areas, such as the Victoria Beach on top right of the map. Therefore, these highly mosquito-populated sites have a relatively high probability of biting an infectious bird, even though there were *not* any

infectious birds/mosquitoes, initially. Later on, at the beginning of week 39, the mosquitoes' activities (and consequently the number of infected mosquitoes) is reduced everywhere due to changes in weather conditions. Figure 3-10 confirms the avian flow of the system is a significant factor in WNV propagation to other locations.

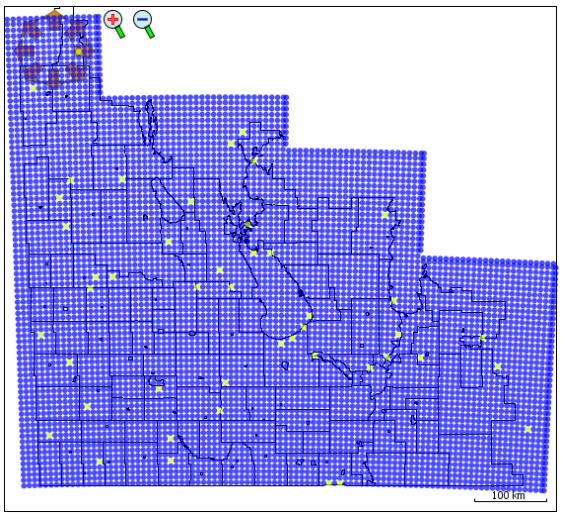


a) Number of total adult and infectious adult mosquitoes on a logarithmic-scale

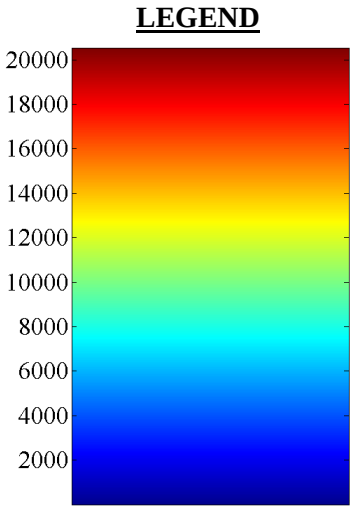


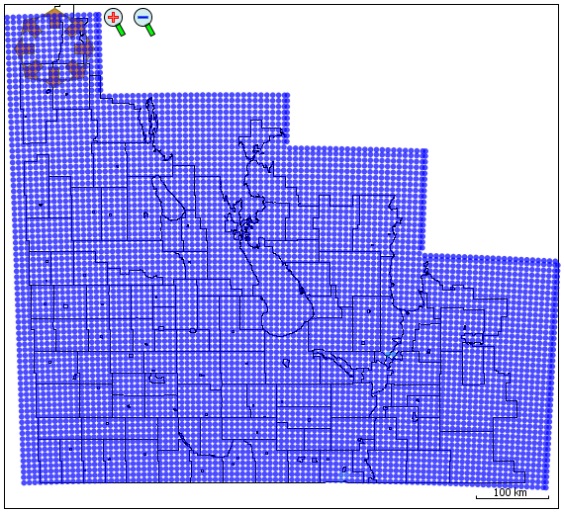
b) Daily values of mean Temperature (°C) and mean Rainfall (mm)

Figure 3-9 Daily weather variations and adult mosquito population dynamics

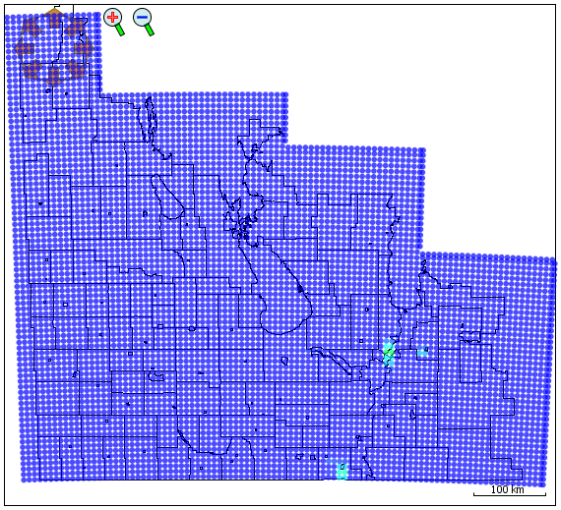


Week 21 (initial conditions)

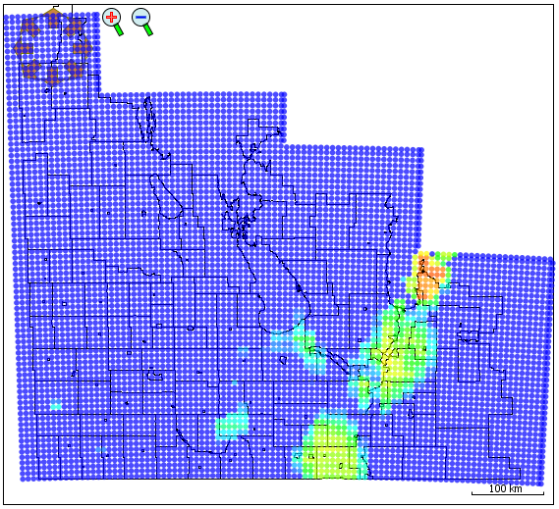




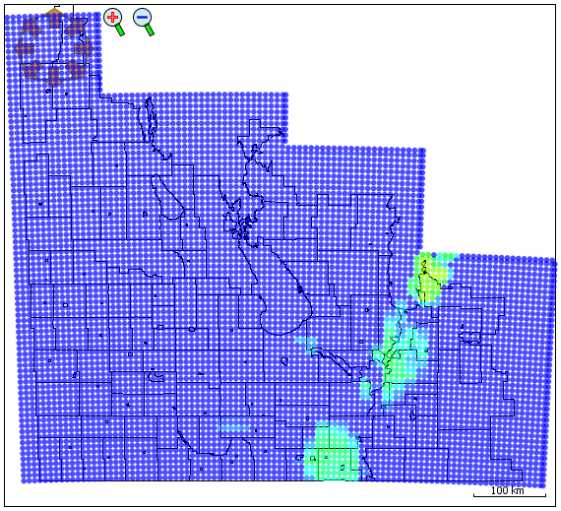
The middle of week 29



Beginning the week 32



The end of week 36



Beginning the week 39

Figure 3-10 Distribution of infected mosquitoes across the province in a number of weeks

3.4.4 Sensitivity Report

The primary findings of the sensitivity analysis were that the output of the CDiffE model is sensitive to all the key inputs (i.e., temperature, rainfall, and land cover factor). Increasing the value of w_L (i.e., landscape impact) yields an increase in the average Pearson correlation between the trap data and simulation output. Increases in daily temperature and rainfall values up to some thresholds also increase mosquito infection rates. Temperature has a more influential impact, however.

Changes of the average of the three evaluation metrics for the balanced-approach solution in response to variations of w_L is plotted in Figure 3-11 for all of the years from 2004 to 2014. Expectedly, the average Pearson correlation gradually increases as the value of w_L goes up to a maximum threshold of approximately 1.3, and then a marginal decreasing trend begins. For the low or even negative values of w_L, the Spearman rank-order correlation remains almost constant at around 50%, followed by a sharp decrease beginning at a positive value of 0.2 for w_L; whereas the population scale has a steady declining trend as the value of w_L increases.

In addition to the three evaluation metrics of mosquito population, a weekly infection ratio is calculated for mosquitoes during the simulations. Recall that weather conditions play a vital role in mosquito dynamics and consequently in the infection ratio of their population. The weekly infection ratio can be used to decide whether an input scenario results in an outbreak depending on the competence-level of target hosts. Investigating the outbreaks, in particular with regards to humans, and possible prevention policies, is an important direction for the future work of this thesis. Figure 3-12 and Figure 3-13 demonstrates changes of the ‘total average’ and ‘average of yearly maxima’ of weekly infection ratios with respect to changes in daily values of mean temperature and rainfall for all the years from 2004 to 2014.

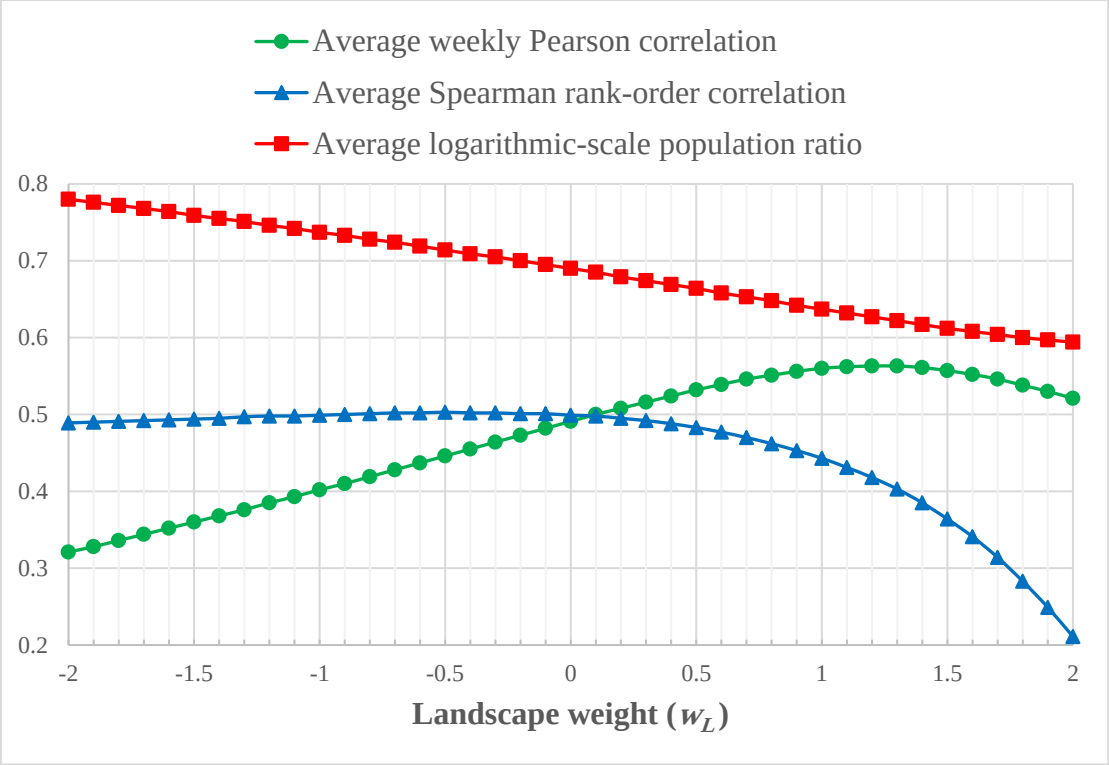


Figure 3-11 Variations of average of the three metrics for 2004 to 2014 with respect to landscape weight (w_L)

For each input value along the horizontal axes in Figure 3-12 and Figure 3-13, the values of Rainfall and Temperature for all days were changed accordingly. For example, an input value of -10 mm in Figure 3-12 means 10 mm is subtracted from the rainfall values of all simulated days (and truncated at zero). Similarly, a value of 10 °C on the horizontal axis in Figure 3-13 means 10 °C is added to the mean daily temperature of all simulated days. The temperature value is also truncated at zero (i.e., it never has a negative value).

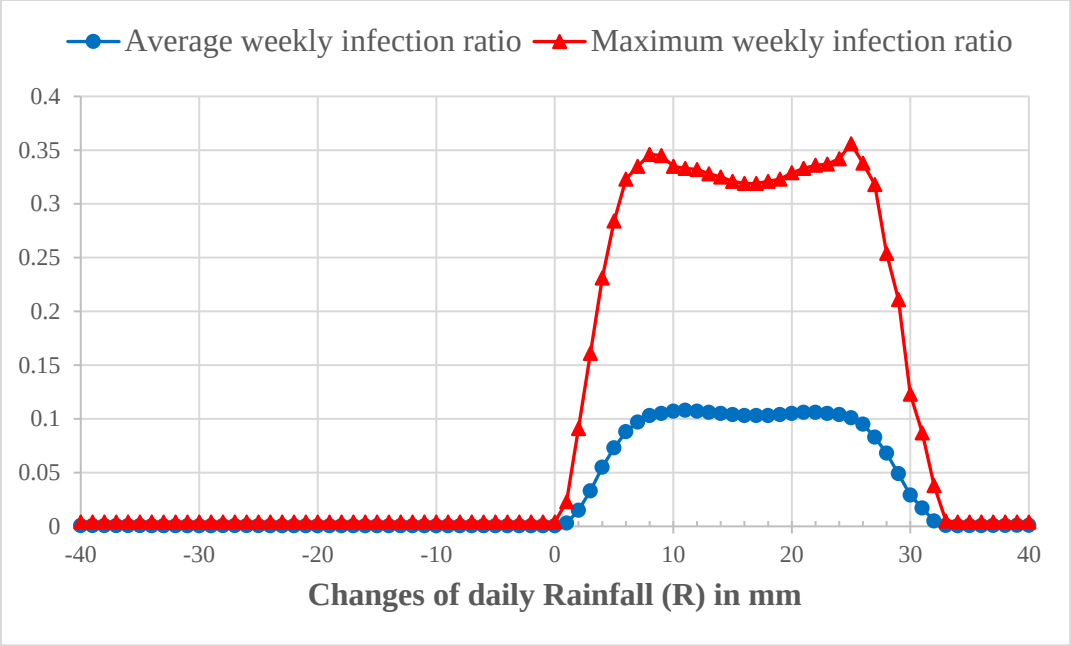


Figure 3-12 Variations of weekly infection ratio, being averaged for 2004 to 2014, with respect to changes in the daily value of Rainfall (R)

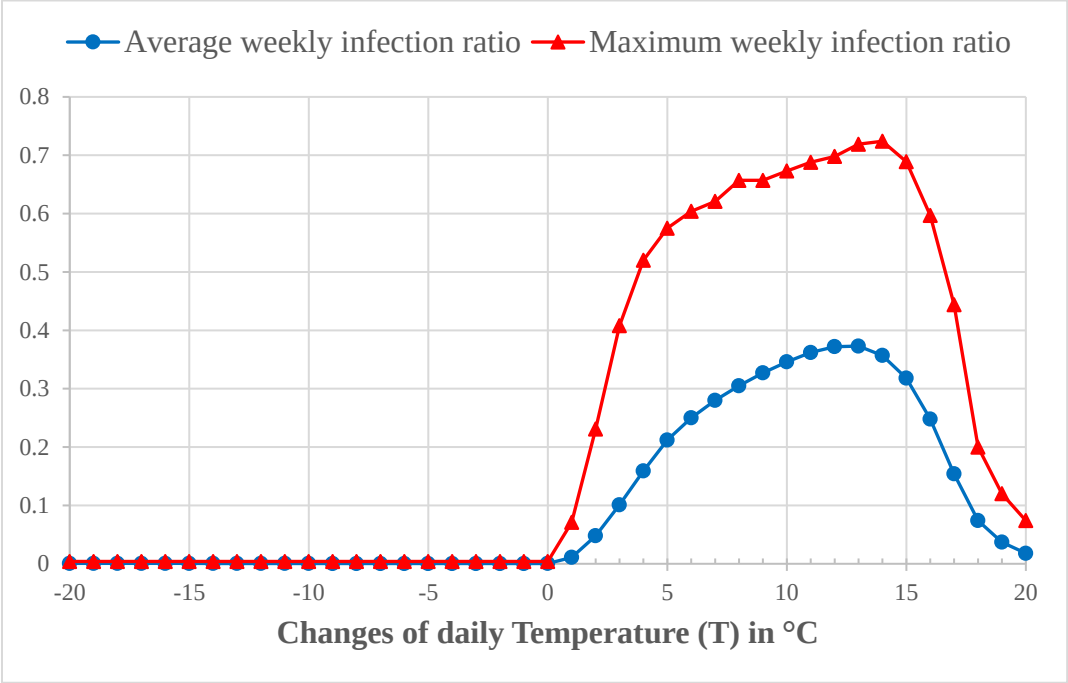


Figure 3-13 Variations of weekly infection ratio, being averaged for 2004 to 2014, with respect to changes in the daily value of mean Temperature (T)

As the initial conditions for these two experiments, half a percent (0.005) of mosquitoes are infected. As shown in subsection 3.4.3, the initially infected mosquitoes could infect a few of birds initially. Then, during the simulation, if the number of mosquitoes grows sufficiently, the susceptible mosquitoes will get infected from those initially infectious birds and the WNV spread grows. Generally, both maximum and average weekly infection ratios have similar trends as shown in Figure 3-12 and Figure 3-13. Overall, the two plots indicate the model is extremely sensitive to weather conditions, as expected.

Recall that the desired level and frequency of rainfall for mosquitoes is controversial [30]–[32]. According to Figure 3-12, if the amount of daily rainfall for each day increases by 2 to 32 mm, a visible increase occurs in weekly infection ratios, compared to a change of zero mm in daily rainfalls. This change has three phases of a sharp positive slope for the values approximately from 2 to 7 mm, relatively constant with some fluctuations for 8 to 25 mm, and a sharp negative slope for 26 to 32 mm. The double-sided effect of rainfall can be explained through mosquito habitat preferences for breeding and a larvae-flushing effect of heavy rainfalls. This trend verifies the fact that small increases in rainfall generally creates more stagnant water for mosquitoes to lay eggs, whereas heavy rainfall flushes and kills them [33]. The other interesting and less intuitive phenomenon was slight changes of infection ratios for negative values of the horizontal axis (i.e., rainfall changes) in Figure 3-12. While it may not be visible in the plot-scale of Figure 3-12, the rainfall change of -40 to -15 mm resulted in a slightly higher infection ratio, compared to rainfall change of -15 to 0 mm. This means the mosquitoes generally may do slightly better in drought conditions compared to normal rainfall conditions.

By comparing the amplitude of infection ratio changes in Figure 3-12 and Figure 3-13, it is evident that mosquito dynamics are considerably more sensitive to temperature than

precipitation. This is confirmed by biological studies, as the temperature affects more mosquito conditions, such as reproduction rate, biting rate and even WNV incubation period [24], [30], [31], [33], [38]–[43]. The general trend of the plot of Figure 3-13 indicates increasing the mean daily temperature generally increases the infection ratio at different rates up to a high temperature where mosquitoes begin to die out due to extremely hot temperatures. Figure 3-13 suggests that if the mean temperature of every day increases by over 20 °C, the death rate of mosquitoes would be so high that infection ratio would probably decline. Similar to rainfall changes, the impact of temperature changes from -20 to 0 °C is not visible in the plot-scale of Figure 3-13. However, interestingly, in consistent cool/cold day conditions (i.e., high negative values on the far left of horizontal axis), the infection ratio is slightly higher than relatively normal cases with temperature change ranges approximately from -5 to 0 °C. Overall, this means lower death rates and biting activities produces more infected mosquitoes than normal death rates and biting activities do. Recall that the simulated temperature was truncated at zero, which means a high value of -20 °C on the horizontal axis forces the simulated temperature to be exactly zero degree Celsius at cooler days (where the actual temperature is 20 °C or less). This result can be explained through the fact that mosquitoes are somewhat resistant to mild-cold temperatures and some hibernate in cold weather conditions.

3.5 Summary of Chapter 3

3.5.1 Conclusions

In this chapter, a cellular difference equation (CDiffE) structure for modelling WNV dynamics is proposed for adoption in data-driven WNV-ABMs. Differential Equations or other aggregative approaches do not have the capacity to capture WNV at a fine-grained scale. Straight

ABMs are too expensive to model virus transmissions at the agent-levels (i.e., triggering rules per every single mosquito). In a hybrid ABM-based approach, difference equations should be exploited as opposed to differential equations. In addition, a cellular map, as opposed to polygon-shaped boundaries [51] or virtual networks [104], makes it easy to effectively analyze spatial aspects of WNV such as human mobility. This ease of usage is verified in Chapter 4. This chapter showed how hybrid methods can effectively capture the spatio-temporal dynamics with an acceptable accuracy and computational cost.

The work here focussed on including as many WNV factors related to mosquito population dynamics as possible. The parameters of the difference equation are temperature- and rainfall-driven during the mosquito season. The impacts of landscape features (i.e., land cover) are included in the equations. Multiple species of local and migratory birds with their weekly migration patterns are considered as the amplifying hosts in the equations. Human agents as the dead-end hosts are included in the set of proposed equations. Twilight times are employed to set the mosquito biting rates. These parameters, specific to southern Manitoba (see Chapter 1), are fed into the simulations for numerical analyses. The parameters regarding the mosquito population development are calibrated based on the trap data available from the Manitoba government. For over 150 bird species, WNV-related parameters, such as competence indices and recovery rates (and probabilities), are estimated based on viremia studies.

As part of model validation, variation of mosquito densities over 15 weeks during the spring and summer for the years 2004 to 2014 is compared against the simulation outputs. Generally, simulations produce trends of mosquito population close to the actual mosquito population as judged via trap data. The system identifies the patterns of mosquito density in approximately 30 communities over 15 weeks relatively well. The average and median system prediction power

among all generated set of solutions are 52% and 56% in terms of correlations between the simulation output and real trap data, respectively.

A key contribution of the work of this chapter is that given a set of input climate data, the system could be used as a tool to detect whether the mosquito population would be increasing, decreasing or staying constant, assuming no unexpected changes to bird conditions and land cover structures. The model output can be used as risk maps for guiding practitioners and public health authorities in Manitoba.

Key findings from the simulations suggest the vicinity of Victoria Beach in Manitoba has a relatively high potential to be a home-base for *Culex* mosquitoes. Installing mosquito traps in its proximity may be an option for the local health authorities, considering the operational expenses. The simulations predict that global warming has a drastic impact on mosquito population if no larviciding program is in effect. Moreover, the simulations exhibit biologically compatible behaviour for extreme weather conditions such as drought where mosquitoes could do relatively well in retaining their population.

The proposed CDiffE model on its own can be seen as a cellular agent-based model. Direct applications to other diseases are possible, in that the CDiffE architecture with some modifications can be utilized to describe dynamics of other mosquito-borne diseases such as Zika and Dengue in other geographical areas with known historical data.

3.5.2 Limitations and Directions for Future Research

Certainly, more simulations can be set up to further investigate the impact of many other parameters, such as bird migration patterns, their competence indices and so on. At this stage, the entire proposed CDiffE model can act as a complex environment for a higher-level agent-based model of WNV propagation, including humans. Such an ABM is the focus of Chapter 4.

The calibration results presented here may not be most optimum outcomes of the system. In this regard, a genetic algorithm with its superior ability to escape from local optima, due to its mutation operator, can be employed in the calibration process. The simulated annealing algorithm with its strengths in combinatorial problems could be also used in the calibration process to find the best combination of equation parameters. Furthermore, upon availability of quantitative data regarding the infected mosquito pools or birds, the infection-related parameters can be tuned. Currently these parameters, including biting rates and transmission parameters, are taken or estimated from other studies.

More recent publications by entomologists and ornithologists on bird viremia and their WNV-competence can always be adopted to further tune the estimates of corresponding parameters. One essential limitation with the proposed model is the implicit assumption of having the same virus transmission probabilities from different bird species to mosquitoes. A meaningful extension of current model is incorporating various transmission probabilities from birds to mosquitoes. Such an extension can be achieved by changing equation (3.9) as follows

$$M_s(t+1) = (1-\mu_m)\left[\prod_j \left(1-\beta_m^j\right)^{\frac{b^j B_i^j(t)}{H_t(t)+\sum_j B_t^j(t)}}\right] M_s(t) + (1-\mu_m)\,\gamma\, M_a(t)$$

where β_m^j is the virus transmission probability from bird species j to mosquitoes. Improvement in initialization methods for the mosquito populations, including the adult-stage and infected mosquitoes could be another possible direction for future work. Such methods may help change the model from a seasonal model to a year-around powerful simulation tool.

Lastly, the model does provide the capability to apply the impact of land cover/use on mosquito breeding rates. However, our collected data regarding the land cover parameter, *L*, may not have appropriate values. In other words, using the reported incident case count for WNV

among humans in North America to indicate the importance of each specific land cover class for mosquito breeding is likely not the best solution. As such, the weight, w_L, of the land cover parameter, L, may not have proper impact to describe the capacity of an environment for mosquito production. This is the primary reason why there is a relatively high number of generated mosquitoes in water regions, which have an unknown value of L in the collected dataset. If suitable data regarding the landscape features are available, the land cover parameter (L) could also directly affect the alpha parameter of the birth rate (α_b) as to indicate the environment carrying capacity of female adult mosquitoes. Alternative solutions to quantitatively map each land cover class to a mosquito density descriptor could be crucial future research. Within this context, simple fuzzy rule-based systems could be used to transfer experts' knowledge of preferred land cover by mosquitoes. Such a fuzzy system would consider the proportion of each land cover class present in a mosquito cell, and then would produce one (or more) descriptor(s) for mosquito birth rate or habitat preference. Once the fuzzy system is constructed, all the w_L values or other similar metrics could be pre-calculated for each mosquito cell. The values would then be loaded when the simulation is started to boost the performance.

Chapter 4: AGENT-BASED MODEL OF PERIPATETIC HUMANS

4.1 Introduction

As mentioned in Chapter 1, a number of Difference Equation (DiffE) or Differential Equation (DE) models has been proposed to capture dynamics of WNV transmission [27], [88]–[90]. However, none has the capacity to model or consider a heterogeneous population of humans and/or their movement. Indeed, straight difference/differential equation or other aggregative approaches are built on the assumption of a homogenous population of agents (i.e., hosts and vectors). ABMs can be utilized to integrate heterogeneity and movement of humans, birds, mosquitoes, as well as their interaction with one another.

This chapter presents a detailed data driven ABM of mobile humans adopting the CDiffE model from Chapter 3 as its environment to simulate WNV transmission, employing various data from southern Manitoba, as outlined in Chapter 2. The CDiffE scheme provides the ABM with fine-grained mosquito cell agents, driven by temperature, rainfall, land cover, daylight, and various bird mobility and distributions.

Despite the importance of human motion in the spread of infectious viruses [160]–[166], often the impacts of human behaviours and mobility are disregarded in many health related ABMs as well as in WNV modeling attempts. In the case of WNV, the reason may be that humans are dead-end hosts in the WNV transmission cycle. In contrast, for some other mosquito-borne diseases such as Malaria, Dengue, and Chikungunya humans are amplifying hosts. As such, there are some studies that focus on the impact of human behaviour in spread of these diseases. A number of these studies were reviewed in Chapter 1. In regard to human movement, all the models reviewed either used simulated displacement through virtual nodes or have adopted some kinds of models/assumptions for human behaviour given the limited data (e.g.,

census and demography) that was available. The proposed ABM in this chapter implements a fine-scale movement trajectory for each human agent according to a fair amount of real-world cellular phone tower connectivity data. Heterogeneous exposure of individuals to mosquitoes and non-random contacts between them are key properties of the human movement component. The proposed ABM assesses whether (fine-scale) human movement is a critical component underlying WNV transmission and infection rates. Simulations are run to examine the reliability of ABM results for counts of infected humans against real WNV data.

4.2 Material and Methods

The ABM consists of the three main agent types, namely, mosquito, bird and human. The transmission cycle of the virus between mosquito and bird is considered to be within the environment of human agents. The environment is validated in Chapter 3. This environment consists a grid of cells, each of which is centered on a certain coordinate within Manitoba. The human agents move through these cells according to trajectories extracted from anonymized cellular phone data, provided by Manitoba Telecom Services (MTS). The following two subsections review the properties of the environment and details of the human agent.

4.2.1 The CDiffE Environment

The environment of the proposed ABM is structured as a cellular difference equation model (CDiffE) as explained in Chapter 3. Recall that in the CDiffE scheme, the region of study is a cellular map where each 5 km x 5 km cell is driven by a difference equation, interacting with one another through bird agents. There are 6067 cells overlaid on southern Manitoba, covering an area of approximately 148,812 km^2.

At hourly time-steps, every cell updates mosquito and bird dynamics according to weather conditions, landscape features, bird migration and the time of day (i.e., sunlight). These hourly dynamics include the total number of different bird species, B^j, the number of infectious mosquitoes, M_i, and their biting rates on humans, b_h. Recall that the CDiffE scheme also proposes a set of equations to capture dynamics of WNV in humans as follows.

$$H_s(t+1) = (1-\beta_h)^{e(t)} H_s(t) \tag{4.1}$$

$$H_i(t+1) = \left(1-(1-\beta_h)^{e(t)}\right) H_s(t) + (1-\zeta_h) H_i(t) \tag{4.2}$$

$$H_r(t+1) = \zeta_h H_i(t) + H_r(t) \tag{4.3}$$

where

$$e(t) = \frac{b_h M_i(t)}{H_t(t)+\sum_j B^j(t)} \tag{4.4}$$

is the expected number of infectious bites that a single human agent is likely to receive. Other parameters used in the above set of equations can be found in Table 3-1 on page 59. The first coefficient in equation (4.2) i.e., $(1-(1-\beta_h)^{e(t)})$ is considered the force of infection for any susceptible human agent at the time-step t. This means this coefficient gives the probability of infection for every human agent in each time-step. Similarly, the first coefficient in equation (4.3) i.e., ζ_h, is the per time-step probability of recovery from the virus for every human agent. The former term is provided by the complex environment for each cell through the CDiffE mechanism, whereas the latter term is static. Both parameters are further altered based on each human agent's immune system.

4.2.2 Human Agents

Two main sources of data for the behaviours of human agents are the census data [36] and trajectories extracted from anonymized cellular phone data provided directly to the author's lab

by a telecommunications service provider [37]. In this section, first the data preprocessing procedure is explained. Secondly, the agent behaviour is discussed.

4.2.2.1 Data Preprocessing

The census data were downloaded from Statistics Canada [167]. Based on the census data, there are 198 population centers (including cities, villages, towns, rural municipalities, etc.), of which 94 are associated with the scope of the model, comprising a total of approximately 1,032,000 people out of the 2011 population of the province of 1,141,000 people. This means, for example, in order to simulate a total population of 300k humans in the province, approximately 271k agents are associated with the scope of model. These agents can then be driven by the ABM. The remainder of the population live in areas of the province which are unknown to the ABM and, in particular, the CDiffE environment.

Providing each human agent with a reasonable and reality-based movement trajectory is the main challenge. Trajectories are extracted from anonymized cellular phone data of a number of users over a limited number of days, provided by Manitoba Telecom Services (MTS). A cellular phone database has many records with the following information: a telecommunications user's anonymized ID, a passive (such as a ping) access time-stamp, and an associated tower ID. The tower IDs can be mapped to actual physical coordinates of a tower within the province. Two different set of cellular phone data from the same provider are combined to create a single trajectory database. The first set of data had information regarding approximately 180,000 distinct users over a span of five days in October 2010. The second database contains information of approximately 80,000 distinct users over a span of 28 days in December 2011. Both datasets had a large number of either missing time-stamps or non-decodable tower IDs.

After pre-processing and re-assembling the data using C#.NET and MySQL, a total of over six million hourly trajectory records were extracted. These records comprise over 258,000 unique users with decodable locations at hourly intervals. The trajectory database has 289 distinct tower coordinates. As the hourly trajectories are extracted to implement mobile human agents, if only a single (tower) location is assigned to a user, the user is disregarded. As such, a total of approximately 71,000 users were removed from the trajectory database. Although this type of data was removed from the trajectory database, it may still provide some indirect input as to actual movement. It is also notable that people who may not move much (i.e., under a radius of two to three km) are already considered in the model due to the structure of the CDiffE environment. In addition, because of a high number of missing access time-stamps, many of the mobile human agents still spend most of their time within a single cell.

As noted, for extracting hourly trajectories, users with missing time-stamps were assumed to remain within the coverage area of the last-connected tower. Moreover, if a user were connected to different tower IDs within an hour, the first connected tower was chosen as the representative location of the user. This is because the ABM and its CDiffE environment have an hourly time-step (or accuracy). Movements of human agents within the hour-long interval were assumed to be within an area of 25 km^2 i.e., within the same CDiffE cell. Recall that the foraging flight of *Cx. tarsalis* mosquitoes, the primary vector of WNV in Manitoba, may be extended well beyond six kilometers, up to 27 km [28], [83].

4.2.2.2 Behaviour Implementation

Mobile human agents are modeled with a simple Susceptible-Infected-Recovered (SIR) state machine (Markov process) with respect to WNV infection. Initially, all people are in the susceptible state with different (random) susceptibility parameters. Therefore, all people in the

same CDiffE cell in the environment may become infected with different probabilities despite having the same force of infection provided by the environment at each time-step. When a person is infected, a recovery count-down timer is turned on for them to keep track of their recovery process. Once the timer hits zero, the infected person goes into the recovered state. Given the time-span of simulations and the fact that recovery from WNV may take several months, implementing the recovery counter may not be necessary. After updating the WNV state machine, each person may move within the environment, if their pre-defined trajectory implies movement. It is assumed the infection does not affect a person's movement. This means human agents follow their assigned trajectories no matter what their WNV states are.

The population of human agents are initially distributed across a number of population centers based upon census data [167]. Considering the computational (time) complexity of the ABM, the human population in the simulations is fixed at 300,000. As a result, there are over 271,000 autonomous human agents with heterogeneous properties (i.e., susceptibility and cellular trajectory) initialized to be in southern Manitoba. During the simulations, each of these mobile agents moves about mostly in southern Manitoba, but they may also temporarily leave southern Manitoba depending on their assigned trajectories. While a human agent is outside the CDiffE environment (i.e., southern Manitoba), they cannot become infected as the environment would be completely unknown to the ABM.

As the population of human agents is distributed according to census data, the trajectories are categorized based on their initial location (tower). This means each population center is assigned a set of trajectories. So, during the initialization, each human agent picks a trajectory from the set of trajectories available in their initial population center rather a completely random

trajectory. It is notable that a number of hourly trajectories were disregarded as their initial location was outside of southern Manitoba. There are a total of over 180,000 unique trajectories.

Using the cellular phone tower locations, the CDiffE environment was partitioned into different telecommunication regions according to a Voronoi diagram [168]. As such, each telecommunication region is considered to be the coverage area of a single telecommunication tower. This means each cellular phone tower location is linked to a unique set of the CDiffE cells. At the end of each time-step, a human user may decide to change their tower location based on their own trajectory. Every time a user changes their tower location, first they remove themselves from their current CDiffE cell. They then pick one random CDiffE cell from the new tower location as their next CDiffE cell to move to. At the end of a trajectory cycle, a human agent remains within the coverage area of the last-known tower until midnight. The agent then moves back to their initial telecommunication tower region at 1:00 am. This is how the trajectory-driven movement algorithm loops through cellular phone trajectories to exhibit more natural and practicable (realistic) movement patterns.

4.3 Results and Simulation Studies

Recall that the WNV data in Manitoba [20] provide weekly mosquito counts from the Centers for Disease Control (CDC) week 22 to CDC week 36 for the years from 2004 to 2014. These data were previously used to validate the CDiffE component in Chapter 3. The WNV data also provide the total number of WNV human cases for each year from 2004 to 2014. To compare the ABM results to the human cases from the WNV data, simulations were set up to run from CDC week 21 until the end of CDC week 39. The CDiffE environment settings are set according to the population-oriented solution from Chapter 3. For each year from 2004 to 2014, simulations were repeated 100 times for a total of 1100 runs. The average of total number of

infected people in each year is shown in Figure 4-1. The figure also shows the total human cases and mosquito count from the real WNV data. Recall that, seroprevalence studies suggest that the number of actual infections is far above the number of reported cases, with many individuals having very limited symptoms. Also, it bears noting that seroprevalence data is worth considering as a data source, as it can be compared with the numbers of recovered individuals.

In general, the output of the ABM produces a proportionately similar trend to the actual number of WNV human cases reported in each year, except for 2010. Statistically speaking, there is an approximately 44% correlation between the output of the ABM and the WNV data over the years. However, if 2010 were excluded, the correlation increases to 84%. In 2010, despite having a relatively high number of mosquitoes (as occurred in the simulation and evidenced by the trap data), surprisingly, no WNV human cases are reported. There are multiple possibilities for this outlier. For example, there might have been a unique larviciding program in effect in 2010 that changed the initial number of infectious mosquitoes. Alternatively, the infectious population of migratory birds may have been different in 2010, or it may simply be that infected people in 2010 did not report to health departments. In any case, the simulation does not distinguish between the initial conditions in different years. The only difference between years is the weather conditions. Additionally, while the mosquito contact rate could have been adjusted by some coefficient to obtain a lower number of infected humans, it was carefully increased not to have zero infected humans in more than a year (i.e., 2004). Such a relatively high contact rate can reveal the impact of different scenarios on the number of infected humans in low-activity years such as 2004 and 2009. It is notable that as the data regarding the number of infectious mosquitoes were not sufficient, no calibration or parameter tuning was performed on any infection-related parameters.

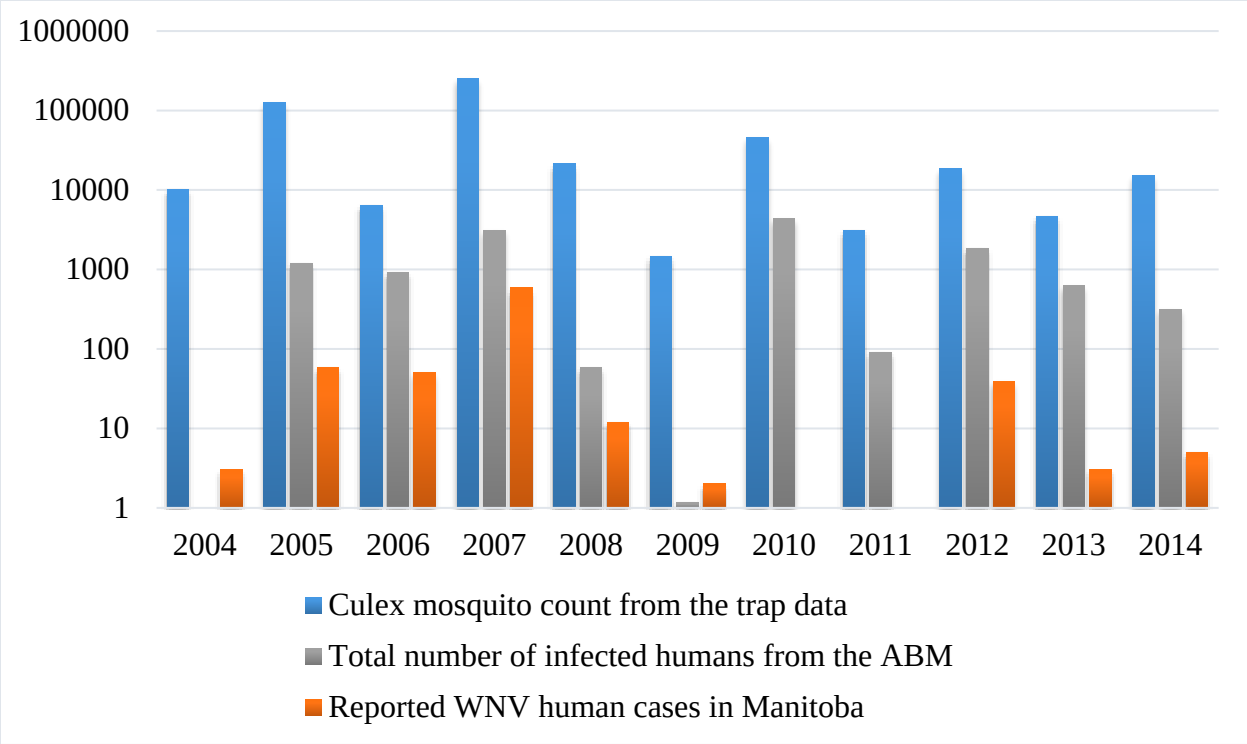


Figure 4-1 Comparison of the ABM output against the WNV data

The weekly human infection counts for all the 1100 runs are shown in Figure 4-2 as a 2D histogram with different colors for each year. The color intensity of each bar corresponds to the relative frequency of each 2D histogram bin within the X-axis interval. Each 2D histogram bin counts the number of runs with a specific interval of infected human count in a certain week as shown on the axes. As expected, the general weekly pattern of WNV incidences in all years is quite similar. Every year, WNV infections begin around week 32 (i.e., beginning towards the middle of August). By the end of week 34 or 35, a substantial increase in the number of infections occurs, and by the end of week 37 it plateaus. Figure 4-3 depicts the same 2D histogram in an aggregate form where all years together consist of a single histogram per each X-axis interval. Also, instead of 2D histogram bins, 2D histogram envelopes are drawn which can be considered as extended boxplots showing empirical quartiles around the median.

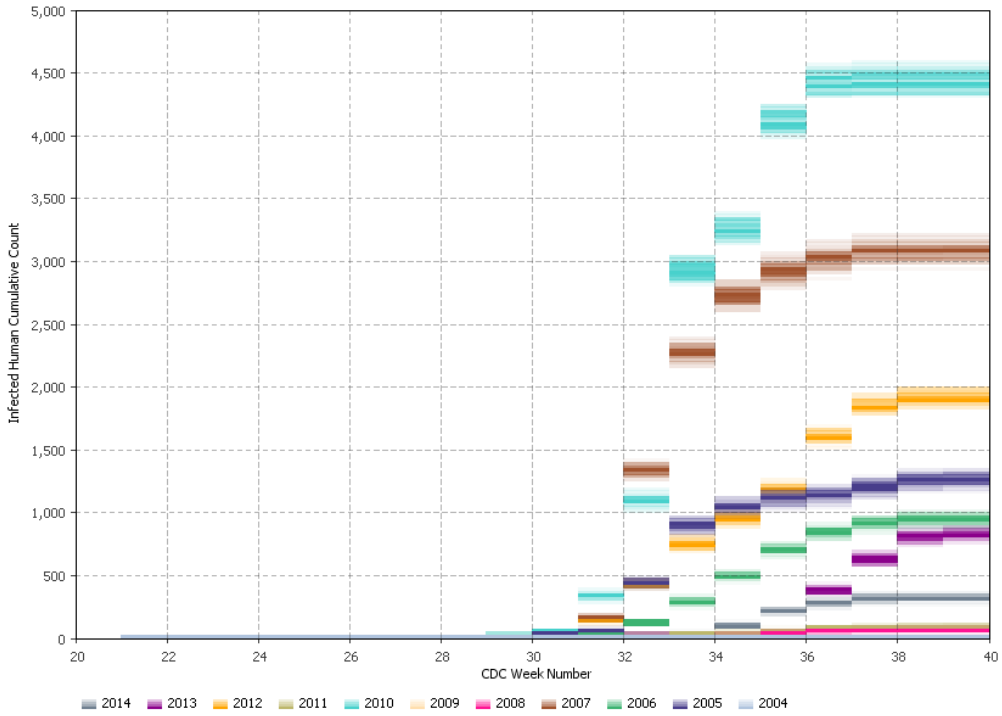


Figure 4-2 Histogram of weekly infected human counts for each year from 2004 to 2014

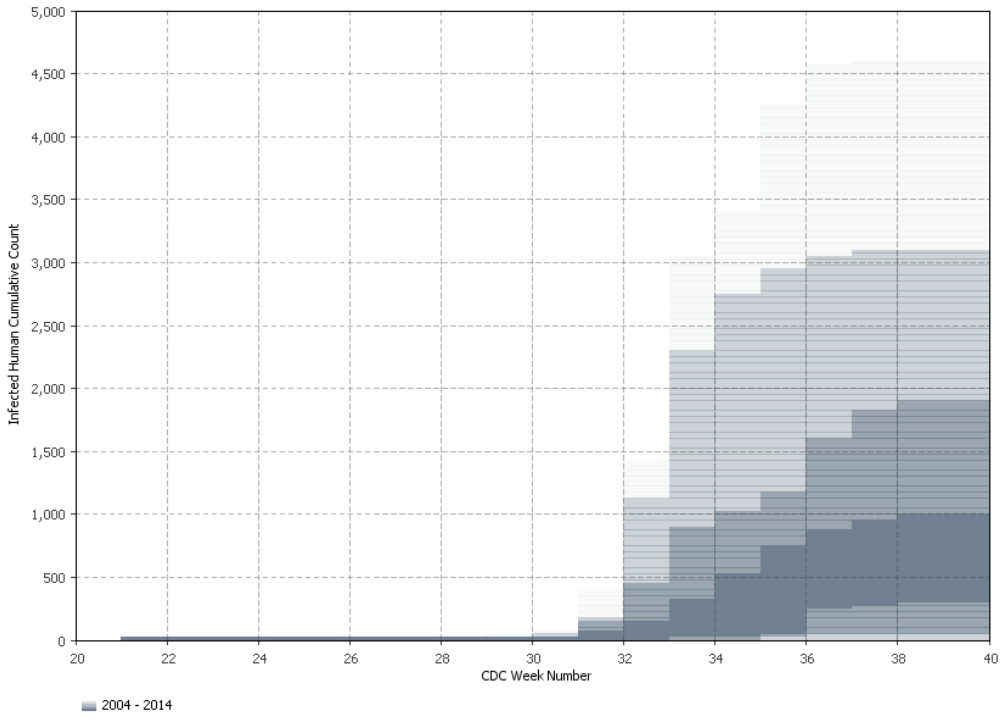
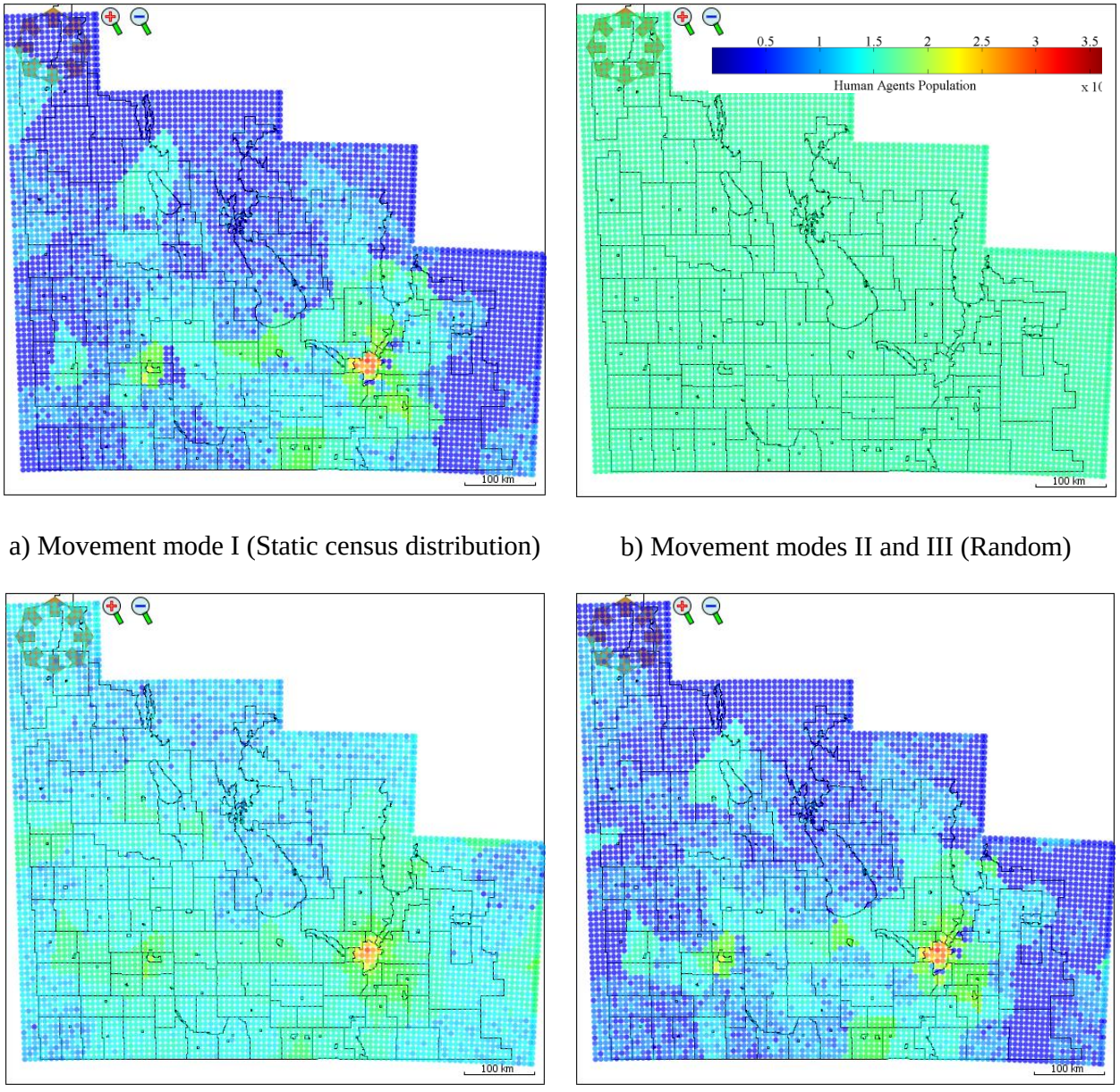


Figure 4-3 Aggregate summary of the simulation runs of all years in the form of a 2D extended boxplot

All the results presented thus far are obtained by employing the regular data-driven movement algorithm explained earlier in sub-section 4.2.2. Now, to determine the impact of mobility in WNV spread, four more movement scenarios are considered. These four scenarios can be characterized into two general categories. First, assume there is no movement at all where the initial distribution of human agents can be either uniformly random or based on the census data. Within this category of movement scenarios, agents remain in the same CDiffE cell for the entire simulation time. The difference between the two scenarios is whether the population density is practicable (realistic). Second, assume human agents move randomly where movement can be either fully random or based on cell tower density. In the fully random movement, there are no restrictions for agents to pick their destination CDiffE cell. However, in the random movement based on cellular tower density, agents first pick a tower coverage region at random. Then, they randomly pick their destination from the associated CDiffE cells of the tower. As such, while agents still move at random, the regions of map with higher densities of cellular towers have higher human population at any given time-step. This approach may be a reasonable measure of movement in some applications when no other form of data is available.

Simulations are set-up to examine each of these movement scenarios in addition to the regular stochastic movement mode. Screenshots of all different movement scenarios in a single time-step are found in Figure 4-4. Due to the stochastic nature of all scenarios, each simulation was repeated 200 times for a total of 1000 runs. The 2D histogram of weekly infection counts for each movement mode is shown in Figure 4-5. As expected, while all the scenarios have similar general patterns of weekly infection prevalence, each and every movement scenario has a quite distinctive number of infected humans at the end of simulation period. The average of the final outcome (i.e., total number of infected humans) for each mode is displayed in Table 4-1 on page

117. Additionally, using a number of two-sample t-tests, the difference between the mean outcome of every pair of scenarios is statistically significant at a very small alpha significance level i.e., around 1E-122.



a) Movement mode I (Static census distribution)

b) Movement modes II and III (Random)

c) Movement mode IV (Towers density based)

d) Movement mode V (Data-driven trajectory)

Figure 4-4 ABM screenshots of different movement modes showing density of human population spread

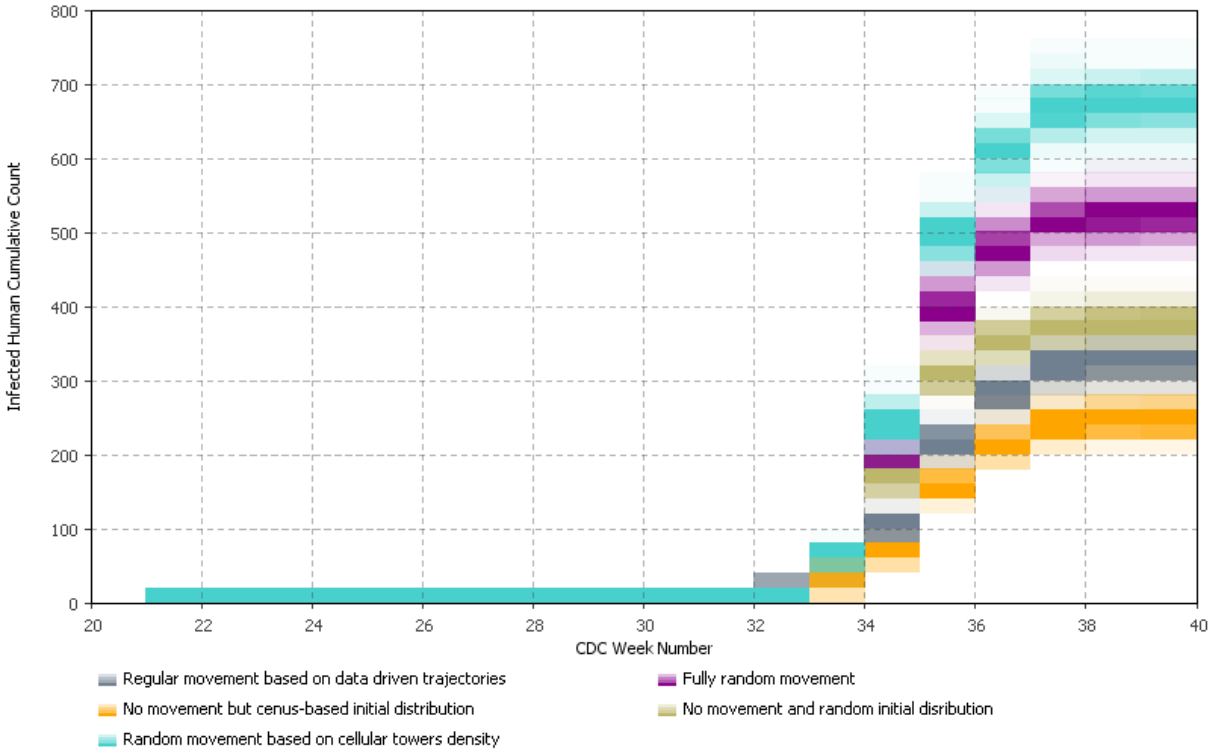


Figure 4-5 Weekly dynamics of infected population under five movement modes in the form of a 2D histogram

Table 4-1 Descriptive statistics of number of infected humans for different movement modes

Movement Mode	Average	Median	Variance
I. Static with Census distribution	241.725	241	243.5672
II. Static with random distribution	375.245	375	260.4975
III. Fully random movement	516.865	516	440.248
IV. Random but towers based movement	664.25	665	613.2035
V. Regular data-driven movement	317.885	317.5	276.1626

4.4 Summary of Chapter 4

4.4.1 Discussion and Limitations

Empirical WNV and mosquito trap data demonstrated the strength of the ABM to predict yearly trends of reported counts of WNV infected humans. However, the ABM, as it is now, cannot produce or justify years with a high number of mosquitoes but few infected people. Other than mosquito abundance, the results suggest that other crucial parameters affecting the transmission of virus to humans are present and must be accounted for. For example, winds may decrease the host-seeking activities of mosquitoes. Wearing protective clothing (such as long-sleeved shirts) can decrease the chances of getting bitten by mosquitoes. Another potentially important consideration could be the time that people spend outside, which also depends on weather. Given the ABM outputs, it may be even that patterns of human movements had slightly changed in some years, which affected the number of human cases. In general, the majority of prevention strategies e.g., applying mosquito repellants, spraying insecticide, larviciding, and minimizing exposure at peak time and places can all be well studied with this particular ABM. Moreover, other human activities with potential indirect impact on WNV transmission (e.g., increased bird reconnaissance) can easily be incorporated into ABMs.

The work here demonstrates that ABMs, in contrast to DE and other aggregative approaches, can be used to model mobility in WNV epidemiology. It was shown that the exact number of people infected is highly sensitive to where and when they spend time. This aspect of the WNV transmission can easily be examined using ABMs. Practitioners can benefit from employing ABMs to guide their decisions regarding prevention and mitigation strategies for humans. Statistical approaches can be employed to obtain general insight through the virus transmission cycle and to identify risk factors. However, in regard to mobile humans who can

move across a town or province, individual-based models can explicitly capture heterogeneity of contacts, resulting in detailed estimates of human infection patterns. Such tools could help identify key potential locations of disease transmission for effective surveillance and control strategies.

The work of this chapter used combinations of census data, cell tower connection times, and tower density as different measures of movements. There are a number of general sources of data that may be used to provide a measure of human mobility at such a large scale. These sources include usage of survey-based information, participatory Global Positioning System (GPS) trackers, and Cell data. Survey-based information needs cooperation of interviewees and may be used for validation of other data types [164]. It can be difficult and expensive to have everyone actively use GPS devices, although many people tend to share their location information, through social media. Cell data, in contrast, is already technically available in many countries, yet with a lower precision compared to GPS. Other than cell tower connection periods, tower traffic may be a reasonable option to be combined with the census distribution. Extracting and recording the aggregated tower traffic is routinely conducted by service providers. This data would improve the precision of cell data trajectories and may be considered as an option for modellers. Service providers are also becoming aware of the utility of providing anonymized or aggregated cell trajectory data as parts of public health responsibilities and individuals are becoming less reticent in providing their location based data for enhanced service or features.

4.4.2 Conclusions

An agent-based model of peripatetic humans is constructed to examine the spread of West Nile Virus infection. The ABM has cells of different and dynamic environmental conditions and habitats for *Cx. tarsalis* mosquitoes in Manitoba. The cellular difference equation at the heart of

the ABM considers multiple species of birds, their roosting locations and migration patterns. Humans with different susceptibility move across the province according to their cell tower trajectories as well as census data. The simulation outcomes have high correlation indices with actual reported WNV human cases data within the province. The importance of fine-grained movement of humans regarding human infection rate is shown through simulations. The work of this chapter suggests a fine-accuracy human movement model is necessary in order to simulate and properly assess non-pharmaceutical prevention strategies as is often unnoticed in WNV models.

Chapter 5: CONCLUDING REMARKS

Chapter one provides an introduction to WNV, its history and symptoms. Chapter one also reviews the current state of agent-based modelling within mosquito-borne diseases, and in particular WNV. These ABMs often capture one or more of mosquito population dynamics, virus transmission cycle, weather and landscape (or other mosquito-habitat) variations, different bird species' movement patterns, and humans' behaviour, outdoor/indoor activities, as well as their mobility. While there are a relatively high number of studies on both ABMs and WNV models on their own, very few WNV-ABMs exist.

Typically, as the number of components of an ABM increases, the individual agents are merged together and modeled as an aggregated agent. ABMs with aggregated agents are often hybrid with some means of controlling the dynamics of the aggregated group (e.g., a DE model). ABMs allow for fine-scale accuracy in prediction as well as less homogenous assumptions. However, these powerful systems are generally computationally expensive. Two other difficulties associated with ABMs are a robust validation procedure and extraction of appropriate and sufficient data from the phenomenon of interest. ABMs in general offer remarkable potential for a detailed understanding of complex systems.

WNV has been able to permanently establish itself in North America as it can infect birds, humans, horses, etc., giving the virus a wide variety of hosts. WNV is also tightly correlated with modelled human movement patterns, weather and habitat for various hosts and mainly vectors. Chapter two elaborates on the integration and compilation of data related to these components in Manitoba. The main technology used in chapter two is based on Anylogic and ArcGIS software. Anylogic simulation software in combination with Esri ArcGIS provides a powerful toolbox for

developers and modellers to simulate almost any GIS-based environment or process. The research of chapter two should be useful to others working on a variety of mosquito-borne diseases, such as Zika, Malaria, Dengue, and Chikungunya by introducing tools and providing a systematic way to extract the required data. In chapter two, different maps are combined together to create a grid land cover map of Manitoba, Canada in a shapefile format compatible with Anylogic in order to modulate mosquito parameters. A significant amount of data regarding 152 birds' species along with their population estimates and locations in Manitoba are gathered and assembled. Municipality shapefile maps are converted to Anylogic built-in GIS regions for better compatibility with census data and initial placement of human agents. Accessing shapefiles and their databases in Anylogic are also discussed.

The primary contribution of this thesis towards an enhanced understanding of WNV transmission is the proposal and validation of a comprehensive and efficient data-driven Cellular Difference Equation (CDiffE) structure for modelling WNV dynamics for adoption in WNV-ABMs, as elaborated upon in chapter three. ABMs are computationally intensive, and applying fast and simple algorithms (within each agent) is a must. Modelling WNV, compared to many other mosquito-transmitted diseases, such as Zika, Malaria, Dengue, and Chikungunya, is more complex as it includes three distinct types of agents: humans, mosquitoes, and various bird species. This makes it even more important to have a computationally efficient ABM for applying an agent-based modelling framework to WNV epidemiology. In chapter three, a computationally fast difference equation is adopted to lay the foundations for the CDiffE scheme and consequently a higher-level hybrid ABM. The CDiffE model on its own can be viewed as an early-stage deterministic ABM. It can be climate-driven and include multiple species of migratory and local birds and humans as hosts for WNV. The geographical output map of the

CDiffE model can be seen as a decision support tool for public health policymakers by providing risk maps of mosquito densities. The mosquito population trends within the CDiffE model over 15 weeks in approximately 30 communities of Manitoba has Pearson and Spearman correlation values of around 60% compared with the real-world trap data. In varying the CDiffE model parameters, some theoretically verified as well as counterintuitive findings were observed. Also, WNV-related parameters for different bird species are estimated according to biological studies on avian viremia in chapter three.

There is a growing interest in modelling infectious disease transmission often in order to ultimately attain better prevention strategies. A primary objective of this study was to design and implement an efficient ABM of WNV spread, considering highly-mobile humans with a high level of heterogeneous properties. The undeniable value of such an ABM lies in simulating and assessing virtually all epidemiological scenarios and strategies (e.g., prevention policies, larviciding practices, etc.) regarding a serious and infectious mosquito-borne disease. Mobile human agents are augmented to the CDiffE model, forming such a hybrid ABM for WNV transmission in chapter four. The proposed (hybrid) ABM was used to evaluate WNV prevalence under different scenarios of human movement patterns. The ABM outcome in chapter four confirms the importance and effects of human movements on WNV spread. The ABM revealed that much more accurate human movement models are required to effectively assess prevention strategies or even to correctly model WNV prevalence among people. Currently human mobility is among the least-concerned issues in WNV modelling studies. The work here clearly demonstrates the need for real-world data regarding human movement. This data is becoming increasingly available and eventually will be integrated into even more sophisticated agent based models for WNV as well as for other mosquito borne diseases.

There are limitations throughout this thesis, each of which was acknowledged and discussed in the corresponding section/chapter(s). The primary limitation of the work here is lack of sufficient and appropriate data in many parts to further tune the system. Despite adoption of a relatively fast difference equation as the core of the ABM, a secondary limitation is still computational limits of current hardware which made the system consider more homogenous assumptions regarding its components. A tertiary limitation is regarding the structure and formulation of CDiffE itself. There may exist better mathematical models to capture the impacts of bird species, weather, habitat, land cover, etc. on the dynamics of mosquitoes and WNV to be utilized at the core of the CDiffE. Alternatively, pre-trained neural networks or other complex function approximators may be adopted at the core of the ABM, trading speed for greater prediction accuracy.

PUBLICATIONS OUT OF THE THESIS

The following manuscripts are directly extracted from the content of this thesis and/or its appendices. At the time of writing this dissertation, some of these papers are under a review process.

1. H. R. Nasrinpour, Marcia R. Friesen and Robert D. McLeod, “Agent Based Modelling and West Nile Virus: A Survey,” in Journal of Medical and Biological Engineering, 2018 [URL: https://doi.org/10.1007/s40846-018-0396-8]
2. H. R. Nasrinpour, Alexander Reimer, Marcia R. Friesen and Robert D. McLeod, “Data Preparation for West Nile Virus Agent-Based Modelling: Protocol for Processing Bird Population Estimates and Incorporating ArcMap in AnyLogic,” in Journal of Medical Internet Research, Research Protocol, Vol 6, No 7 July 17, 2017 [URL: http://www.researchprotocols.org/2017/7/e138/]
3. H. R. Nasrinpour, A. Gumel, Marcia R. Friesen and Robert D. McLeod “Multi-species, Climate-driven and Landscape-based Cellular Difference Equation Model for West Nile Virus: Toward an Agent-Based Model of West Nile Virus Spread,” [under review]
4. H. R. Nasrinpour, Marcia R. Friesen and Robert D. McLeod, "Agent-Based Modelling of West Nile Virus Spread among Peripatetic Humans," [under review]
5. H. R. Nasrinpour, Marcia Friesen and Robert McLeod, “An Agent-Based Model of Message Propagation in the Facebook Electronic Social Network,” available in arXiv:1611.07454 [cs.SI] [URL: https://arxiv.org/abs/1611.07454]

REFERENCES

[1] R. Axelrod, *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration*, vol. 1. Princeton University Press, 1997.

[2] E. Bonabeau, "Agent-based modeling: methods and techniques for simulating human systems.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 99 Suppl 3, no. suppl 3, pp. 7280–7, May 2002.

[3] C. Macal and M. North, "Introductory tutorial: Agent-based modeling and simulation," in *Proceedings - Winter Simulation Conference*, 2015, vol. 2015–Janua, pp. 6–20.

[4] M. Gardner, "Mathematical games: The fantastic combinations of John Conway's new solitaire game 'life,'" *Sci. Am.*, 1970.

[5] B. C. P. Demianyk, "Development of Agent-Based Models for Healthcare: Applications and Critique," University of Manitoba, 2015.

[6] T. C. Schelling, "Dynamic Models of Segregation," *J. Math. Sociol.*, vol. 1, pp. 143–186, 1971.

[7] N. Gilbert, "Agent-Based Models," *SAGE Publ.*, vol. 153, no. 153, p. 98, 2008.

[8] D. Helbing, *Social Self-Organization: Agent-Based Simulations and Experiments to Study Emergent Social Behavior*. 2008.

[9] J. M. Epstein, *Generative Social Science: Studies in Agent-Based Computational Modeling*. Princeton: Princeton University Press, 2012.

[10] R. Axelrod and L. Tesfatsion, "Appendix A A Guide for Newcomers to Agent-Based Modeling in the Social Sciences," 2006, pp. 1647–1659.

[11] D. Helbing, Ed., *Agent-Based Simulations and Experiments to Study Emergent Social Behavior*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.

[12] T. Filatova, P. H. Verburg, D. C. Parker, and C. A. Stannard, "Spatial agent-based models for socio-ecological systems: Challenges and prospects," *Environ. Model. Softw.*, vol. 45, pp. 1–7, Jul. 2013.

[13] S. C. Bankes, "Agent-based modeling: a revolution?," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 99 Suppl 3, no. suppl 3, pp. 7199–200, May 2002.

[14] H. A. Hahn, "The Conundrum of Verification and Validation of Social Science-based Models," *Procedia Comput. Sci.*, vol. 16, pp. 878–887, 2013.

[15] C. M. Macal, "Everything you need to know about agent-based modelling and simulation," *J. Simul.*, vol. 10, no. 2, pp. 144–156, May 2016.

[16] K. C. Smithburn, T. P. Hughes, a W. Burke, J. H. Paul, and A. African, "A neurotropic virus isolated from the blood of a native of uganda1," *Am. J. Trop. Med. Hyg.*, vol. s1-20, pp. 471–492, 1940.

[17] H. G. Zeller and I. Schuffenecker, "West Nile virus: An overview of its spread in Europe and the Mediterranean basin in contrast to its spread in the Americas," *Eur. J. Clin. Microbiol. Infect. Dis.*, vol. 23, no. 3, pp. 147–156, 2004.

[18] D. Nash *et al.*, "The Outbreak of West Nile Virus Infection in the New York City Area in 1999," *N. Engl. J. Med.*, vol. 344, no. 24, pp. 1807–1814, 2001.

[19] K. Tachiiri, B. Klinkenberg, S. Mak, and J. Kazmi, "Predicting outbreaks: a spatial risk assessment of West Nile virus in British Columbia.," *Int. J. Health Geogr.*, vol. 5, p. 21, 2006.

[20] Government of Manitoba, "Surveillance for West Nile virus in Manitoba," *Manitoba Health, Seniors and Active Living*. [Online]. Available:

http://www.gov.mb.ca/health/wnv/stats.html. [Accessed: 15-Apr-2015].
[21] Health Canada, "West Nile virus," 10-Oct-2012. [Online]. Available: http://healthycanadians.gc.ca/diseases-conditions-maladies-affections/disease-maladie/wnv-vno-eng.php. [Accessed: 15-Apr-2015].
[22] Government of Manitoba, "West Nile virus fact sheet." [Online]. Available: http://www.gov.mb.ca/health/publichealth/factsheets/target.pdf. [Accessed: 15-Apr-2015].
[23] L. D. Kramer, L. M. Styer, and G. D. Ebel, "A global perspective on the epidemiology of West Nile virus.," *Annu. Rev. Entomol.*, vol. 53, pp. 61–81, 2008.
[24] American Mosquito Control Association, "Mosquito Life Cycle." [Online]. Available: http://www.mosquito.org/life-cycle. Archived at: http://www.webcitation.org/6iJntDZde. [Accessed: 15-Apr-2015].
[25] L. B. Goddard, A. E. Roth, W. K. Reisen, and T. W. Scott, "Vertical transmission of West Nile Virus by three California Culex (Diptera: Culicidae) species.," *J. Med. Entomol.*, vol. 40, no. 6, pp. 743–746, 2003.
[26] M. J. Turell, M. L. O'Guinn, D. J. Dohm, and J. W. Jones, "Vector competence of North American mosquitoes (Diptera: Culicidae) for West Nile virus," *J. Med. Entomol.*, vol. 38, no. 2, pp. 130–134, 2001.
[27] C. Bowman, A. B. Gumel, P. van den Driessche, J. Wu, and H. Zhu, "A mathematical model for assessing control strategies against West Nile virus.," *Bull. Math. Biol.*, vol. 67, no. 5, pp. 1107–33, Sep. 2005.
[28] M. J. Turell, D. J. Dohm, M. R. Sardelis, M. L. Oguinn, T. G. Andreadis, and J. a Blow, "An update on the potential of north American mosquitoes (Diptera: Culicidae) to transmit West Nile Virus.," *J. Med. Entomol.*, vol. 42, pp. 57–62, 2005.
[29] J. H. Rappole, B. W. Compton, P. Leimgruber, J. Robertson, D. I. King, and S. C. Renner, "Modeling movement of West Nile virus in the western hemisphere," *Vector-Borne Zoonotic Dis.*, vol. 6, pp. 128–139, 2006.
[30] V. Chevalier, A. Tran, and B. Durand, "Predictive modeling of west nile virus transmission risk in the mediterranean basin: How far from landing?," *Int. J. Environ. Res. Public Health*, vol. 11, pp. 67–90, 2013.
[31] M. O. Ruiz *et al.*, "Local impact of temperature and precipitation on West Nile virus infection in Culex species mosquitoes in northeast Illinois, USA.," *Parasit. Vectors*, vol. 3, no. 1, p. 19, Jan. 2010.
[32] S. H. Paull *et al.*, "Drought and immunity determine the intensity of West Nile virus epidemics and climate change impacts," *Proc. R. Soc. B Biol. Sci.*, vol. 284, no. 1848, p. 20162078, Feb. 2017.
[33] H. Dieng *et al.*, "The effects of simulated rainfall on immature population dynamics of Aedes albopictus and female oviposition," *Int. J. Biometeorol.*, vol. 56, no. 1, pp. 113–120, 2012.
[34] W. H. Cooke, K. Grala, and R. C. Wallis, "Avian GIS models signal human risk for West Nile virus in Mississippi.," *Int. J. Health Geogr.*, vol. 5, p. 36, 2006.
[35] P. R. Epstein and C. Defilippo, "West Nile Virus and drought," *Glob. Chang. Hum. Heal.*, vol. 2, no. 2, pp. 105–107, 2001.
[36] J. Shaman, J. F. Day, and M. Stieglitz, "Drought-induced amplification and epidemic transmission of West Nile virus in southern Florida.," *J. Med. Entomol.*, vol. 42, pp. 134–141, 2005.
[37] P. P. Marra *et al.*, "West Nile Virus and Wildlife," *Bioscience*, vol. 54, no. 5, p. 393, 2004.

[38] A. J. Cornel, P. G. Jupp, and N. K. Blackburn, “Environmental temperature on the vector competence of Culex univittatus (Diptera: Culicidae) for West Nile virus.,” *J. Med. Entomol.*, vol. 30, no. 2, pp. 449–56, Mar. 1993.
[39] D. J. Dohm, M. L. O’Guinn, and M. J. Turell, “Effect of environmental temperature on the ability of Culex pipiens (Diptera: Culicidae) to transmit West Nile virus.,” *J. Med. Entomol.*, vol. 39, no. 1, pp. 221–5, Jan. 2002.
[40] W. K. Reisen, D. Cayan, M. Tyree, C. M. Barker, B. Eldridge, and M. Dettinger, “Impact of climate variation on mosquito abundance in California,” *J. Vector Ecol.*, vol. 33, no. 1, pp. 89–98, Jun. 2008.
[41] W. K. Reisen, Y. Fang, and V. M. Martinez, “Effects of temperature on the transmission of west nile virus by Culex tarsalis (Diptera: Culicidae).,” *J. Med. Entomol.*, vol. 43, no. 2, pp. 309–317, Mar. 2006.
[42] B. L. Dodson, L. D. Kramer, and J. L. Rasgon, “Effects of larval rearing temperature on immature development and West Nile virus vector competence of Culex tarsalis,” *Parasit. Vectors*, vol. 5, no. 1, p. 199, 2012.
[43] C. C. Chen, T. Epp, E. Jenkins, C. Waldner, P. S. Curry, and C. Soos, “Modeling monthly variation of Culex tarsalis (Diptera: Culicidae) abundance and West Nile virus infection rate in the Canadian prairies,” *Int. J. Environ. Res. Public Health*, vol. 10, pp. 3033–3051, 2013.
[44] A. T. Ciota, A. C. Matacchiero, A. M. Kilpatrick, and L. D. Kramer, “The effect of temperature on life history traits of Culex mosquitoes.,” *J. Med. Entomol.*, vol. 51, no. 1, pp. 55–62, Jan. 2014.
[45] L. Zou, S. N. Miller, and E. T. Schmidtmann, “A GIS tool to estimate west nile virus risk based on a degree-day model,” *Environ. Monit. Assess.*, vol. 129, pp. 413–420, 2007.
[46] R. Bargaoui, S. Lecollinet, and R. Lancelot, “Mapping the serological prevalence rate of west nile fever in equids, Tunisia.,” *Transbound. Emerg. Dis.*, vol. 62, no. 1, pp. 55–66, Feb. 2015.
[47] A. Jutla, A. Huq, and R. R. Colwell, “Diagnostic Approach for Monitoring Hydroclimatic Conditions Related to Emergence of West Nile Virus in West Virginia,” *Front. Public Heal.*, vol. 3, p. 10, Feb. 2015.
[48] T.-W. Chuang, C. W. Hockett, L. Kightlinger, and M. C. Wimberly, “Landscape-Level Spatial Patterns of West Nile Virus Risk in the Northern Great Plains,” *Am. J. Trop. Med. Hyg.*, vol. 86, no. 4, pp. 724–731, Apr. 2012.
[49] F. Ahmadnejad *et al.*, “Spread of West Nile virus in Iran: a cross-sectional serosurvey in equines, 2008-2009.,” *Epidemiol. Infect.*, vol. 139, no. 10, pp. 1587–93, Oct. 2011.
[50] M. P. Ward, “Equine West Nile virus disease occurrence and the Normalized Difference Vegetation Index.,” *Prev. Vet. Med.*, vol. 88, no. 3, pp. 205–12, Mar. 2009.
[51] M. Bouden, B. Moulin, and P. Gosselin, “The geosimulation of West Nile virus propagation: a multi-agent and climate sensitive tool for risk management in public health.,” *Int. J. Health Geogr.*, vol. 7, p. 35, 2008.
[52] Z. Li, J. Hayse, I. Hlohowskyj, K. Smith, and R. Smith, “Agent-based model for simulation of West Nile virus transmission,” in *The agent 2005 conference on generative social processes, models, and mechanisms*, 2005, pp. 459–472.
[53] E. B. Hayes, N. Komar, R. S. Nasci, S. P. Montgomery, D. R. O’Leary, and G. L. Campbell, “Epidemiology and transmission dynamics of West Nile virus disease.,” *Emerg. Infect. Dis.*, vol. 11, no. 8, pp. 1167–73, Aug. 2005.

[54] T. M. Colpitts, M. J. Conway, R. R. Montgomery, and E. Fikrig, “West Nile virus: Biology, transmission, and human infection,” *Clin. Microbiol. Rev.*, vol. 25, no. 4, pp. 635–648, 2012.
[55] P. Curry, “Saskatchewan Mosquitoes and West Nile Virus,” *Blue Jay*, no. 62, pp. 104–111, 2004.
[56] A. M. Kilpatrick, L. D. Kramer, S. R. Campbell, E. O. Alleyne, A. P. Dobson, and P. Daszak, “West Nile virus risk assessment and the bridge vector paradigm,” *Emerg. Infect. Dis.*, vol. 11, no. 3, pp. 425–429, 2005.
[57] M. J. Turell, M. R. Sardelis, M. L. O’Guinn, and D. J. Dohm, “Potential vectors of West Nile virus in North America.,” *Curr. Top. Microbiol. Immunol.*, vol. 267, pp. 241–52, 2002.
[58] G. L. Hamer *et al.*, “Culex pipiens (Diptera: Culicidae): a bridge vector of West Nile virus to humans.,” *J. Med. Entomol.*, vol. 45, no. 1, pp. 125–8, Jan. 2008.
[59] a. M. Kilpatrick, “Globalization, Land Use, and the Invasion of West Nile Virus,” *Science (80-.).*, vol. 334, no. 6054, pp. 323–327, 2011.
[60] G. L. Hamer *et al.*, “Host selection by Culex pipiens mosquitoes and west nile virus amplification,” *Am. J. Trop. Med. Hyg.*, vol. 80, no. 2, pp. 268–278, 2009.
[61] a M. Kilpatrick, P. Daszak, M. J. Jones, P. P. Marra, and L. D. Kramer, “Host heterogeneity dominates West Nile virus transmission.,” *Proc. Biol. Sci.*, vol. 273, no. May, pp. 2327–2333, 2006.
[62] G. L. Hamer *et al.*, “Fine-scale variation in vector host use and force of infection drive localized patterns of West Nile virus transmission,” *PLoS One*, vol. 6, no. 8, 2011.
[63] R. Kent, L. Juliusson, M. Weissmann, S. Evans, and N. Komar, “Seasonal Blood-Feeding Behavior of Culex tarsalis (Diptera: Culicidae) in Weld County, Colorado, 2007,” *J. Med. Entomol*, vol. 46, no. 2, pp. 380–390, 2009.
[64] G. Molaei *et al.*, “Host feeding pattern of Culex quinquefasciatus (Diptera: Culicidae) and its role in transmission of West Nile virus in Harris County, Texas,” *Am. J. Trop. Med. Hyg.*, vol. 77, no. 1, pp. 73–81, 2007.
[65] G. Molaei *et al.*, “Vector-host interactions governing epidemiology of West Nile virus in southern California,” *Am. J. Trop. Med. Hyg.*, vol. 83, no. 6, pp. 1269–1282, 2010.
[66] M. P. Ward, A. Raim, S. Yaremych-Hamer, R. Lampman, and R. J. Novak, “Does the roosting behavior of birds affect transmission dynamics of West Nile virus?,” *Am. J. Trop. Med. Hyg.*, vol. 75, no. 2, pp. 350–5, Aug. 2006.
[67] B. L. Krebs *et al.*, “Host group formation decreases exposure to vector-borne disease: a field experiment in a ‘hotspot’ of West Nile virus transmission,” *Proc. R. Soc. London B Biol. Sci.*, vol. 281, no. 1796, 2014.
[68] W. M. Janousek, P. P. Marra, and A. M. Kilpatrick, “Avian roosting behavior influences vector-host interactions for West Nile virus hosts.,” *Parasit. Vectors*, vol. 7, no. 1, p. 399, Aug. 2014.
[69] G. Molaei, T. G. Andreadis, P. M. Armstrong, J. F. Anderson, and C. R. Vossbrinck, “Host Feeding Patterns of Culex Mosquitoes and West Nile Virus Transmission , Northeastern United States,” vol. 12, no. 3, 2006.
[70] J. Figuerola *et al.*, “Size matters: West Nile Virus neutralizing antibodies in resident and migratory birds in Spain.,” *Vet. Microbiol.*, vol. 132, no. 1–2, pp. 39–46, Nov. 2008.
[71] T. C. Thiemann, D. a Lemenager, S. Kluh, B. D. Carroll, H. D. Lothrop, and W. K. Reisen, “Spatial variation in host feeding patterns of Culex tarsalis and the Culex pipiens

complex (Diptera: Culicidae) in California.," *J. Med. Entomol.*, vol. 49, no. 4, pp. 903–16, 2012.
[72] R. Campbell, T. C. Thiemann, D. Lemenager, and W. K. Reisen, "Host-Selection Patterns of Culex tarsalis (Diptera : Culicidae) Determine the Spatial Heterogeneity of West Nile Virus Enzootic Activity in Northern California Host-Selection Patterns of Culex tarsalis (Diptera : Culicidae) Determine the Spatial Heter," *J. Med. Entomol.*, vol. 50, no. 6, pp. 1303–9, Nov. 2013.
[73] J. E. Simpson *et al.*, "Vector host-feeding preferences drive transmission of multi-host pathogens: West Nile virus as a model system," *Proc. R. Soc. B Biol. Sci.*, vol. 279, no. August 2011, pp. 925–933, 2012.
[74] A. M. M. Kilpatrick, S. L. S. L. LaDeau, and P. P. P. Marra, "Ecology of West Nile virus transmission and its impact on birds in the western hemisphere," *Auk*, vol. 124, no. 4, pp. 1121–1136, 2007.
[75] L. F. Chaves, L. C. Harrington, C. L. Keogh, A. M. Nguyen, and U. D. Kitron, "Blood feeding patterns of mosquitoes: random or structured?," *Front. Zool.*, vol. 7, no. 1, p. 3, Jan. 2010.
[76] J. Goddard, "Setting up a mosquito control program," *Bureau of General Environmental Services*, Mississippi State Department of Health, pp. 1–42, 2003.
[77] A. R. Barr, "Occurrence and distribution of the Culex pipiens complex.," *Bull. World Health Organ.*, vol. 37, no. 2, pp. 293–296, 1967.
[78] C. S. Apperson *et al.*, "Host-Feeding Habits of *Culex* and Other Mosquitoes (Diptera: Culicidae) in the Borough of Queens in New York City, with Characters and Techniques for Identification of *Culex* Mosquitoes," *J. Med. Entomol.*, vol. 39, no. 5, pp. 777–785, 2002.
[79] North Dakota Grand Forks Public Health, "Local Mosquito Species." [Online]. Available: http://www.gfmosquito.com/local-mosquito-species/. [Accessed: 21-Apr-2015].
[80] N. W. Yiannakoulias, D. P. Schopflocher, and L. W. Svenson, "Modelling Geographic Variations in West Nile Virus," *Can. J. Public Heal.*, vol. 97, pp. 374–378, 2006.
[81] A. C. Chen *et al.*, "Predicting Weekly Variation of Culex tarsalis (Diptera : Culicidae) West Nile Virus Infection in a Newly Endemic Region , the Canadian Prairies Predicting Weekly Variation of Culex tarsalis (Diptera : Culicidae) West Nile Virus Infection in a Newly En," 2012.
[82] W. C. Reeves, R. E. Bellamy, C. H. Tempelis, and M. F. Lofy, "A Three-Year Study of the Feeding Habits of Culex tarsalis in Kern County, California," *Am. J. Trop. Med. Hyg.*, vol. 14, no. 1, pp. 170–177, Jan. 1965.
[83] W. Reisen, "The western encephalitis mosquito, Culex tarsalis," *Wing Beats*, vol. 4, no. 2, p. 16, 1993.
[84] E. Hearle, "The Mosquitoes of the Lower Fraser Valley, British Columbia, and their Control.," *Mosquitoes Low. Fraser Val. Br. Columbia, their Control.*, no. 17, 1926.
[85] J. McLintock, "The Mosquitoes of the Greater Winnipeg area," *Can. Entomol.*, vol. 76, no. 05, pp. 89–104, May 1944.
[86] V. Rodríguez-Prieto, B. Martínez-López, M. Martínez, M. J. Muñoz, and J. M. Sánchez-Vizcaíno, "Identification of suitable areas for West Nile virus outbreaks in equid populations for application in surveillance plans: the example of the Castile and Leon region of Spain.," *Epidemiol. Infect.*, vol. 140, no. 9, pp. 1617–31, Sep. 2012.
[87] T. Balenghien, F. Fouque, P. Sabatier, and D. J. Bicout, "Theoretical formulation for

mosquito host-feeding patterns: application to a West Nile virus focus of southern France.," *J. Med. Entomol.*, vol. 48, no. 5, pp. 1076–90, Sep. 2011.
[88] D. M. Thomas and B. Urena, "A model describing the evolution of West Nile-like encephalitis in New York City," *Math. Comput. Model.*, vol. 34, no. 7–8, pp. 771–781, 2001.
[89] M. J. Wonham, T. De-Camino-Beck, and M. a Lewis, "An epidemiological model for West Nile virus: invasion analysis and control applications.," *Proc. Biol. Sci.*, vol. 271, no. October 2003, pp. 501–507, 2004.
[90] G. Cruz-Pacheco, L. Esteva, J. A. Montaño-Hirose, and C. Vargas, "Modelling the dynamics of West Nile Virus," *Bull. Math. Biol.*, vol. 67, no. 6, pp. 1157–1172, Nov. 2005.
[91] S. Funk, M. Salathé, and V. A. A. Jansen, "Modelling the influence of human behaviour on the spread of infectious diseases: a review," *J. R. Soc. Interface*, vol. 7, no. 50, 2010.
[92] M. Salathé and J. H. Jones, "Dynamics and Control of Diseases in Networks with Community Structure," *PLoS Comput. Biol.*, vol. 6, no. 4, p. e1000736, Apr. 2010.
[93] M. R. Friesen and R. D. McLeod, "A Survey of Agent-Based Modeling of Hospital Environments," *IEEE Access*, vol. 2, pp. 227–233, 2014.
[94] M. Laskowski, B. C. P. Demianyk, J. Witt, S. N. Mukhi, M. R. Friesen, and R. D. McLeod, "Agent-based modeling of the spread of influenza-like illness in an emergency department: a simulation study.," *IEEE Trans. Inf. Technol. Biomed.*, vol. 15, no. 6, pp. 877–89, Nov. 2011.
[95] R. Neighbour, L. Oppenheimer, S. N. Mukhi, M. R. Friesen, and R. D. McLeod, "Agent based modeling of 'crowdinforming' as a means of load balancing at emergency departments.," *Online J. Public Health Inform.*, vol. 2, no. 3, 2010.
[96] B. G. Silverman, N. Hanrahan, G. Bharathy, K. Gordon, and D. Johnson, "A systems approach to healthcare: agent-based modeling, community mental health, and population well-being.," *Artif. Intell. Med.*, vol. 63, no. 2, pp. 61–71, Feb. 2015.
[97] B. Cummins, R. Cortez, I. M. Foppa, J. Walbeck, and J. M. Hyman, "A spatial model of mosquito host-seeking behavior," *PLoS Comput. Biol.*, vol. 8, no. 5, 2012.
[98] S. J. de Almeida, R. P. Martins Ferreira, Á. E. Eiras, R. P. Obermayr, and M. Geier, "Multi-agent modeling and simulation of an Aedes aegypti mosquito population," *Environ. Model. Softw.*, vol. 25, no. 12, pp. 1490–1507, 2010.
[99] D. L. Chao, S. B. Halstead, M. E. Halloran, and I. M. Longini, "Controlling Dengue with Vaccines in Thailand," *PLoS Negl. Trop. Dis.*, vol. 6, no. 10, p. e1876, Oct. 2012.
[100] P. A. Eckhoff, "A malaria transmission-directed model of mosquito life cycle and ecology," *Malar. J.*, vol. 10, no. 1, p. 303, 2011.
[101] S. Arifin, Y. Zhou, G. J. Davis, J. E. Gentile, G. R. Madey, and F. H. Collins, "An agent-based model of the population dynamics of Anopheles gambiae," *Malar. J.*, vol. 13, no. 1, p. 424, 2014.
[102] C. Gunaratne, M. I. Akbas, I. Garibay, and O. Ozmen, "Evaluation of Zika Vector Control Strategies Using Agent-Based Modeling," 2016.
[103] S. M. Mniszewski, C. A. Manore, C. Bryan, S. Y. Del Valle, and D. Roberts, "Towards a Hybrid Agent-based Model for Mosquito Borne Disease," in *Summer Computer Simulation Conference (SCSC 2014)*, 2014, p. 10.
[104] C. A. Manore *et al.*, "A network-patch methodology for adapting agent-based models for directly transmitted disease to mosquito-borne disease," *J. Biol. Dyn.*, vol. 9, no. 1, pp.

52–72, Jan. 2015.
[105] B. Adams and D. D. Kapan, “Man Bites Mosquito: Understanding the Contribution of Human Movement to Vector-Borne Disease Dynamics,” *PLoS One*, vol. 4, no. 8, p. e6763, Aug. 2009.
[106] T. A. Perkins, T. W. Scott, A. Le Menach, and D. L. Smith, “Heterogeneity, Mixing, and the Spatial Scales of Mosquito-Borne Pathogen Transmission,” *PLoS Comput. Biol.*, vol. 9, no. 12, p. e1003327, Dec. 2013.
[107] H. Padmanabha, D. Durham, F. Correa, M. Diuk-Wasser, and A. Galvani, “The Interactive Roles of Aedes aegypti Super-Production and Human Density in Dengue Transmission,” *PLoS Negl. Trop. Dis.*, vol. 6, no. 8, p. e1799, Aug. 2012.
[108] J. Régnière, R. St-Amant, and A. Béchard, “BioSIM 10 – User’s manual,” *Natl. Resour. Canada, Can. For. Serv., Info Rep LAU-X-155*, 2014.
[109] D. J. Madder, G. A. Surgeoner, and B. V Helson, “Number of generations, egg production, and developmental time of Culex pipiens and Culex restauns (Diptera: Culicidae) in southern Ontario.,” *J. Med. Entomol.*, vol. 20, no. 3, pp. 275–87, May 1983.
[110] J. R. Sauer, J. E. Hines, K. L. Fallon, D. J. Pardieck, J. Ziolkowski, and W. A. Link., “The North American Breeding Bird Survey, Results and Analysis 1966 - 2013,” *Version 01.30.2015 USGS Patuxent Wildlife Research Center, Laurel, MD*, 2014.
[111] W. T. Reeves, “Particle systems---a technique for modeling a class of fuzzy objects,” *ACM SIGGRAPH Comput. Graph.*, vol. 17, no. 2, pp. 359–375, 1983.
[112] A. Borshchev, *The Big Book of Simulation Modeling: Multimethod Modeling with AnyLogic 6*. AnyLogic North America, 2013.
[113] K. R. Dahal and T. E. Chow, “A GIS toolset for automated partitioning of urban lands,” *Environ. Model. Softw.*, vol. 55, pp. 222–234, May 2014.
[114] D. Fecht, L. Beale, and D. Briggs, “A GIS-based urban simulation model for environmental health analysis,” *Environ. Model. Softw.*, vol. 58, pp. 1–11, Aug. 2014.
[115] G. Formetta, A. Antonello, S. Franceschi, O. David, and R. Rigon, “Hydrological modelling with components: A GIS-based open-source framework,” *Environ. Model. Softw.*, vol. 55, pp. 190–200, May 2014.
[116] L. Laranjo, D. Rodrigues, A. M. Pereira, R. T. Ribeiro, and J. M. Boavida, “Use of Electronic Health Records and Geographic Information Systems in Public Health Surveillance of Type 2 Diabetes: A Feasibility Study,” *JMIR Public Heal. Surveill.*, vol. 2, no. 1, p. e12, Mar. 2016.
[117] P. M. Frew *et al.*, “An Integrated Service Delivery Model to Identify Persons Living with HIV and to Provide Linkage to HIV Treatment and Care in Prioritized Neighborhoods: A Geotargeted, Program Outcome Study.,” *JMIR public Heal. Surveill.*, vol. 1, no. 2, p. e16, 2015.
[118] Government of Manitoba, “Land Use / Land Cover, Manitoba Land Initiative.” [Online]. Available: http://mli2.gov.mb.ca/landuse/index.html. Archived at: http://www.webcitation.org/6lCu8bYBT. [Accessed: 29-Sep-2015].
[119] Manitoba Breeding Bird Atlas, “Manitoba Breeding Bird Atlas (2010-2014): raw breeding evidence,” *Bird Studies Canada*, 2014. [Online]. Available: http://www.birdatlas.mb.ca/index_en.jsp. [Accessed: 17-May-2016].
[120] T. D. Rich *et al.*, “Partners in Flight North American Landbird Conservation Plan (PIF Population Estimates Database),” 2004.
[121] Bird Studies Canada and NABCI, “Bird Conservation Regions,” 2014.

[122] B. Carey *et al.*, *Finding Birds In Southern Manitoba*. Manitoba Naturalists Society, 2006.
[123] BAM Project Team, "BAM Effective Detection Radius," *Boreal Avian Modelling Project*, 2010. [Online]. Available: http://www.borealbirds.ca/index.php/bam_edr Archived at: http://www.webcitation.org/6lCvGYU0G. [Accessed: 17-May-2016].
[124] Manitoba Breeding Bird Atlas, "Guide For Point Counters," Manitoba Breeding Bird Atlas, Winnipeg, 2010.
[125] P. Rodewald, Ed., *The Birds of North America Online*. Ithaca, NY: Cornell Laboratory of Ornithology, 2015.
[126] D. Murray-Rust, D. T. Robinson, E. Guillem, E. Karali, and M. Rounsevell, "An open framework for agent based modelling of agricultural land use change," *Environ. Model. Softw.*, vol. 61, pp. 19–38, Nov. 2014.
[127] C. Downes and B. T. Collins, "The Canadian Breeding Bird Survey, 1966-94," *Can. Wildl. Serv. Prog. Note No. 210.*, p. 36, 1996.
[128] D. A. Kirk, Antony W Diamond, Alan R Smith, George E Holland, and Paul Chytyk, "Population Changes in Boreal Forest Birds in Saskatchewan and Manitoba," *Wilson Bull.*, vol. 109, no. 1, pp. 1–27, 1997.
[129] M. S. Majumder, M. Santillana, S. R. Mekaru, D. P. McGinnis, K. Khan, and J. S. Brownstein, "Utilizing Nontraditional Data Sources for Near Real-Time Estimation of Transmission Dynamics During the 2015-2016 Colombian Zika Virus Disease Outbreak," *JMIR Public Heal. Surveill.*, vol. 2, no. 1, p. e30, Jun. 2016.
[130] D. A. Ewing, C. A. Cobbold, B. V Purse, M. A. Nunn, and S. M. White, "Modelling the effect of temperature on the seasonal population dynamics of temperate mosquitoes.," *J. Theor. Biol.*, vol. 400, pp. 65–79, Jul. 2016.
[131] G. Marini, P. Poletti, M. Giacobini, A. Pugliese, S. Merler, and R. Rosà, "The Role of Climatic and Density Dependent Factors in Shaping Mosquito Population Dynamics: The Case of Culex pipiens in Northwestern Italy.," *PLoS One*, vol. 11, no. 4, p. e0154018, 2016.
[132] A. Abdelrazec and A. B. Gumel, "Mathematical Assessment of the Role of Temperature and Rainfall on Mosquito Population Dynamics," *J. Math. Biol.*, p. 42, Sep. 2016.
[133] K. Okuneye and A. B. Gumel, "Analysis of a temperature- and rainfall-dependent model for malaria transmission dynamics," *Math. Biosci.*, Apr. 2016.
[134] M. A. Lewis, J. Renclawowicz, P. Van Den Driessche, and M. Wonham, "A comparison of continuous and discrete-time West Nile virus models," *Bull. Math. Biol.*, vol. 68, no. 3, pp. 491–509, 2006.
[135] S. E. Bowden, K. Magori, and J. M. Drake, "Regional differences in the association between land cover and West Nile virus disease incidence in humans in the United States," *Am. J. Trop. Med. Hyg.*, vol. 84, no. 2, pp. 234–238, 2011.
[136] J. R. Anderson, E. E. Hardy, J. T. Roach, and R. E. Witmer, "A land use and land cover classification system for use with remote sensor data," *Prof. Pap.*, 1976.
[137] N. Komar *et al.*, "Experimental infection of North American birds with the New York 1999 strain of West Nile virus," *Emerg. Infect. Dis.*, vol. 9, no. 3, pp. 311–322, 2003.
[138] T. C. Thiemann, S. S. Wheeler, C. M. Barker, and W. K. Reisen, "Mosquito host selection varies seasonally with host availability and mosquito density.," *PLoS Negl. Trop. Dis.*, vol. 5, no. 12, p. e1452, Dec. 2011.
[139] W. K. Reisen, H. D. Lothrop, and T. Thiemann, "Host selection patterns of Culex tarsalis (Diptera: Culicidae) at wetlands near the Salton Sea, Coachella Valley, California, 1998-

2002.,” *J. Med. Entomol.*, vol. 50, no. 5, pp. 1071–6, Sep. 2013.
[140] A. M. Kilpatrick *et al.*, “Predicted and observed mortality from vector-borne disease in small songbirds.,” *Biol. Conserv.*, vol. 165, pp. 79–85, Sep. 2013.
[141] W. K. Reisen and D. C. Hahn, “Comparison of immune responses of brown-headed cowbird and related blackbirds to west Nile and other mosquito-borne encephalitis viruses.,” *J. Wildl. Dis.*, vol. 43, no. 3, pp. 439–49, Jul. 2007.
[142] W. K. Reisen, Y. Fang, and V. M. Martinez, “Avian host and mosquito (Diptera: Culicidae) vector competence determine the efficiency of West Nile and St. Louis encephalitis virus transmission.,” *J. Med. Entomol.*, vol. 42, no. 3, pp. 367–75, May 2005.
[143] N. Nemeth, D. Gould, R. Bowen, and N. Komar, “Natural and experimental West Nile virus infection in five raptor species.,” *J. Wildl. Dis.*, vol. 42, no. 1, pp. 1–13, Jan. 2006.
[144] Y. Fang and W. K. Reisen, “Previous infection with West Nile or St. Louis encephalitis viruses provides cross protection during reinfection in house finches.,” *Am. J. Trop. Med. Hyg.*, vol. 75, no. 3, pp. 480–5, Sep. 2006.
[145] W. K. Reisen, Y. Fang, and V. Martinez, “Is nonviremic transmission of West Nile virus by Culex mosquitoes (Diptera: Culicidae) nonviremic?,” *J. Med. Entomol.*, vol. 44, no. 2, pp. 299–302, Mar. 2007.
[146] N. Komar *et al.*, “Avian hosts for West Nile virus in St. Tammany Parish, Louisiana, 2002.,” *Am. J. Trop. Med. Hyg.*, vol. 73, no. 6, pp. 1031–7, Dec. 2005.
[147] S. A. Langevin, A. C. Brault, N. A. Panella, R. A. Bowen, and N. Komar, “Variation in virulence of West Nile virus strains for house sparrows (Passer domesticus).,” *Am. J. Trop. Med. Hyg.*, vol. 72, no. 1, pp. 99–102, Jan. 2005.
[148] W. K. Reisen *et al.*, “Ability of transstadially infected Ixodes pacificus (Acari: Ixodidae) to transmit West Nile virus to song sparrows or western fence lizards.,” *J. Med. Entomol.*, vol. 44, no. 2, pp. 320–7, Mar. 2007.
[149] L. Clark *et al.*, “Susceptibility of greater sage-grouse to experimental infection with West Nile virus.,” *J. Wildl. Dis.*, vol. 42, no. 1, pp. 14–22, Jan. 2006.
[150] Nicholas Komar, “West Nile Virus: Epidemiology and Ecology in North America,” *Adv. Virus Res.*, vol. 61, pp. 185–234, 2003.
[151] M. S. Godsey Jr., K. Burkhalter, M. Delorey, and H. M. Savage, “Seasonality and time of host-seeking activity of Culex tarsalis and floodwater Aedes in northern Colorado, 2006-2007,” *J Am Mosq Control Assoc*, vol. 26, no. 2, pp. 148–159, 2010.
[152] W. K. Reisen, H. D. Lothrop, and R. P. Meyer, “Time of host-seeking by Culex tarsalis (Diptera:Culicidae) in California.,” *J. Med. Entomol.*, vol. 34, no. 4, pp. 430–7, Jul. 1997.
[153] H. R. Nasrinpour, A. Massah Bavani, and M. Teshnehlab, “Grouped Bees Algorithm: A Grouped Version of the Bees Algorithm,” *Computers*, vol. 6, no. 1, 2017.
[154] M. Laguna, “Optimization of complex systems with OptQuest,” *Univ. Color.*, 1997.
[155] W. K. Reisen, M. M. Milby, and R. P. Meyer, “Population dynamics of adult Culex mosquitoes (Diptera: Culicidae) along the Kern River, Kern County, California, in 1990.,” *J. Med. Entomol.*, vol. 29, no. 3, pp. 531–43, May 1992.
[156] L. Pan, L. Qin, S. X. Yang, and J. Shuai, “A Neural Network-Based Method for Risk Factor Analysis of West Nile Virus,” *Risk Anal.*, vol. 28, no. 2, pp. 487–496, Apr. 2008.
[157] A. N. Clements, *The Biology of Mosquitoes: Sensory, reception, and behaviour*. Chapman & Hall, 1992.
[158] Prairie Pest Monitoring Network, “Drought Watch - Map Archive,” *Agriculture and Agri-Food Canada*. [Online]. Available: http://www.agr.gc.ca/DW-GS/historical-

historiques.jspx?lang=eng&jsEnabled=true.
[159] J. C. Allen, “A Modified Sine Wave Method for Calculating Degree Days,” *Environ. Entomol.*, vol. 5, no. 3, 1976.
[160] M. C. González, C. A. Hidalgo, and A.-L. Barabási, “Understanding individual human mobility patterns,” *Nature*, vol. 453, no. 7196, pp. 779–782, Jun. 2008.
[161] V. Colizza, A. Barrat, M. Barthelemy, A.-J. Valleron, and A. Vespignani, “Modeling the Worldwide Spread of Pandemic Influenza: Baseline Case and Containment Interventions,” *PLoS Med.*, vol. 4, no. 1, p. e13, Jan. 2007.
[162] L. Hufnagel, D. Brockmann, and T. Geisel, “Forecast and control of epidemics in a globalized world.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, no. 42, pp. 15124–9, Oct. 2004.
[163] S. Eubank *et al.*, “Modelling disease outbreaks in realistic urban social networks,” *Nature*, vol. 429, no. 6988, pp. 180–184, May 2004.
[164] S. T. Stoddard *et al.*, “The Role of Human Movement in the Transmission of Vector-Borne Pathogens,” *PLoS Negl. Trop. Dis.*, vol. 3, no. 7, p. e481, Jul. 2009.
[165] R. C. Reiner, S. T. Stoddard, and T. W. Scott, “Socially structured human movement shapes dengue transmission despite the diffusive effect of mosquito dispersal,” *Epidemics*, vol. 6, pp. 30–36, 2014.
[166] P. Martens and L. Hall, “Malaria on the move: human population movement and malaria transmission,” *Emerg. Infect. Dis.*, vol. 6, no. 2, pp. 103–9, 2000.
[167] S. C. Government of Canada, “Statistics Canada: 2011 Census Profile.”
[168] F. Aurenhammer and Franz, “Voronoi diagrams---a survey of a fundamental geometric data structure,” *ACM Comput. Surv.*, vol. 23, no. 3, pp. 345–405, Sep. 1991.

APPENDICES

Appendix A: Sample Java Codes and Bird Species Data

A.1 Sample Codes

A.1.1 Shapefile Database Connection String

The correct connection string for connecting to the database of a shapefile depends on the OS and the dBase associated drivers installed on it. In the example here, the OS is a 64-bit Microsoft Windows with a 64-bit Access Database Engine installed on it. If Anylogic is running on a 32-bit Java virtual machine, the required database driver must exist in the 32-bit version of ODBC data sources, and vice-versa for the 64-bit versions. This driver could be either MS Access dBase drive or FoxPro driver. The following Java example code is using the MS Access driver.

```
try {
    Class.forName("sun.jdbc.odbc.JdbcOdbcDriver");
    String connString = "jdbc:odbc:Driver={Microsoft Access dBASE Driver
(*.dbf, *.ndx, *.mdx)}; DefaultDir=D:\\";
    // D:\ is the database location
    // Microsoft Access dBASE Driver (*.dbf, *.ndx, *.mdx) exists in Data
Sources (ODBC)
    java.sql.Connection conn =
java.sql.DriverManager.getConnection(connString);
    String sql="SELECT * from Grid"; // Grid.dbf is the file name
    java.sql.Statement stmt=conn.createStatement();
    java.sql.ResultSet resultSet=stmt.executeQuery(sql);
    while (resultSet!= null & resultSet.next())
        traceln(resultSet.getString(1));

    traceln("Done!");
 }
catch (ClassNotFoundException e)
 {
    e.printStackTrace();
 }
 catch (SQLException e) {
  e.printStackTrace();
 }
```

Another connection string sample for C#.NET is as follows.

```
System.Data.OleDb.OleDbConnection connTest = new
System.Data.OleDb.OleDbConnection();
connTest.ConnectionString = @"Provider=Microsoft.Jet.OLEDB.4.0;Data
Source=D:\;Extended Properties=dBASE IV;User ID=Admin;Password=;";
connTest.Open();
OleDbCommand com = new OleDbCommand("SELECT * FROM Grid",connTest);
```

A.1.2 Conversion of Shapefile Polygon Features to Anylogic GIS Regions

Polygon features in a shapefile are known as *PoliticalArea(s)* in Anylogic. The code below demonstrates how to extract the coordinates of all political areas in a shapefile, and store them in binary file using the libraries available in Anylogic 7.0.3. First a Java class called *GISPolygon* needs to be defined with the minimal functions. The coordinates of shapefile polygons are then stored in an instance of *GISPolygon*, and is saved as a binary file on the hard disk.

```
java.util.LinkedHashMap <Integer, GISPolygon> ShapefilePolygons = new
java.util.LinkedHashMap<Integer, GISPolygon>();
// Let's assume there's a GIS map component called map, and //
// the first shapefile on the map is the shapefile of our interest //
Object[] politicalAreasObjects =
map.getLayers()[0].getPoliticalAreas().toArray();
PoliticalArea polArea;
GISPolygon gisPolygon;
int featureID=-1; // FID in shapefiles
for (int i=0;i<politicalAreasObjects.length;i++)
{
    polArea = (PoliticalArea) politicalAreasObjects[i];
    featureID = (int) Float.parseFloat(polArea.name); // The name coloumn
index of the shapefile on the map is set to refer to the FID
    OMGeometryList gmList = polArea.getGeometry();
    OMGraphicList grList = (OMGraphicList) gmList;
    Object parent = grList.getOMGraphicAt(0);
    Object child = ((OMGraphicList) parent).getOMGraphicAt(0);
    // There may be a number of nested layers of features in the shapefile//
    while (child.getClass().equals(OMGraphicList.class)) {
        child = ((OMGraphicList) child).getOMGraphicAt(0);
        parent = ((OMGraphicList) parent).getOMGraphicAt(0);
    }
    int polyCount = ((OMGraphicList) parent).size(); // The number of
political areas under the same FID
    gisPolygon = new GISPolygon(featureID);
    for(int j=0;j<polyCount;j++)
    {
        OMPoly poly = (OMPoly) ((OMGraphicList) parent).getOMGraphicAt(j);
        double[] polyCoords =  poly.getLatLonArray();
```

```
        ProjMath.arrayRadToDeg(polyCoords);
        gisPolygon.AddLatLonArray(polyCoords);
    }
    ShapefilePolygons.put(featureID, gisPolygon);
}
gisPolygon.SaveShapefilePolgons(ShapefilePolygons, "D:\\MyShapefile.dat"); //
Store on hard disk as a binray file
```

The source code for the *GISPolygon* class is as follows.

```
public class GISPolygon implements Serializable {
    public ArrayList<double[]> latLonArrayList;
    public int npolygons=0;
    public int featureID;
    public GISPolygon() {
        this.featureID = -1; this.npolygons=0;
    }
    public GISPolygon(int featureID) {
        this.npolygons = 0; this.featureID = featureID;
    }
    public void AddLatLonArray(double[] latLonArray) {
        if (this.latLonArrayList == null)
                    this.latLonArrayList = new ArrayList<double[]>();

            this.latLonArrayList.add(latLonArray);
            this.npolygons ++;
    }
    public void SaveShapefilePolgons(Object shapefilePolygons, String
filePath) {
        try {
            FileOutputStream fout = new FileOutputStream(filePath);
            ObjectOutputStream oos = new ObjectOutputStream(fout);
            oos.writeObject(shapefilePolygons);
            oos.flush(); oos.close();
            System.out.println("File saved!");
        }
        catch(Exception ex) {
            traceln(ex.toString());
        }
    }
    private static final long serialVersionUID = 1L;
}
```

Some functions and libraries used above, such as casting from *OMGeometryList* to *OMGraphicList*, may not be available in all Anylogic versions. However, the coordinate extraction procedure is still similar, and could be adopted by developers.

Finally, the code below shows how to restore saved coordinates, and display them as Anylogic *GISregion(s)* on a map. It is notable that the *GISregion* component is not available in Anylogic 7.0.3, as such the code below was tested in Anylogic 7.2.0 PLE.

```
LinkedHashMap <Integer ,GISPolygon> loadedPolygon = new
java.util.LinkedHashMap<Integer, GISPolygon>();
LinkedHashMap <Integer ,GISMultiRegion> MultiGISRegionsList = new
LinkedHashMap<Integer, GISMultiRegion>();
try {
    ObjectInputStream in = new ObjectInputStream(new
FileInputStream("D:\\MyShapefile.dat"));
    loadedPolygon = (LinkedHashMap<Integer, GISPolygon>) in.readObject();
    in.close();
}
catch(Exception ex) {
    traceln(ex.toString());
}
GISRegion gisRegion;
GISMultiRegion gisMulti;
for (Integer entryKey : loadedPolygon.keySet()){
    GISPolygon poly = loadedPolygon.get(entryKey);
    gisMulti = new GISMultiRegion(entryKey.toString());
    for (int i=0; i<poly.latLonArrayList.size();i++) {
        gisRegion = new GISRegion(map, poly.latLonArrayList.get(i));
        gisMulti.add(gisRegion);
        map.add(gisRegion);
    }
    MultiGISRegionsList.put(entryKey, gisMulti);
}
```

A.2 Birds Species Data

The collected data on bird species can be found in Table A.I below.

Table A.I Birds Species Properties

Family and Species	Common Name	Home Range [a]	Flight Speed [b]	Roosting [c]	Breeding Months	Sources
Accipitridae						
Accipiter cooperii	Cooper's Hawk	905	9	Solitary	4 - 7	[1]–[3]
Accipiter striatus	Sharp-shinned Hawk	918	7	Solitary	4 - 8	[1], [4]
Buteo jamaicensis	Red-tailed Hawk	1163	9	Solitary	2 - 9	[5]–[8]
Buteo platypterus	Broad-winged Hawk	583	11	Solitary	5 - 8	[9], [10]
Buteo regalis	Ferruginous Hawk	1652	16	Solitary	4 - 8	[11]–[13]
Buteo swainsoni	Swainson's Hawk	2249	7	Communal	4 - 8	[11], [14], [15]
Circus cyaneus	Northern Harrier	910	9	Solitary	4 - 9	[6], [16], [17]
Haliaeetus leucocephalus	Bald Eagle	2622	13	Solitary	4 - 9	[18]–[20]
Alaudidae						
Eremophila alpestris	Horned Lark	127	11	Flocking	3 - 8	[21]–[23]
Alcedinidae						
Megaceryle alcyon	Belted Kingfisher	1609	8	Solitary	4 - 8	[24]–[26]
Apodidae						
Chaetura pelagica	Chimney Swift	4000	13	Communal	6 - 8	[27], [28]
Bombycillidae						
Bombycilla cedrorum	Cedar Waxwing	36	9	Communal	6 - 10	[29]–[31]
Calcariidae						
Calcarius ornatus	Chestnut-collared Longspur	112	9	Flocking	5 - 8	[32], [33]
Caprimulgidae						
Chordeiles minor	Common Nighthawk	523	10	Flocking	6 - 9	[11], [34]
Cardinalidae						
Passerina cyanea	Indigo Bunting	160	9	Solitary	6 - 9	[11], [35]
Pheucticus ludovicianus	Rose-breasted Grosbeak	140	8	Solitary	6 - 8	[36]–[38]
Piranga olivacea	Scarlet Tanager	199	8	Solitary	6 - 8	[37], [39]
Spiza americana	Dickcissel	390	11	Flocking	5 - 8	[11], [40], [41]
Cathartidae						
Cathartes aura	Turkey Vulture	12657	13	Communal	5 - 9	[42]–[44]
Certhiidae						
Certhia americana	Brown Creeper	500	7	Solitary	5 - 8	[37], [45]
Columbidae						
Columba livia	Rock Pigeon	5300	16	Communal	4 - 11	[46], [47]
Zenaida macroura	Mourning Dove	4000	17	Communal	3 - 10	[27], [48], [49]
Corvus brachyrhynchos	American Crow	1555	11	Flocking	3 - 6	[50]–[52]
Corvidae						

Corvus corax	Common Raven	3590	11	Communal	3 - 7	[50], [53]
Cyanocitta cristata	Blue Jay [d]	103	9	Solitary	4 - 6	[22], [54]–[56]
Perisoreus canadensis	Gray Jay [d]	455	9	Solitary	3 - 6	[55], [57]
Pica hudsonia	Black-billed Magpie	126	8	Flocking	4 - 6	[58], [59]
Cuculidae						
Coccyzus erythropthalmus	Black-billed Cuckoo	305	10	Solitary	6 - 10	[11], [60], [61]
Emberizidae						
Ammodramus bairdii	Baird's Sparrow	200	7	Flocking	6 - 9	[62], [63]
Ammodramus leconteii	Le Conte's Sparrow	200	7	Solitary	6 - 9	[62], [64]
Ammodramus savannarum	Grasshopper Sparrow	75	7	Solitary	6 - 8	[62], [65], [66]
Chondestes grammacus	Lark Sparrow	139	12	Flocking	5 - 8	[11], [31], [67]
Junco hyemalis	Dark-eyed Junco	82	8	Flocking	5 - 9	[26], [37], [68]
Melospiza georgiana	Swamp Sparrow	113	13	Flocking	5 - 8	[11], [69], [70]
Melospiza lincolnii	Lincoln's Sparrow	100	13	Solitary	6 - 8	[11], [71]
Melospiza melodia	Song Sparrow	113	13	Solitary	3 - 9	[11], [72], [73]
Passerculus sandwichensis	Savannah Sparrow	170	16	Flocking	6 - 9	[22], [74], [75]
Pipilo erythrophthalmus	Eastern Towhee	300	8	Solitary	5 - 8	[37], [76]
Pipilo maculatus	Spotted Towhee	157	8	Solitary	4 - 8	[37], [77]–[79]
Pooecetes gramineus	Vesper Sparrow	142	7	Flocking	5 - 9	[11], [80], [81]
Spizella pallida	Clay-colored Sparrow	99	9	Flocking	6 - 8	[11], [38], [82]
Spizella passerina	Chipping Sparrow	99	9	Flocking	4 - 9	[11], [38], [73], [83]
Zonotrichia albicollis	White-throated Sparrow	93	8	Flocking	6 - 8	[11], [84]
Falconidae						
Falco columbarius	Merlin	2523	14	Solitary	3 - 9	[85], [86]
Falco sparverius	American Kestrel	671	10	Solitary	4 - 8	[6], [87], [88]
Fringillidae						
Carpodacus mexicanus	House Finch	1500	6	Communal	4 - 8	[89]–[91]
Carpodacus purpureus	Purple Finch	380	6	Solitary	4 - 9	[90], [92], [93]
Coccothraustes vespertinus	Evening Grosbeak	359	17	Flocking	5 - 8	[94]–[96]
Loxia curvirostra	Red Crossbill	254	9	Communal	1 - 10	[22], [37], [70], [97], [98]
Loxia leucoptera	White-winged Crossbill	1000	9	Communal	1 - 11	[37], [99]
Spinus pinus	Pine Siskin	113	15	Communal	4 - 8	[70], [100], [101]
Spinus tristis	American Goldfinch	800	7	Communal	7 - 9	[22], [102], [103]
Hirundinidae						
Hirundo rustica	Barn Swallow	600	7	Communal	5 - 9	[104]–[106]
Petrochelidon pyrrhonota	Cliff Swallow	1500	8	Communal	5 - 8	[107], [108]
Progne subis	Purple Martin	2871	12	Communal	5 - 8	[22], [49], [109]–[111]
Riparia riparia	Bank Swallow	800	14	Communal	4 - 8	[16], [106], [112]
Stelgidopteryx serripennis	Northern Rough-winged Swallow	500	7	Solitary	5 - 7	[105], [113]
Tachycineta bicolor	Tree Swallow	4000	7	Communal	5 - 7	[49], [106], [114], [115]
Icteridae						
Agelaius phoeniceus	Red-winged Blackbird	1609	7	Communal	4 - 8	[116]–[119]
Dolichonyx oryzivorus	Bobolink	90	7	Flocking	6 - 8	[119]–[121]

Euphagus carolinus	Rusty Blackbird	345	8	Communal	5 - 8	[122]
Euphagus cyanocephalus	Brewer's Blackbird	1600	12	Communal	2 - 8	[123]–[125]
Icterus galbula	Baltimore Oriole	100	11	Solitary	5 - 7	[11], [70], [126]
Icterus spurius	Orchard Oriole	113	11	Communal	6 - 8	[11], [31], [127]
Molothrus ater	Brown-headed Cowbird	1186	13	Communal	4 - 8	[128], [129]
Quiscalus quiscula	Common Grackle	12921	14	Communal	4 - 7	[27], [116], [130], [131]
Sturnella neglecta	Western Meadowlark	149	9	Flocking	4 - 9	[70], [132]
Xanthocephalus xanthocephalus	Yellow-headed Blackbird	1600	10	Communal	5 - 8	[26], [133], [134]
Laniidae						
Lanius ludovicianus	Loggerhead Shrike	207	13	Solitary	3 - 7	[6], [88], [135], [136]
Mimidae						
Dumetella carolinensis	Gray Catbird	37	7	Solitary	5 - 8	[11], [70], [137]
Mimus polyglottos	Northern Mockingbird	102	8	Solitary	3 - 8	[6], [37], [70], [138], [139]
Toxostoma rufum	Brown Thrasher	113	10	Solitary	4 - 8	[37], [140], [141]
Motacillidae						
Anthus spragueii	Sprague's Pipit	143	6	Solitary	5 - 8	[142], [143]
Pandionidae						
Pandion haliaetus	Osprey	14000	13	Solitary	5 - 9	[27], [144]
Paridae						
Poecile atricapillus	Black-capped Chickadee	216	5	Communal	4 - 7	[6], [145]–[147]
Poecile hudsonicus	Boreal Chickadee	216	5	Flocking	5 - 9	[146], [148], [149]
Parulidae						
Cardellina canadensis	Canada Warbler	80	7	Flocking	6 - 8	[37], [150]–[152]
Cardellina pusilla	Wilson's Warbler	213	7	Solitary	6 - 8	[37], [152]–[154]
Geothlypis philadelphia	Mourning Warbler [d]	113	7	Solitary	6 - 9	[6], [37], [70], [155], [156]
Geothlypis trichas	Common Yellowthroat	96	7	Solitary	6 - 8	[37], [70], [157], [158]
Mniotilta varia	Black-and-white Warbler	145	7	Solitary	5 - 8	[37], [152], [159], [160]
Oporornis agilis	Connecticut Warbler	39	7	Solitary	6 - 8	[37], [152], [159], [161]
Oreothlypis celata	Orange-crowned Warbler	80	7	Solitary	4 - 8	[37], [152], [162], [163]
Oreothlypis peregrina	Tennessee Warbler	124	7	Flocking	7 - 9	[31], [37], [152], [164], [165]
Oreothlypis ruficapilla	Nashville Warbler [d]	113	7	Solitary	6 - 8	[37], [70], [152], [166]
Parkesia noveboracensis	Northern Waterthrush	206	8	Solitary	6 - 8	[37], [167]
Seiurus aurocapilla	Ovenbird	98	8	Solitary	5 - 8	[6], [37], [168]
Setophaga americana	Northern Parula	45	7	Solitary	4 - 8	[37], [70], [169]
Setophaga castanea	Bay-breasted Warbler	106	7	Solitary	6 - 8	[37], [70], [170]
Setophaga coronata	Yellow-rumped Warbler	169	7	Flocking	6 - 8	[37], [171], [172]
Setophaga fusca	Blackburnian Warbler	59	7	Flocking	6 - 8	[6], [37], [173]
Setophaga magnolia	Magnolia Warbler	113	7	Solitary	6 - 9	[6], [37], [70], [174]
Setophaga palmarum	Palm Warbler	110	7	Solitary	5 - 8	[37], [175]
Setophaga pensylvanica	Chestnut-sided Warbler	95	7	Flocking	6 - 9	[31], [37], [176]
Setophaga petechia	Yellow Warbler	490	11	Flocking	6 - 8	[152], [177]
Setophaga pinus	Pine Warbler	99	7	Flocking	5 - 7	[37], [178], [179]
Setophaga ruticilla	American Redstart	45	6	Solitary	6 - 8	[6], [37], [70], [180]

Setophaga striata	Blackpoll Warbler	117	6	Flocking	6 - 8	[37], [181], [182]
Setophaga tigrina	Cape May Warbler	56	7	Solitary	6 - 8	[37], [183]
Setophaga virens	Black-throated Green Warbler	113	7	Flocking	5 - 8	[6], [37], [70], [184]
Vermivora chrysoptera	Golden-winged Warbler	138	7	Solitary	5 - 8	[37], [164], [185]
Passeridae						
Passer domesticus	House Sparrow	1500	13	Communal	4 - 9	[27], [49], [186], [187]
Phasianidae						
Bonasa umbellus	Ruffed Grouse	164	8	Solitary	4 - 11	[188], [189]
Meleagris gallopavo	Wild Turkey	3586	14	Communal	4 - 11	[22], [190]
Perdix perdix	Gray Partridge	623	13	Communal	4 - 7	[191]–[193]
Phasianus colchicus	Ring-necked Pheasant	339	12	Flocking	4 - 10	[194]–[196]
Tympanuchus phasianellus	Sharp-tailed Grouse	608	13	Communal	5 - 8	[11], [197]
Picidae						
Colaptes auratus	Northern Flicker	282	7	Solitary	5 - 7	[198], [199]
Dryocopus pileatus	Pileated Woodpecker	896	7	Solitary	5 - 7	[56], [70], [199], [200]
Melanerpes erythrocephalus	Red-headed Woodpecker	164	5	Solitary	5 - 9	[201], [202]
Picoides arcticus	Black-backed Woodpecker	594	7	Solitary	6 - 8	[199], [201]
Picoides dorsalis	American Three-toed Woodpecker	608	7	Solitary	6 - 8	[199], [201], [203]
Picoides pubescens	Downy Woodpecker	126	6	Solitary	4 - 7	[199], [204]
Picoides villosus	Hairy Woodpecker	460	8	Solitary	4 - 8	[199], [205]–[208]
Regulidae						
Regulus calendula	Ruby-crowned Kinglet	138	6	Solitary	5 - 8	[37], [209]
Regulus satrapa	Golden-crowned Kinglet	89	6	Solitary	6 - 9	[37], [70], [210]
Sittidae						
Sitta canadensis	Red-breasted Nuthatch	178	5	Solitary	5 - 8	[70], [211], [212]
Sitta carolinensis	White-breasted Nuthatch	219	5	Solitary	5 - 7	[70], [146], [213]
Strigidae						
Asio flammeus	Short-eared Owl	511	12	Flocking	4 - 7	[11], [214], [215]
Bubo virginianus	Great Horned Owl	981	17	Solitary	3 - 4	[214], [216]–[218]
Strix varia	Barred Owl	689	6	Solitary	2 - 10	[219], [220]
Sturnidae						
Sturnus vulgaris	European Starling	7500	11	Communal	4 - 7	[27], [49], [116], [221], [222]
Trochilidae						
Archilochus colubris	Ruby-throated Hummingbird	183	13	Solitary	5 - 10	[223]–[226]
Troglodytidae						
Cistothorus palustris	Marsh Wren	56	7	Solitary	4 - 9	[227]–[229]
Cistothorus platensis	Sedge Wren	139	7	Solitary	6 - 9	[37], [230]–[232]
Troglodytes aedon	House Wren	75	7	Solitary	5 - 9	[6], [37], [233], [234]
Troglodytes hiemalis	Winter Wren	138	7	Solitary	5 - 9	[37], [235]
Turdidae						
Catharus fuscescens	Veery [d]	113	15	Flocking	6 - 7	[70], [236], [237]
Catharus guttatus	Hermit Thrush	103	15	Communal	5 - 9	[70], [237], [238]
Catharus ustulatus	Swainson's Thrush	128	15	Solitary	6 - 9	[237], [239], [240]

Sialia currucoides	Mountain Bluebird	147	8	Flocking	4 - 9	[11], [241], [242]
Sialia sialis	Eastern Bluebird	247	8	Communal	3 - 9	[6], [11], [243]
Turdus migratorius	American Robin	400	9	Communal	4 - 8	[27], [37], [244]–[247]
Tyrannidae						
Contopus cooperi	Olive-sided Flycatcher	219	7	Solitary	6 - 9	[248], [249]
Contopus sordidulus	Western Wood-Pewee	82	7	Solitary	5 - 9	[250]–[252]
Contopus virens	Eastern Wood-Pewee	118	7	Solitary	5 - 9	[38], [250], [253]
Empidonax alnorum	Alder Flycatcher	98	13	Solitary	6 - 8	[70], [254], [255]
Empidonax flaviventris	Yellow-bellied Flycatcher	113	13	Flocking	6 - 8	[70], [255], [256]
Empidonax minimus	Least Flycatcher	60	13	Communal	6 - 8	[6], [70], [255], [257], [258]
Empidonax traillii	Willow Flycatcher	81	13	Solitary	6 - 9	[255], [259], [260]
Myiarchus crinitus	Great Crested Flycatcher	101	10	Solitary	5 - 8	[70], [248], [255], [261]
Sayornis phoebe	Eastern Phoebe [e]	95	10	Solitary	4 - 8	[70], [262], [263]
Sayornis saya	Say's Phoebe [e]	94	10	Solitary	5 - 8	[264], [265]
Tyrannus tyrannus	Eastern Kingbird	212	10	Flocking	6 - 8	[11, p. 67], [38], [263], [266]
Tyrannus verticalis	Western Kingbird	219	8	Solitary	5 - 7	[11, p. 67], [22], [267], [268]
Vireonidae						
Vireo flavifrons	Yellow-throated Vireo	100	8	Solitary	6 - 8	[37], [269]
Vireo gilvus	Warbling Vireo	100	8	Solitary	5 - 8	[37], [270]–[272]
Vireo olivaceus	Red-eyed Vireo	109	8	Flocking	5 - 8	[37], [270], [273]
Vireo philadelphicus	Philadelphia Vireo	113	8	Solitary	6 - 8	[37], [274]
Vireo solitarius	Blue-headed Vireo	100	8	Solitary	6 - 8	[37], [275]

[a.] Home range values are the ceiling for the average radius of an estimated circular home range area in meters.

[b.] Flight speed values are reported in meters per second. As it was not clear as to the type of speed that different papers reported, it is not recommended to compare the birds' speed by these values.

[c.] Flocking under the Roosting column means that the species roost individually or in pairs during the breeding season, and then form flocks for migration in the fall.

[d.] It is notable that for the Mourning Warbler, Veery, and Nashville Warbler, the sources of data were not clear as to the type of roosting that they did. The entries given represent a best guess, but further data would be needed for greater precision. Specifically, for the Veery, no mention of communal roosting was found in literature, but no mention of any other type was found. The Blue Jay and Gray Jay were difficult to properly categorize due to ambiguity in literature regarding their roosting behaviour.

[e.] For the flight speeds of Eastern Phoebe and Say's Phoebe species, the average speed of the Tyrannidae birds in the dataset is used.

A.3 References for Appendix A

[1] M. B. B. V. Goodwin, “Flight-Speeds of Hawks and Crows,” *Auk*, vol. 60, no. 4, pp. 487–492, 1943.

[2] O. E. Curtis, R. N. Rosenfield, and J. Bielefeldt., “Cooper’s Hawk (Accipiter cooperii),” *Birds North Am. Online*, 2006.

[3] C. Polite and L. Kiff, “COOPER’S HAWK Accipiter cooperii,” in *California Wildlife Habitat Relationships (CWHR)*, Vol I-III., D. C. Zeiner, W. F. L. Jr., K. E. Mayer, and M. White, Eds. Sacramento, California: California Depart. of Fish and Game, 1990.

[4] K. L. Bildstein and K. Meyer, “Sharp-shinned Hawk (Accipiter striatus),” *Birds North Am. Online*, 2000.

[5] S. Johnson, “Red-Tailed Hawks (Buteo jamaicensis),” 2011. [Online]. Available: http://beautyofbirds.com/redtailedhawks.html. Arhived at: http://www.webcitation.org/6lCvnOQPe. [Accessed: 05-Nov-2015].

[6] T. W. Schoener, “Sizes of Feeding Territories among Birds Ecological Society of America,” *Ecol. Soc. Am. Stable*, vol. 49, no. 1, pp. 123–141, 1968.

[7] C. Polite, J. Pratt, and S. Bailey, “RED-TAILED HAWK Buteo jamaicensis,” in *California Wildlife Habitat Relationships (CWHR)*, Vol I-III., D. C. Zeiner, W. F. L. Jr., K. E. Mayer, and M. White, Eds. Sacramento, California: California Depart. of Fish and Game, 1990.

[8] C. R. Preston and R. D. Beane, “Red-tailed Hawk (Buteo jamaicensis),” *Birds North Am. Online*, 2009.

[9] L. J. Goodrich, S. T. Crocoll, and S. E. Senner, “Broad-winged Hawk (Buteo platypterus),” *Birds North Am. Online*, 2014.

[10] P. H. Bloom, M. D. Mccrary, and M. J. Gibson, “Red-Shouldered Hawk Home-Range and Habitat Use in Southern California,” *J. Wildl. Manage.*, vol. 57, no. 2, pp. 258–265, 1993.

[11] P. Bumstead, *Canadian Feathers : a Loon-atics Guide to Anting, Mimicry and Dump-nesting*. Simply Wild Publications, 2001.

[12] G. S. Dwight and J. R. Murphy, “Breeding ecology of raptors in the eastern Great Basin of Utah,” *Brigham Young University Science Bulletin - Biological Series*, vol. 18, no. 3. pp. 1–76, 01-Jun-1973.

[13] M. J. Bechard and J. K. Schmutz, “Ferruginous Hawk (Buteo regalis),” *Birds North Am. Online*, 1995.

[14] M. J. Bechard, C. S. Houston, J. H. Sarasola, and A. S. England, “Swainson’s Hawk (Buteo swainsoni),” *Birds North Am. Online*, 2010.

[15] C. Polite and L. Kiff, “SWAINSON’S HAWK Buteo swainsoni,” in *California Wildlife Habitat Relationships (CWHR)*, Vol I-III., D. C. Zeiner, W. F. L. Jr., K. E. Mayer, and M. White, Eds. Sacramento, California: California Depart. of Fish and Game, 1990.

[16] T. Alerstam, M. Rosén, J. Bäckman, P. G. P. Ericson, and O. Hellgren, “Flight Speeds among Bird Species: Allometric and Phylogenetic Effects,” *PLoS Biol.*, vol. 5, no. 8, p. e197, Jul. 2007.

[17] K. G. Smith, S. R. Wittenberg, R. B. Macwhirter, and K. L. Bildstein, “Northern Harrier (Circus cyaneus),” *Birds North Am. Online*, 2011.

[18] A. Travsky and D. G. P. Equvais, “Species Assessment for Bald Eagle (Haliaeetus leucocephalus) in Wyoming,” 2004.
[19] H. Rutledge, “Bald Eagle Description Page 2,” *American Bald Eagle Information*. [Online]. Available: http://www.baldeagleinfo.com/eagle/eagle8.html Archived at: http://www.webcitation.org/6lD5OKOFU. [Accessed: 05-Nov-2015].
[20] D. A. Buehler, “Bald Eagle (Haliaeetus leucocephalus),” *Birds North Am. Online*, 2000.
[21] R. J. Cannings and W. Threlfall, “Horned Lark Breeding Biology at Cape St. Mary’s, Newfoundland,” *Wilson Bull.*, vol. 93, no. 4, p. 12, 1981.
[22] M. T. Cooke, *Flight speed of birds*. Washington, D.C. : U.S. Dept. of Agriculture, 1937.
[23] R. C. Beason, “Horned Lark (Eremophila alpestris),” *Birds North Am. Online*, 1995.
[24] J. F. Kelly, E. S. Bridge, and M. J. Hamas, “Belted Kingfisher (Ceryle alcyon),” *Birds North Am. Online*, 2009.
[25] M. Green, L. Mewaldt, R. Duke, and D. Winkler, “BELTED KINGFISHER Megaceryle alcyon,” in *California Wildlife Habitat Relationships (CWHR)*, Vol I-III., D. C. Zeiner, W. F. L. Jr., K. E. Mayer, and M. White, Eds. Sacramento, California: California Depart. of Fish and Game, 1990.
[26] H. B. Wood, “Flight Speed of Some Birds,” *Auk*, vol. 50, no. 4, pp. 452–453, 1933.
[27] G. D. Schnell and J. J. Hellack, “Flight Speeds of Brown Pelicans, Chimney Swifts, and Other Birds,” *Bird-Banding*, vol. 49, no. 2, pp. 108–112, 1978.
[28] T. K. Steeves, S. B. Kearney-McGee, M. A. Rubega, C. L. Cink, and C. T. Collins, “Chimney Swift (Chaetura pelagica),” *Birds North Am. Online*, 2014.
[29] M. C. Witmer, D. J. Mountjoy, and L. Elliot, “Cedar Waxwing (Bombycilla cedrorum),” *Birds North Am. Online*, 2014.
[30] D. R. Khanna, *Biology of Birds*. Discovery Publishing House, 2005.
[31] S. W. Gillihan, *Bird Conservation on Golf Courses: A Design and Management Manual*, vol. 8. John Wiley & Sons, 2000.
[32] D. J. Hussell and R. Montgomerie, “Lapland Longspur (Calcarius lapponicus),” *Birds North Am. Online*, 2002.
[33] B. Bleho, K. Ellison, D. P. Hill, and L. K. Gould, “Chestnut-collared Longspur (Calcarius ornatus),” *Birds North Am. Online*, 2015.
[34] R. M. Brigham, J. Ng, R. G. Poulin, and S. D. Grindal, “Common Nighthawk (Chordeiles minor),” *Birds North Am. Online*, 2011.
[35] R. B. Payne, “Indigo Bunting (Passerina cyanea),” *Birds North Am. Online*, no. 4, 2006.
[36] V. E. Wyatt and C. M. Francis, “Rose-breasted Grosbeak (Pheucticus ludovicianus),” *Birds North Am. Online*, 2002.
[37] S. A. Cabrera-Cruz, T. J. Mabee, and R. V. Patraca, “Using Theoretical Flight Speeds to Discriminate Birds from Insects in Radar Studies,” *Condor*, vol. 115, no. 2, pp. 263–272, May 2013.
[38] E. P. Odum and E. J. Kuenzler, “Measurement of Territory and Home Range Size in Birds,” *Auk*, vol. 72, no. 2, pp. 128–137, Apr. 1955.
[39] T. B. Mowbray, “Scarlet Tanager (Piranga olivacea),” *Birds North Am. Online*, 1999.
[40] S. Temple, “Dickcissel (Spiza americana),” *Birds North Am. Online*, 2002.

[41] K. M. S. Wells, J. J. Millspaugh, M. R. Ryan, and M. W. Hubbard, “Factors Affecting Home Range Size and Movements of Post-Fledging Grassland Birds,” *Wilson J. Ornithol.*, vol. 120, no. 1, pp. 120–130, Mar. 2008.
[42] N. J. Buckley, “Food finding and the influence of information, local enhancement, and communal roosting on foraging success of North American vultures,” *Auk*, vol. 113, no. 2, pp. 473–488, 1996.
[43] C. S. Houston, G. L. Holroyd, B. Terry, M. Blom, and M. J. Stoffel, “Tracking Saskatchewan Nestling Turkey Vultures,” *Blue Jay*, vol. 65, no. 4, pp. 201–207, 2007.
[44] D. A. Kirk and M. J. Mossman, “Turkey Vulture (Cathartes aura),” *Birds North Am. Online*, 1998.
[45] J.-F. Poulin *et al.*, “Brown Creeper (Certhia americana),” *Birds North Am. Online*, 2013.
[46] P. E. Lowther and R. F. Johnston, “Rock Pigeon (Columba livia),” *Birds North Am. Online*, 2013.
[47] N. E. Baldaccini, D. Giunchi, E. Mongini, and L. Ragionieri, “Foraging flights of wild rock doves (Columba l. livia) : a spatio-temporal analysis,” *Ital. J. Zool.*, vol. 67, no. 4, pp. 371–377, 2000.
[48] D. L. Otis, J. H. Schulz, D. Miller, R. E. Mirarchi, and T. S. Baskett, “Mourning Dove (Zenaida macroura),” *Birds North Am. Online*, 2008.
[49] D. F. Caccamise and J. Fischl, “Patterns of Association of Secondary Species in Roosts of European Starlings and Common Grackles,” *Wilson Bull.*, vol. 97, no. 2, pp. 173–182, 1985.
[50] J. M. Marzluff and T. Angell, *In the Company of Crows and Ravens*. Yale University Press, 2008.
[51] M. P. Ward, A. Raim, S. Yaremych-Hamer, R. Lampman, and R. J. Novak, “Does the roosting behavior of birds affect transmission dynamics of West Nile virus?,” *Am. J. Trop. Med. Hyg.*, vol. 75, no. 2, pp. 350–5, Aug. 2006.
[52] N. A. Verbeek and C. Caffrey, “American Crow (Corvus brachyrhynchos),” *Birds North Am. Online*, 2002.
[53] W. I. Boarman and B. Heinrich, “Common Raven (Corvus corax),” *Birds North Am. Online*, 1999.
[54] K. G. Smith, K. A. Tarvin, and G. E. Woolfenden, “Blue Jay (Cyanocitta cristata),” *Birds North Am. Online*, 2013.
[55] Texas Parks and Wildlife Department, “Animal Speeds.” [Online]. Available: http://tpwd.texas.gov/publications/nonpwdpubs/young_naturalist/animals/animal_speeds/index.phtml. Archived at: http://www.webcitation.org/6lCxGHXZ7. [Accessed: 06-Nov-2015].
[56] J. Bowman, “Is dispersal distance of birds proportional to territory size?,” *Can. J. Zool.*, vol. 81, no. 2, pp. 195–202, Feb. 2003.
[57] D. Strickland and H. Ouellet, “Gray Jay (Perisoreus canadensis),” *Birds North Am. Online*, 2011.
[58] B. W. Tobalske, N. E. Olson, and K. P. Dial, “Flight style of the black-billed magpie: variation in wing kinematics, neuromuscular control, and muscle composition.,” *J. Exp. Zool.*, vol. 279, no. 4, pp. 313–29, Nov. 1997.
[59] C. H. Trost, “Black-billed Magpie (Pica hudsonia),” *Birds North Am. Online*, 1999.
[60] J. M. Hughes, “Yellow-billed Cuckoo (Coccyzus americanus),” *Birds North Am. Online*, 2015.
[61] J. M. Hughes, “Black-billed Cuckoo (Coccyzus erythropthalmus),” *Birds North Am. Online*, 2001.

[62] P. E. 2005 Lowther, “Le Conte’s Sparrow (Ammodramus leconteii),” *Birds North Am. Online*, 2005.
[63] M. T. Green, P. E. Lowther, S. L. Jones, S. K. Davis, and B. C. Dale, “Baird’s Sparrow (Ammodramus bairdii),” *Birds North Am. Online*, 2002.
[64] H. Q. Baldwin, C. W. Jeske, M. A. Powell, P. C. Chadwick, and W. C. Barrow, “Home-Range Size and Site Tenacity of Overwintering Le Conte’s Sparrows in a Fire Managed Prairie,” *Wilson J. Ornithol.*, vol. 122, no. 1, pp. 139–145, Mar. 2010.
[65] M. F. . Delany, C. T. . Moore, and J. M. . Hamblen, “Florida grasshopper sparrow management needs,” *Final report to Florida Game and Fresh Water Fish Commission; Tallahassee, Florida*, 1992. [Online]. Available: http://s3.amazonaws.com/file-storage.INDIVIDUAL-ACTIVITIES-CooperativeResearchUnits.digitalmeasures.usgs.edu/cmoore/tech_publications/Delany et al (1992) FGS Final Report-1.pdf. Archived at: http://www.webcitation.org/6lCxp46eH. [Accessed: 06-Nov-2015].
[66] P. D. Vickery, “Grasshopper Sparrow (Ammodramus savannarum),” *Birds North Am. Online*, 1996.
[67] J. W. Martin and J. R. Parrish, “Lark Sparrow (Chondestes grammacus),” *Birds North Am. Online*, 2000.
[68] V. Nolan, Jr. *et al.*, “Dark-eyed Junco (Junco hyemalis),” *Birds North Am. Online*, 2002.
[69] T. B. Mowbray, “Swamp Sparrow (Melospiza georgiana),” *Birds North Am. Online*, 1997.
[70] R. M. DeGraaf, *Technical Guide to Forest Wildlife Habitat Management in New England*. UPNE, 2006.
[71] E. M. Ammon, “Lincoln’s Sparrow (Melospiza lincolnii),” *Birds North Am. Online*, 1995.
[72] P. Arcese, M. K. Sogge, A. B. Marr, and M. A. Patten, “Song Sparrow (Melospiza melodia),” *Birds North Am. Online*, 2002.
[73] R. M. DeGraaf and M. Yamasaki, *New England Wildlife: Habitat, Natural History, and Distribution*. UPNE, 2001.
[74] D. L. Ginter and M. J. Desmond, “Influence of foraging and roosting behavior on home-range size and movement patterns of Savannah Sparrows wintering in south Texas,” *Wilson Bull.*, vol. 117, no. 1, pp. 63–71, Mar. 2005.
[75] N. T. Wheelwright and J. D. Rising, “Savannah Sparrow (Passerculus sandwichensis),” *Birds North Am. Online*, 2008.
[76] J. S. Greenlaw, “Eastern Towhee (Pipilo erythrophthalmus),” *Birds North Am. Online*, 2015.
[77] S. Bartos Smith and J. S. Greenlaw, “Spotted Towhee (Pipilo maculatus),” *Birds North Am. Online*, 2015.
[78] D. Dobkin, L. Mewaldt, R. Duke, and S. Granholm, “SPOTTED TOWHEE Pipilo maculatus,” in *California Wildlife Habitat Relationships (CWHR)*, Vol I-III., D. C. Zeiner, W. F. L. Jr., K. E. Mayer, and M. White, Eds. Sacramento, California: California Depart. of Fish and Game, 1990.
[79] M. C. Baker and L. R. Mewaldt, “The use of space by white-crowned sparrows: Juvenile and adult ranging patterns and home range versus body size comparisons in an avian granivore community,” *Behav. Ecol. Sociobiol.*, vol. 6, no. 1, pp. 45–52, 1979.

[80] A. Macías-Duarte and A. O. Panjabi, “Home range and habitat use of wintering Vesper Sparrows in grasslands of the Chihuahuan Desert in Mexico,” *Wilson J. Ornithol.*, vol. 125, no. 4, pp. 755–762, Dec. 2013.
[81] S. L. Jones and J. E. Cornely, “Vesper Sparrow (Pooecetes gramineus),” *Birds North Am. Online*, 2002.
[82] T. A. Grant and R. W. Knapton, “Clay-colored Sparrow (Spizella pallida),” *Birds North Am. Online*, 2012.
[83] A. L. Middleton, “Chipping Sparrow (Spizella passerina),” *Birds North Am. Online*, 1998.
[84] J. B. Falls and J. G. Kopachena, “White-throated Sparrow (Zonotrichia albicollis),” *Birds North Am. Online*, 2010.
[85] D. M. Becker and C. H. Sieg, “Home range and habitat utilization of breeding male Merlins, Falco columbarius, in southeastern Montana,” *Can. Field-Naturalist*, vol. 101, pp. 398–403, 1987.
[86] I. G. Warkentin, N. S. Sodhi, R. H. M. Espie, A. F. Poole, L. W. Oliphant, and P. C. James, “Merlin (Falco columbarius),” *Birds North Am. Online*, 2005.
[87] J. A. Smallwood and D. M. Bird, “American Kestrel (Falco sparverius),” *Birds North Am. Online*, 2002.
[88] R. Wauer, “How Fast Do Birds Fly?,” *The Nature Writers of Texas*, 2007. [Online]. Available: http://texasnature.blogspot.ca/2007/02/how-fast-do-birds-fly-by-ro-wauer-on.html. Archived at: http://www.webcitation.org/6lD0TD9vl. [Accessed: 05-Nov-2015].
[89] A. A. Dhondt, M. J. L. Driscoll, and E. C. H. Swarthout, “House Finch Carpodacus mexicanus roosting behaviour during the non-breeding season and possible effects of mycoplasmal conjunctivitis,” *Ibis (Lond. 1859).*, vol. 149, no. 1, pp. 1–9, 2007.
[90] B. Tobalske, W. Peacock, and K. Dial, “Kinematics of flap-bounding flight in the zebra finch over a wide range of speeds,” *J. Exp. Biol.*, vol. 202 (Pt 13, pp. 1725–39, Jul. 1999.
[91] A. V. Badyaev, V. Belloni, and G. E. Hill, “House Finch (Carpodacus mexicanus),” *Birds North Am. Online*, 2012.
[92] J. Herman and M. McGarry, “Examination of home range in male purple finches (Carpodacus purpureus),” 2011.
[93] J. T. Wootton, “Purple Finch (Carpodacus purpureus),” *Birds North Am. Online*, 1996.
[94] S. W. Gillihan and B. Byers, “Evening Grosbeak (Coccothraustes vespertinus),” *Birds North Am. Online*, 2001.
[95] D. Dobkin, S. Granholm, L. Mewaldt, and R. Duke, “EVENING GROSBEAK Coccothraustes vespertinus,” in *California Wildlife Habitat Relationships (CWHR)*, Vol I-III., D. C. Zeiner, W. F. L. Jr., K. E. Mayer, and M. White, Eds. Sacramento, California: California Depart. of Fish and Game, 1990.
[96] R. Meinertzhagen, “Speed and Altitude of Bird Flight (With Notes on Other Animals),” *Ibis (Lond. 1859).*, vol. 97, no. 1, pp. 81–117, Apr. 2008.
[97] C. S. Adkisson, “Red Crossbill (Loxia curvirostra),” *Birds North Am. Online*, 1996.
[98] H. Ellegren, “Speed of migration and migratory flight lengths of passerine birds ringed during autumn migration in Sweden,” *Ornis Scand.*, vol. 24, no. 3, pp. 220–228, 1993.
[99] C. W. Benkman, “White-winged Crossbill (Loxia leucoptera),” *Birds North Am. Online*, 2012.

[100] W. R. Dawson, “Pine Siskin (Carduelis pinus),” *Birds North Am. Online*, 2014.
[101] J. J. Videler, *Avian Flight*. 2005.
[102] E. L. Coutlee, “Agonistic Behavior in the American Goldfinch,” *Wilson Bull.*, vol. 79, no. 1, pp. 89–109, 1967.
[103] K. J. McGraw and A. L. Middleton, “American Goldfinch (Carduelis tristis),” *Birds North Am. Online*, 2009.
[104] C. Brown and M. B. Brown, “Barn Swallow (Hirundo rustica),” *The Birds of North America Online*. 1999.
[105] R. W. Blake, R. Kolotylo, and H. de la Cueva, “Flight speeds of the barn swallow, Hirundo rustica,” *Can. J. Zool.*, vol. 68, no. 1, pp. 1–5, Jan. 1990.
[106] A. Heagy *et al.*, “Recovery Strategy for the Barn Swallow (Hirundo rustica) in Ontario,” Peterborough, Ontario, 2014.
[107] R. M. Shelton, B. E. Jackson, and T. L. Hedrick, “The mechanics and behavior of cliff swallows during tandem flights,” *J. Exp. Biol.*, vol. 217, no. 15, pp. 2717–2725, Aug. 2014.
[108] C. R. Brown and M. B. Brown, “Cliff Swallow (Petrochelidon pyrrhonota),” *Birds North Am. Online*, 1995.
[109] K. R. Russell and S. a Gauthreaux, “Spatial and temporal dynamics of a Purple Martin pre-migratory roost,” *Wilson Bull.*, vol. 111, no. 3, pp. 354–362, 1999.
[110] Purple Martin Conservation Association, “Purple Martin Terminology.” [Online]. Available: https://www.purplemartin.org/purple-martins/biology/41/terminology/ Archived at: http://www.webcitation.org/6lD1ZVpja. [Accessed: 23-Jan-2016].
[111] S. Tarof and C. R. Brown, “Purple Martin (Progne subis),” *Birds North Am. Online*, 2013.
[112] B. A. 1999 Garrison, “Bank Swallow (Riparia riparia),” *Birds North Am. Online*, 1999.
[113] M. J. De Jong, “Northern Rough-winged Swallow (Stelgidopteryx serripennis),” *Birds North Am. Online*, 1996.
[114] D. W. Winkler, K. K. Hallinger, D. R. Ardia, R. J. Robertson, B. J. Stutchbury, and R. R. Cohen, “Tree Swallow (Tachycineta bicolor),” *Birds North Am. Online*, 2011.
[115] R. A. Kolotylo, “Flight Speeds and Energetics of Seven Bird Species,” University of British Columbia, 1989.
[116] B. Meanley, “The Roosting Behavior of the Red-Winged Black-Bird in the Southern United States,” *Wilson Bull.*, vol. 77, no. 3, pp. 218–228, 1965.
[117] Illinois Natural History Survey, “Red-winged blackbird.” [Online]. Available: http://wwx.inhs.illinois.edu/collections/birds/ilbirds/56/ Archived at: http://www.webcitation.org/6lD21SkZV. [Accessed: 06-Nov-2015].
[118] K. Yasukawa and W. A. Searcy, “Red-winged Blackbird (Agelaius phoeniceus),” *Birds North Am. Online*, 1995.
[119] O. S. Pettingill, *Ornithology in Laboratory and Field*. Elsevier, 2013.
[120] R. Renfrew, A. M. Strong, N. G. Perlut, S. G. Martin, and T. A. Gavin, “Bobolink (Dolichonyx oryzivorus),” *Birds North Am. Online*, 2015.

[121] J. A. Wiens, “An Approach to the Study of Ecological Relationships among Grassland Birds,” *Ornithol. Monogr.*, no. 8, pp. 1–93, Jan. 1969.
[122] M. L. Avery, “Rusty Blackbird (Euphagus carolinus),” *Birds North Am. Online*, 2013.
[123] L. Williams, “Breeding Behavior of the Brewer Blackbird,” *Condor*, vol. 54, no. 1, pp. 3–47, 1952.
[124] S. G. Martin, “Brewer’s Blackbird (Euphagus cyanocephalus),” *Birds North Am. Online*, 2002.
[125] S. Granholm, L. Mewaldt, and R. Duke, “BREWER’S BLACKBIRD Euphagus cyanocephalus,” in *California Wildlife Habitat Relationships (CWHR)*, Vol I-III., D. C. Zeiner, W. F. L. Jr., K. E. Mayer, and M. White, Eds. Sacramento, California: California Depart. of Fish and Game, 1990.
[126] J. D. Rising and N. J. Flood., “Baltimore Oriole (Icterus galbula),” *Birds North Am. Online*, 1998.
[127] W. C. Scharf and J. Kren, “Orchard Oriole (Icterus spurius),” *Birds North Am. Online*, 2010.
[128] S. I. Rothstein, J. Verner, and E. Steven, “Radio-Tracking Confirms a Unique Diurnal Pattern of Spatial Occurrence in the Parasitic Brown-Headed Cowbird,” *Ecology*, vol. 65, no. 1, pp. 77–88, 1984.
[129] P. E. Lowther, “Brown-headed Cowbird (Molothrus ater),” *Birds North Am. Online*, 1993.
[130] O. E. Bray, W. C. J. Royall, J. L. Guarino, and R. E. Johnson, “Activities of radio-equipped Common Grackles during fall migration,” *Wilson Bull.*, vol. 91, no. 1, pp. 78–87, 1979.
[131] B. D. Peer and E. K. Bollinger, “Common Grackle (Quiscalus quiscula),” *Birds North Am. Online*, 1997.
[132] S. K. Davis and W. E. Lanyon, “Western Meadowlark (Sturnella neglecta),” *Birds North Am. Online*, 2008.
[133] D. J. Twedt and R. D. Crawford., “Yellow-headed Blackbird (Xanthocephalus xanthocephalus),” *Birds North Am. Online*, 1995.
[134] S. Granholm, L. Mewaldt, and R. Duke, “YELLOW-HEADED BLACKBIRD Xanthocephalus xanthocephalus,” in *California Wildlife Habitat Relationships (CWHR)*, Vol I-III., D. C. Zeiner, W. F. L. Jr., K. E. Mayer, and M. White, Eds. Sacramento, California: California Depart. of Fish and Game, 1990.
[135] S. Granholm, L. Mewaldt, and R. Duke, “LOGGERHEAD SHRIKE Lanius ludovicianus,” in *California Wildlife Habitat Relationships (CWHR)*, Vol I-III., D. C. Zeiner, W. F. L. Jr., K. E. Mayer, and M. White, Eds. Sacramento, California: California Depart. of Fish and Game, 1990.
[136] R. Yosef, “Loggerhead Shrike (Lanius ludovicianus),” *Birds North Am. Online*, 1996.
[137] R. J. Smith, M. I. Hatch, D. A. Cimprich, and F. R. Moore, “Gray Catbird (Dumetella carolinensis),” *Birds North Am. Online*, 2011.
[138] G. Farnsworth, G. A. Londono, J. U. Martin, K. C. DERRICKSON, and R. BREITWISCH, “Northern Mockingbird (Mimus polyglottos),” *Birds North Am. Online*, no. 7, 2011.
[139] D. Dobkin, L. Mewaldt, and R. Duke, “NORTHERN MOCKINGBIRD Mimus polyglottos,” in *California Wildlife Habitat Relationships (CWHR)*, Vol I-III., D. C. Zeiner, W. F. L. Jr., K. E. Mayer, and M. White, Eds. Sacramento, California: California Depart. of Fish and Game, 1990.
[140] J. F. Cavitt and C. A. Haas, “Brown Thrasher (Toxostoma rufum),” *Birds North Am. Online*, 2014.
[141] J. Eastman, *The Eastman Guide to Birds: Natural History Accounts for 150 North American Species*. Stackpole Books, 2000.

[142] M. B. Robbins, “Display Behavior of Male Sprague’s Pipits,” *Wilson Bull.*, vol. 110, no. 3, pp. 435–438, 1998.
[143] S. K. Davis, M. B. Robbins, and B. C. Dale, “Sprague’s Pipit (Anthus spragueii),” *Birds North Am. Online*, 2014.
[144] A. F. Poole, R. O. Bierregaard, and M. S. Martell, “Osprey (Pandion haliaetus),” *Birds North Am. Online*, 2002.
[145] J. R. Foote, D. J. Mennill, L. M. Ratcliffe, and S. M. Smith, “Black-capped Chickadee (Poecile atricapillus),” *Birds North Am. Online*, 2010.
[146] C. H. Greenewalt, “The flight of the Black-capped Chickadee and the White-breasted Nuthatch,” *Auk*, vol. 72, pp. 1–5, 1955.
[147] E. P. Odum, “Annual Cycle of the Black-Capped Chickadee: 3,” *Auk*, vol. 59, no. 4, pp. 499–531, Oct. 1942.
[148] A. Hadley and A. Desrochers, “Winter Habitat Use by Boreal Chickadee Flocks in a Managed Forest,” *Wilson J. Ornithol.*, vol. 120, no. 1, pp. 139–145, 2008.
[149] M. S. Ficken, M. A. Mclaren, and J. P. Hailman, “Boreal Chickadee (Poecile hudsonica),” *Birds North Am. Online*, 1996.
[150] C. Savignac, “COSEWIC assessment and status report on the Canada warbler, Wilsonia canadensis, in Canad,” *Library and Archives Canada Electronic Collection*, 2008. [Online]. Available: http://epe.lac-bac.gc.ca/100/200/301/environment_can/cws-scf/cosewic-cosepac/canada_warbler-e/CW69-14-548-2008E.pdf Archived at: http://www.webcitation.org/6lD2aBxNQ. [Accessed: 17-Nov-2015].
[151] L. Reitsma, M. Goodnow, M. T. Hallworth, and C. J. Conway, “Canada Warbler (Wilsonia canadensis),” *Birds North Am. Online*, 2010.
[152] P. Berthold, E. Gwinner, and E. Sonnenschein, *Avian Migration*. Springer Science & Business Media, 2013.
[153] E. M. Ammon and W. M. Gilbert., “Wilson’s Warbler (Wilsonia pusilla),” *Birds North Am. Online*, 1999.
[154] M. Green, L. Mewaldt, R. Duke, and D. Winkler, “WILSON’S WARBLER Wilsonia pusilla,” in *California Wildlife Habitat Relationships (CWHR)*, Vol I-III., D. C. Zeiner, W. F. L. Jr., K. E. Mayer, and M. White, Eds. Sacramento, California: California Depart. of Fish and Game, 1990.
[155] G. W. Cox, “A Life History of the Mourning Warbler,” *Wilson Bull.*, vol. 72, no. 1, pp. 5–28, 1960.
[156] J. Pitocchelli, “Mourning Warbler (Oporornis philadelphia),” *Birds North Am. Online*, 1993.
[157] M. Green, L. Mewaldt, R. Duke, and D. Winkler, “COMMON YELLOWTHROAT Geothlypis trichas,” in *California Wildlife Habitat Relationships (CWHR)*, Vol I-III., D. C. Zeiner, W. F. L. Jr., K. E. Mayer, and M. White, Eds. Sacramento, California: California Depart. of Fish and Game, 1990.
[158] M. J. Guzy and G. Ritchison., “Common Yellowthroat (Geothlypis trichas),” *Birds North Am. Online*, 1999.
[159] R. D. James, “Habitat management guidelines for birds of Ontario wetlands, including marshes, swamps and fens or bogs of various types,” 1985.

[160] J. C. Kricher, “Black-and-white Warbler (Mniotilta varia),” *Birds North Am. Online*, 2014.
[161] J. Pitocchelli, J. Jones, D. Jones, and J. Bouchie, “Connecticut Warbler (Oporornis agilis),” *Birds North Am. Online*, 2012.
[162] W. M. Gilbert, M. K. Sogge, and C. Van Riper III, “Orange-crowned Warbler (Vermivora celata),” *Birds North Am. Online*, 2010.
[163] M. Green, L. Mewaldt, R. Duke, and D. Winkler, “ORANGE-CROWNED WARBLER Vermivora celata,” in *California Wildlife Habitat Relationships (CWHR)*, Vol I-III., D. C. Zeiner, W. F. L. Jr., K. E. Mayer, and M. White, Eds. Sacramento, California: California Depart. of Fish and Game, 1990.
[164] K. Adlinger, M. Bakermans, J. Larkin, J. Lehman, and A. Tisdale, “Monitoring and Evaluating Golden-winged Warbler Use of Breeding Habitat,” 2014.
[165] C. C. Rimmer and K. P. Mcfarland., “Tennessee Warbler (Vermivora peregrina),” *Birds North Am. Online*, 2012.
[166] P. E. Lowther and J. M. Williams, “Nashville Warbler (Vermivora ruficapilla),” *Birds North Am. Online*, 2011.
[167] D. M. Whitaker and S. W. Eaton, “Northern Waterthrush (Seiurus noveboracensis),” *Birds North Am. Online*, 2014.
[168] P. Porneluzi, M. A. Van Horn, and T. M. Donovan, “Ovenbird (Seiurus aurocapilla),” *Birds North Am. Online*, no. 88, 2011.
[169] R. R. Moldenhauer and D. J. Regelski, “Northern Parula (Parula americana),” *Birds North Am. Online*, 2012.
[170] L. Venier, S. Holmes, and J. M. Williams, “Bay-breasted Warbler (Dendroica castanea),” *Birds North Am. Online*, 2011.
[171] D. L. Slager and P. G. Rodewald, “Disjunct Nocturnal Roosting by a Yellow-rumped Warbler (Setophaga coronata) during Migratory Stopover,” *Wilson J. Ornithol.*, vol. 127, no. 1, pp. 109–114, Mar. 2015.
[172] P. D. Hunt and D. J. Flaspohler, “Yellow-rumped Warbler (Dendroica coronata),” *Birds North Am. Online*, 1998.
[173] D. H. Morse, “Blackburnian Warbler (Dendroica fusca),” *Birds North Am. Online*, 2004.
[174] E. Dunn and G. A. Hall, “Magnolia Warbler (Dendroica magnolia),” *Birds North Am. Online*, 1994.
[175] J. Wilson and W. Herbert., “Palm Warbler (Dendroica palmarum),” *Birds North Am. Online*, 1996.
[176] B. E. Byers, M. Richardson, and D. W. Brauning, “Chestnut-sided Warbler (Dendroica pensylvanica),” *Birds North Am. Online*, 2013.
[177] P. E. Lowther, C. Celada, N. K. Klein, C. C. Rimmer, and D. A. Spector, “Yellow Warbler (Dendroica petechia),” *Birds North Am. Online*, 1999.
[178] S. C. . Latta and M. L. Sondreal, “Observations on the Abundance, Site Persistence, Home Range, Foraging, and Nesting of the Pine Warbler on Hispaniola, and First Record of Ground Nesting for This Species,” *Ornitol. Neotrop.*, vol. 10, no. 1, pp. 43–54, 1999.

[179] P. G. Rodewald, J. H. Withgott, and K. G. Smith, “Pine Warbler (Dendroica pinus),” *Birds North Am. Online*, 1999.
[180] T. W. Sherry and R. T. Holmes, “American Redstart (Setophaga ruticilla),” *Birds North Am. Online*, no. 277, 1997.
[181] B. Morse, “Weather and Migrating Fall Warblers in Eastern Washington,” *Washingt. Ornithol. Soc. News*, no. 74, 2001.
[182] W. DeLuca, R. Holberton, P. D. Hunt, and B. C. Eliason, “Blackpoll Warbler (Dendroica striata),” *Birds North Am. Online*, 2013.
[183] M. E. Baltz and S. C. Latta, “Cape May Warbler (Dendroica tigrina),” *Birds North Am. Online*, 1998.
[184] D. H. Morse and A. F. Poole, “Black-throated Green Warbler (Dendroica virens),” *Birds North Am. Online*, 1993.
[185] J. L. Confer, P. Hartman, and A. Roth, “Golden-winged Warbler (Vermivora chrysoptera),” *Birds North Am. Online*, 2011.
[186] P. E. Lowther and C. L. Cink., “House Sparrow (Passer domesticus),” *Birds North Am. Online*, 2006.
[187] S. Granholm, L. Mewaldt, and R. Duke, “HOUSE SPARROW Passer domesticus,” in *California Wildlife Habitat Relationships (CWHR)*, Vol I-III., D. C. Zeiner, W. F. L. Jr., K. E. Mayer, and M. White, Eds. Sacramento, California: California Depart. of Fish and Game, 1990.
[188] G. Ahlborn, N. Johnson, and G. Ahlborn, “RUFFED GROUSE Bonasa umbellus,” in *California Wildlife Habitat Relationships (CWHR)*, Vol I-III., D. C. Zeiner, W. F. L. Jr., K. E. Mayer, and M. White, Eds. Sacramento, California: California Depart. of Fish and Game, 1990.
[189] D. H. Rusch, S. Destefano, M. C. Reynolds, and D. Lauten, “Ruffed Grouse (Bonasa umbellus),” *Birds North Am. Online*, 2000.
[190] J. T. McRoberts, M. C. Wallace, and S. W. Eaton, “Wild Turkey (Meleagris gallopavo),” *Birds North Am. Online*, 2014.
[191] C. Novoa, S. Dumas, and J. Resseguier, “Home-range size of Pyrenean grey partridges Perdix perdix hispaniensis during the breeding season,” *Wildlife Biology*, vol. 12, no. 1. pp. 11–18, 2006.
[192] Iowa Department of Natural Resources, “Gray (Hungarian) partridge (Perdix perdix),” 2001. [Online]. Available: http://www.iowadnr.gov/Portals/idnr/uploads/education/Species/birds/Hun.pdf. Archived at: http://www.webcitation.org/6lCyR61f6. [Accessed: 06-Nov-2015].
[193] J. P. Carroll, “Gray Partridge (Perdix perdix),” *Birds North Am. Online*, 1993.
[194] New World Encyclopedia contributors, “Pheasant,” *New World Encyclopedia*, 2009. [Online]. Available: http://www.newworldencyclopedia.org/entry/Pheasant Archived at: http://www.webcitation.org/6lD38ZdUX. [Accessed: 26-Nov-2015].
[195] G. Ahlborn, N. Johnson, and G. Ahlborn, “RING-NECKED PHEASANT Phasianus colchicus,” in *California Wildlife Habitat Relationships (CWHR)*, Vol I-III., D. C. Zeiner, W. F. L. Jr., K. E. Mayer, and M. White, Eds. Sacramento, California: California Depart. of Fish and Game, 1990.
[196] J. H. Giudice and J. T. Ratti, “Ring-necked Pheasant (Phasianus colchicus),” *Birds North Am. Online*, 2001.

[197] J. W. Connelly, M. W. Gratson, and K. P. Reese, “Sharp-tailed Grouse (Tympanuchus phasianellus),” *Birds North Am. Online*, 1998.
[198] K. L. Wiebe and W. S. Moore, “Northern Flicker (Colaptes auratus),” *Birds North Am. Online*, 2008.
[199] B. W. Tobalske, “Scaling of Muscle Composition, Wing Morphology, and Intermittent Flight Behavior in Woodpeckers,” *Auk*, vol. 113, no. 1, pp. 151–177, Jan. 1996.
[200] E. L. Bull and J. A. Jackson, “Pileated Woodpecker (Dryocopus pileatus),” *Birds North Am. Online*, 1995.
[201] K. T. Vierling, V. A. Saab, and B. W. Tobalske, “Lewis’s Woodpecker (Melanerpes lewis),” *Birds North Am. Online*, 2013.
[202] B. Frei, K. G. Smith, J. H. Withgott, and P. G. Rodewald, “Red-headed Woodpecker (Melanerpes erythrocephalus),” *Birds North Am. Online*, 2015.
[203] D. L. Leonard, Jr., “American Three-toed Woodpecker (Picoides dorsalis),” *Birds North Am. Online*, 2001.
[204] J. A. Jackson and H. R. Ouellet., “Downy Woodpecker (Picoides pubescens),” *Birds North Am. Online*, 2002.
[205] L. Kilham, “Courtship and Territorial Behaviour of Hairy Woopeckers,” *Auk*, vol. 77, no. 3, pp. 259–270, 1960.
[206] K. A. Covert-Bratland, W. M. Block, and T. C. Theimer, “Hairy woodpecker winter ecology in ponderosa pine forests representing different ages since wildfire,” *J. Wildl. Manage.*, vol. 70, no. 5, pp. 1379–1392, 2006.
[207] J. A. Jackson, H. R. Ouellet, and B. J. Jackson, “Hairy Woodpecker (Picoides villosus),” *Birds North Am. Online*, 2002.
[208] L. Kilham, “Reproductive Behavior of Hairy Woodpeckers I. Pair Formation and Courtship,” *Wilson Bull.*, vol. 78, no. 3, pp. 251–265, 1966.
[209] D. L. Swanson, J. L. Ingold, and G. E. Wallace, “Ruby-crowned Kinglet (Regulus calendula),” *Birds North Am. Online*, 1994.
[210] D. L. Swanson, J. L. Ingold, and R. Galati, “Golden-crowned Kinglet (Regulus satrapa),” *Birds North Am. Online*, 2012.
[211] C. K. Ghalambor and T. E. Martin, “Red-breasted Nuthatch (Sitta canadensis),” *Birds North Am. Online*, 1999.
[212] S. C. Ball, “Migration of Red-Breasted Nuthatches in Gaspe,” *Ecol. Monogr.*, vol. 17, no. 4, pp. 501–533, Feb. 1947.
[213] T. C. Grubb, Jr. and V. V. Pravosudov, “White-breasted Nuthatch (Sitta carolinensis),” *Birds North Am. Online*, 2008.
[214] Simply Wild Canada, “Owls of Canada.” [Online]. Available: http://www.simplywildcanada.com/wild-species/birds-of-canada/owls-of-canada/ Archived at: http://www.webcitation.org/6lD3RVJm6. [Accessed: 05-Nov-2015].
[215] D. A. Wiggins, D. W. Holt, and S. M. Leasure, “Short-eared Owl (Asio flammeus),” *Birds North Am. Online*, 2006.

[216] J. R. Bennett and P. H. Bloom, “Range and Habitat Use by Great Horned Owls (Bubo virginianus) in Southern California,” *J. Raptor Res.*, vol. 39, no. 2, pp. 119–126, 2005.
[217] C. Artuso, C. S. Houston, D. G. Smith, and C. Rohner, “Great Horned Owl (Bubo virginianus),” *Birds North Am. Online*, 2014.
[218] A. M. Makarieva, V. G. Gorshkov, and B. L. Li, “Why do population density and inverse home range scale differently with body size?: Implications for ecosystem stability,” *Ecol. Complex.*, vol. 2, no. 3, pp. 259–271, 2005.
[219] K. M. Mazur and P. C. James, “Barred Owl (Strix varia),” *Birds North Am. Online*, 2000.
[220] R. A. Kroeger, H. D. Grushka, and T. C. Helvey, “Low Speed Aerodynamics for Ultra-Quiet Flight,” Mar. 1972.
[221] P. R. Cabe, “European Starling (Sturnus vulgaris),” *Birds North Am. Online*, 1993.
[222] S. Granholm, L. Mewaldt, and R. Duke, “EUROPEAN STARLING Sturnus vulgaris,” in *California Wildlife Habitat Relationships (CWHR)*, Vol I-III., D. C. Zeiner, W. F. L. Jr., K. E. Mayer, and M. White, Eds. Sacramento, California: California Depart. of Fish and Game, 1990.
[223] S. Weidensaul, T. R. Robinson, R. R. Sargent, and M. B. Sargent, “Ruby-throated Hummingbird (Archilochus colubris),” *Birds North Am. Online*, 2013.
[224] L. Chambers, “Ruby-throated Hummingbird.” [Online]. Available: http://www.hummingbirds.net/rubythroated.html. Archived at: http://www.webcitation.org/6lD3wmHf8. [Accessed: 13-Nov-2015].
[225] F. Rousseu, Y. Charette, and M. Bélisle, “Resource defense and monopolization in a marked population of ruby-throated hummingbirds (Archilochus colubris),” *Ecol. Evol.*, vol. 4, no. 6, pp. 776–793, 2014.
[226] Texas Parks and Wildlife Department, “Urban Wildlife Fact Sheet Set.” [Online]. Available: https://tpwd.texas.gov/publications/pwdpubs/media/pwd_lf_k0700_0167.pdf Archived at: http://www.webcitation.org/6lCxcCrDf. [Accessed: 13-Nov-2015].
[227] D. E. Kroodsma and J. Verner, “Marsh Wren (Cistothorus palustris),” *Birds North Am. Online*, 2014.
[228] K. J. Gutzwiller and S. H. Anderson, “Habitat Suitability Index Models: Marsh Wren,” *Biol. Rep.*, vol. 82, no. 10.139, p. 13, 1987.
[229] C. A. Hartman, J. T. Ackerman, G. Herring, J. Isanhart, and M. Herzog, “Marsh wrens as bioindicators of mercury in wetlands of Great Salt Lake: do blood and feathers reflect site-specific exposure risk to bird reproduction?,” *Environ. Sci. Technol.*, vol. 47, no. 12, pp. 6597–605, Jun. 2013.
[230] J. T. Burns, “Nests, Territories, and Reproduction of Sedge Wrens (Cistothorus platensis),” *Wilson Bull.*, vol. 94, no. 3, pp. 338–349, 1982.
[231] J. R. Herkert, D. E. Kroodsma, and J. P. Gibbs, “Sedge Wren (Cistothorus platensis),” *Birds North Am. Online*, 2001.
[232] A. Fujikawa, “Home range of Coryphaspiza melanotis and Cistothorus platensis in the central Brasil and a review of home ranges and territories of birds in the Neotropics,” Universidade de São Paulo, 2012.

[233] D. E. Kroodsma, “Coexistence of Bewick’s Wrens and House Wrens in Oregon,” *Auk*, vol. 90, no. 2, pp. 341–352, 1973.
[234] L. S. Johnson, “House Wren (Troglodytes aedon),” *Birds North Am. Online*, 2014.
[235] S. J. Hejl, J. A. Holmes, and D. E. Kroodsma, “Winter Wren (Troglodytes troglodytes),” *Birds North Am. Online*, 2002.
[236] L. R. Bevier, A. F. Poole, and W. Moskoff, “Veery (Catharus fuscescens),” *Birds North Am. Online*, 2005.
[237] N. J. Bayly, C. Gómez, and K. A. Hobson, “Energy reserves stored by migrating Gray-cheeked Thrushes Catharus minimus at a spring stopover site in northern Colombia are sufficient for a long-distance flight to North America,” *Ibis (Lond. 1859).*, vol. 155, no. 2, pp. 271–283, Apr. 2013.
[238] R. Dellinger, P. B. Wood, P. W. Jones, and T. M. Donovan, “Hermit Thrush (Catharus guttatus),” *Birds North Am. Online*, 2012.
[239] T. Rinaldi and M. Worland, “Conservation Assessment for Swainson’s Thrush (Catharus ustulatus),” *USDA Forest Service, Eastern Region*, 2004. [Online]. Available: http://www.fs.usda.gov/Internet/FSE_DOCUMENTS/fsm91_054321.pdf. Archived at: http://www.webcitation.org/6lCyjP0XW. [Accessed: 18-Nov-2015].
[240] D. E. Mack and W. Yong, “Swainson’s Thrush (Catharus ustulatus),” *Birds North Am. Online*, 2000.
[241] H. W. Power and M. P. Lombard, “Mountain Bluebird (Sialia currucoides),” *Birds North Am. Online*, 1996.
[242] D. Gaines, L. Mewaldt, and R. Duke, “MOUNTAIN BLUEBIRD Sialia currucoides,” in *California Wildlife Habitat Relationships (CWHR)*, Vol I-III., D. C. Zeiner, W. F. L. Jr., K. E. Mayer, and M. White, Eds. Sacramento, California: California Depart. of Fish and Game, 1990.
[243] P. A. Gowaty and J. H. Plissner, “Eastern Bluebird (Sialia sialis),” *Birds North Am. Online*, no. 381, 2015.
[244] D. H. Hirth, A. E. Hester, and F. Greeley, “Dispersal and Flocking of Marked Young Robins (Turdus M. Migratorius) after Fledging,” *Bird-Banding*, vol. 40, no. 3, p. 208, 1969.
[245] N. Vanderhoff, R. Sallabanks, and F. C. James, “American Robin (Turdus migratorius),” *Birds North Am. Online*, 2014.
[246] M. A. Diuk-Wasser, G. Molaei, J. E. Simpson, C. M. Folsom-O’Keefe, P. M. Armstrong, and T. G. Andreadis, “Avian communal roosts as amplification foci for West Nile virus in urban areas in Northeastern United States,” *Am. J. Trop. Med. Hyg.*, vol. 82, no. 2, pp. 337–343, 2010.
[247] S. Granholm, L. Mewaldt, and R. Duke, “AMERICAN ROBIN Turdus migratorius,” in *California Wildlife Habitat Relationships (CWHR)*, Vol I-III., D. C. Zeiner, W. F. L. Jr., K. E. Mayer, and M. White, Eds. Sacramento, California: California Depart. of Fish and Game, 1990.
[248] F. T. Muijres, M. S. Bowlin, L. C. Johansson, and A. Hedenström, “Vortex wake, downwash distribution, aerodynamic performance and wingbeat kinematics in slow-flying pied flycatchers.,” *J. R. Soc. Interface*, vol. 9, no. 67, pp. 292–303, Feb. 2012.
[249] B. Altman and R. Sallabanks, “Olive-sided Flycatcher (Contopus cooperi),” *Birds North Am. Online*, 2012.

[250] P. Otto, “What’s a stranded wood-pewee to do?,” *Kane County Chronicle*, 2012. [Online]. Available: http://www.kcchronicle.com/mobile/article.xml/articles/2012/11/02/r_6zeodgkeslmsrmtheo2qvg/index.xml Archived at: http://www.webcitation.org/6lD4PICRO. [Accessed: 18-Nov-2015].

[251] C. Bemis and J. D. Rising, “Western Wood-Pewee (Contopus sordidulus),” *Birds North Am. Online*, 1999.

[252] R. C. Eckhardt, “The Adaptive Syndromes of Two Guilds of Insectivorous Birds in the Colorado Rocky Mountains,” *Ecol. Monogr.*, vol. 49, no. 2, pp. 129–149, 1979.

[253] J. P. McCarty, “Eastern Wood-Pewee (Contopus virens),” *Birds North Am. Online*, 1996.

[254] P. E. Lowther, “Alder Flycatcher (Empidonax alnorum),” *Birds North Am. Online*, 1999.

[255] B. Dunson, “The Incredible Flight of a Willow Flycatcher,” *The Galax Gazette*, 2014. [Online]. Available: http://www.galaxgazette.com/content/incredible-flight-willow-flycatcher Archived at: http://www.webcitation.org/6lD4ZehQb. [Accessed: 18-Nov-2015].

[256] D. A. Gross and P. E. Lowther, “Yellow-bellied Flycatcher (Empidonax flaviventris),” *Birds North Am. Online*, 2011.

[257] F. Doyon, P. E. Higgelke, and H. L. MacLeod, “Least Flycatcher (Empidonax minimus),” *Prepared for Millar Western Forest Products’ Biodiversity Assessment Project - ISFORT*, 2000. [Online]. Available: http://isfort.uqo.ca/sites/isfort.uqo.ca/files/fichiers/publications_ISFORT/least_flycatcher_hsm.pdf. Archived at: http://www.webcitation.org/6lD4d2d7K. [Accessed: 18-Nov-2015].

[258] S. Tarof and J. V. Briskie, “Least Flycatcher (Empidonax minimus),” *Birds North Am. Online*, 2008.

[259] S. N. Cardinal and E. H. Paxton, “Home Range, movement, and habitat use of the Southwestern Willow Flycatcher, Roosevelt Lake, AZ - 2004,” *U.S. Geological Survey report to the U.S. Bureau of Reclamation, Phoenix*, 2005. [Online]. Available: http://sbsc.wr.usgs.gov/cprs/research/projects/swwf/Reports/telemetry2004report.pdf. [Accessed: 18-Nov-2015].

[260] J. A. Sedgwick, “Willow Flycatcher (Empidonax traillii),” *Birds North Am. Online*, 2000.

[261] K. E. Miller and W. E. Lanyon, “Great Crested Flycatcher (Myiarchus crinitus),” *Birds North Am. Online*, 1997.

[262] H. P. Weeks Jr., “Eastern Phoebe (Sayornis phoebe),” *Birds North Am. Online*, 2011.

[263] D. W. Johnston, “Niche Relationships among Some Deciduous Forest Flycatchers,” *Auk*, vol. 88, no. 4, pp. 796–804, 1971.

[264] D. Gaines, L. Mewaldt, and R. Duke, “SAY’S PHOEBE Sayornis saya,” in *California Wildlife Habitat Relationships (CWHR)*, Vol I-III., D. C. Zeiner, W. F. L. Jr., K. E. Mayer, and M. White, Eds. Sacramento, California: California Depart. of Fish and Game, 1990.

[265] J. M. Schukman and B. O. Wolf, “Say’s Phoebe (Sayornis saya),” *Birds North Am. Online*, 1998.

[266] M. T. Murphy, “Eastern Kingbird (Tyrannus tyrannus),” *Birds North Am. Online*, 1996.

[267] L. R. Gamble and T. M. Bergin, “Western Kingbird (Tyrannus verticalis),” *Birds North Am. Online*, 2012.

[268] H. A. Hespenheide, “Competition and the Genus Tyrannus,” *Wilson Bull.*, vol. 76, no. 3, pp. 265–281, 1964.
[269] P. G. Rodewald and R. D. James., “Yellow-throated Vireo (Vireo flavifrons),” *Birds North Am. Online*, 2011.
[270] R. D. James, “Foraging Behavior and Habitat Selection of Three Species of Vireos in Southern Ontario,” *Wilson Bull.*, vol. 88, no. 1, pp. 62–75, 1976.
[271] R. Sterling, “Vireo gilvus: eastern warbling-vireo,” *Animal Diversity Web*, 2011. [Online]. Available: http://www.biokids.umich.edu/critters/Vireo_gilvus/ Archived at: http://www.webcitation.org/6lD4knNii. [Accessed: 01-Dec-2015].
[272] T. Gardali and G. Ballard, “Warbling Vireo (Vireo gilvus),” *Birds North Am. Online*, 2000.
[273] D. A. Cimprich, F. R. Moore, and M. P. Guilfoyle, “Red-eyed Vireo (Vireo olivaceus),” *Birds North Am. Online*, 2000.
[274] W. Moskoff and S. K. Robinson, “Philadelphia Vireo (Vireo philadelphicus),” *Birds North Am. Online*, 2011.
[275] E. Morton and R. D. James, “Blue-headed Vireo (Vireo solitarius),” *Birds North Am. Online*, 2014.

Appendix B: Supplementary Figures

Daily mosquito counts in selected Manitoba communities, and average (per-community) daily temperature and rainfall values for a number of years are provided in figures below.

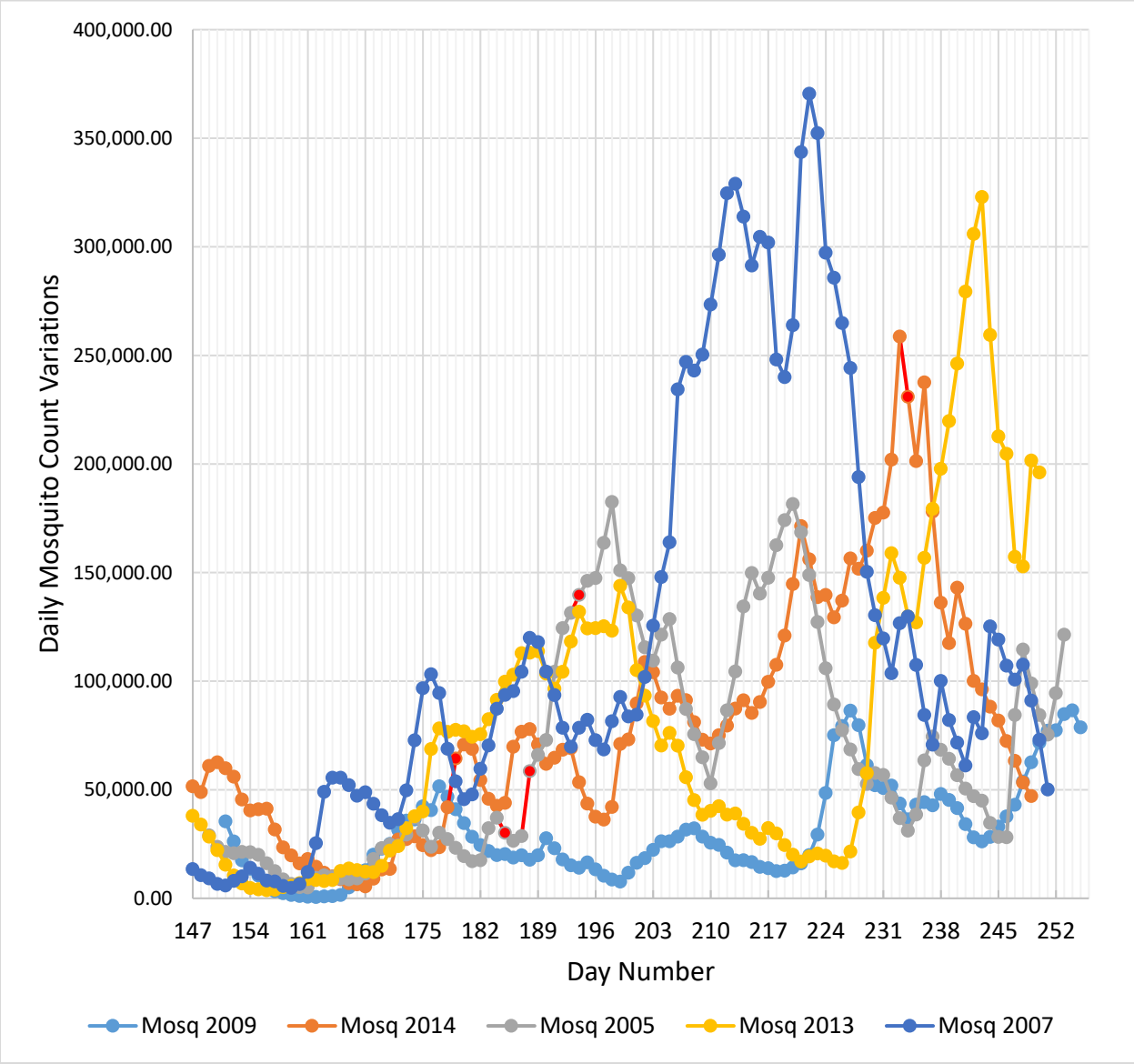


Figure B-1 Total number of adult mosquitoes in selected communities per day for years 2005, 2009, 2007, 2013, and 2014

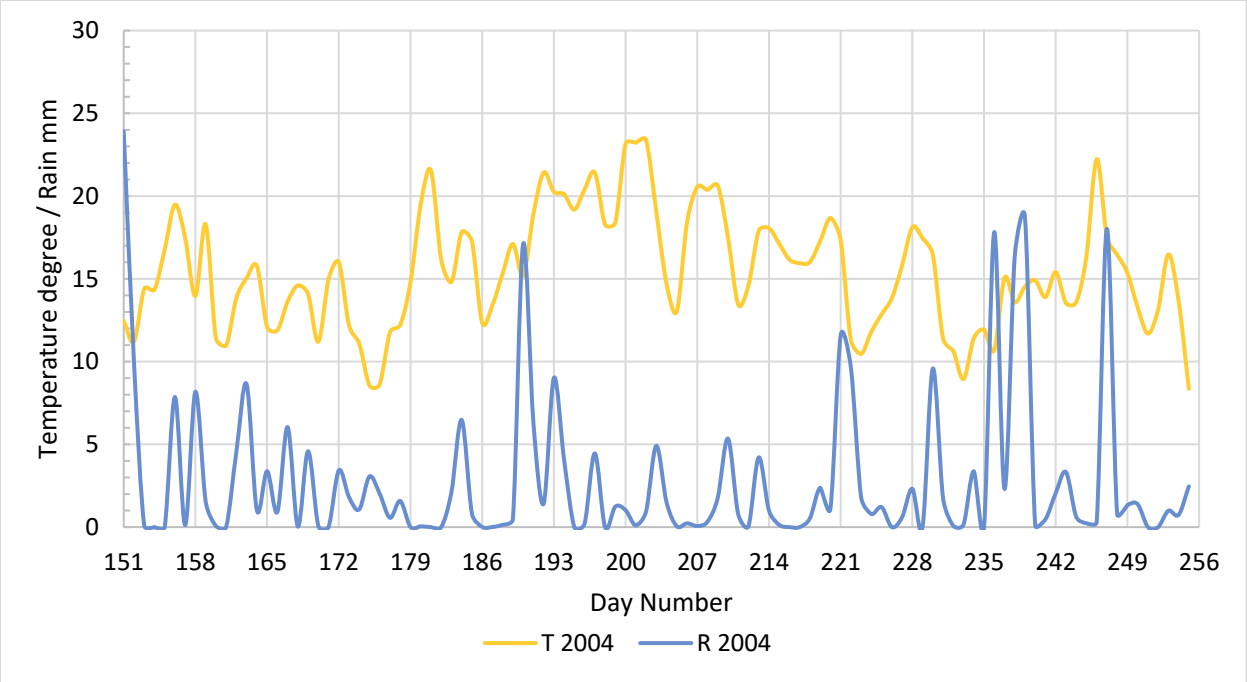


Figure B-2 Temperature and rainfall daily variations in 2004; CDC week 22 began on day 151 i.e., Sunday May 30, 2004

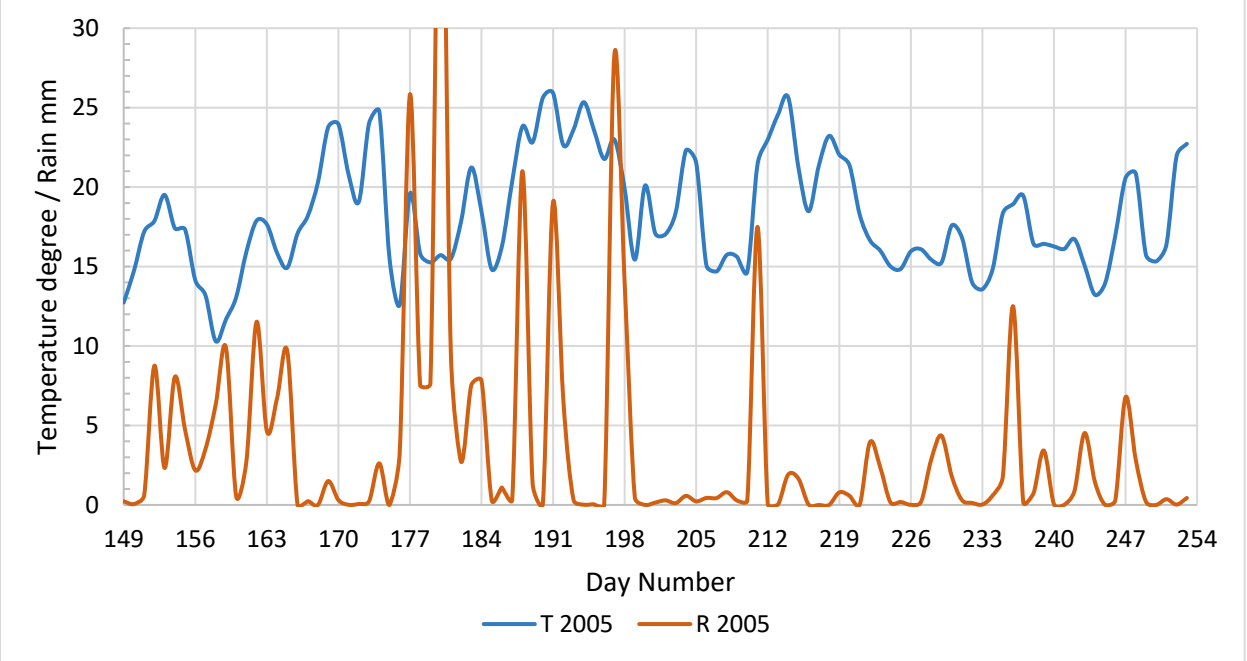


Figure B-3 Temperature and rainfall daily variations in 2005; CDC week 22 began on day 149 i.e., Sunday May 29, 2005

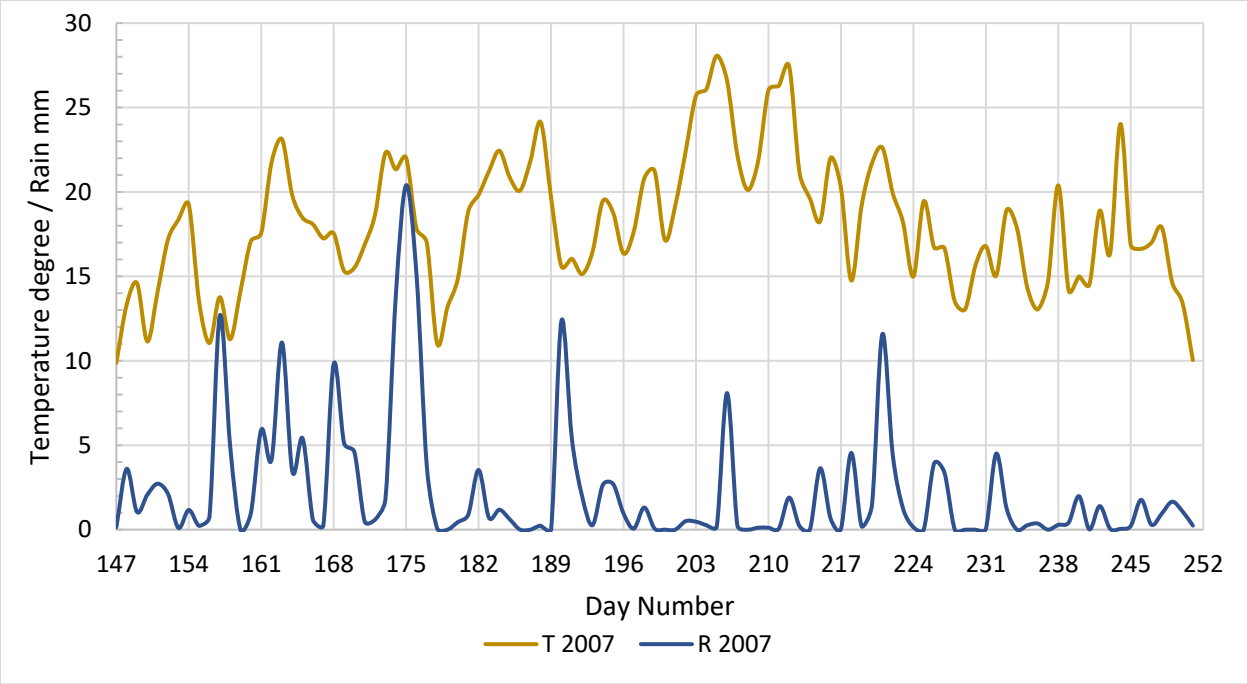


Figure B-4 Temperature and rainfall daily variations in 2007; CDC week 22 began on day 147 i.e., Sunday May 27, 2007

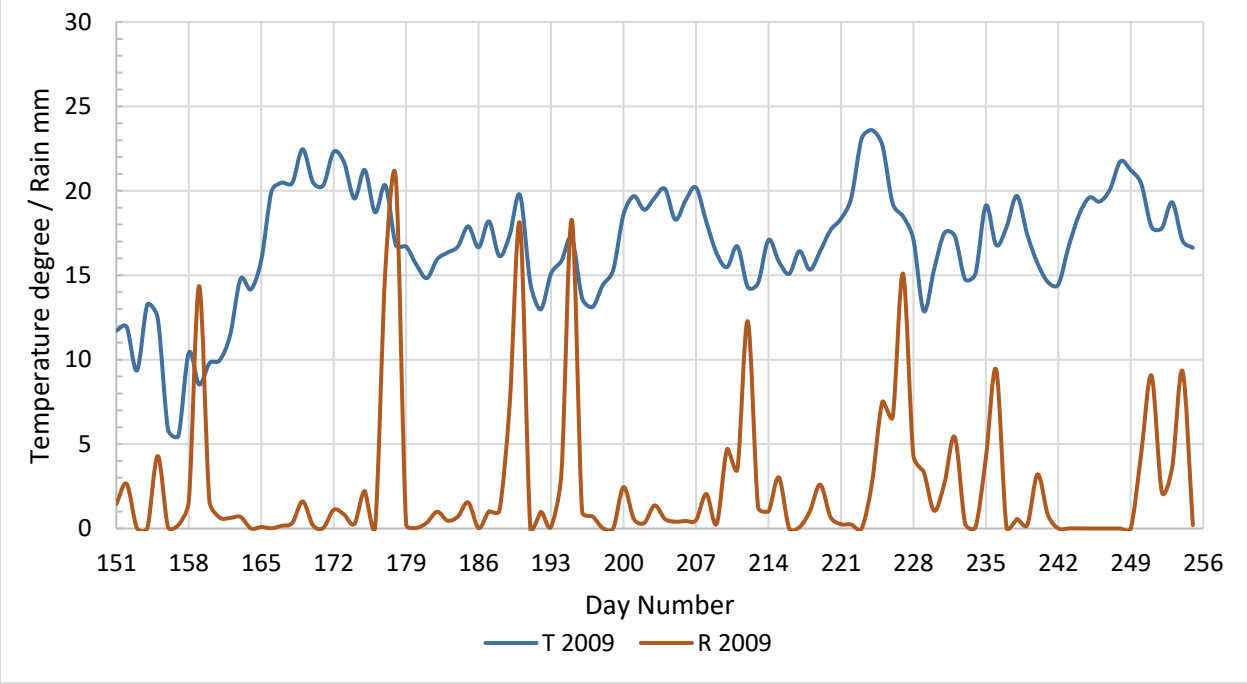


Figure B-5 Temperature and rainfall daily variations in 2009; CDC week 22 began on day 151 i.e., Sunday May 31, 2009

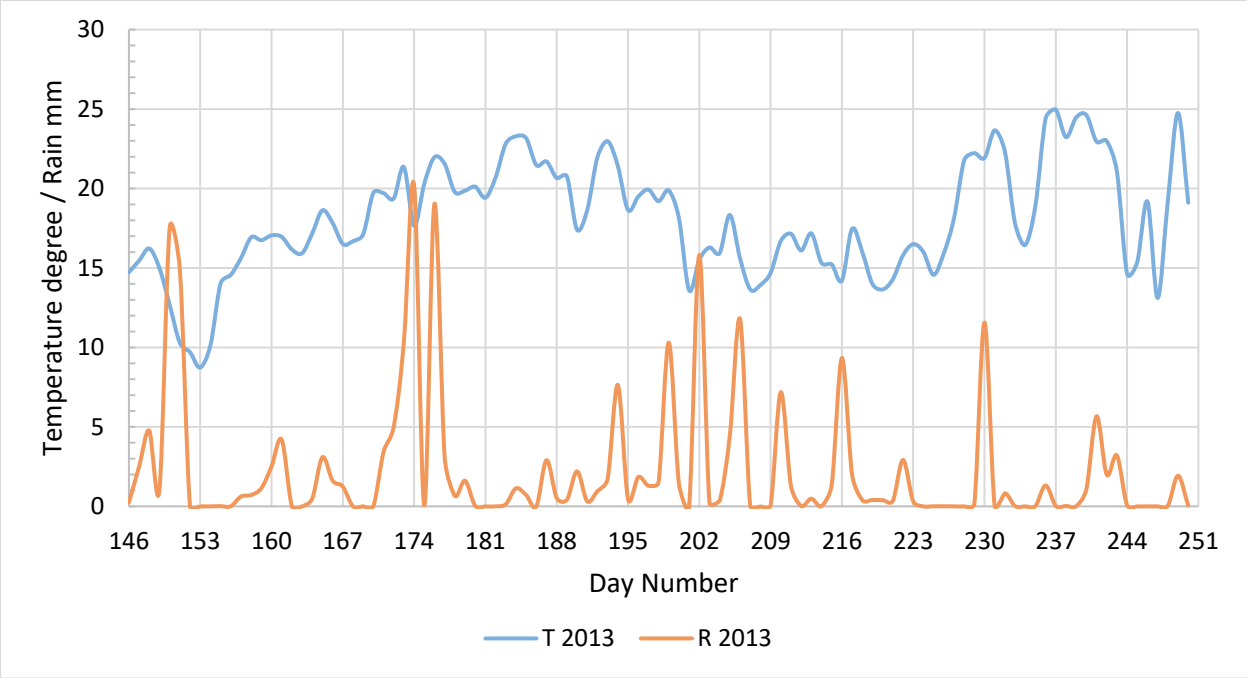


Figure B-6 Temperature and rainfall daily variations in 2013; CDC week 22 began on day 146 i.e., Sunday May 26, 2013

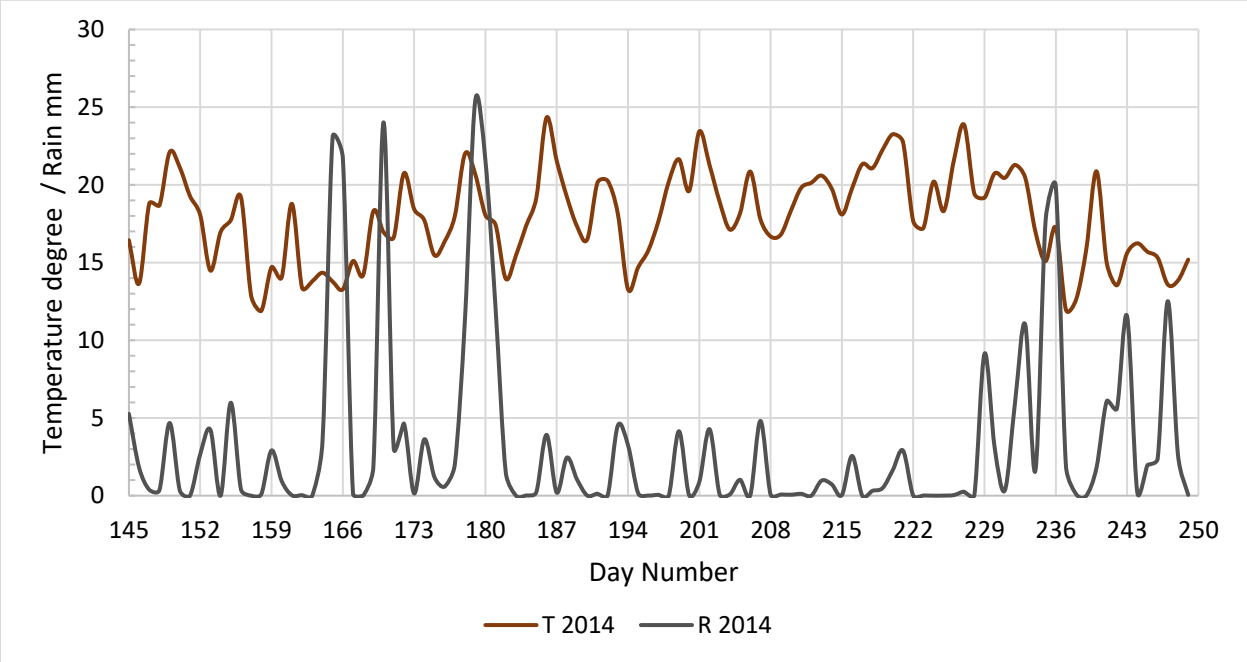
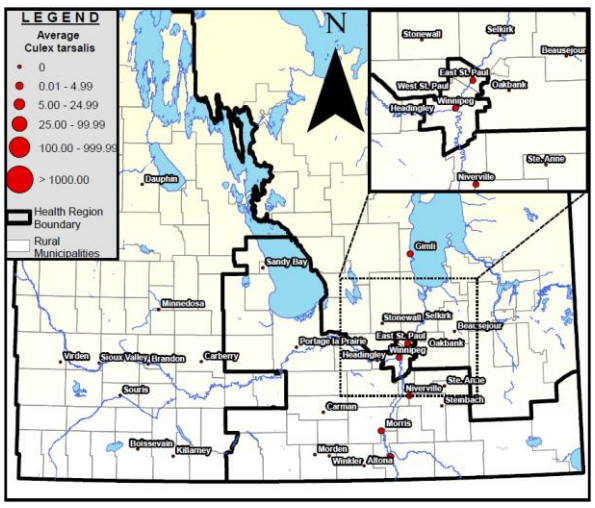
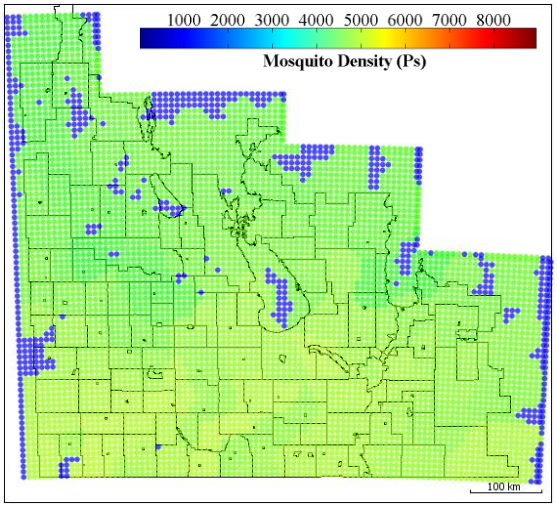
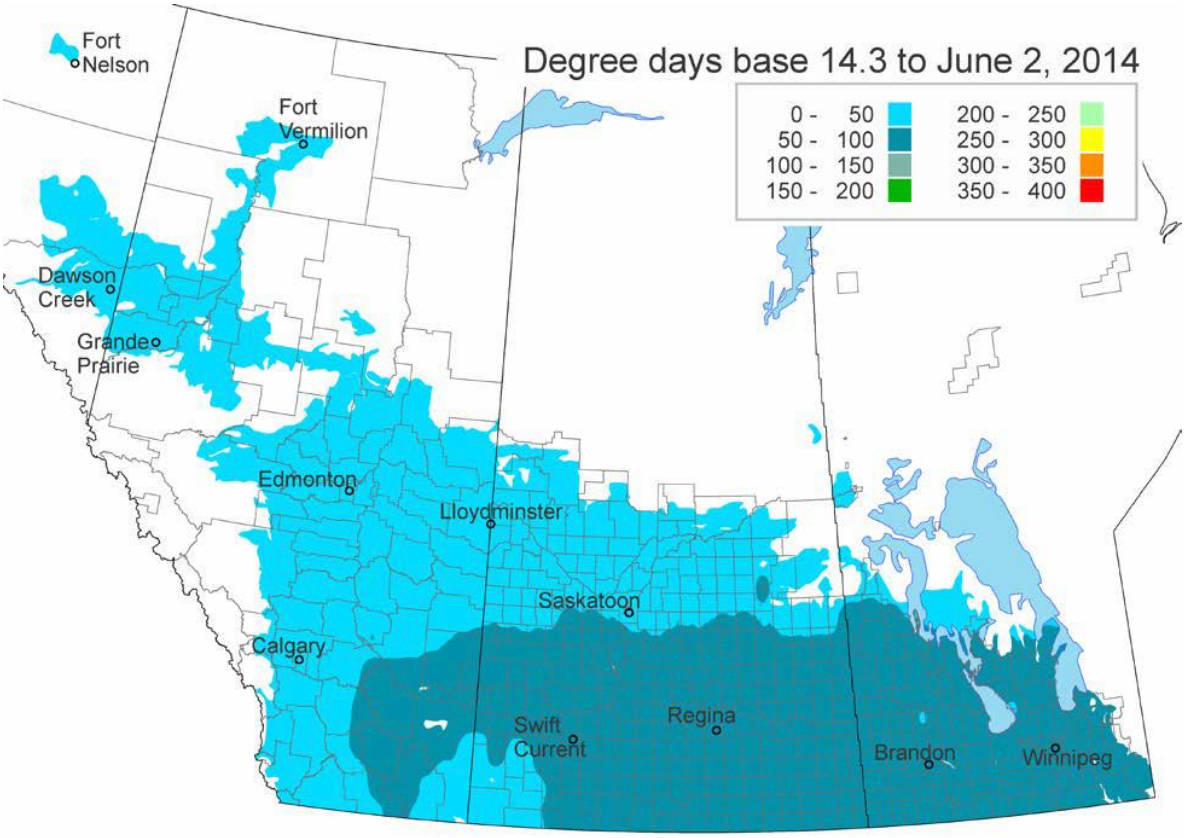


Figure B-7 Temperature and rainfall daily variations in 2014; CDC week 22 began on day 145 i.e., Sunday May 25, 2014

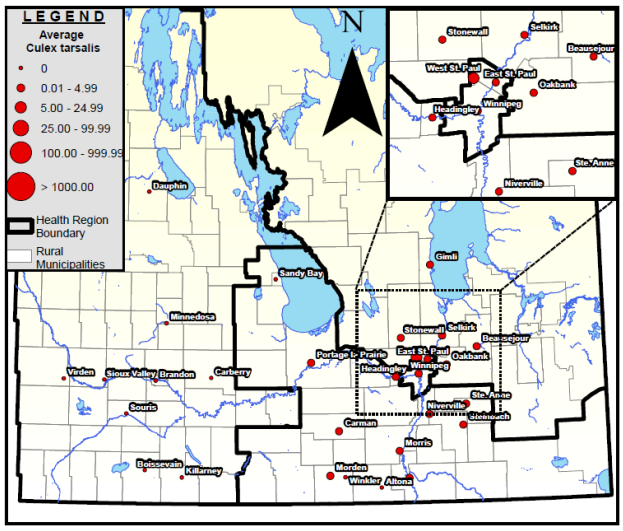
Comparison of weekly distribution of mosquitoes across the province according to Manitoba trap data [20] and simulation, and the map of cumulative degree-days [20], [158] from week 22 to week 37 of 2014 are given in figures below.

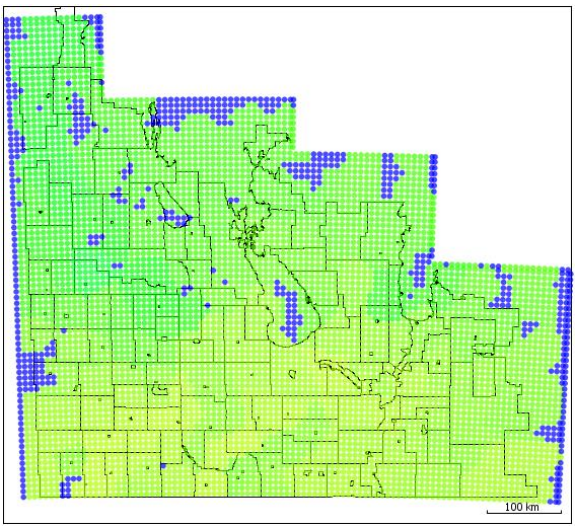


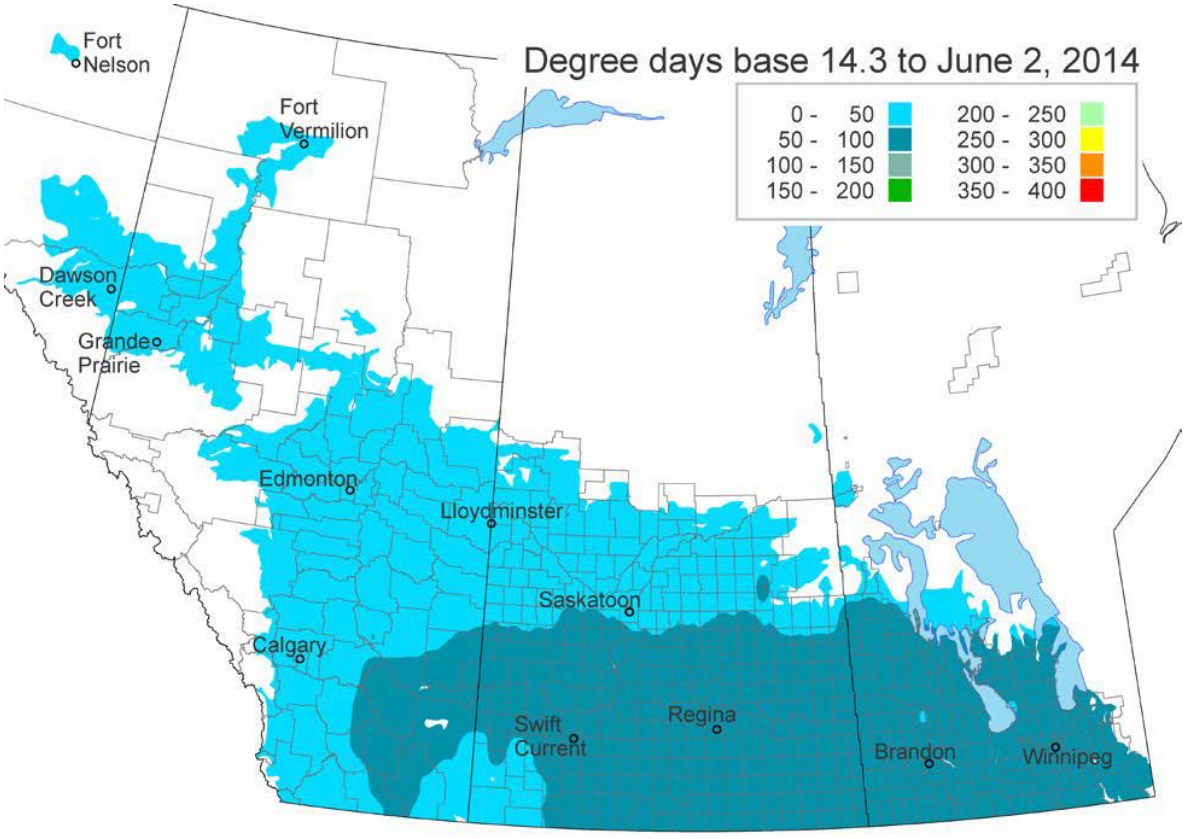




Week 22 (Trap data and degree-days image sources: [20], [158])

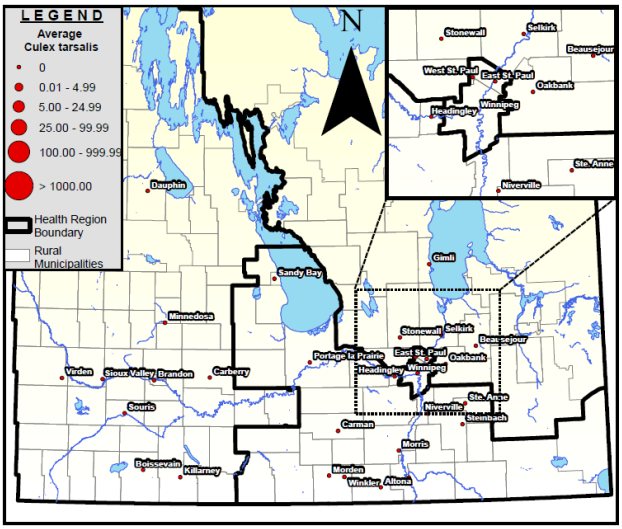


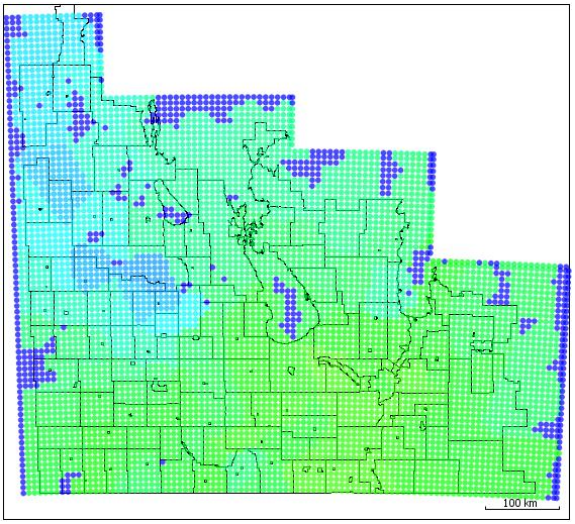


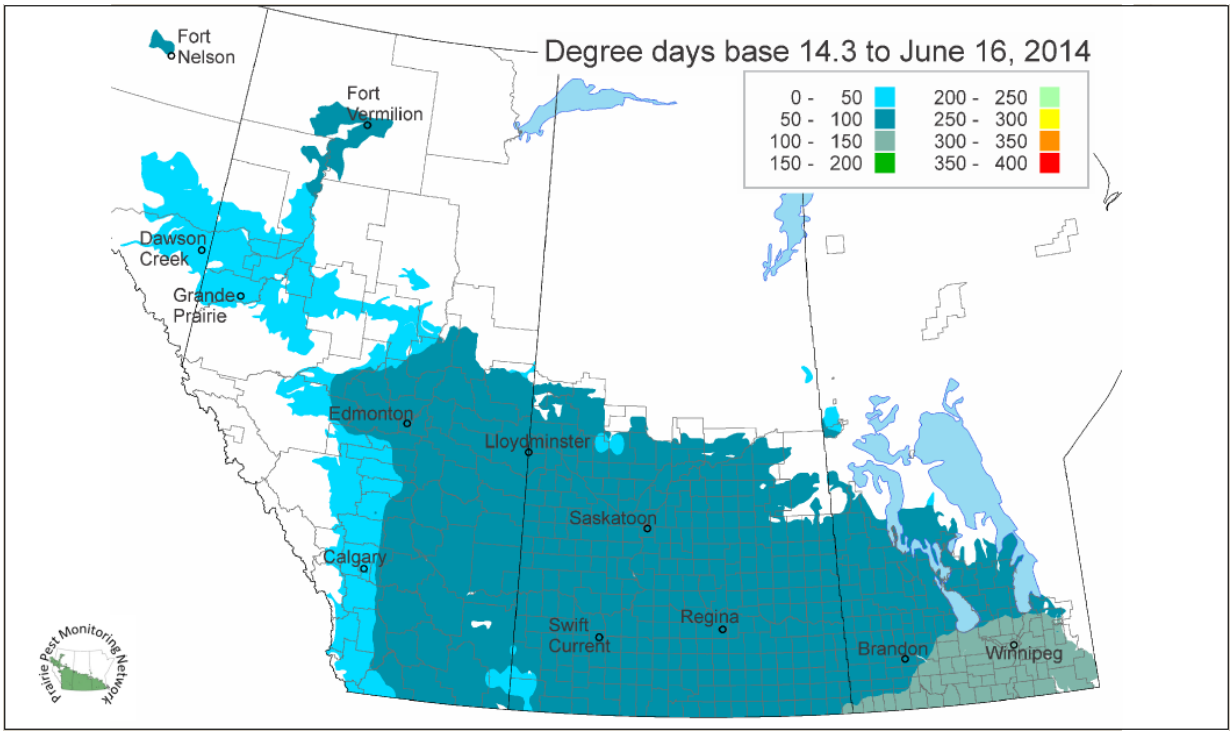


Week 23

(Trap data and degree-days image sources: [20], [158])

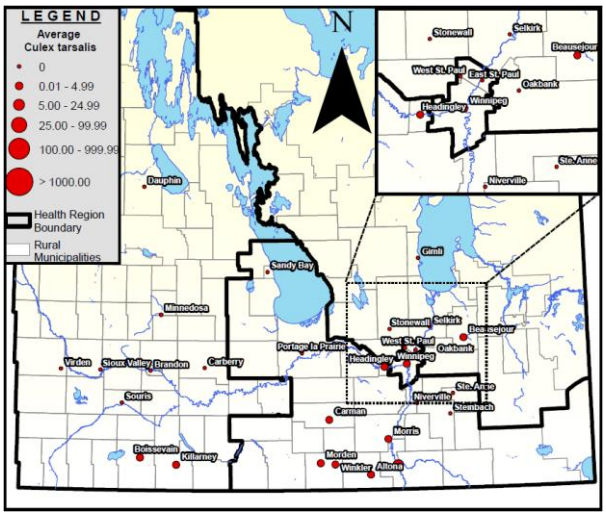


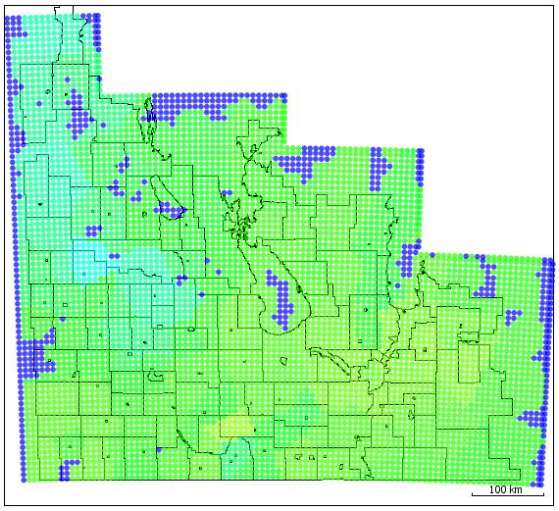


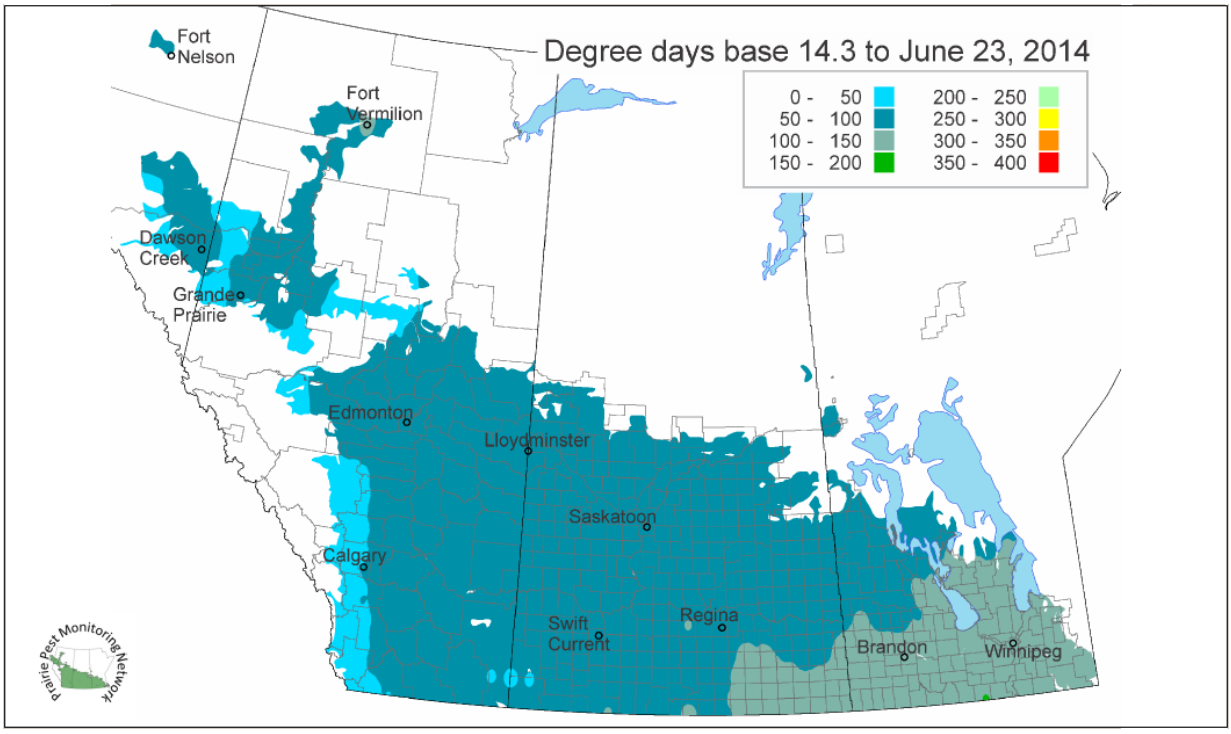


Week 24

(Trap data and degree-days image sources: [20], [158])

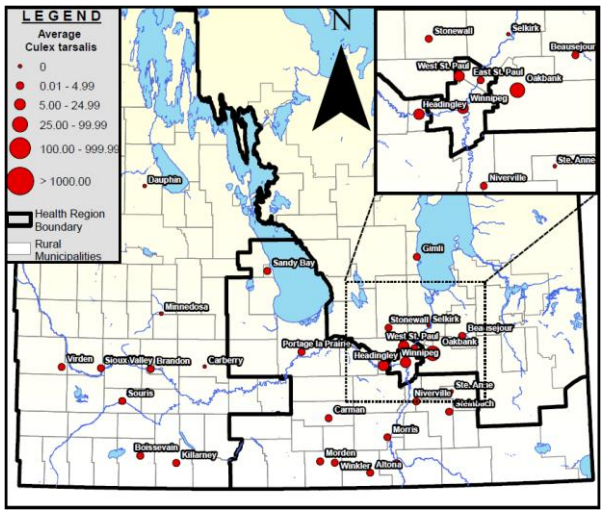


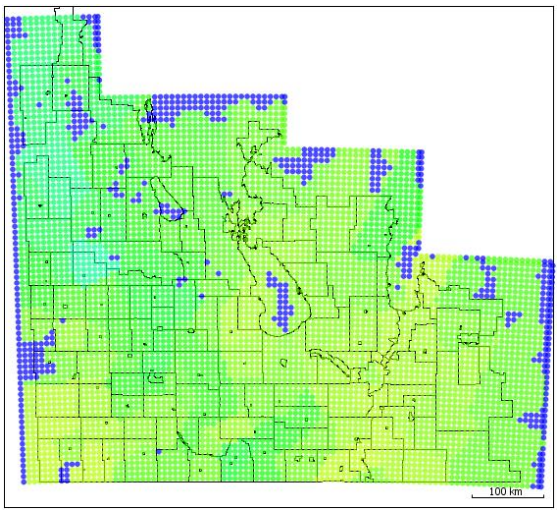


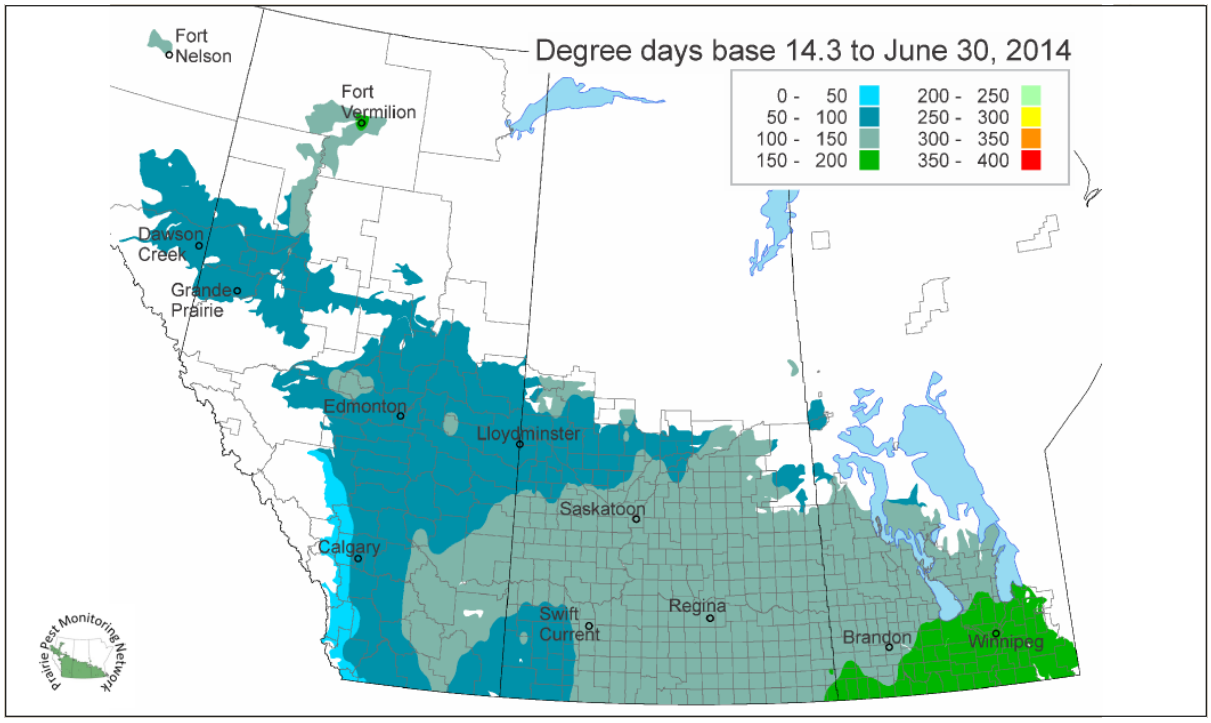


Week 25

(Trap data and degree-days image sources: [20], [158])

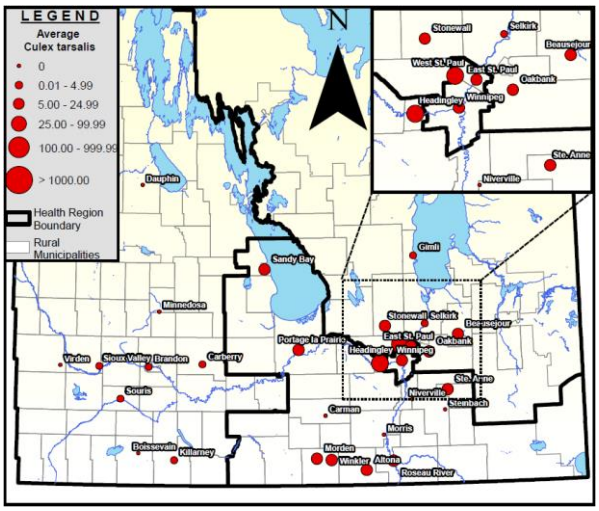


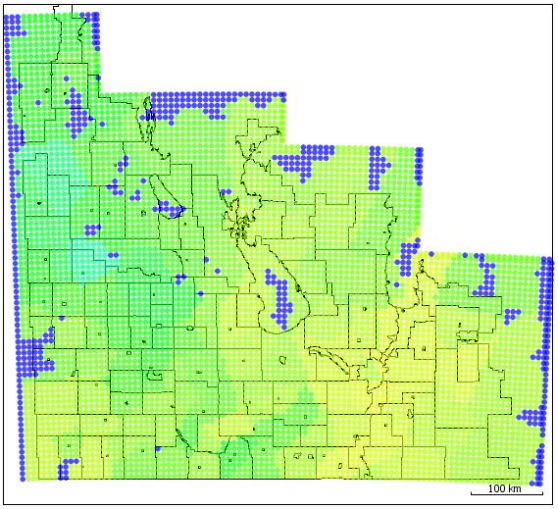


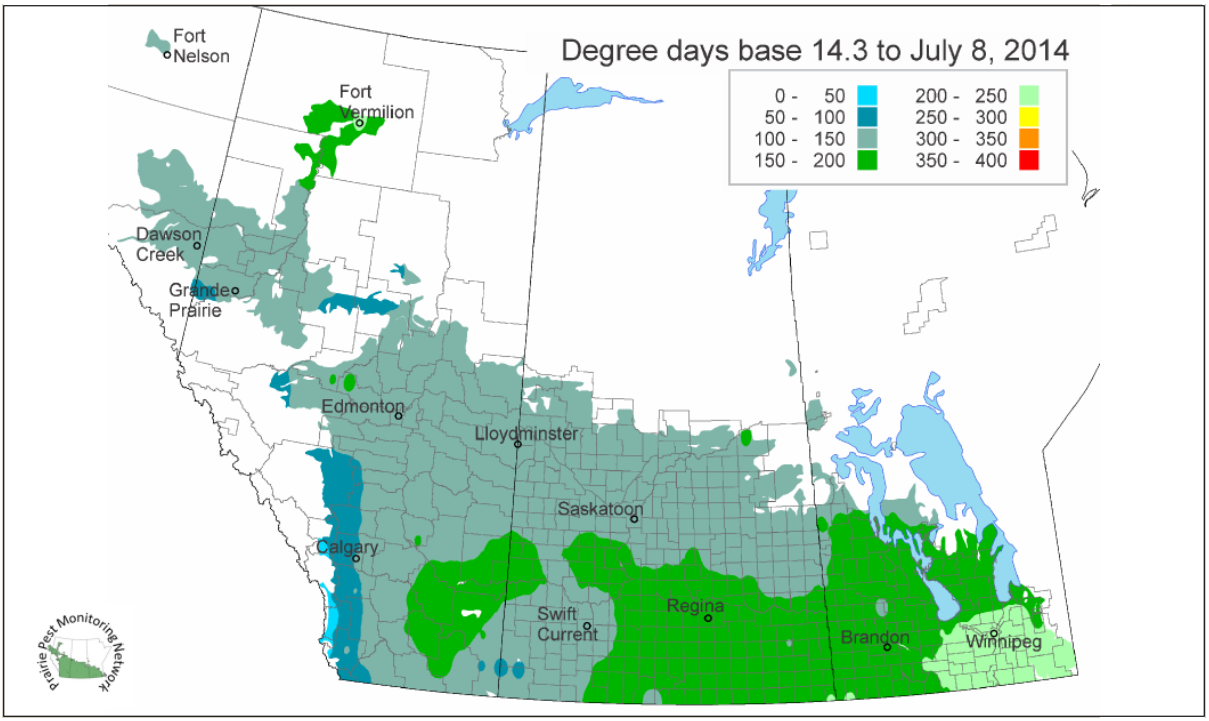


Week 26

(Trap data and degree-days image sources: [20], [158])

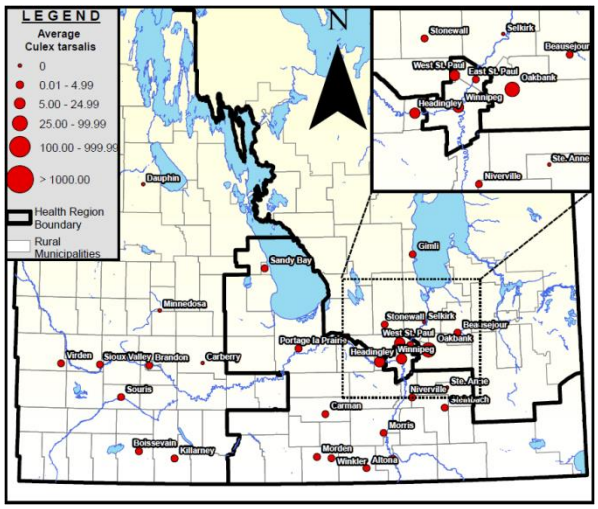


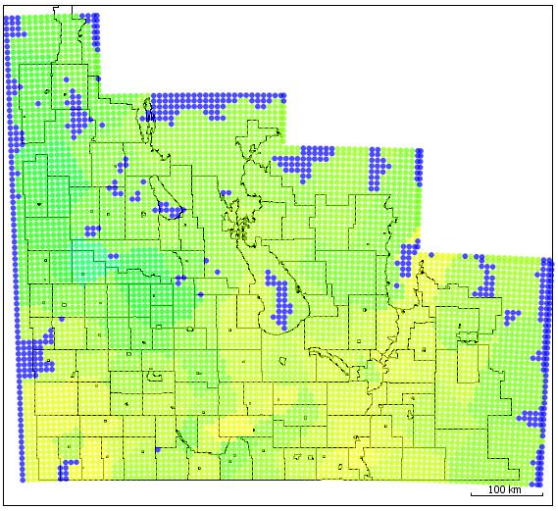


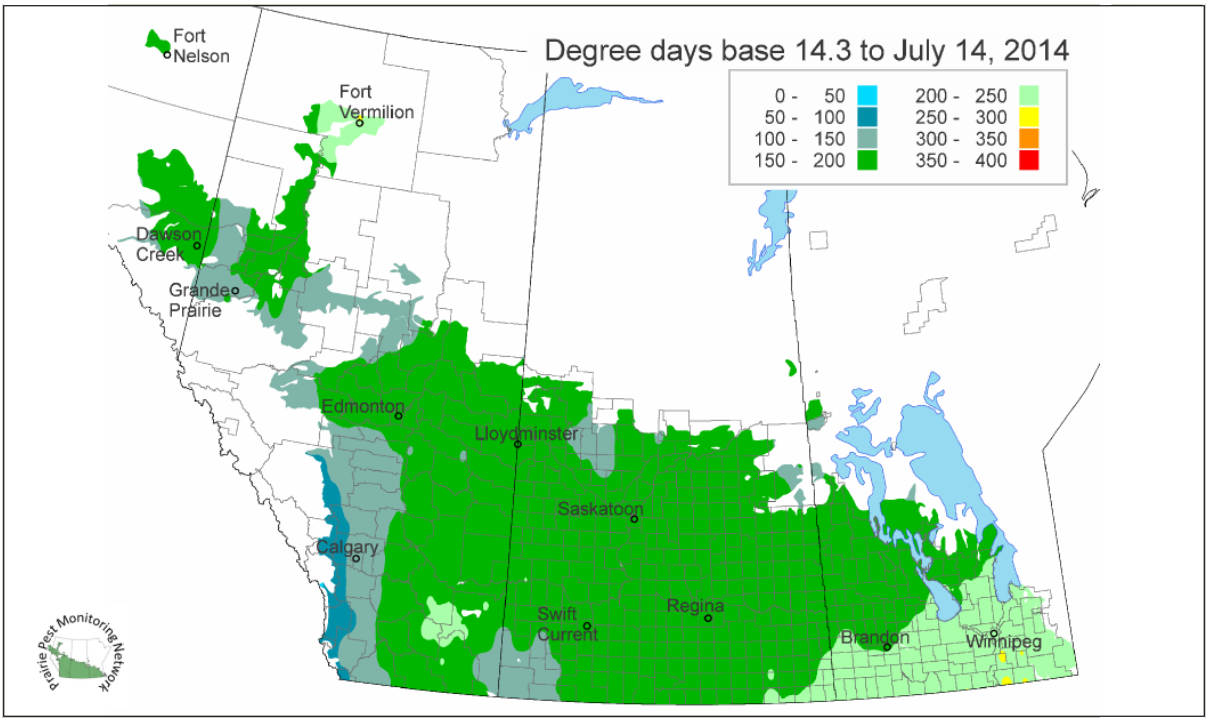


Week 27

(Trap data and degree-days image sources: [20], [158])

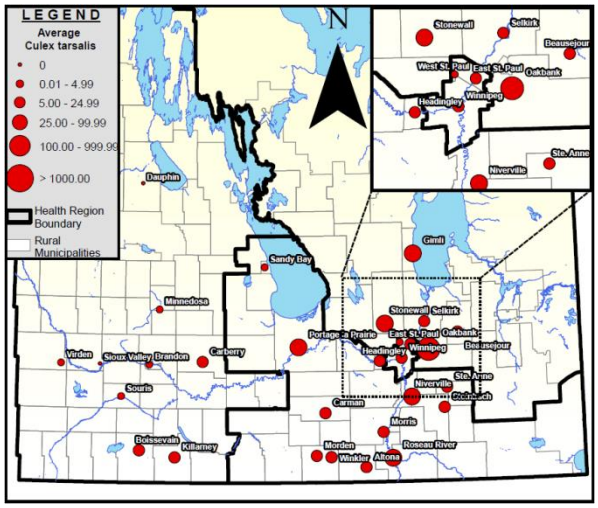


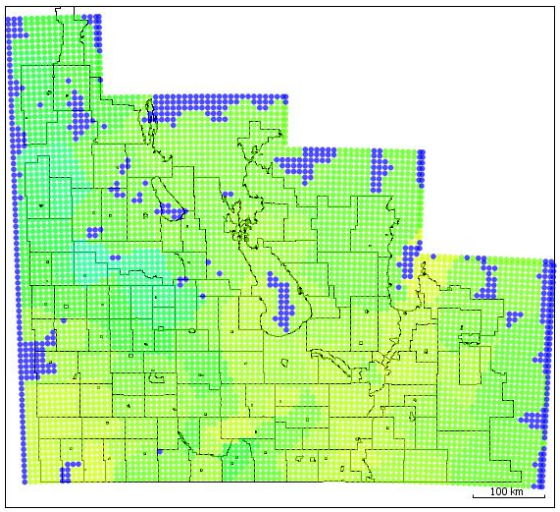


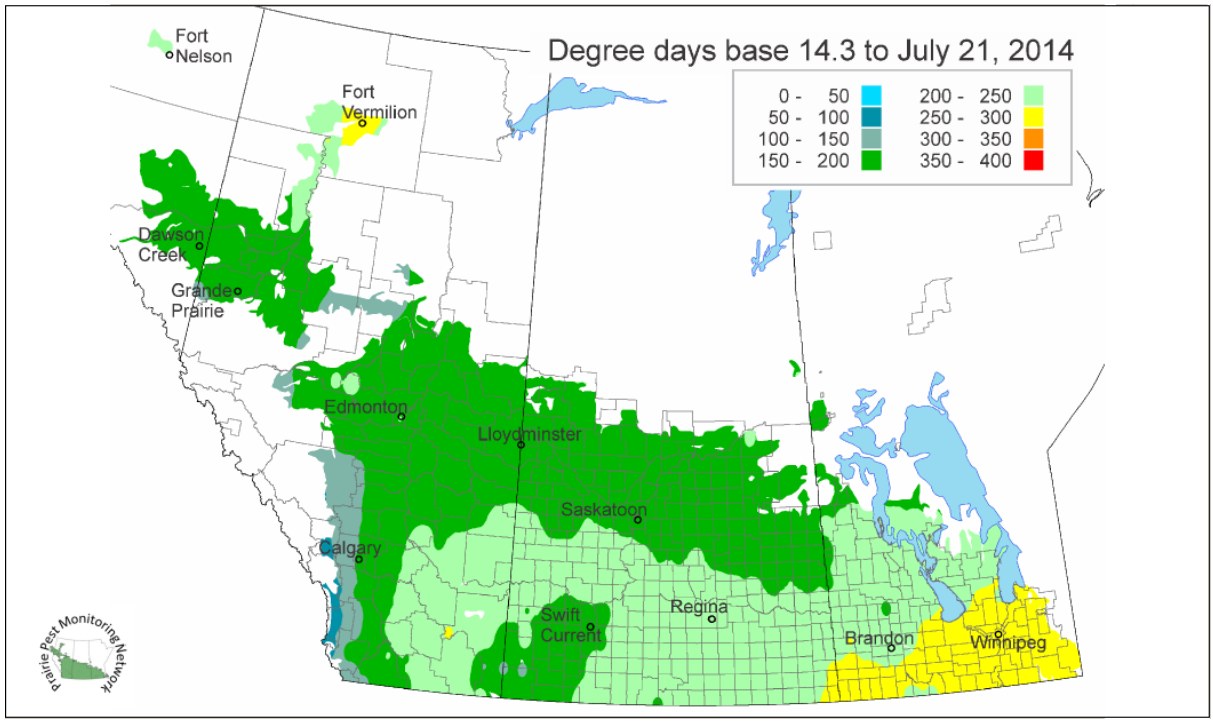


Week 28

(Trap data and degree-days image sources: [20], [158])

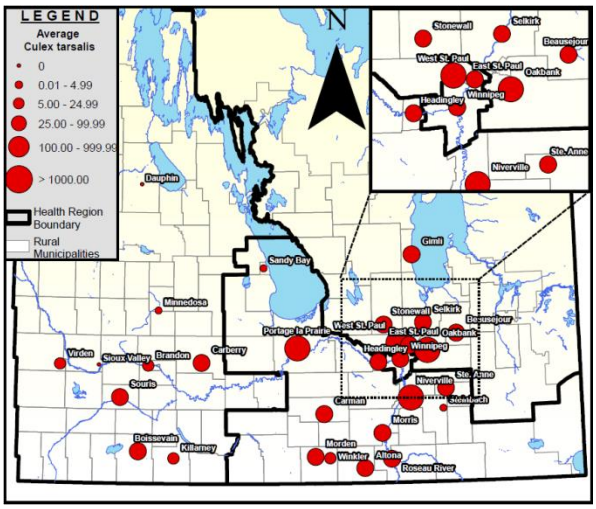


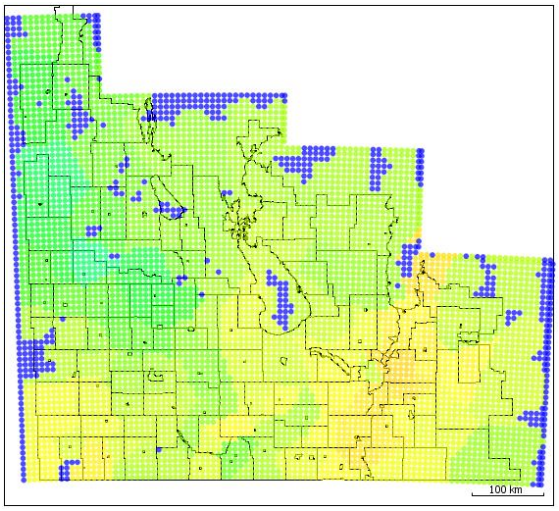


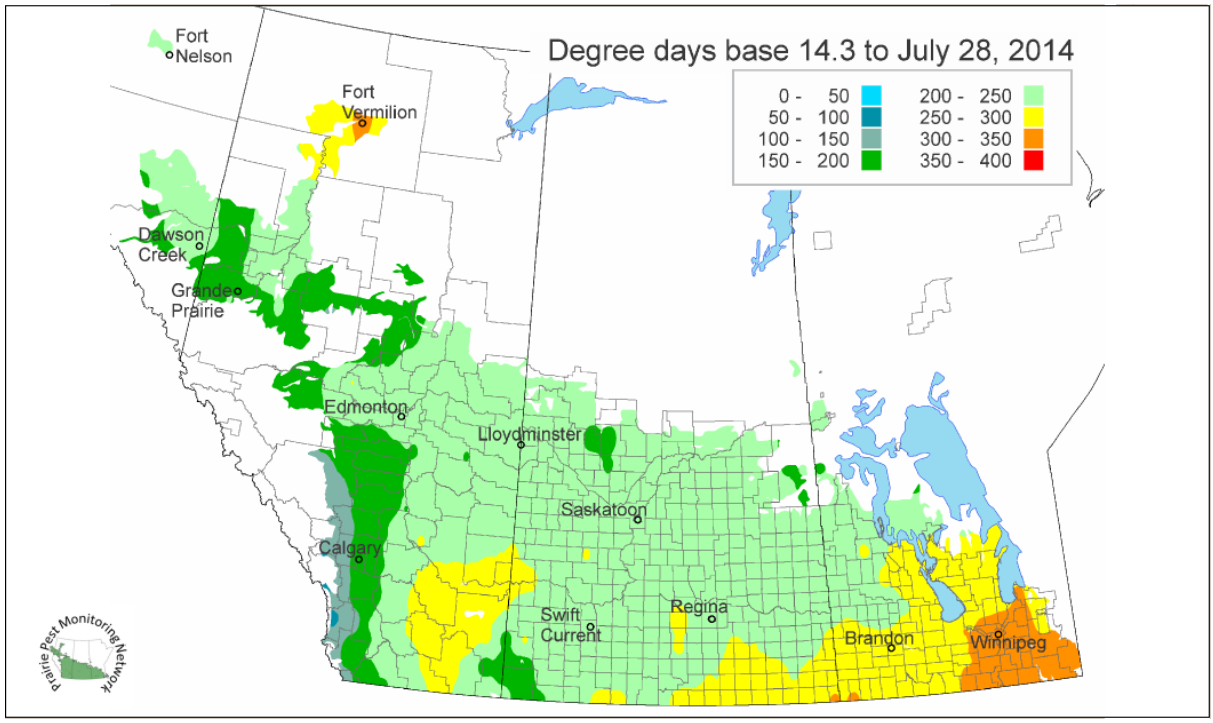


Week 29

(Trap data and degree-days image sources: [20], [158])

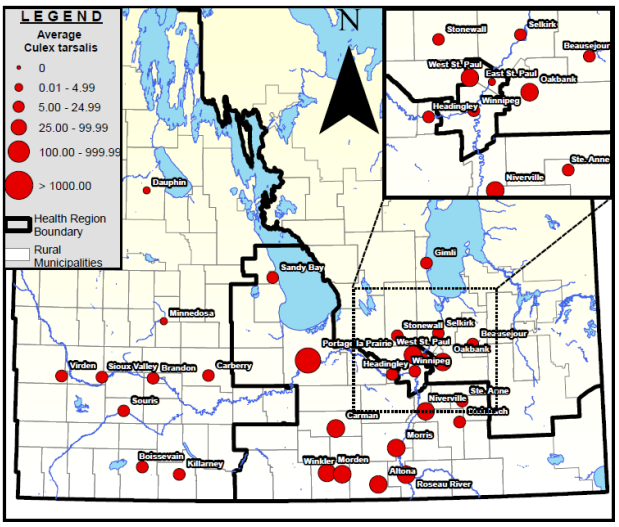


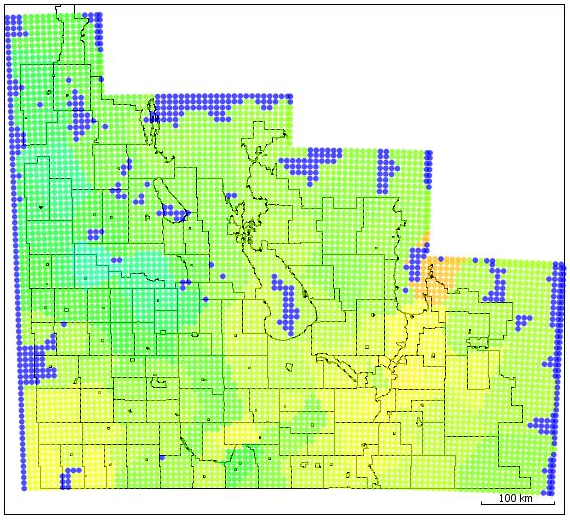


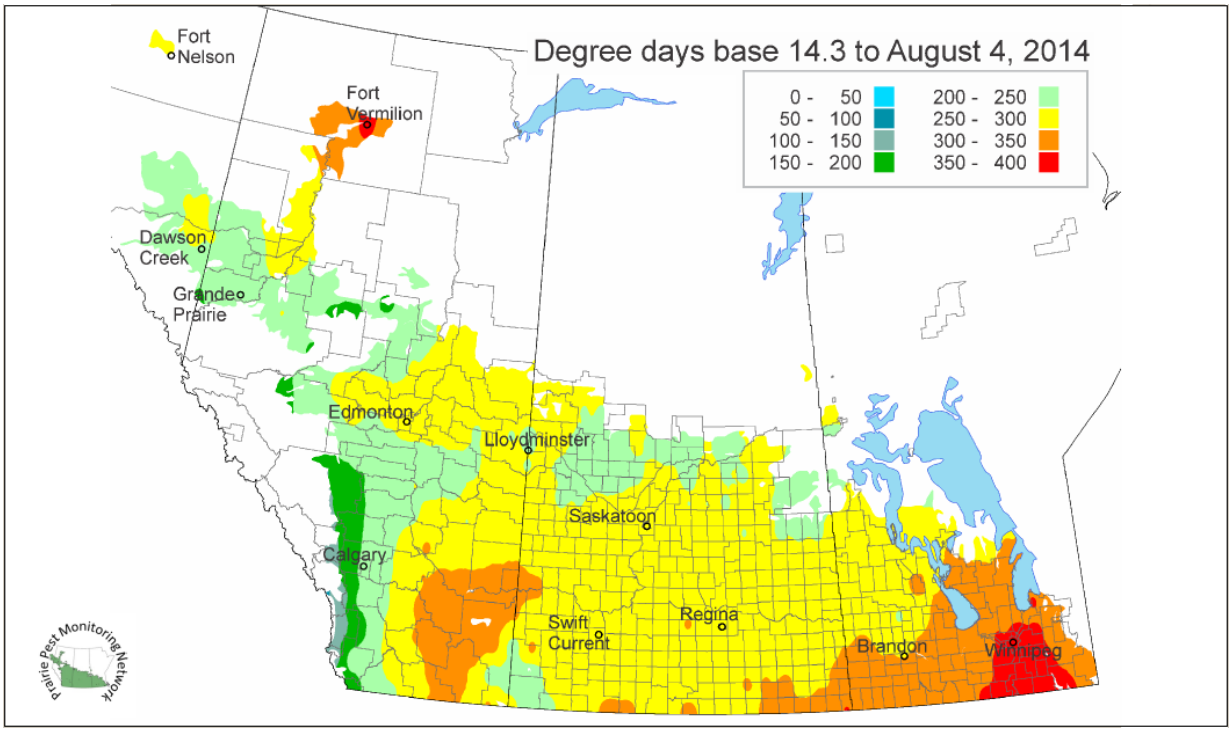


Week 30

(Trap data and degree-days image sources: [20], [158])

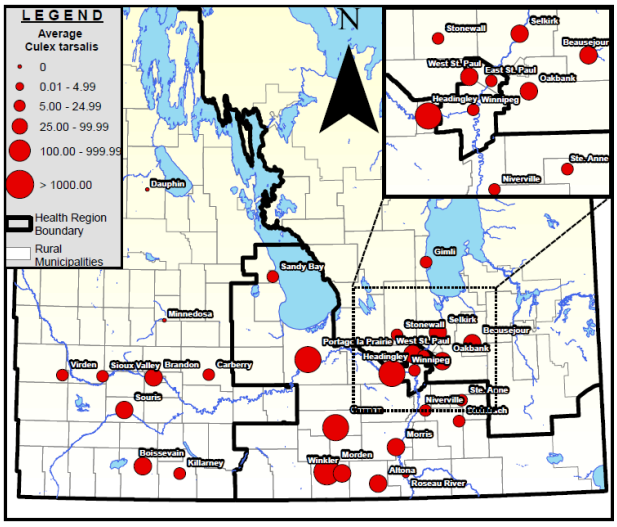


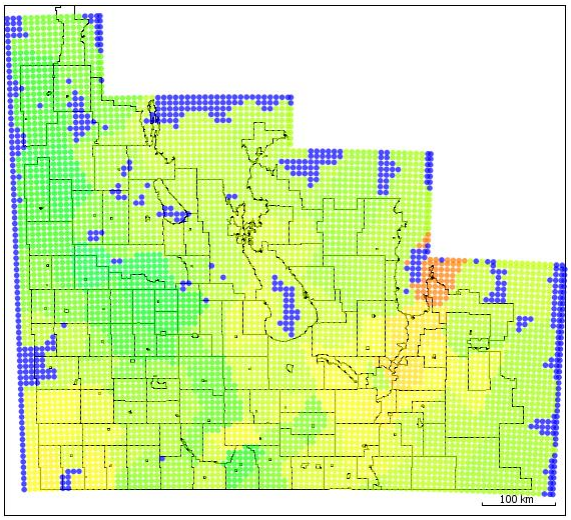


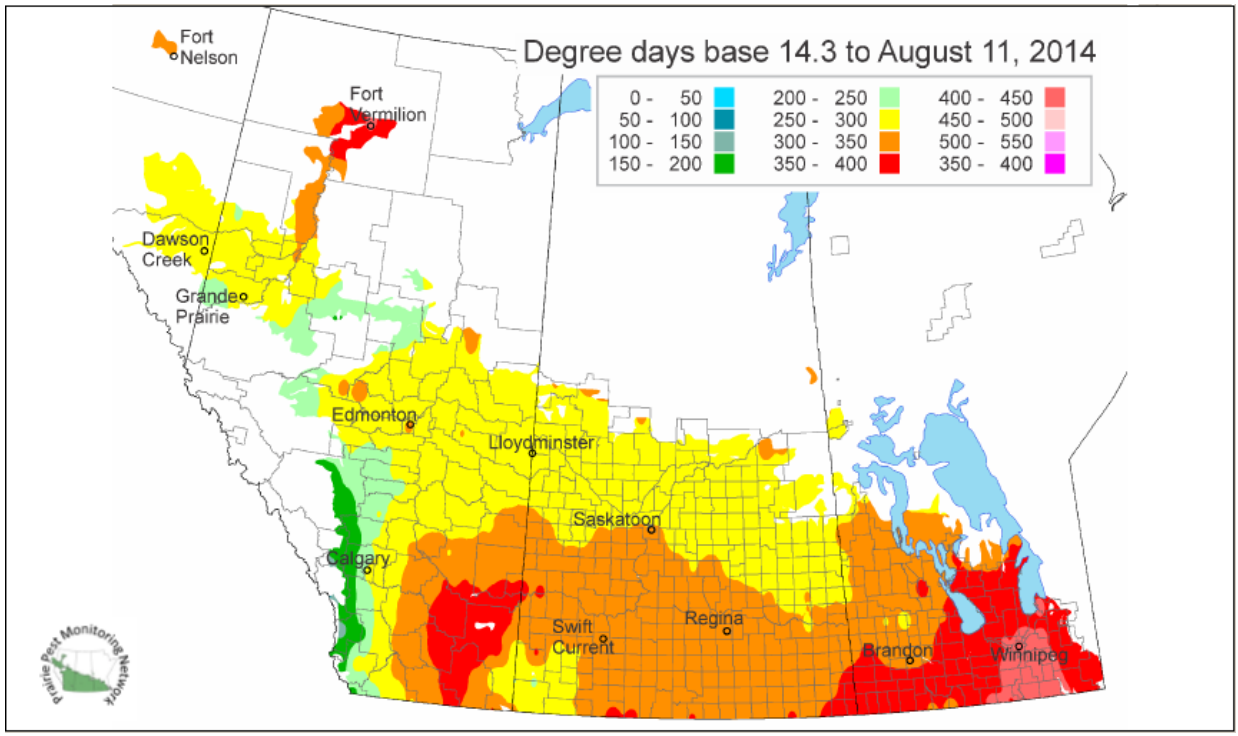


Week 31

(Trap data and degree-days image sources: [20], [158])

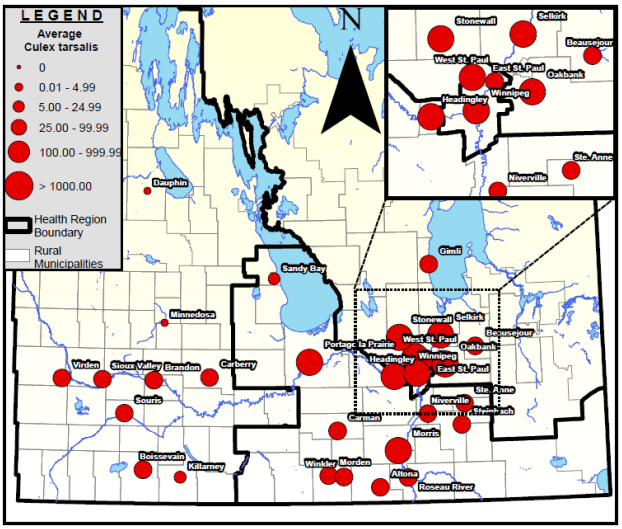


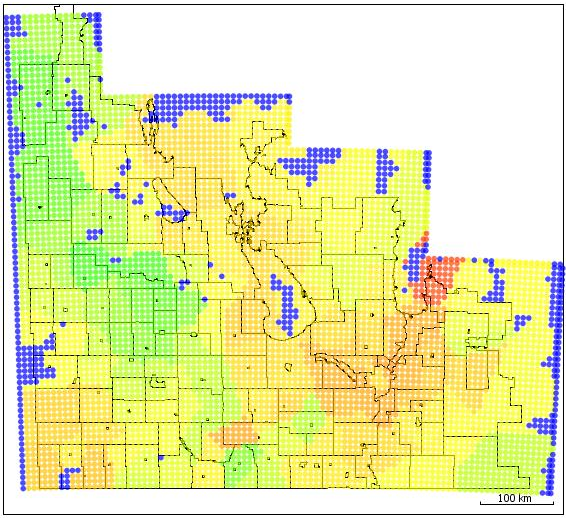


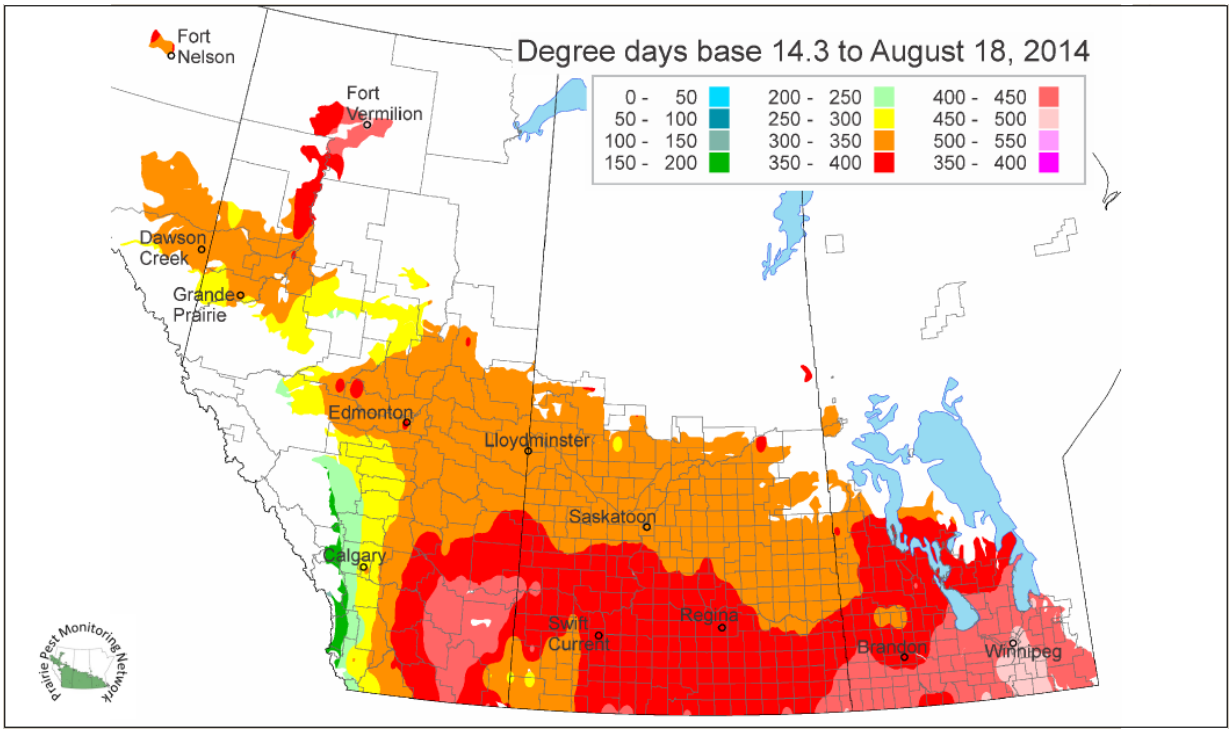


Week 32

(Trap data and degree-days image sources: [20], [158])

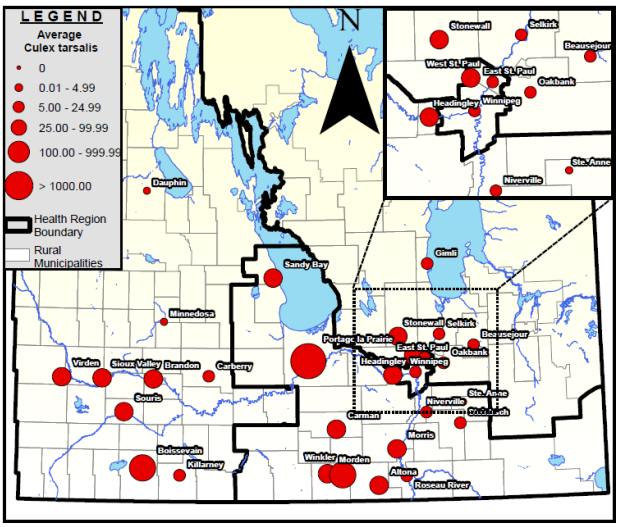


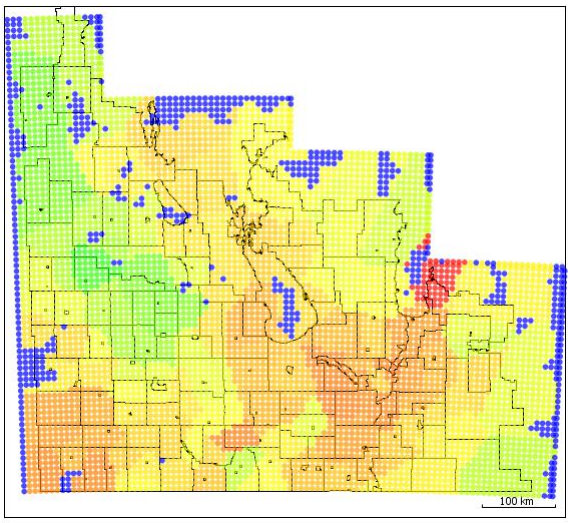


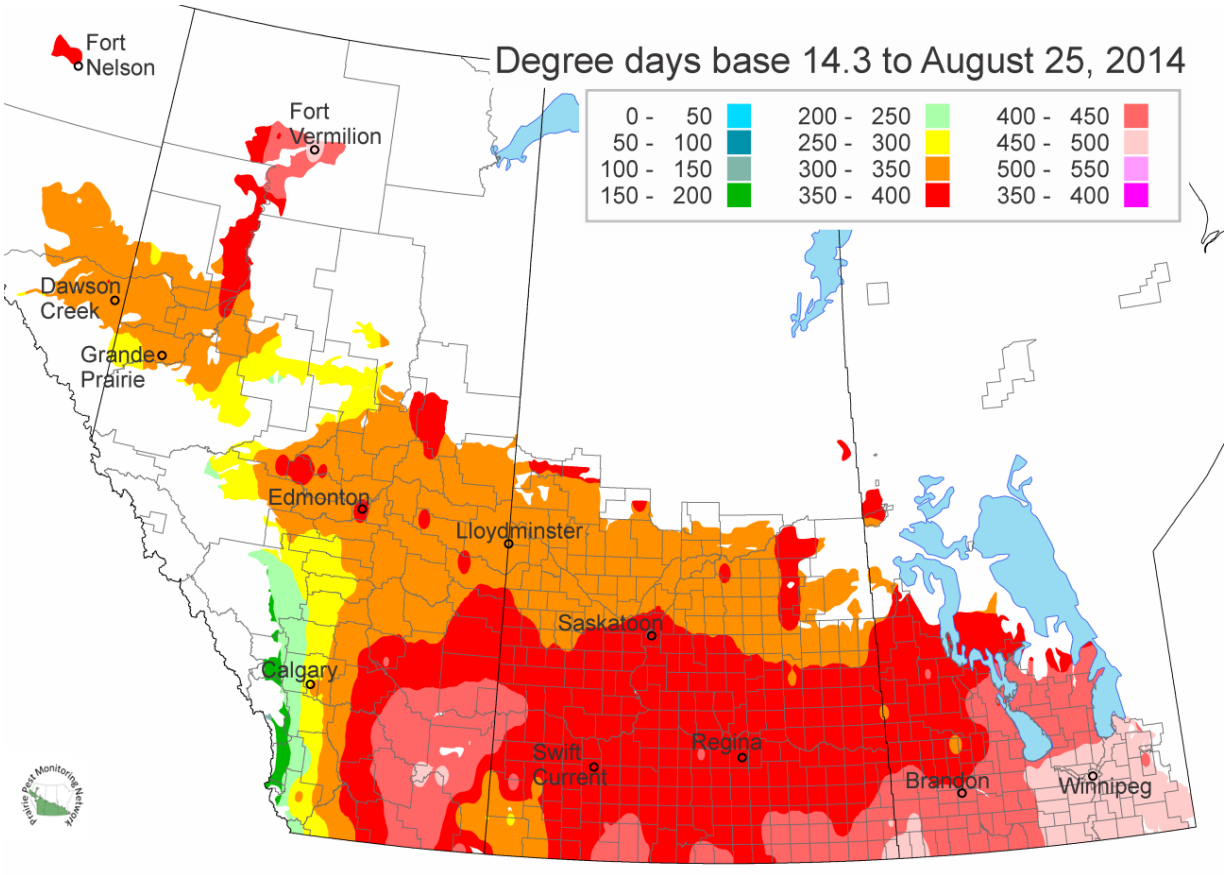


Week 33

(Trap data and degree-days image sources: [20], [158])

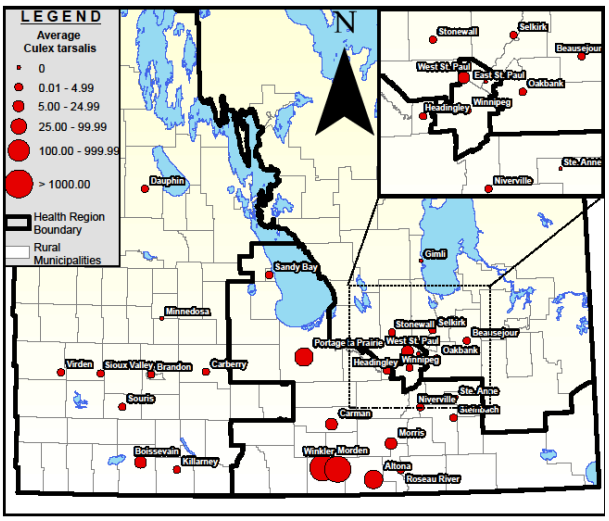


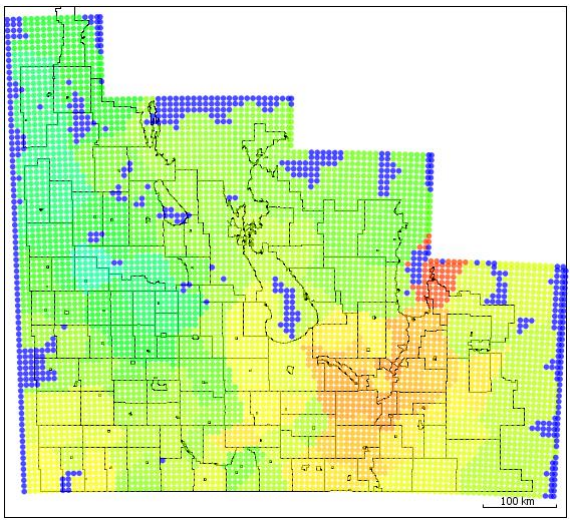


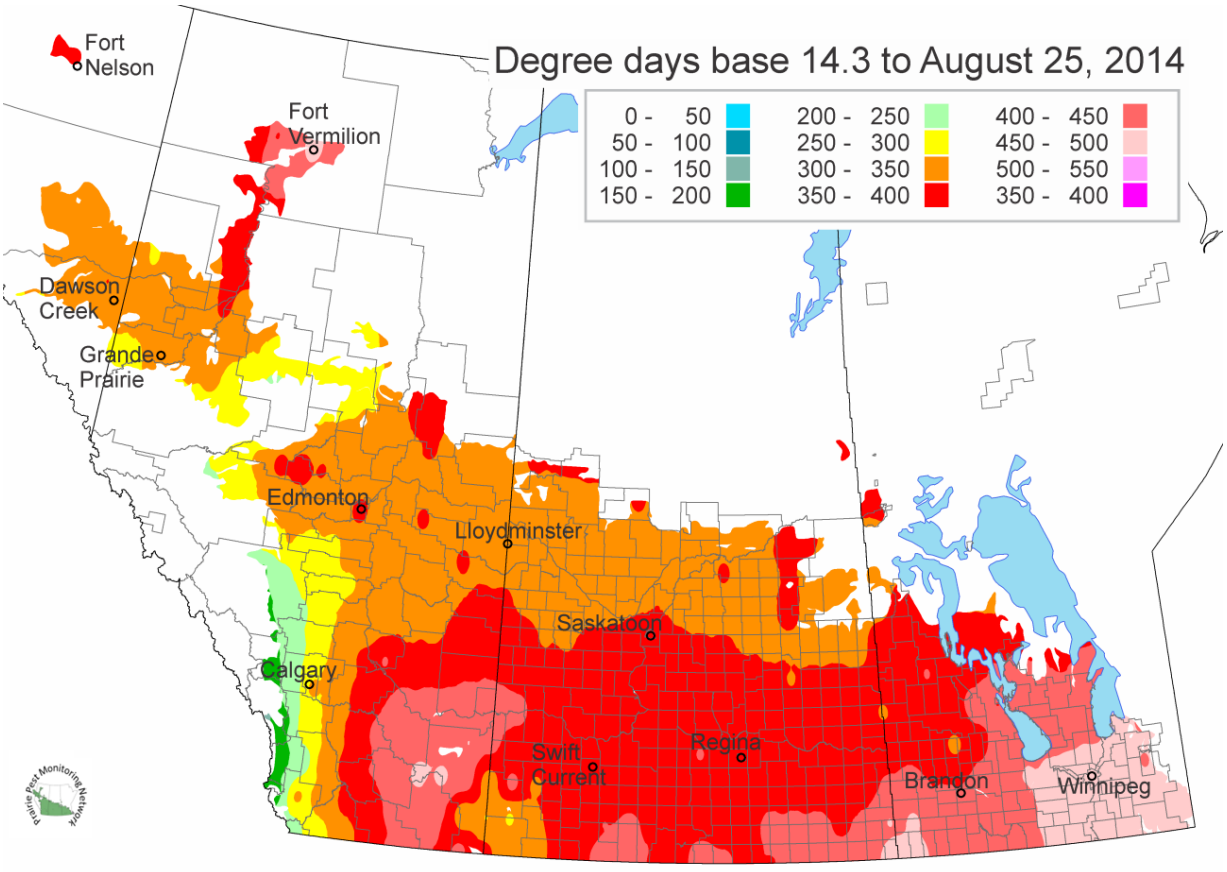


Week 34

(Trap data and degree-days image sources: [20], [158])

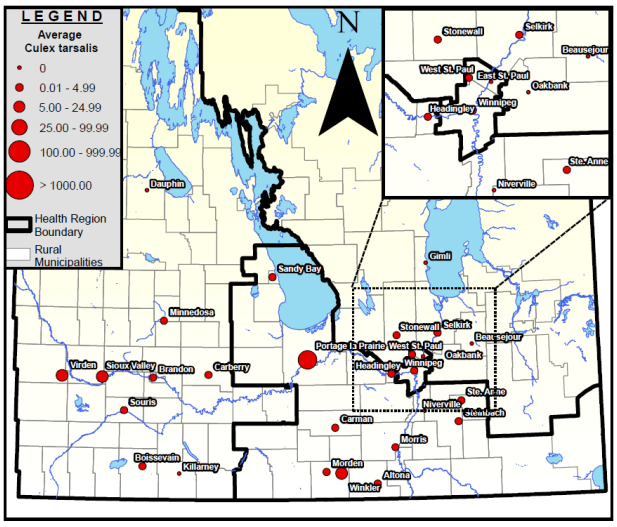


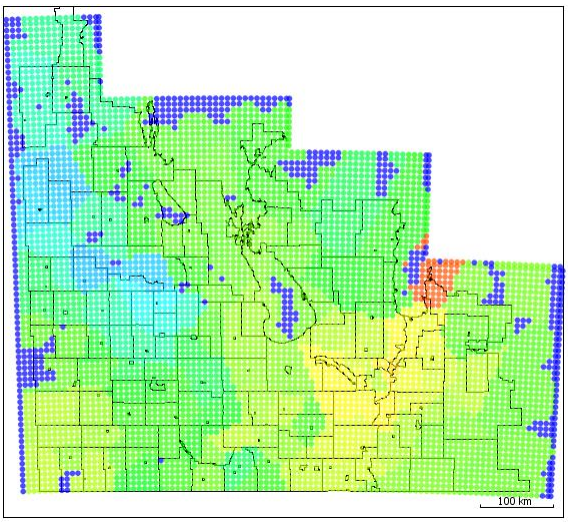


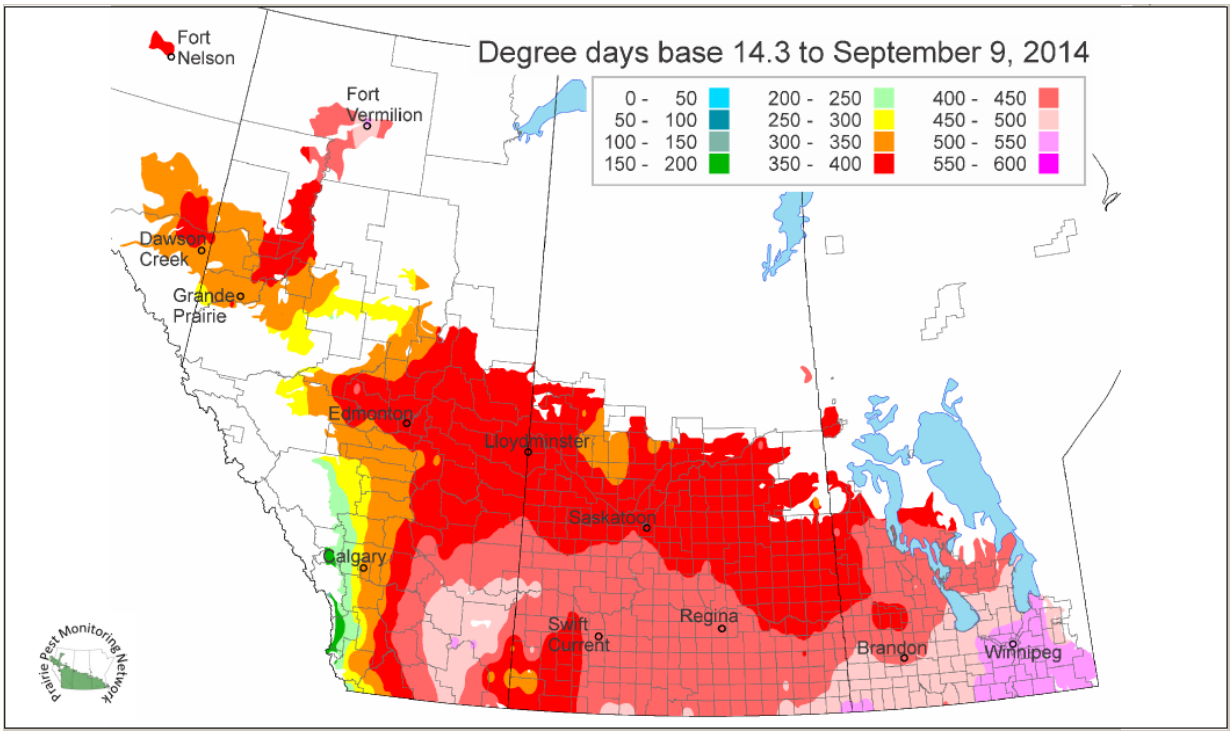


Week 35

(Trap data and degree-days image sources: [20], [158])

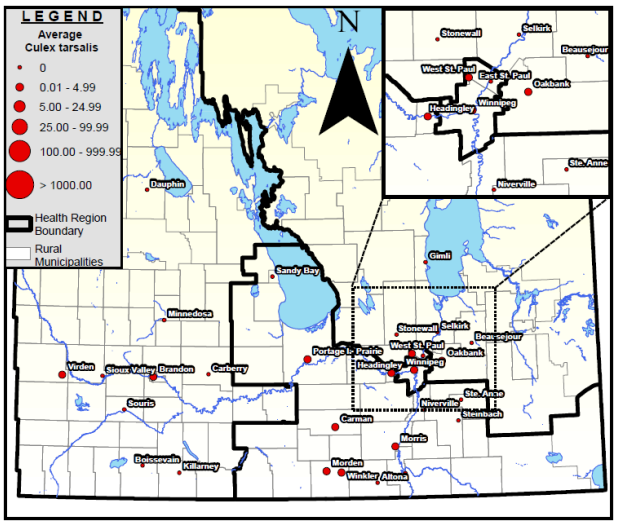


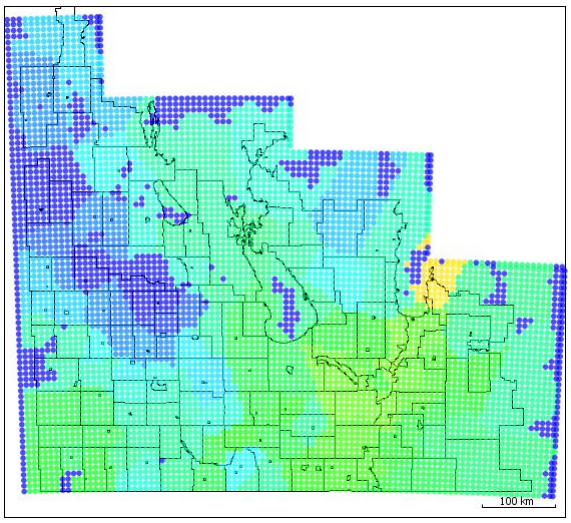


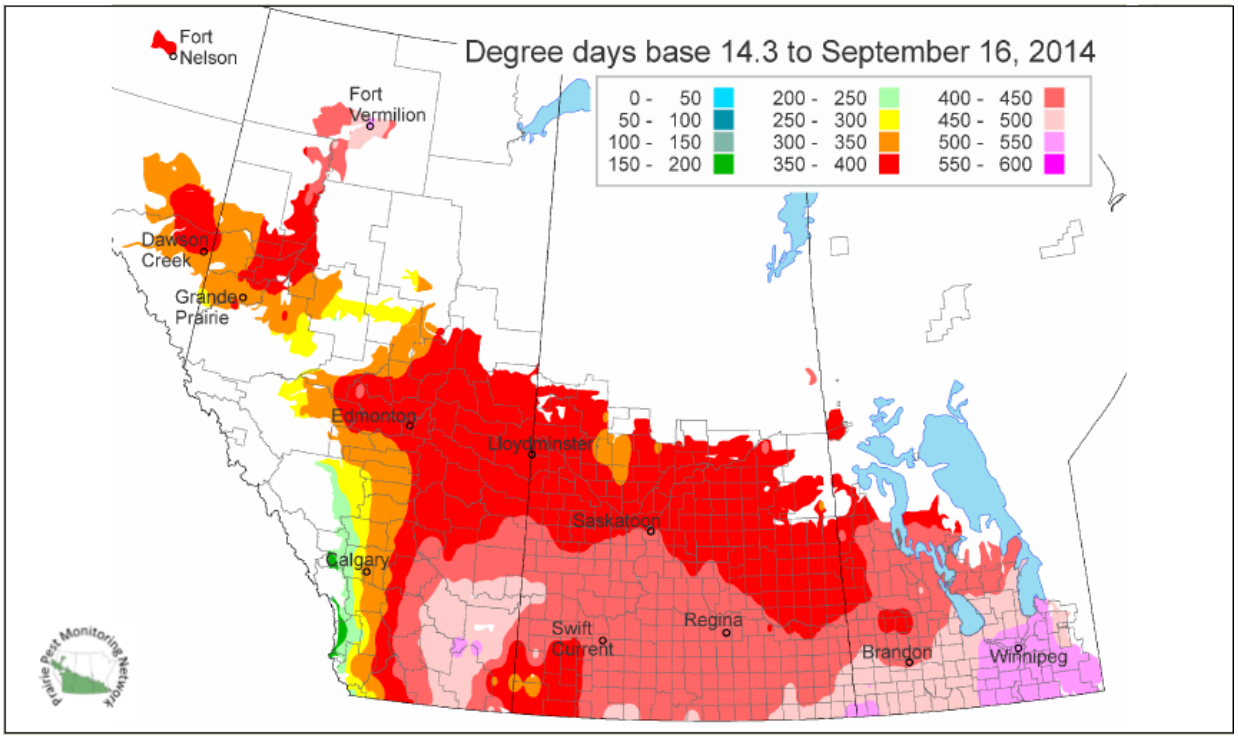


Week 36

(Trap data and degree-days image sources: [20], [158])







Week 37

(Trap data and degree-days image sources: [20], [158])

Appendix C: Agent-Based Model of Facebook Post Propagation

This appendix demonstrates the author's depth of knowledge in ABMs by outlining an ABM which simulates the spread of a meme, the associated difficulties with using real data, and the difficulties associated with verification and validation of a non-trivial ABM. All of these priorities similarly apply to the WNV model which is the focus of the thesis, although the WNV model is considerably more complex than the Facebook meme model. This appendix may be omitted without detracting from the thesis.

A large scale agent-based model of common Facebook users was designed to develop an understanding of the underlying mechanism of information diffusion within online social networks at a micro-level analysis. The agent-based model network structure is based on a sample from Facebook. Using an erased configuration model and the idea of common neighbours, a new correction procedure was investigated to overcome the problem of missing graph edges to construct a representative sample of the Facebook network graph. The model parameters are based on assumptions and general activity patterns (such as posting rate, time spent on Facebook etc.) taken from general data on Facebook. Using the agent-based model, the impact of post length, post score and publisher's friend count on the spread of wall posts in several scenarios was analyzed. Findings indicated that post content has the highest impact on the success of post propagation. However, amusing and absorbing but lengthy posts (e.g., a funny video) do not spread as well as short but unremarkable ones (e.g., an interesting photo). In contrast to product adoption and disease spread propagation models, the absence of a similar "epidemic" threshold in Facebook post diffusion is observed.

C.1 Introduction

Marketers have widely accepted the importance of Word-Of-Mouth (WOM) for a product success [1], [2]. For example, Philips in 2006, Hewlett-Packard (HP) in 2008, Microsoft in 2009 and Ford in 2009 all developed different types of word-of-mouth seed marketing campaigns to promote their sales [3]. The objective is to seed a marketing campaign with the intention of fostering message propagation or spread. This kind of viral marketing is not necessarily only for promoting a product; it can also help in acquiring new members or broadcasting a message or trends in general. In comparison with traditional marketing, WOM has a longer lasting impression to a new member [4]. Moreover, word-of-mouth can work well in cyberspace such as online communities, emails, product ratings, blogs, forums and electronic social networks. In particular, online WOM could be more attractive to companies because of associated lower costs, wider accessibility, immediate distribution, ease of use [5], and better tracking analytics. The importance of online social networks for venders to advertise their products can also be verified by considering the numerous blogs and publications regarding online social networks in marketing science [6], [7].

Among online social networks, Facebook is currently the most well-known social network on the internet. However, the answer to the mysterious question of how to make a post go viral on Facebook still remains somewhat of an enigma. There are many recommendations, anecdotes and hints on different blogs to help one reach maximum influence. Apart from the complex news feed algorithm of Facebook, including a very large network of different people with complex psychologies and so many soft factors, a better understanding of message propagation mechanisms through this type of online social network is needed. Some detailed studies of the Digg social news website [8] and Twitter social network [9] exist that pay specific attention to

the rules and structure of those social networks. However, many of the other works in the context of social networks and diffusion are either too generic or at a macro-level aggregate, without utilizing a detailed model of an online social network (see section C.4).

In this appendix, the Facebook message propagation process is examined at a micro-level. We focus on two properties of a Facebook post (post length and post interest) and one attribute of the sender (friend count within Facebook). As an instance, an attempt is made to determine which of the two properties of post length and post content (interest) plays a more impactful role in the success of post diffusion. By inspecting Facebook at the level of individuals, a better understanding of the underlying dynamics of message propagation process in Facebook is hopefully obtained. Moreover, ultimately the goal is to understand the similarities and differences between the propagation of relatively intangible entities such as messages and memes and the propagation of tangible entities such as infectious diseases within a population (see section C.4).

An Agent-Based Model (ABM) of aspects of the Facebook social network was created to delineate and understand the patterns of message propagation. Agent-based modeling [10], [11] is a natural way of simulating systems where individual agents (e.g., people) play significant roles. In this bottom-up approach, the system contains a set of autonomous individuals, i.e., agents, interacting based on a set of rules within an environment. From the micro-level interactions of reading a friend's post and sharing the post among friends, the macro-level patterns of diffusion of the post evolves. This makes ABMs a suitable method of analyzing Facebook where both micro-level and macro-level analyses are of interest. In addition, we have a heterogeneous population of Facebook users, which is inherently suitable to an agent-based model where each user can have their own profile, in this case, preferences of when to sign in or share a post.

ABMs are generally well suited to model social networks when either the agents or the topology of the interactions is heterogeneous or complex [12].

The idea of deploying an ABM is not only for analysing the impact of post length versus post score, but rather for making a tool potentially capable of including a variety of features to set up different experiments that empirical work cannot address. Cellular Automata as a limited form of ABM has previously been applied for modeling a seeding program [3]. In our study, similar to cellular automata, agents have an internal state machine but are not modeled as (limited) cells. The agents are people modeled as nodes within a graph. Each has its own internal parameters such as the number of friends within the social network. They can asynchronously and independently act (see section C.2). As an example, agents can decide when to log in/out, read and share a note among friends. The detailed behaviour of agents is explained in section C.2. ABMs are also inherently extendable to almost infinite levels of details, and in this case the ABM can be extended to study other possible actions by Facebook users (see section C.5). Depending on the scope of the research, one can add or remove rules and states to or from agents. An advantage of ABM over Differential Equation (DE) or Statistical models is its lack of complex math, which means we do not have to understand relatively complex model formulations [12]. In other words, we just need to be able to describe the system and agent behaviour in detail with a set of "what-if" rules, in a problem-specific and natural lexicon. These characteristics make ABMs well suited to problems that are computationally irreducible [13]. This has however left ABMs open to critique as being more difficult to validate.

Obviously, there are also other limitations within ABMs. ABMs can easily be slow and computationally intensive. In fact, the speed performance was one of the main obstacles

encountered here in obtaining the results which are presented in section C.3. These limitations are acknowledged in section C.5.

The remainder of this appendix addresses the following:

- A realistic agent-based model of Facebook posts diffusion was created. This framework can easily be extended to include more features of Facebook and its users.
- In an initial set of simulations, the relative importance of each input factor such as post score, post length and publisher's friend count is compared.
- A second set of simulations explores the impact of the details of post score versus post length particularly for shorter posts like URLs or photos more specifically.
- A third set of simulations sheds light on two seeding strategies relative to a mass of users versus a few hub users.
- Generally, post content has the highest impact for information propagation within the electronic social network; however, among the posts which spread fairly well through the network, post length is of more importance than the post content (interest) and the initial seeder's friend count.
- Surprisingly, it is shown that there is no tipping point [14] for post diffusion analogous to the transition to epidemic spread observed in infectious diseases. On average, the moment a post is submitted is when it reaches its peak of the probability of being shared or read by friends.

It is also shown, unlike product adoption or disease spread, that it is unlikely for a Facebook post to go viral and reach a fair percentage of the entire network. In this case (like other celebrity phenomena), the fact that some posts obviously do go viral may skew a typical Facebook user's perception of the probability of their own post doing so.

C.2 ABM Architecture

The ABM is implemented in the Java-Based educational version of the Anylogic software toolkit, which supports Agent-Based, Discrete Event and System Dynamics Modeling. In this section, the agent-based model, agents' properties, the structure of their environment, and the governing rules are explained in detail.

C.2.1 The Big Picture

People using Facebook can either visit the webpage on their browser or use the Facebook application on their mobile device. In either case, once you open your Facebook profile, you may receive a list of notifications of what has previously happened since your last login. The difference is that in the second case, you can stay signed into your Facebook profile with your Facebook phone or tablet application, which results in receiving notifications when they occur. Once a person truly decides to check their Facebook profile, they usually go through the notifications and then generally switch to their news feed (home) page to see the activities and posts from friends or other Facebook pages/groups to which they are affiliated. At any time during the visit on Facebook, a user might decide to post a text note or upload a photo/video on their Facebook (wall/timeline) page. They might also copy a post previously shared by a friend and paste it on their own page in order to share it with their own friends. Through this feature, a post would spread over the network. The other common way to interact with a post is to "like" which is invoked by clicking a like button below a post. Currently, there are also many other features available on Facebook such as private messaging and applications that all act like incoming stimuli to a Facebook user to draw their attention. The current scope of this study is general posts by users on their own wall page. We recognize that this is a simplification of actual social networking via Facebook. The simplification was necessary in the first attempt at creating

a model. Detailed simplifying assumptions are presented in subsection C.2.4 where the rules of behaviors by agents are explained.

Herein, akin to reality, time passes continuously in minutes and seconds. Agents (Facebook users) are connected to one another within a virtual social network. Each agent, independently from all the other events, decides when to log in and when to log off. During the interval they are logged into the system, they go and check their friends' Facebook wall pages, each post one by one, until they decide to switch to another friend's page. Each post has some interest score and requires its own unique time to be read. Once an agent finds out something interesting on a friend's page, they might decide to share it again on their own page. Also at any time when an agent is online, they can publish a new post of their own. This agent is denoted the initial seeder/publisher of the post. Agents - when they are online - are able to receive notification of recent activities from their immediate friends. In the current agent-based model, this activity only includes the case where a friend shares a post on their own wall page. In this agent-based model, similar to the real Facebook where users check their notifications, upon receiving a notification by an online agent, they go through the notification and read the post shared by their friend.

C.2.2 Agents and Parameters

The agent-based model consists of only one type of agent which is a Facebook user or individual. Agents can create and publish different posts with two important properties of *Post Length* and *Post Score*. Each agent has a set of internal parameters including *Activity*, *Average Login Time*, *Average Post Rate*, *Friend Time* and *Friend Count*.

Wherever possible, published reports on Facebook user statistics were used to set agent parameters, and where a given parameter was reported in several references, reasonable inferences and consolidations were made [15], [16], etc.

The Activity parameter is the main parameter that characterizes the heterogeneity in individual preferences for posting/sharing notes. In other words, each user posts new notes at a given rate, which depends on the user's Activity parameter. Also, the chance of sharing a post already shared by a friend is related to this parameter. This parameter is assigned a uniform random value from 0% to 100%. Obviously, the higher the value, the more posts the user generates. Generally, the more active users (users who post more often) are also online more often. The exact association between the frequency of activity and online time is described below.

The Average Login Time parameter indicates how many minutes a user is online in a day, on average. This parameter has a normal distribution and is used to calculate the Login Time parameter by the following formula: $Login\ Time = 2 \times Activity \times Average\ Login\ Time$. This means that a typical user, whose activity parameter is 50%, has a Login Time parameter value around the value of Average Login Time parameter. Also the formula states that active users spend more time online on average. Furthermore, the login time follows a long tail distribution that also accounts for 'lurkers' (high online time but low Activity).

The Average Login Time parameter has a normal distribution. There are different choices for its normal distribution properties. In 2011, Facebook Press Room reported the average Facebook user spends more than 11 hours per month on Facebook [17]. Assuming 30-day months, this means over 22 minutes per day. In addition, [18] claims the "average user spends an average 15 hours and 33 minutes on Facebook per month," which equates to roughly 31 minutes per day. There are also some reports on mobile usage, such as an average of 441 minutes per visitor in each month (i.e., 14.7 minutes per day) reported in [19]. Consolidating these sources, the average online time was set at 23 minutes per day, which may change in the future work.

Considering the fact that most obsessed Facebook users spend daily average of 8 hours on the site [20], the average online time is bounded by a maximum of 10 hours per day.

Assuming a day is 1440 minutes, a login time of 23 minutes results in having a logout time of 1440 – 23 = 1417 minutes. This means if the login time for a user is calculated to be 23 minutes, they will be online for 23 minutes and will be offline for 1417 minutes in a day. However, since most people tend to check their Facebook profile more than once a day (e.g., It is reported that "on average, [a person] visits the Facebook app/site 13.8 times during the day, for two minutes and 22 seconds each time" [21]) these 23 and 1417 minutes have to be divided into different intervals. These intervals are drawn from an exponential distribution with the mean value of 1440/1417 and 1440/23 minutes per a day (1440 minutes), for the login and logout times respectively. The choice of exponential distributions for login/logout rates where the probability to login is the highest immediately after logging out may sound irrelevant. However, similar to working out at gym, people are likely to check their social profiles within a certain period of day (e.g., it is reported that "peak Facebook time is during the evening, just before bed" [21]). Although for some smartphone users with Facebook application installed on their phone, this period might be from 6 am to 11 pm. In either case, it would be safe to assume that people would not check their Facebook profiles when they are asleep! In the agent-based model, the login intervals should not be too far away from one another, and there has to be a limit to control when people log into their profile. The choice of exponential distributions attempts to keep the online intervals close to each other. In addition, a login rate of an average of 23 minutes/day compared to a logout rate of an average of 1417 minutes/day is low enough to span the daytime. The choice of the exponential distribution is also related to performance issues. Since it is the default distribution for rate triggers in Anylogic, deploying another distribution would

significantly have decreased the speed of simulation. Part of the reason may be that the triggers scheduler in Anylogic can be set ahead in outer code loops of the program.

The Average Post Rate parameter defines the average number of new notes published per month by each user. It has a truncated normal distribution based on the parameters shown in Table I. The actual distribution of the number of posts is calculated as: *Post Rate* = 2 × *Activity* × *Average Post Rate*. This distribution is plotted in Fig. 1. Assuming a month is 30 days, the exact time when an agent publishes a new note is drawn from an exponential distribution every time the user logs in with a rate of 30/*Post Rate*. There are several different reports on the rate of different posts in Facebook. Additionally, these numbers are intuitively known to keep changing. However, we had to adopt one of these reports, which is that in every 20 minutes, over 1 million links are shared, 1.8 million statuses are updated and 2.7 million photos are uploaded [22]. This results in having 11.9 billion posts per month. Add to that, the knowledge that Facebook had 845 million monthly active users as of December 31, 2011 [16], implies that there are approximately 14 posts per month for each person. This justifies the choice of an average of 13.8 posts per month for each user.

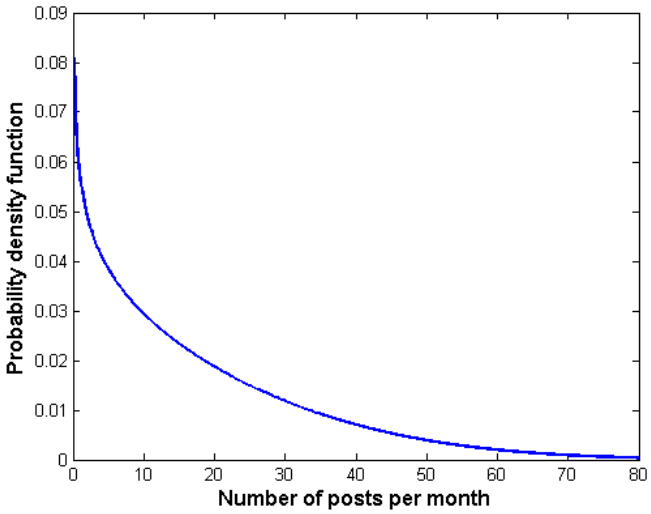


Fig. 1. Post Rate density function

Each user is connected to some other users known as friends. The number of friends a user has is controlled by the Friend Count parameter. The number of seconds dedicated for each friend to check their posts is set by the Friend Time parameter. Every time a user begins checking their friend's posts, this timeout value is generated based on the associated normal distribution. Once the time is up, the user leaves the current friend and looks at another friend's recent posts. Complementary to this parameter, the Post Length parameter dictates how many seconds are required to completely read a certain post. When a user publishes a new note, this value and the Post Score parameter are calculated and assigned to the post. The Post Score parameter represents how interesting and appealing a post is.

Table I Simulation parameters and their distributions

Parameter	Distribution	Values			
		Mean	Std. Dev.	Min	Max
Activity	Uniform	0.5	$1/\sqrt{12}$	0	1
Average Login Time* (minutes/day)	Normal	23	120	0	1080
Average Post Rate* (times/month)	Normal	13.8	13.8	0	300
Friend Time (seconds)	Normal	30	30	10	600
Post Length (seconds)	Normal	30	30	10	600
Post Score	Uniform	0.5	$1/\sqrt{12}$	0	1

* Multiplied by 2 times *Activity* for calculating the associated rates

Table I summarizes the model parameters and their distribution properties. As shown in the table, the Activity and Post Score parameter are drawn from standard uniform distributions between 0 and 1. For the other parameters, sampling from truncated normal distributions in Anylogic is employed. This kind of distribution is essentially the standard normal distribution which is stretched by the *Mean* coefficient, then shifted to the right by *Std. Dev.*, after that it is

truncated to fit in [*Min, Max*] interval. Truncation is performed by discarding every sample outside this interval and taking a subsequent try.

C.2.3 Environment

Akin to online social networks, the environment is a graph where each node represents a user/agent whose social friends are neighbour nodes in the graph. In this subsection, we describe this graph and its properties in detail.

The Facebook network graph can be viewed as a small-world network [23], [24] as most nodes can be reached from every other by a small number of hops. Generally, however, scale-free networks are a better choice to model a social network graph, as they have a more realistic degree for the power law distribution. In fact, it is shown that scale-free networks themselves are ultra-small worlds, where the shortest paths become even smaller [25]. Yet the strict power-law distribution is not accurate enough to represent Facebook's degree distribution [26]. More precisely, if the power-law distribution of $P(k) = k^{-\alpha}$ is accepted to be the degree distribution of nodes, two sections for $1 \leq k < 300$ and $300 \leq k \leq 5000$ can be approximated by a power law with exponents $\alpha_{k<300} = 1.32$ and $\alpha_{k\geq300} = 3.38$, respectively [15]. Therefore it was decided to synthesize the graph by directly sampling from the Facebook graph. Ideally one should generate multiple sample graphs and run the experiments on all the sampled graphs to ensure the results are not specific to just one graph. But due to processing time-limitations, only a few graphs were sampled; as all the sampled graphs shared comparable properties, one representative graph was selected for the basis of the agent-based model.

A dataset of the Facebook social graph released by the Networking Group of the University of California, Irvine [15] was used. Two datasets of uniform sampling and Metropolis-Hastings Random Walks (MHRW) were available. The Metropolis-Hastings algorithm is a Markov Chain

Monte Carlo (MCMC) method to simulate a complex distribution from which direct sampling is difficult. The MHRW option was chosen, as the Facebook IDs within this dataset are consecutive numbers, which makes it easier to construct the graph. Gjoka *et al.* obtained this dataset by 28 Facebook-wide independent MHRWs in April of 2009 [15]. The dataset contains the number of friends and their Facebook IDs for approximately 957K unique users. It was not possible to directly build our network structure on the dataset itself, as it was missing a large number of edges. Strictly speaking, the dataset has a uniform random sample of users in Facebook. Here, however, a dataset was required which would densely cover only some small region of the Facebook graph. Such a dataset could be obtained by a Breadth First Search (BFS) crawling method. Hence the following approach was taken:

1) The first 500 sampled Facebook user IDs were assigned to the 500 primary agents in the model.
2) Each primary agent was connected to a number of new secondary agents according to the number of friends of their corresponding Facebook user in the dataset. This resulted in a network of total 89,977 agents, but the number of edges was not sufficient.
3) Extra links between the secondary agents were inserted based on a custom distribution of all the sampled users in the dataset. This resulted in a network with a total of 7,528,164 edges.

The network was created and used (saved) for further experiments. After the second step, each node on average had only one connection, which is not at all the case in the real-world Facebook graph. The reason is that the dataset tells us the node x is connected to $x_1,x_2,\ldots,x_n$; but hardly ever is any information available for each of x_i's. As a result, the third step is necessary. The first and second steps are quite straightforward. The challenge is in the third, where creating

an undirected graph of n nodes according to a certain degree distribution (here denoted distribution F) is desired. In this case n = 89,477 and the probability distribution F is given by the dataset.

Furthermore, it was desirable to create the graph in such a way that nodes with more mutual friends have higher probability of making friends than that of total random nodes. Therefore, the approach is based on the idea of common neighbours as well as the Erased Configuration Model [27]. It is defined as follows. For each node n_i, draw a degree d_i according to the probability distribution F. Make d_i half-links (or stubs) and connect them to the node n_i. When all the stubs are created, start the following loop procedure through the stub pool:

1) Choose three distinct random stubs d_1, d_2 and d_3 and remove them from the stub pool. Notice that each stub is already connected to a node.
2) Among these three stubs, find the two stubs d_i and d_j with more mutual neighbours between their corresponding connected nodes in such a way that they neither make multiple edges between the nodes n_i and n_j nor make a loop on the node n_i or n_j.
3) If the step 2)2 was successful, join the two stubs d_i and d_j to make a connection between the two different nodes of n_i and n_j; otherwise return one of the stubs (e.g., d_1) to the stub pool.

The above procedure removes two stubs from the stub pool at each repetition. Therefore, after exactly $\left\lfloor \sum_{i=1}^{i=n} d_i/2 \right\rfloor$ iterations it terminates; it could be that one stub is left behind in the stub pool. This is quite natural in the Erased Configuration Model, and the proof of its convergence to the desired degree distribution can be found in [27].

To ensure that the graph represents the ‘six degrees of separation’ phenomenon, the diameter of the synthesized graph (i.e., the longest shortest path) was estimated. To measure such a

statistic, 6K nodes (out of roughly 90K nodes) were randomly selected and the Breadth-First Search (BFS) algorithm was run for each node to count the number of reachable nodes at each hop. It was found that all the nodes after three to five hops reach 100% connectivity to all the rest. The percentage of reachable pairs within a certain distance is shown in Fig. 2, which is similar to the graph of degrees of separation of Facebook reported in [26]. The average distance was calculated to be 4.0074, which is comparable to 4.7 and 4.3 of the global and U.S. population of Facebook users in May 2011, respectively [26], [28].

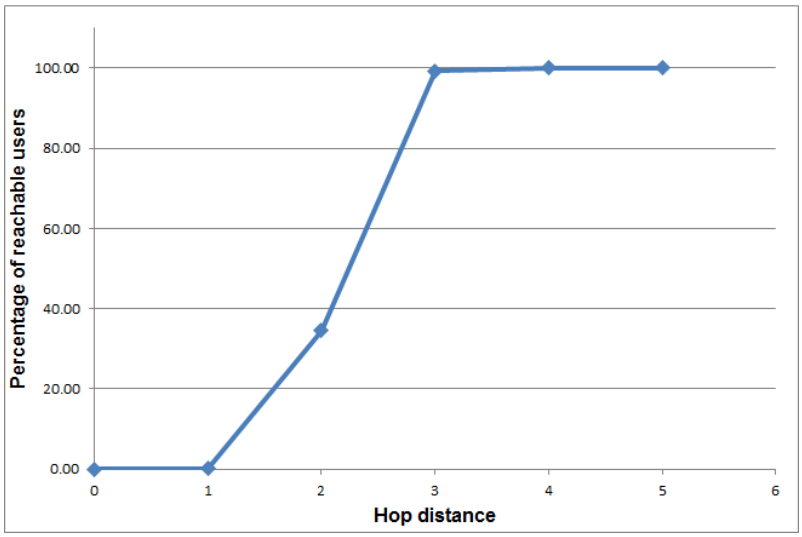


Fig. 2. Degrees of Separation: Percentage of user pairs within x hops of each other

The other statistics regarding the constructed graph, the dataset [15] and the Facebook graph [26] are illustrated in Table II, demonstrating that the dataset with a smaller number of nodes was sampled successfully. As the constructed graph has a much smaller number of nodes (but retains the same number for maximum possible number of friends), it is much less sparse than the real Facebook graph. This is part of the reason the mean distance has been reduced, compared to Facebook. Most conservatively, the synthesized graph can be considered as an acceptable example of Facebook connectivity structures of a sub-region of Facebook network graph. Nevertheless, 90K is almost the maximum possible number of users to handle during the

simulations. The reason for this is that all the nodes have their own autonomous behaviour and processes, and their interaction with each other is a function of the number of edges in the graph. The time complexity of only constructing the network graph itself is O(|E|), where |E| is on the order of a million.

Finally, the complementary cumulative degree distribution function (CCDF) of the dataset and our graph is illustrated in Fig. 3, displayed on a log-log scale. As can be seen in this line graph, the distributions do not strictly follow power-law distributions, which are straight lines on a log-log plot.

Table II Network graph statistical properties

	Facebook [26]	**Dataset [15]**	**Our Graph**
Mean Degree	190	168	169
Median Degree	99	110	111
Min Degree	0	0	1
Max Degree	5000	4979	3734
No. of Nodes	721 M	957 K	90 K
Mean Distance	4.74	N/A	4.0072

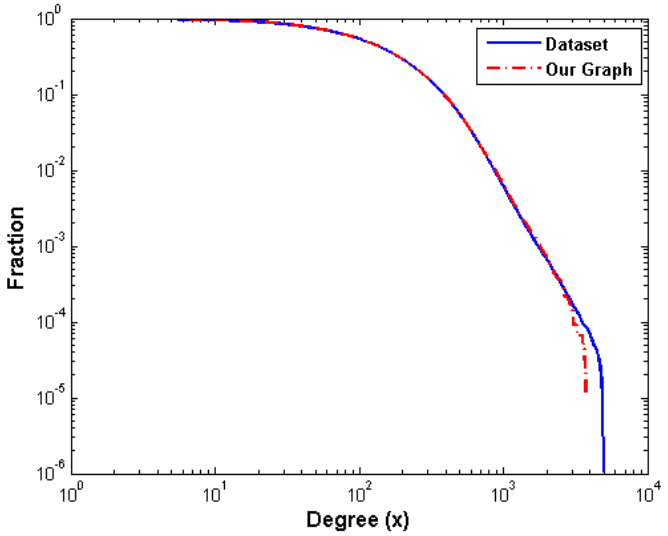


Fig. 3. Degree Distribution: The fraction of users who have degree x or greater

C.2.4 Rules

In delineating the scope of the model, the following assumptions have to be made:

1) As we know, there are many different pieces of content, such as photos, videos, links, status updates, event invites, notes and etc. [29] that can be published on a Facebook wall/timeline page. However, throughout the rest of this appendix the word "post" or "note" is used to indicate generic content that is posted on a wall page. This is because all content shared on a Facebook page act as incoming stimuli to a Facebook user. The objective in this study does not aim to cover the detailed properties of a successful or uneventful post. Regardless of the specific type of note, each post requires some time to be noticed by user agents. This is why the post length is of interest to us. In addition, each post is trying to convey a message to the viewer/reader. The message could be a warning, funny picture, personal news, an inspirational quote, amazing fact, etc. Each might be appealing to specific types of people. However, to keep the model simple, we considered a single scalar value to represent the general interest in a post as the post score. Each user, depending on their activity level, may share an interesting post.
2) Users cannot send private messages to one another. They can neither make comments on any post nor share any note on somebody else's wall except for their own wall; in other words, they are only able to share something on their own wall. Adding the private messaging property could have only increased the complexity of the model as the model is limited to public means of sharing a post over the network. Posting notes directly on someone else's page is not at all as common as sharing posts on personal wall pages. Incorporating these features would have demanded obtaining more statistical data about

the activity behaviour of people on Facebook. Here an effort was made to not increase the number of model parameters, as tuning these parameters is one of the most challenging parts of designing an agent-based model. Even at the current model, values of some parameters are based on intuitive rational assumptions rather than actual data. Commenting on posts is a very important feature of Facebook social network. Through comments or lack thereof, a post can stay alive or die. Comments and Likes in Facebook have a direct relationship with the news feed algorithm of Facebook. In simple words, users are more likely to receive more (recent) posts on their home page from those friends to whom they had most interaction in the past. One way of making interaction with friends is via liking or commenting on their posts. The exact Facebook algorithm to rank the news feed page is unknown. In our model, we let each user agent choose which friends to have interaction with. The commenting feature is removed from the model to decrease its complexity. Adding these features back into the model could be a very nice extension of current model.

3) The only possible relationship between users is bidirectional friendship. This means subscriptions to Pages or Groups held in common between two or more agents, and any other similar features are ignored in the model. Page and groups can be considered as normal user agents in the graph with higher number of friends and activity level. In social science, they are referred to as hubs. So there is no need to distinguish pages from people in our model. Subscription to a Facebook user is similar to a unidirectional friendship. This feature was added later to Facebook. One can think of the current agent-based model as a model of Facebook in its first years without this feature.

4) Similar to the notification feature in Facebook, when a user is online, if a friend of theirs shares a note, the user will receive a notification message within the model. As a result, considering assumption 1) above, there is no way to receive or potentially save a notification for an offline user. In this model, when users go online they begin to check recent posts by friends; as such, adding a notification feature to the model would not have made a significant change to the results; but it would make a difference in time complexity of the model.
5) The network is static. This means users cannot make or remove any friendship connections so as not to change the reality-based network graph. The scope of the current agent-based model does not include effects of dynamics of the network graph. There is a viral marketing study where the evolution of network graph and changes in preferences of users for different subjects have been analyzed [30]. This heterogeneity of preferences in different topics is controlled by the activity parameter in our agent-based model. Changing a user's friends does not significantly change chances of repost. Because the activity parameter is uniformly distributed and there is no similarity between friends' activities. One might say more active people are more likely to be friends with one another. But there are many more important demographics characteristics between friends such as nationality and age which trends to be similar. Thus, the activity parameter cannot be considered as a significant factor in this list.

The agent-based model keeps an inner state for each user, controlling their behaviour. This stochastic hierarchical state machine is shown in Fig. 4. All the users are in one of the two general states of *Online* or *Offline*. Every time the model restarts, the login and logout rate for switching between these two general states are assigned to each user. The intervals when a user

is in the *Online*/*Offline* state are controlled by login and logout rate. When the login/logout transition is triggered, they log into/log out of the system. Login transition times are drawn by an exponential distribution with the mean value of $\frac{Login\ Time}{1440}$ minutes per a day (1440 minutes). This was explained in more detail in section C.2.1.

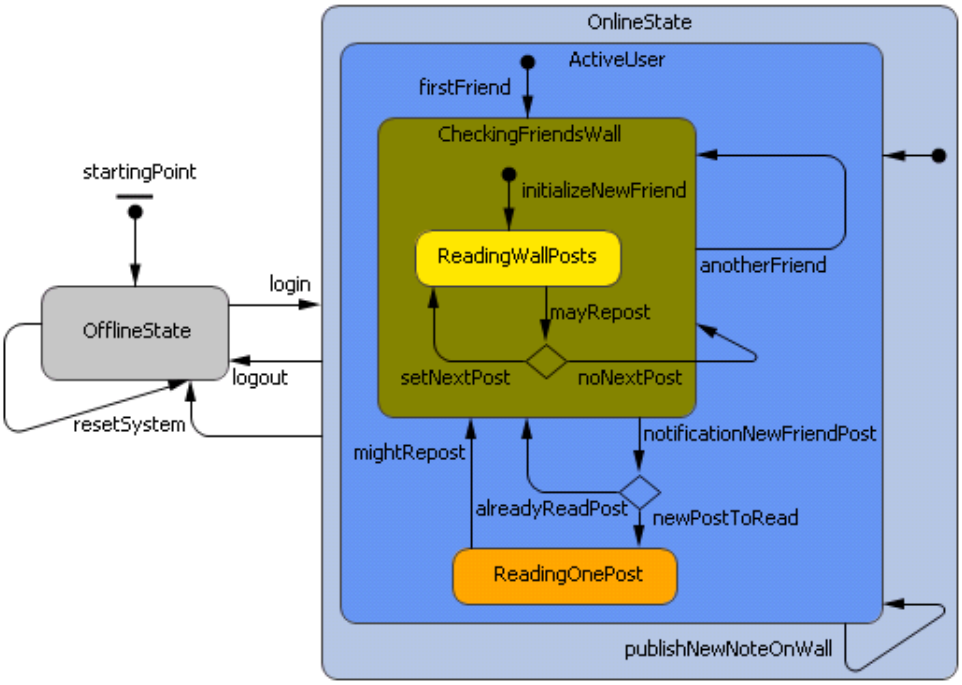


Fig. 4. User state machine

As seen in Fig. 4, all the interactions and events for a user happen when they are in the O*nline* state. Users in the *Online* state can either produce a note and share it or repost a note which was shared before by another user (which then appears on their own wall, in keeping with assumption 1) above). As mentioned earlier, each user posts new notes with a given rate based on the user's activity.

During these intervals between publishing a new note, users check their friends' wall pages. It was possible to employ a stack to sort recently posted notes for each user. However, in order to increase the level of autonomy of users, the freedom to choose which friend to see their posts is given to users. In the preliminary builds of the model, users sorted their friends based on

nationality and differences in their ages. As we did not aim to analyze clustering effects of nationality and ages, in the final version, they randomly select friends to review their most-recent wall posts to speed up the simulation. This is also accordant with the Facebook policy to make a ranked list of friends' posts for each user in their homepage (News Feed). This feed is an algorithmically ranked list of friends' posts based on a number of optimization criteria [31]. In its simplest case, the feed contains recent posts of those friends with higher probability of having interaction with the user. For each friend, a normally distributed random value is given as the maximum time threshold to check their wall page. When the time is passed, the user selects next friend to check their wall posts. This loop continues until one of high-level timeouts, such as logging off or publishing a new post, happens.

As shown in Fig. 4, when a user is checking friends' wall posts, if one of their friends publishes or reposts a note, they will be notified. Consequently, the user stops their current task of checking friends' page and begins reading the new shared note, which may be considered a result of human curiosity. Users are able to recognize a duplicate post and skip it in one second. Otherwise they would spend time in the amount of the post's length to read it.

When a user completely reads a note, they may decide to copy the note from the friend's wall page and repost it. The likelihood of this decision increases with the product of the post score and the user's activity. As an instance, a super active user with the activity value equal to 100% would definitely share an interesting post with the score of one (with probability $1 \times 100\% = 1$). One could have picked other formulas to define the probability of sharing. Multiplication, however, was found to be the simplest (and fastest) to proceed with the simulation. By this mechanism, posts propagate over the network, which is the subject of this study. It was also possible to let users share the post before they themselves read it completely. This scenario

usually happens when people see a public warning that they think it might be of use to others. Again, adding a new parameter and possibility into the system would have decreased the speed of simulation. It was decided to keep the model as simple as possible in the current version.

One might ask if it is not more likely that the marginal score of a post also increases as the post is shared over and over by users. This is the case for product adoption models, mostly inspired by the classical Bass diffusion model [32] or disease epidemic modeling approaches such as the Susceptible-Infected-Removed (SIR) models [33]. In the product adoption models, it makes sense that as the number of people/friends using a certain product grows, others would become more interested in purchasing the product as well. Also, in epidemic models, as the number of infected people with a certain virus increases, the chance of transmission of the virus to others increases. However, propagation of a meme through a social network might be different. Similar to product adoption, some individuals might find sharing a popular post very cool or helpful, such as posts asking to unify people on some belief or giving alerts. On the other hand, some may consider repeated exposure to a specific peace of content as being boring or démodé. The second case is verified in Digg social news website where people have a tendency of not sharing repeated news [34]. Thus, as herein all various types of content are modeled as posts with different scores, we neither increase nor decrease post score as it disseminates through the users; each user, depending on their activity level, might pick a post and share it. A post score, which is fixed, can be determined by external factors which are not part of this study. For example, assuming a post is a product advert, a combination of the market conditions and psychological and sociological characteristics of consumers would determine this factor for each person. There are different studies concerning this aspect such as the decoy effect in [35]. Having implemented an agent-based model, they showed how an individual consumer's

judgment on purchasing a product changes from brand A toward brand B after the introduction of a decoy brand, in a competitive market.

In the current simulations, for a given post, the post length time must be passed until we can conclude that the user has read the post. For instance, in the situation where a user decides to log out while reading a post, the post will not be checked as a read post.

C.2.5 Verification and Validation

As per Rand and Rust [36], verification of a model ensures the simulated model matches the conceptual model. This procedure is mainly through documentation, program testing and test cases [36]. In the present case, the exact written assumptions and rules of behaviour defined above were coded. The programming code is also extensively commented. Each function and module of the agent-based model was solely tested on a small network of users to receive a known output for a given input. After all the debugging was done, corner cases with extreme values (such as full connectivity, no connectivity, zero activity-level users, one activity-level user, zero post score, one post score, etc.) were run.

As per Rand and Rust [36], the four major steps of ABM validation were followed. These are micro-face validation, macro-face validation, empirical input validation and empirical output validation. At the micro-face level, actions of users are a limited form of a real Facebook user's possible actions. Also, the mechanism through which a post propagates is a type of cascade model corresponding to the real world. This paradigm has been validated by various studies over the past. At the macro-face level, it is observed that most posts will not be shared by a friend, which is the same scenario for a typical Facebook user. Also, our aggregate pattern of the post share distribution on-face is similar to Facebook and other social network statistics. At the empirical input level, the ranges of all the parameters are drawn from either Facebook network

statistics or reasonable assumptions. Further explanation of the input parameters is discussed in section C.5.

Relative to empirical output validation, it is not possible to exactly validate the agent-based model against reality at this stage, nor is this the claim in this work. The correct way is to run our experiments on Facebook; firstly however, it is not clear how to set a post's probability of being shared (i.e., its score), therefore it cannot be validated in this way in practice, apart from the fact we should be able to monitor exactly how many users have indeed read a post completely. Again this cannot be measured in practice. We may be able to only distinguish if someone has seen the post for at least a few seconds. Unless Facebook or similar social media service providers were using, for example, built-in counters to calculate how many seconds each user spends on a specific post, which is very unlikely for non-video posts; and if there exists such a counter, it would only work when you actually open a post (e.g., a photo) but could not be calculated when you are at your Facebook home page displaying more than one post, as it is not known which post you are exactly looking at. Having said that, it is still possible to demonstrate that real world data are possible outputs of our agent-based model, meaning that our average results match average results in reality.

A set of experiments was performed in order to validate the results here against the Facebook statistics reported in [31]. According to the statistics for users' most recent post, "the median post reached 24% of a user's friends (mean = 24%, SD = 10%)," provided that the most recent post was at least 48 hours old. Their population size is 589 different users with the median friend count of 335 (mean = 457, SD = 465).

To do such an experiment, 20 unique users were selected such that their friend counts were drawn from a normal distribution with the same mean and standard deviation. As a result, the

population of our selected users had a median friend count of 335 (mean = 464, SD = 337) which is quite close to the sample they used in [31].

Then their Activity parameter was set to be 1%, to prevent them from publishing any other note, since we wanted to perform the test for their most recent post; after that a post with score of zero and a short length of 10 seconds was published by the user. The score is set to zero to make sure the post does not spread over the network as we are interested in the number of immediate friends who read the post. The length was set to 10 seconds as this is our minimum post length at the current model. As mentioned above, the statistics reported by Facebook is in fact the number of people who have seen the post, although not necessarily read a long post completely. Then we ran the model. The model ran for 24 simulated hours as an initial warm up stage. Then a chosen user published a specific post. The simulation ran for another 48 simulated hours, after that the result was saved. The whole experiment for all 20 users was repeated two times, for a total of 40 runs. The results were very close in both runs for each user. Real time computation for each run was 9-12 hours on workstations equipped with an Intel Xenon CPU W3679 @3.2GHz with 16 GB RAM or higher configuration.

We observed that the median perceived audience size was 19% of a user's friends (mean = 25%, SD = 18%), which is near although slightly less than the expected 24% median of Facebook [31]. One difference in our experiment and Facebook statistics is that our results are the statistics after exactly 48 hours, whereas the Facebook ones are the statistics after at least 48 hours. So it is reasonable to reach a lower percent of immediate friends. Strictly speaking, in our agent-based model, the post life time is defined to be the last moment when the post is read. The average and median post life time in this set of experiments was 47.08 and 46.81 hours, respectively.

C.3 Simulation Studies

In this section, the agent-based model was run with different input settings to explore the impacts of each factor in the post diffusion process.

In the current model, the targeted input parameters include: (1) Post length; (2) Post score; (3) Friend count which is the number of friends of the first publisher of the note. As each minute in the simulated world takes 7-10 seconds to run computationally in the real world, we have limited our initial exploration to these three parameters. Future simulations will explore the impact of other parameters, including the day/time to disseminate the note over the graph.

The following statistics as the model output were recorded: (1) The number of users who have read the note; (2) The number of users who have reposted the note; (3) The times when the note was read / reposted; (4) The last time when the note was read as its life time.

The Analysis Of Variance (ANOVA) procedure between different scenarios is employed to find the importance of each parameter relative to the spread of a message within the online social network. In some cases, multivariate regression was run to test the magnitude of each factor. However as the outputs of an agent-based simulation should not be interpreted quantitatively [11], [12], the numerical values of a linear regression coefficients are interpreted qualitatively. This means even if we obtained numbers as the magnitude of importance of each parameter, they have to be discussed at the qualitative level.

C.3.1 Study 1: General Insight on Input Parameters

Within the first study, a total of 19 unique scenarios (simulations) were set up and each simulation was repeated 20 times, for a total of 380 runs for eight simulated days each. The real-world computing time required for each run was one to two days. The total computing time required for these simulations was over 380 days. In each scenario, our chosen user publishes a

certain post after one simulated day of warmup phase, then one week after the spread, the outputs are saved and the simulation ends.

The results of the first set of simulations are illustrated in Table III. According to the first (six) rows of the table, it is immediately clear that as long as the number of repost is near zero, the post score does not have much impact on the number read. Because the post is not shared by anyone except for the first publisher, the number read is directly related to the number of immediate friends of the publisher, which is trivial. For the simulation IDs 1-6 plus IDs 9-11, with the fixed post length of 30 seconds, where the post score is relatively small, Analysis Of Variance (ANOVA) yielded a significant effect of friend count ($F_{7.53} = 25.07$, $p < 0.01$), no significant effect of post score ($F_{7.53} = 0.22$, $p = 0.8$) and no significant interaction effect on friend count × post score ($F_{5.12} = 0.23$, $p = 0.9$).

Table III Results in first set of simulations after one week

ID	Post Score	Post Length	Friend Count	No. of Reads			No. of Reposts		
				Mean	Median	Std. Dev.	Mean	Median	Std. Dev.
1	0.001	30	9	2.3	2	1.4	0	0	0
2	0.01	30	9	1.7	2	0.7	0	0	0
3	0.1	30	9	2.0	1	1.2	0.2	0	0.4
4	0.001	30	139	13.0	10	10.3	0	0	0
5	0.01	30	139	10.5	8	7.7	0	0	0
6	0.1	30	139	17.1	13	16.5	1.30	1	1.9
7	0.3	30	139	33.0	25	28.8	6.2	5	6.5
8	0.5	30	139	101.1	13	341.7	32.1	4	110.2
9	0.001	30	530	45.8	28	46.6	0	0	0
10	0.01	30	530	57.0	41	45.9	0.3	0	0.5
11	0.1	30	530	55.9	42	50.6	4.2	2	4.8
12	0.3	30	530	73.3	50	68.0	14.0	8	13.6
13	0.5	30	530	2075.5	1588	2109.5	702.1	532	724.1
14	0.1	60	530	30.4	20	24.1	2.3	2	2.2
15	0.3	60	530	33.4	24	25.9	6.1	6	4.6
16	0.5	60	530	41.6	23	52.7	13.8	7	18.5
17	0.1	90	530	12.3	10	8.2	0.6	0	1.1
18	0.3	90	530	12.8	10	11.7	2.6	2	3.1
19	0.5	90	530	14.1	11	12.5	3.7	2	3.8

Among these simulations with almost no reposts, higher variances in the number of reads are observed for the experiment IDs of 9-11. This is in line with the Facebook observation that "a post produced by a user with many friends has more variability in the audience size than one produced by a user with few friends" [31]. Here, part of the reason is due to the differences in the first publisher's activity parameter. For example, in simulation #9 with zero reposts, it was observed that one repetition of the experiment had a very low activity parameter of 5%; for this reason, the post remained as the most recent post of the publisher and did not slide down the wall page. Consequently, more friends had chance to read this post. In this simulation, we had the highest number of reads which in turn increased the variance of the number of reads.

Also, the time of publishing a note relative to other events at the time is another important factor to receive a high number of reads. In fact, the activity parameter and timing both represent the complexity and heterogeneity of users and their interaction within the system. We did not explicitly analyze the impact of the activity parameter in the number of reads. However we did observe that for a certain post length in cases where the post score is low, the friend count parameter has the greatest impact and the publisher's activity is the second dominant parameter in determining the audience size. The reason is that, if the post is not interesting enough to be shared by others, it would only be read by the immediate friends of the publisher. So in order to increase the number of reads in this case, a higher friend count would help. Secondly, a lower activity level by the publisher keeps the post recent and top on their wall page. This in turn increases chances of being seen by others.

The initial implication is that if one cannot make an attractive post with high interest, at minimum, one needs to have it posted by a user with large number of friends in order to reach its maximum audience.

Of further interest is knowing more about the importance of a post score versus a publisher's friend count. Assuming the note is interesting, the question is whether one should focus on finding a hub user with many friends to post it or should one improve the quality of note as much as possible. Consider the simulation IDs 6-8 and 11-13, all of which have the same post length but relatively high scores published by users with different friend counts. Among these experiments, ANOVA yielded a major effect of post score ($F_{4.8} = 19.04$, $p < 0.01$), a bit weaker but still strong effect of friend count ($F_{6.86} = 18.43$, $p < 0.01$) and major interaction effect on post score × friend count ($F_{4.8} = 16.36$, $p < 0.01$). We can conclude that in cases where the post interest and the friend count are large enough, the former parameter is more influential than the latter. Keep in mind that friend count is still important, and one needs to consider the combined effect of both together. However, if one is able to find a user with an acceptable number of friends, it is recommended to focus more on post content rather that necessarily finding a hub user with many friends. This phenomenon implies that having a good seeder may help reach/saturate a local cluster of the network faster, but ultimately a higher post score is needed to reach further regions of the network.

The next question is the trade-off between post quality and post length when we have a well-connected user with relatively high number of friends to publish the desired post. To compare the impact of post score versus post length, for simulation IDs 11-19, the number of reads is shown as a heat-map in Fig. 5. According to the figure, post length dominates post score within all the range of post scores and lengths simulated. This means in order to reach a maximum audience, keeping the message brief is more significant rather than making an impressive but lengthy one. For example, by comparing the simulations #11 and #16, both of which have similar number of reads, it is observed that for a long note, more reposts (and consequently more time) is required

to reach a similar audience size of a short note with a lesser score. Statistically speaking, for the simulation IDs 11-19, ANOVA yielded a strong effect of post score ($F_{4.73} = 18.33$, $p < 0.01$), a bit stronger effect of post length ($F_{4.73} = 20.35$, $p < 0.01$) and a significant interaction effect on post score × post length ($F_{3.43} = 18.03$, $p < 0.01$). Therefore both post length and post score are important properties, and their combined effect has to be considered when making a post; yet, depending on the situation, the post length can be considered of more importance as if the length extends over a certain threshold it severely affects the post reachability no matter what the post score is. The reason is that users generally do not spend much time to judge a post. For example, a post might be very amusing, but as it is lengthy a typical user never spend sufficient time to recognize true score of the post.

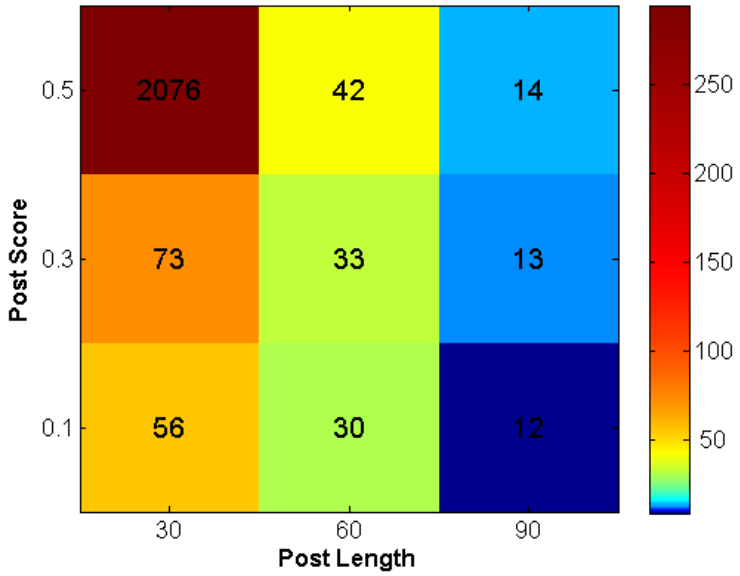


Fig. 5. Impact of Post Score versus Post Length on number of reads (medium to long notes)

C.3.2 Study 2: Importance of Score versus Length in Short Posts

After inferring that a post should not be too long if it is to propagate, in the second study, 12 more scenarios were simulated to test a broad scale of different post scores versus relatively short post lengths. In this set of simulations, the user who publishes the post for first time is a

new user with a different position in the network and a subsequently changed local network structure. This ensures our results are not specific only to some local region of the network structure.

This time, a total of 12 unique simulations were set up and each simulation was repeated 10 times, for a total of 120 runs for two simulated days each. In each simulation, the post spread began after a 24-hour initial warm-up phase, and then one day later the outputs were saved and simulation ended. In these experiments, smaller post lengths of 10 and 30 seconds with higher chances of sharing were tested. The post score changed from 0.001 to 1 to study a broader scale of scores. The results of our second set of experiments are shown in Table IV.

Table IV Results in second set of experemints after one day

ID	Post Score	Post Length	Friend Count	No. of Reads			No. of Reposts		
				Mean	Median	Std. Dev.	Mean	Median	Std. Dev.
20	0.001	30	319	33.8	28	25.4	0	0	0
21	0.1	30	319	35.2	19	32.9	2.6	2	2.5
22	0.25	30	319	48.4	41	26.4	7.6	7	3.9
23	0.5	30	319	1857.2	58	2465.4	635.8	17	852.7
24	0.75	30	319	7651.8	6670	8085.1	3815.3	3324	4035.3
25	1	30	319	23559	23468	507.6	15325.6	15213	340.8
26	0.001	10	319	46.7	46	12.3	0	0	0
27	0.1	10	319	92.2	101	55.2	5.1	4	4.7
29	0.25	10	319	120.7	91	95.7	18.1	10	18.9
30	0.5	10	319	12449.9	16964	8,608	4077.9	5565	2,825.3
31	0.75	10	319	32473.5	32297	1,007.3	15375	15256	457
32	1	10	319	41165.8	41036	1071.6	25424.1	25330	587.5

According to Table IV, as a post score increases from 0.25 to 0.5, a significant change in the number of reads is observed. Obviously, post score should likely have the greatest impact in general. For a more detailed comparison, the simulation is split into two subsections of low and high post scores. One heat-map for each part is shown in Fig. 6. From the heat-map on the left corresponding to low score posts, it can be seen that post length dominates the post score. On

average, all the scenarios of shorter (i.e., 10 seconds) posts reach comparable size or larger audiences than longer (i.e., 30 seconds) posts with any scores, as long as post score is not very high (less than 25%). This verifies the previous result with a different seeder and time to collect the result. However, for the heat-map on right, this inference does not hold true any longer. For higher score posts, shorter posts do not necessarily reach more users, and it depends on both post score and post length together. For example, simulation #25 with a post length of 30 seconds and score of 100% has reached a larger audience than simulation #30 with a shorter post length of 10 seconds but lower score of 50%. Yet simulation #31 with a short post length of 10 seconds and score of 75% has larger audience than simulation #30. From an ANOVA perspective, a significant effect of post length ($F_{6.88} = 190.99$, $p < 0.01$), a more significant effect of post score ($F_{3.19} = 293.77$, $p < 0.01$) and a significant interaction effect on post score × post length ($F_{3.37} = 46$, $p < 0.01$) is observed. Therefore, generally both post score and post length are very important in the success of a post propagating. However, the relationship with these two parameters and audience size is not linear.

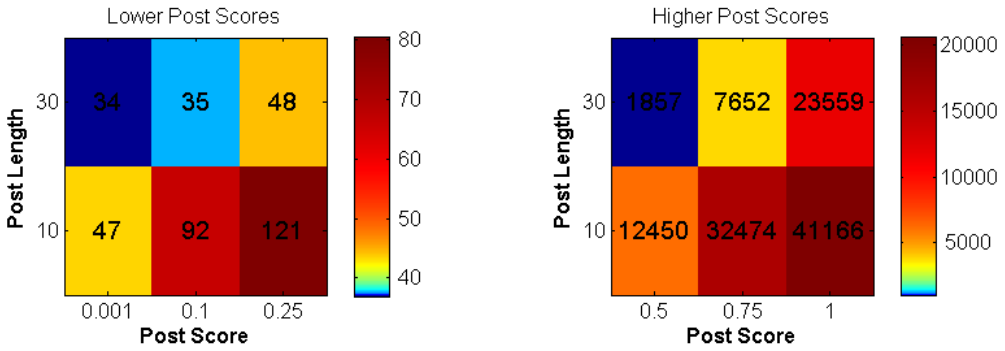


Fig. 6. Impact of Post Score versus Post Length on number of reads (short notes)

Assuming it would be desirable to explain impacts of each factor in the number of reads with a linear model, Table V shows the result of fitting a linear model to our second study. The linear regression model has an adjusted R^2 of 76%. However, expectedly, the plot of residuals does not suggest a linear model as being a suitable model for these scenarios. The coefficients

statistics confirm our previous inference using ANOVA and heat-maps. According to Table V**,** considering a α significance level of 1%, the activity parameter does not have any significant impact on the number of reads with a p-value of 0.4. Both post score and post length have significant impacts. Post score has higher influence within all the simulations studied in the second set of simulations together.

In order to obtain a better model of number of reads, each of the following (five) functions was applied to our model inputs: Inverse, Logarithm, Square root, Square and Cube. So instead of three inputs of Post Score, Post Length and Activity, there are now 3×6 = 18 inputs to choose from. The best linear model based on all these inputs was found to be a linear model of post score squared and inverse of post length. The new model has a better adjusted R^2 of 81% with the coefficients shown in Table VI. Properties of over one million posts were saved after two days from the starting point of simulations in different runs for another linear regression. The basic inputs were post score, post length and friend count. However, once again the five functions of Inverse, Logarithm, Square root, Square and Cube were applied to each of the basic inputs to find the best combination of inputs for the regression model. The best model was found to have an adjusted R^2 of only 32%. Subsequently, posts with zero repost were excluded from our input data to have more coherent input data, which resulted in almost half a million posts remaining. With the objective to have only one function of each of the basic inputs as an input to our linear model, the best fitted model was found to be a linear model of post score squared, logarithm of post length and square root of friend count. This model has an adjusted R^2 of 54%. The coefficients of the model are shown in Table VII**.** The scatterplot of standardized residuals versus standardized predicted value is shown in Fig. 7. Ideally it would be desirable to have a

uniform scattering of points around the zero reference line; but as the agent-based model is highly non-linear, a better linear model could not be fit to its outcome.

Table V Coefficients of linear model of number of reads fitted to second study

Model input	Coefficients	Std. Error	Standardized Coefficients	t Stat	P-value
Intercept (constant)	3552.302	1981.203		1.793	0.076
Post Score*	33077.74	1788.742	0.820	18.492	2.2E-36
Post Length*	-441.586	63.51394	-0.308	-6.953	2.25E-10
Activity	1923.83	2441.076	0.035	0.788	0.432

* Significant at an alpha level of 0.01

Table VI Coefficients of transformed inputs of linear model of number of reads fitted to second study

Model input	Coefficients	Std. Error	Standardized Coefficients	t Stat	P-value
Intercept (constant)	-9029.287	1353.794		-6.670	≤0.001
(Post Score)^2	3.396	0.158	0.849	21.443	≤0.001
1 / (Post Length)	132908.250	17021.300	0.309	7.808	≤0.001

Table VII Coefficients of general model of number of reads for those posts for which number of reposts ≥ 1

Model input	Coefficients	Std. Error	Standardized Coefficients	t Stat	P-value
Intercept (constant)	27952.530	69.463		402.410	
√ (Friend Count)	149.849	1.636	.094	91.583	≤0.001
Log (Post Length)	-21236.924	42.333	-.510	-501.659	≤0.001
(Post Score)^2	1.905	.004	.553	542.076	≤0.001

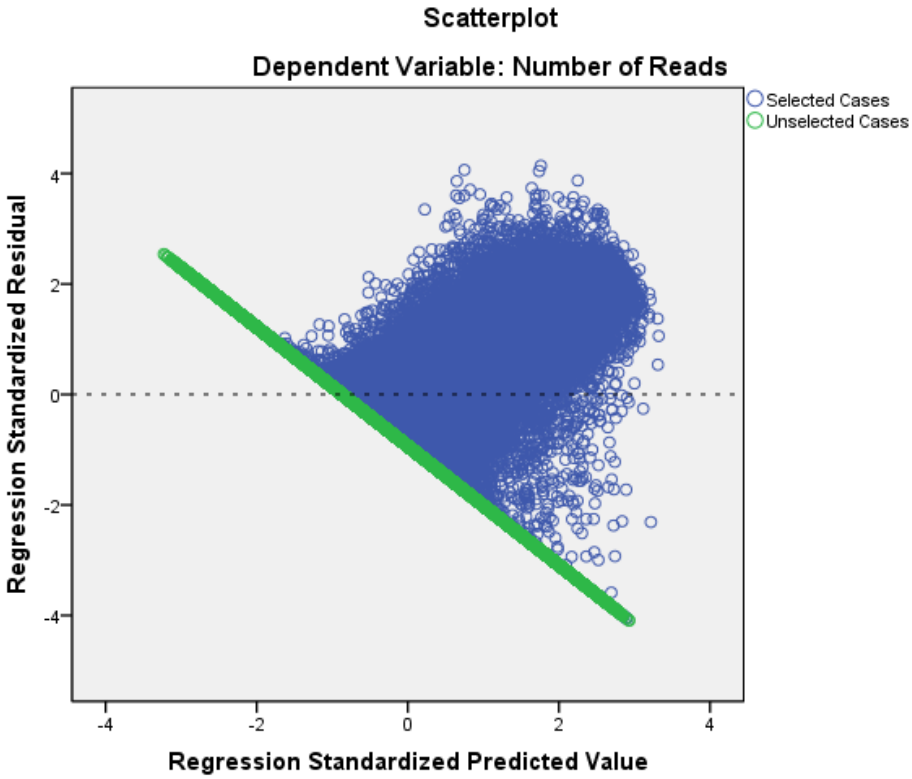


Fig. 7. Scatterplot of standardized residuals of linear model versus standardized predicted value of number of reads

All the inputs have coefficients significantly different from zero. Additionally, by looking at the standardized coefficients, it is clear that generally both ‘logarithm of post length’ and ‘post score squared’ have significant impacts on number read. Yet post score squared is a stronger predictor.

According to the second study, one needs to avoid lengthy posts, and the post still needs to have some minimum score. There are certain thresholds for post lengths and post scores that post properties should lie within. Initially, it is better to focus on post score to gain some interest and chances of sharing, then one should try to shorten the post length considering the limited time of

users. Lastly, when the message is brief enough, fine-tuning the post score can achieve better results than slightly reducing of post length.

There is an uncommon difference between the median and average number of reads in simulation #23. The low median number of reads can suggest that a typical user may not read a long post. However if, with the help of good timing, users read a long (but interesting) post, the post would propagate very well through the network increasing the total number of reads. Therefore in simulation #23 when the timing is matched, we observe a high number of reads increasing the average number of reads, and when timing does not cooperate well, users show little interest in the post.

C.3.2 Study 3: Comparison of two Seeding Strategies

In the third and last study, the objective was to gain insight by comparing a classical seeding strategy of a small number of hub users with many friends versus seeding a large of number users with few friends. First, a special post was published by four randomly chosen users in the network, each of which has exactly 50 friends. Then, the simulation was repeated for two randomly chosen users with 100 friends each. The details are described below.

All four users had a fixed activity level of 50%. At a certain time after the initial warm-up phase, they all shared a unique short post with 50% score and 10 seconds length. Then 24 simulation hours after the spread, the numbers of reads were collected and saved. The simulation was repeated 10 times. Then the whole simulation was repeated for two more sets of four random users with exactly 50 friends, for a total of 30 runs for the 4×50-friend case. The same simulation was carried out for three different sets of two users with exactly 100 users each, for a total of 30 runs for the 2×100-friend case. The average number read for the 4×50-friend case and

the 2×100-friend case is 13300.7 and 9540.1, respectively. The statistics are shown in Table VIII.

Table VIII Statistics of number of reads in third study

Case	Mean	Std. Deviation	Std. Error Mean
4×50-friend	13300.73	8228.25	1502.27
2×100-friend	9540.10	9096.65	1660.81

The average results suggest that a mass of small seeders may broadcast a certain post better than a few hub users. However, assuming the more general case of not-equal variances for these two cases, one cannot reject the null hypothesis of having equal means in these two cases. Technically, the null cannot be rejected by t-test with a t stat of 1.679 and the degrees of freedom of 57.426. This is equivalent to a non-significant p-value of 0.099. In other words, the difference between the two cases is not statistically significantly different.

To see the trend of message propagation, Fig. 8 plots the number of reads and number of reposts versus time for a run of simulation #13 on a log-log scale. All other simulations where the post is shared by some users have a similar trend. The only time required for the outbreak to propagation is the seconds required to read the post. Unlike other diffusion patterns such as product adoption or the spreads of infectious disease, this trend is not s-shaped for Facebook posts. In other words, there is no classical tipping point or epidemic threshold for post propagation after which we could expect an outbreak in the number read. The reasons are discussed in section C.4. This trend may sound surprising at first, yet it is consistent with real observations of Facebook. According to a select group of brand posts data of Facebook in November 2012, each post, on average, reached half of their target audience within 30 minutes

after publication [37]. Also a quite similar trend is reported for the number of retweets in Twitter [38].

Another surprising result is that the median number of reposts is zero. This means it is quite common that posts by a typical user on Facebook do not attain even a single repost. We confirm that there exist posts with high number of reposts in both Facebook and our agent-based model. These highly shared contents are mostly (high scored) posts published by popular pages, which can be thought of as hub users in our agent-based model. Yet many of the posts, especially the ones submitted by typical users receive few or very limited reposts. Most of the published results about social networks are generally focused on successful posts and their properties, and as such, statistics regarding the failed ones submitted by random usual users could not be found. The total distribution of number of reposts is shown in Fig. 9. This long-tailed distribution indicates only a few posts gain a huge number of reads. The distribution is also consistent on the surface with that of retweets (popularity) for twitter reported by [38]. Section C.4 discusses this consistency.

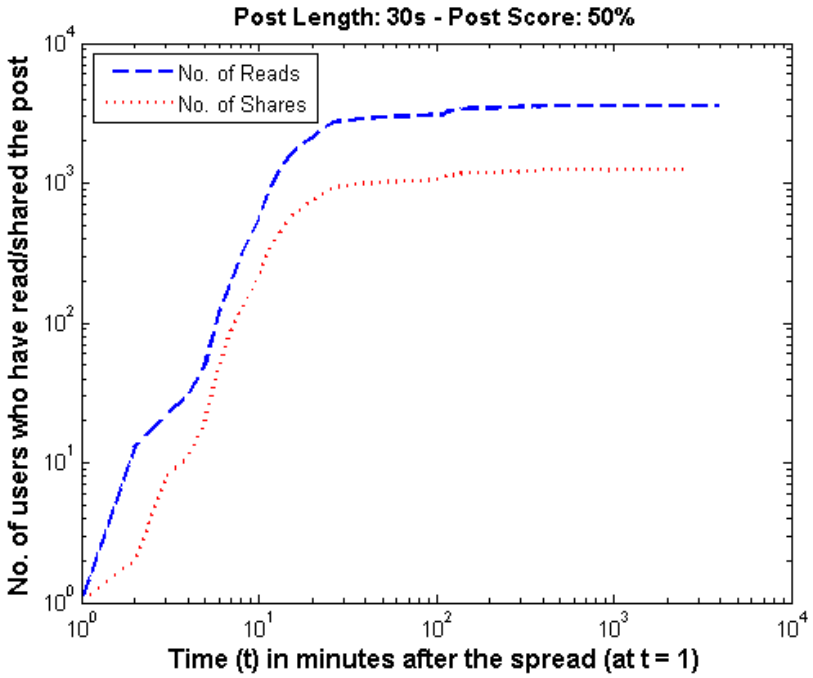


Fig. 8. Trend of message propagation

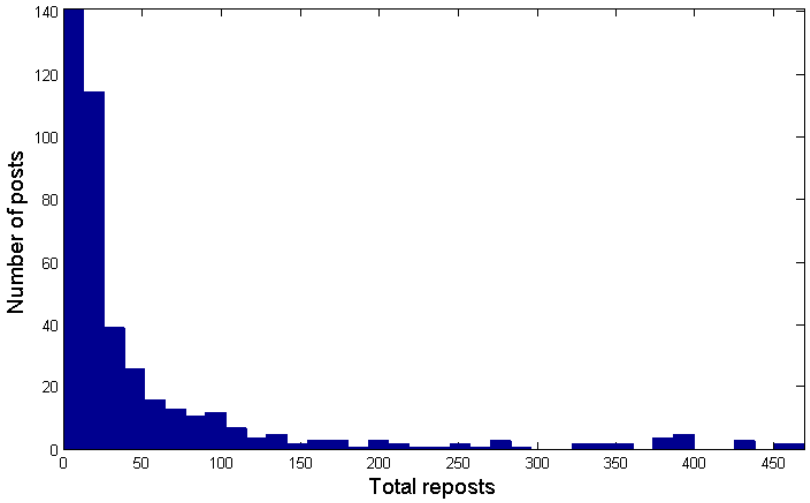


Fig. 9. Reposts distribution

Initially, we expected that having a score of 10% would be enough for a post to be broadcast and seen by everybody. Firstly, in this case, the average probability of reposting is 1/20. Secondly, when the post is published for the first time, only the immediate friends will see the post. The immediate friends should be online in a certain time period to catch the post. Therefore, if the friend count is not sufficient, the post will never have any chances to spread further. In other words, if the network were fully connected so that users observe the whole population, the reposting probability of 1/20 might have been adequate, however this is not the case in this agent-based model.

C.4 Related Work and Discussion

Diffusion is a core process in many areas such as energy flows in physics, disease spread in biology, behavioral contagion in sociology and product adoption in economics. Several researchers have studied aspects of this phenomenon in human systems. In human systems, a general way of social contagion is through the word-of-mouth (WOM) mechanism which is quite similar to spread of an infectious disease within a population. Below, the current simulation

results are discussed in relation to other common diffusion processes of product adoption (marketing), disease spread (epidemiology), and electronic networks.

Within product adoption research, Moldovan *et al.* studied the importance of WOM to new product success [1]. In their research, a product score is divided into two dimensions of originality and usefulness of a product, and the number of online reviews is used as a proxy for the amount of WOM. Reference [3], using an ABM, studies the actual value of a seeding program for WOM in terms of market expansion and purchase acceleration for a certain product. Reference [30] attempted to predict users' adoption of a given product on the Digg social news website. They studied the effect of network-level dynamics and changes in individual preferences for different topics, and proposed a viral marketing strategy which was tested with an agent-based model. Reference [39] studies social commerce where individual sellers are connected to each other through an online social network. Using time series analysis at the marketplace level and Bayesian statistical analysis at the shop level, they explore whether connecting the sellers to each other increase the sale, and how the position of a seller within the networks influences their value. Reference [40] studies impact of different connection patterns among individuals on the diffusion process in a European online social network. Using a hazard-rate model, they investigate characteristics and local structure of potential adopters and their neighbours.

Tirunillai and Tellis also confirm the importance of user-generated content as a type of WOM and in particular study the impact of online product reviews and ratings on stock market performance using multivariate time-series models [5]. Similarly, [41] finds that consumer generated online product ratings has a direct relation with the product sales; although the previously submitted ratings affect the future ratings. Reference [42] also investigates the

evolution of online reviews of books over time and sequence. Weblogs as a part of the larger set of online social media can also influence purchase/adoption of a product. An analytical model for a blogger is made and studied in [43]. In particular, they explain why blogs may link to rivals, and what the relative benefits and costs of such linking are.

Goldenberg *et al.* study the role of hubs in the diffusion process of products over the Cyworld social networking site in Korea [44]. They define two types of innovator hubs who need little exposure to adopt a product and follower (social) hubs who are well-connected. Using an ABM, they analyzed the impact of each type on the market eventual size and the speed of adoption.

Cellular automata modeling and aggregate level modeling together are used in [45] to study growth rate of a new product. By separating network externalities effects such as mass media and advertising from internal interactions (i.e., word of mouth through the network), they found a chilling effect on growth rate of a product. This is a "wait-and-see" state for a product when potential consumers wait for others to adopt a certain product and then decide whether to purchase or not. This partially explains why product adoption has an s-shaped growth rate. Thus growth of products can follow a two-stage process which includes a slow start due to the chilling effect and then a fast growth because of the bandwagon effect. However, a similar concept for post-sharing does not exist. One does not need to see if others have shared a post to decide whether one should or not; basically because one is not purchasing a post with real currency.

In the context of contagious disease, [46] studies SIR models of epidemic disease spread in networks. Rahmandad and Sterman compare AB and DE models of epidemic spread on different networks [47]. There are some similarities and differences between the spread of an electronic post and an infectious disease. The propagation is similar in the way in which it cascades through

the network. Unlike posts and similar to products, disease has an s-shaped spread rate (for basic review, see [48]). In a disease epidemic, the rate of people entering from a susceptible into the infected state depends on the number of people in other states (such as the infective state). As people become more and more infected, chance of transmission of disease is increasing until the epidemic threshold is passed and the disease rapidly spreads throughout the population. In our agent-based model, it is assumed that a post score, corresponding to chances of reposting, is constant. In epidemiology, a quicker recovery rate makes a smaller epidemic; it takes a longer time to happen as people recover faster and the number of infected in the beginning of the process is not enough to form an outbreak. In our agent-based model, users need to read a post completely (i.e., become infected) then may spread it. With a short post, a shorter time is expected for the outbreak to occur, and it has a higher peak, as the time needed to read the post is smaller and more people are likely to read it. During the infection period of the disease epidemic case, the chance of transmission of the disease is constantly present; so the epidemics with a slower recovery (removal) rate (i.e., a long infected period) that slowly kill people are more dangerous to the populations having more deaths at the end (for discussion see [48] pages 21-22). In social networks, users generally share a post only once and their friends usually read the post once and decide whether to share or not. As the assumption is people do not share a post while reading it, a longer post (unlike a longer infected period of a disease) does not provide a better opportunity to spread the post. Lastly, the ultimate objective in disease modelling is mitigating against the spread of the disease such as targeted vaccinations as a means of achieving herd immunity; but in social networks, more spread and penetration are desired.

Diffusion of applications in Facebook is studied in [49] using a customized commercial application about the movie industry to track user behaviour. Using hazard modeling, they test

effectiveness of passive-broadcast messaging versus active-personalized messaging. In another study on Facebook, susceptibility of various type of users (e.g., young, women, married etc.) to influence their adoption decisions is measured [50]. Application adoption and social influence in Facebook is also studied in [51] using fluctuation scaling (FS) method. They track popularity of a set of applications among all users in their dataset collected in 2007. Their observation is limited to the adoption of an application, and not necessarily the usage. Similarly, sharing a post needs an increased level of engagement, rather than simply reading it.

Reference [52] analyzed the prevalence data reported by a computer virus for a time window of 50 months. They found the absence of an epidemic threshold for virus spread on scale-free networks due to an infinite connectivity effect phenomenon in large scale-free networks. This effect happens because of a heterogeneous rising and falling in the number of links of nodes in the scale-free graph. This factor is also applicable to the current agent-based model as the synthesized graph shares this feature of scale-free networks.

There are various and sometimes conflicting recommendations regarding choosing the optimal set of users to publish a message (see [53] for summary of previous research). Reference [53] analyzes four different seeding strategies for both messages and products using two small-scale field experiments and one real-life viral marketing campaign. Their four seeding strategies are: 1) seeding well-connected hub users; 2) seeding low-degree members, called the fringes; 3) seeding high-betweenness users, called the bridges that connect different sub-networks; and 4) random seeding. They find high-degree and high-centrality strategies are preferable in general. In contrast, [54] finds large cascades are driven by critical mass of individuals, but not necessarily influentials (hubs). In the current third simulation, some evidence in favor of this strategy of mass of individuals also emerged. However, the test was limited and the results were not

significantly different from seeding a few hub users. In summary, it is safe to conclude each strategy may work well under certain conditions.

Furthermore, in the current agent-based model, by investigating properties of the most successful posts, it was found that as long as a post is not qualified enough to be reposted by typical users, it does not have any chance to diffuse across the whole network. In other words, relying only on reposts by active users (with high tendency to share) does not guarantee message propagation. Because wall pages of active users who repost many notes tend to be crowded with different notes, the note would only have a minimal chance to be read by someone else. Nevertheless, if the note is reposted by a selective hub user with many connections but low activity to post other notes, a significant growth in the number of reads and reposts will be achieved. As the post remains as the most recent post of the hub, it gains maximum exposure to the hub's friends, although one still needs the post to be highly scored to be ultimately reposted by a critical mass of users. Therefore, to achieve a maximum audience, it is recommended to produce the desired post and begin its publishing by different hubs who have many connections but (who) are quite discerning about their willingness to share other notes.

Reference [55] analyzes the impact of the degree distribution of a network on adoption process, using an agent-based model. They discovered that while most researchers simply assume the adoption process propagates over the entire visible (overt) network, the actual active subset of the network over which the propagation occurs may have quite different properties from the entire network. They provided evidence that the degree distribution of the active network is generally different from that of the entire network. The degree distribution has a significant effect on contagion properties of nodes within the network.

In an inspiring paper by Lerman and Ghosh, the active network of users on Digg and Twitter is extracted, and then it is studied to see how the structure of the (active) network affects the dynamics of information flow on each network [38]. The general mechanism of the spread of information in both sites are similar to Facebook, where users watch their friends' activity, and they may share/tweet/vote for a post to make the post visible to their own friends/fans/followers. The precise underlying details vary from case to case though. As mentioned in section C.3, the evolution of number of tweets received by each post in Twitter reported in [38] is utterly similar to that of reposts in the current work as displayed in Fig. 8; (successful) posts display a burst of growth at the beginning, and the growth saturates after a while. The point at which growth saturates is the cascade size indicating how far and wide the post has penetrated though the network. Reference [38] reports that Digg's (active) social network of their dataset has a larger clustering coefficient than Twitter's, meaning that its network is denser. As such, initially posts spread easier in the more highly interconnected Digg's network, but they eventually spread more distance in Twitter's less densely connected network. The distribution reported for the number of retweets (or reposts) in [38] is somewhat analogous to the distribution of number of reposts in the current agent-based model displayed in Fig. 9. Retweets distribution in Twitter has a small number of posts (or tweets) with almost zero retweets, followed by an exponential peak and then a gradual decrease of retweets to zero again, creating a long tail. This means a majority of posts gain a few (but non-zero) retweets in Twitter and very few posts exist with many retweets. Fig. 9 states the same fact finding, except that it claims that the majority of posts in Facebook have almost zero reposts. The reason is that the Twitter dataset used in [38] does not include any non-popular tweets (with low score). They collected only frequently-retweeted posts on Twitter using the Tweetmeme.com site at the time. In addition, tweets can only have a limited number of

characters. As such, they all can be considered as short-length posts as in the current agent-based model, with higher chances of being reposted. However the distribution reported in Fig. 9 corresponds to all the posts with various length and score simulated within the agent-based model.

Reference [56] employs an agent-based model to simulate reposting behaviour of users in the Twitter social network. They explicitly model the competition for humans' limited attention among different posts (memes), and how it affects memes' popularity. They developed a memory mechanism to reflect users' past behaviour to what they would share in the future, as users are likely to share posts similar to what they shared before. Similar to the result we presented in this appendix, long-tailed distributions for memes' popularity and lifetime are reported. They found extremely heterogeneous behaviour; a few memes are extremely successful while most of them die out quickly. Reference [57] proposes a coupled hidden Markov model to capture neighbour users' influence on users' posting activity. Their model is also tested on Twitter.

An interesting study on information cascades on Digg by Steeg *et al.* revealed that a high level of clustering structure of the Digg network limits the overall growth of cascade [34]. In highly clustered networks, people are usually exposed to a certain post multiple times through multiple friends, which in turn lowers the epidemic threshold, speeding the spread up initially. However, the surprising effect reported by [34] is that repeated exposure to the same post does not encourage people to repost it. This is a fundamental difference between the spread of information and classical spread of products or diseases. Reference [34] shows this effect drastically limits the overall cascade/epidemic size. As an example, while many posts with a fast starting spread exist in their dataset, only one post about Michael Jackson's death reached a

significant fraction of 5% of active Digg users. They observed that the effective number of people who have not been exposed to a post is gradually decreasing. In addition, other effects such as decay of visibility and novelty could be other reasons why the epidemic stops [34]. In one more study on Twitter, rapid decay of visibility combined with the limited attention of users are determined to be the primary reasons for preventing the growth of propagation of online information [9].

The way the current model network described in this appendix is synthesized does not produce a high (and desired) level of clustering coefficient. This means that the number of reads for successful posts may have been overestimated. That being said, most successful posts in the current simulations end up reaching around 30% of the entire network. The percentage of viral posts is less than 2.5% of all the posts generated within the model. Moreover, it was discovered that all these posts had extremely high scores of mostly over 90% and short lengths of mostly 10-15 seconds published by users with various friend counts. Precisely speaking, these viral posts have scores ranged from 70% to 100% with a mean and median of 89% and 91%, respectively. The post lengths are varied from 10 to 25 seconds with a mean and median of 15 and 14.5 seconds, respectively. It implies that viral messages can be published by a user with low number of friends, but certainly various users including hubs would have to repost it during the spread. More importantly, the post must be brief while extremely highly qualified. We used a singular numerical value to represent a post score, but a post can have different aspects to be engaging to different kinds of people. There are various categories of posts: promotional offers/deals, advice and tips, warnings, amusing video clips, amazing pictures, personal news, motivating speeches, campaigns recalls etc. Thus, a score of 90% in our simulations represents a high quality post in a variety of features. But in reality, it is nearly impossible to make such a universally fascinating

post. Certainly, viral posts exhibit a variety of features. For example, death news of a famous celebrity has a wide range of viral features making it likely to go viral. Such a post has information, novelty, may contain stimulating quotes or represent a group of people's mourning or respect.

C.5 Conclusion and Limitations

There has been a great deal of work on dissemination of information in online social networks especially on Twitter and Digg. Studies on application adoption in Facebook also exist. The work of this appendix presents a large scale stochastic agent-based approach for modeling wall post propagation within the Facebook network. Network and other input parameters have been drawn from and tuned to published sources of Facebook statistics. Other studies on the realm of social media diffusion have confirmed the importance of various factors including underlying network structure, local network structure of following spreaders, influence degree and activity level of each spreader, type and novelty of a post and people's response to repeated exposure. At the current stage, analyzing all these factors in a single study would be too complex or too generic. The agent-based platform created here is potentially capable of testing all factors by some code modifications. In some cases, one would have to insert a few more parameters into the system, such as a social influence parameter for each person for example.

In this study, various scenarios have been explored to investigate the impact of post length, post score, and the post publisher's friend count on post diffusion. It is observed that posts with small scores hardly ever spread farther than the immediate friends of the publisher, meaning that higher friend count for the publisher has a stronger impact on audience size than higher post score does. However, beyond a certain level of message quality, the post score has a larger influence on message propagation through the network than publisher's friend count. This

intuitively means the content of a message is more important than who has delivered it. In cases with relatively medium post interests and medium friend counts where there definitely exists some reposting, the post length is the most influential parameter followed by the post score and the friend count as the second and third influential factors, respectively. This implies that creating a long post makes it boring and significantly affects its chances of getting shared by others. Whereas for relatively short posts, increasing the quality of post contributes more to the audience size than cutting the post length any shorter, which does not necessarily boost the growth of number of reads. The intuition behind this result is that people spend a minimum amount of time on each post, and once a post length is below that minimum length, there would be no need to make the post any shorter. Keeping these hints in mind could help marketers to find a balance between length and content of post making their ideal post advert for example.

Adjustments are needed to study other online social networks using this agent-based model. For example, reconstructing the network, limiting post lengths and the way users look through the posts from friends have to be changed in order to study the Twitter social network. However, the results reported here may not be limited only to the Facebook online social network. Less intuitive findings about the dynamic of post spread mechanism such as lack of epidemic threshold in the propagation of information have been previously confirmed in other online social networks. Along with other work on seeding strategies, it is observed that both hub-seeding and having a large number of individual seeders could result in having a viral post reaching an epidemic portion of population of at least a sub-region of the network graph. The simulations performed to compare these two strategies were limited to conclude a general statement though.

Achieving an ‘epidemic reach’ of the entire Facebook network is nearly impossible. People are online at different times and it is vital for a post to catch their attention when they are online. Assuming the timing can be handled through the interface design of online social networking websites, a viral post still needs to be highly scored in a variety of features. Each person is likely to become engaged in a certain category of posts. Modeling a post score based on a single scalar value may be an over simplistic assumption. A much more realistic way to define a post score is using a vector where each component describes post content in a different perspective. Then users’ activity parameters also need to be vectorized to capture the heterogeneity of population in different directions. Add to this the fact that certain categories of contents (e.g., politics, fashions, and sports) are being shared more on certain social networking sites (e.g., Twitter, Pinterest, and Facebook) [58]. The limitation in the current study is that an active user is likely to share any type/category of posts, and also a high-scored post has all the features of all sorts of appealing posts which cannot be true in practice.

One more realistic extension of current research would be the insertion of dynamic scores for post. There are different ways and reasons to change a post’s score during the spread. A rational reason is that after users spend a few seconds on a post, they get a better idea of how interesting the post is. So they may decide to continue reading/watching the post or disregard it. This is especially true for video posts. Once the dynamic scoring feature is added into the model, one can let users share a post even before they read it completely.

There are other limitations with the current study which bring opportunities for further extensions of this research. Sensitivity analyses for input parameters such as average post length or login rate were limited only to the primary stages of making the agent-based model. The parameters and distributions were tuned on the basis of a smaller network and Facebook’s known

statistics. The parameters were fixed once it was observed that the main agent-based model of the larger network had rational functionality in line with the conceptual model. Testing sensitivity of all the outputs to all input parameters on such a large set of data would require a lot of time and effort. Such reports could offer insights on what to expect if people begin spending twice as much time on Facebook for example, or investigate robustness of results across a range of parameters and distributions. In addition, the exact correlations, threshold values and interactive effects of simulation inputs (post length, post score and publisher's friend count) still need to be determined through more simulations.

One crucial direction for future work is a more comprehensive way of modeling the Facebook news feed ranking algorithm, which used to be called the EdgeRank algorithm. Our current model of EdgeRank emphasizes Recency of posts too highly, meaning that if a post is recently published, it has the highest chance to be seen by users. Facebook tries to identify those who have most interaction with a user. Currently the algorithm is considering numerous factors (including recency) to decide which posts to show for each person in their home news page. For example, if one likes or comments on posts by a person/group, chances of receiving more posts from those persons/groups will be increased. So, modeling these two Facebook features could be an essential feature to complement the current agent-based model.

C.6 Acknowledgment for Appendix C

The author would like to thank Professor Miroslaw Pawlak for his input and guidance for the mathematical analysis of the work.

C.7 References for Appendix C

[1] S. Moldovan, J. Goldenberg and A. Chattopadhyay, "The different roles of product originality and usefulness in generating word-of-mouth," *Intern. J. of Research in Marketing*, vol. 28, pp. 109-119, 2011.

[2] D. Godes and D. Mayzlin, "Using online conversations to study word-of-mouth," *Marketing Science*, vol. 23, no. 4, pp. 545-560, 2004.

[3] B. Libai, E. Muller and R. Peres, "Decomposing the value of Word-of-Mouth seeding programs: acceleration versus expansion," *Journal of Marketing Research*, vol. 50, no. 2, pp. 161-176, 2013.

[4] M. Trusov, R. E. Bucklin and K. Pauwels, "Effects of word-of-mouth versus traditional marketing: Findings from an internet social networking site," *Journal of Marketing*, vol. 73, no. 5, pp. 90-102, 2009.

[5] S. Tirunillai and G. J. Tellis, "Does chatter really matter? Dynamics of user-generated content and stock performance," *Marketing Science*, vol. 31, no. 2, pp. 198-215, 2012.

[6] A. Patino, D. A. Pitta and R. Quinones, "Social media's emerging importance in market research," *Journal of Consumer Marketing*, vol. 29, no. 3, pp. 233-237, 2012.

[7] H. Park and H. Cho, "Social network online communities: information sources for apparel shopping," *Journal of Consumer Marketing*, vol. 29, no. 6, pp. 400-411, 2012.

[8] T. Hogg and K. Lerman, "Stochastic models of user-contributory web sites," *Proc. of the 3rd Intl Conf on Weblogs and Social Media (ICWSM2009)*, pp. 50-57, 2009.

[9] N. O. Hodas and K. Lerman, "How visibility and divided attention constrain social contagion," in *Proceedings of the 2012 ASE/IEEE International Conference on Social Computing (SocialCom-2012)*, Washington, DC, 2012.

[10] R. Axelrod, The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration, Princeton, NJ: Princeton University Press, 1997.

[11] E. Bonabeau, "Agent-based modeling: methods and techniques for simulating human systems," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. Suppl 3, pp. 7280-7287, 2002.

[12] R. Garcia, "Uses of Agent-based modeling in innovation/new product development research,"

Journal of Product Innovation Management, vol. 22, no. 5, p. 380–398, 2005.

[13] T. Rowland, "Computational Irreducibility," [Online]. Available: http://mathworld.wolfram.com/ComputationalIrreducibility.html. [Accessed 23 Dec 2014].

[14] M. Gladwell, The Tipping Point: How Little Things Can Make a Big Difference, Little, Brown and Company, 2000.

[15] M. Gjoka, M. Kurant, C. T. Butts and A. Markopoulou, "Walking in Facebook: a case study of unbiased sampling of OSNs," in *INFOCOM'10 Proceedings of the 29th conference on Information communications*, 2010.

[16] D. A. Ebersman, "Amendment No. 3 to REGISTRATION STATEMENT on Form S-1," Facebook Inc., Menlo Park, 2012.

[17] M. Kagan, "12 Essential Facebook Stats [Data]," 31 May 2011. [Online]. Available: http://blog.hubspot.com/blog/tabid/6307/bid/14715/12-Essential-Facebook-Stats-Data.aspx. [Accessed 30 March 2013].

[18] K. Burbary, "Facebook demographics revisited – 2011 statistics," 7 March 2011. [Online]. Available: http://socialmediatoday.com/kenburbary/276356/facebook-demographics-revisited-2011-statistics. [Accessed 31 March 2013].

[19] S. Radwanick, "comScore Introduces Mobile Metrix 2.0, Revealing that Social Media Brands Experience Heavy Engagement on Smartphones," 7 May 2012. [Online]. Available: http://www.comscore.com/Insights/Press_Releases/2012/5/Introducing_Mobile_Metrix_2_Insight_into_Mobile_Behavior. [Accessed 31 March 2013].

[20] B. London, "Virtual vanity: Nation's most obsessed Facebook users spend a staggering EIGHT hours a day on the site," 10 September 2012. [Online]. Available: http://www.dailymail.co.uk/femail/article-2200962/Nations-obsessed-Facebook-users-spend-staggering-hours-day-site.html. [Accessed 31 March 2013].

[21] F. Keating, "Got a smartphone? You probably check Facebook FOURTEEN times a day," 2013. [Online]. Available: http://www.dailymail.co.uk/sciencetech/article-2300466/Smartphone-users-check-Facebook-14-times-day-admit-looking-movies.html. [Accessed 18 9 2014].

[22] C. Pring, "100 social media statistics for 2012," 11 January 2012. [Online]. Available: http://thesocialskinny.com/100-social-media-statistics-for-2012/. [Accessed 31 March 2013].

[23] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature,* vol.

393, no. 6684, pp. 440-442, 1998.

[24] S. Schnettler, "A structured overview of 50 years of small-world research," *Social Networks,* vol. 31, no. 3, pp. 165-178, 2009.

[25] R. Cohen and S. Havlin, "Scale-free networks are ultrasmall," *Physics Review Letter,* vol. 90, no. 5, 2003.

[26] J. Ugander, B. Karrer, L. Backstrom and C. Marlow, "The anatomy of the Facebook social graph," *Arxiv preprint arXiv:11114503,* 2011.

[27] T. Britton, M. Deijfen and A. Martin-Lof, "Generating simple random graphs with prescribed degree distribution," *Journal of Statistical Physics,* vol. 124, no. 6, September 2006.

[28] L. Backstrom, P. Boldi, M. Rosa, J. Ugander and S. Vigna, "Four degrees of separation," *arXiv:1111.4570,* 2012.

[29] S. Gunelius, "25 Basic Facebook Terms You Need to Know," 29 Dec. 2011. [Online]. Available: http://sproutsocial.com/insights/2011/12/facebook-terms-definitions/. [Accessed 26 03 2013].

[30] H. Sharara, W. Rand and L. Getoor, "Differential adaptive diffusion: understanding diversity and learning whom to trust in viral marketing," in *5th International AAAI Conference on Weblogs and Social Media*, 2011.

[31] M. S. Bernstein, E. Bakshy, M. Burke and B. Karrer, "Quantifying the invisible audience in social networks," 2013.

[32] F. M. Bass, "A new product growth for model consumer durables," *Management Science,* vol. 15, no. 5, pp. 215-227, January 1969.

[33] W. O. Kermack and A. G. McKendrick, "A Contribution to the mathematical theory of epidemics," *Proceedings of the Royal Society of London. Series A,* vol. 115, no. 772, pp. 700-721, 1927.

[34] G. V. Steeg, R. Ghosh and K. Lerman, "What stops social epidemics?," in *5th International AAAI Conference on Weblogs and Social Media*, Barcelona, 2011.

[35] T. Zhang and D. Zhang, "Agent-based simulation of consumer purchase decision-making and the decoy effect," *Journal of Business Research,* vol. 60, p. 912–922, 2007.

[36] W. M. Rand and R. T. Rust, "Agent-based modeling in marketing: guidelines for rigor,"

International Journal of Research in Marketing, 2011.

[37] MarketingCharts, "Facebook Posts Get Half Their Reach Within 30 Minutes of Being Published," 2012. [Online]. Available: http://www.marketingcharts.com/online/facebook-posts-get-half-their-reach-within-30-minutes-of-being-published-24453/. [Accessed 16 Oct 2014].

[38] K. Lerman and R. Ghosh, "Information Contagion: an empirical study of the spread of news on Digg and Twitter social networks," in *Proceedings of 4th International Conference on Weblogs and Social Media (ICWSM-10)*, 2010.

[39] A. T. Stephen and O. Toubia, "Deriving value from social commerce networks," *Journal of Marketing Research*, vol. 47, no. 2, pp. 215-228, 2010.

[40] Z. Katona, P. P. Zubcsek and M. Sarvary, "Network effects and personal influences: the diffusion of an online social network," *Journal of Marketing Research*, vol. 48, no. 3, pp. 425-443, 2011.

[41] W. W. Moe and M. Trusov, "The value of social dynamics in online product ratings forums," *Journal of Marketing Research*, vol. 48, no. 3, pp. 444-456, 2011.

[42] D. Godes and J. C. Silva, "Sequential and temporal dynamics of online opinion," *Marketing Science*, vol. 31, no. 3, pp. 448 - 473, 2011.

[43] D. Mayzlin and H. Yoganarasimhan, "Link to success: How blogs build an audience by promoting rivals," *Management Science*, vol. 58, no. 9, pp. 1651 - 1668, 2012.

[44] J. Goldenberg, S. Han, D. R. Lehmann and J. W. Hong, "The role of hubs in the adoption process," *Journal of Marketing*, vol. 73, no. 2, pp. 1-13, 2009.

[45] J. Goldenberg, B. Libai and E. Muller, "The chilling effect of network externalities," *International Journal of Research in Marketing*, vol. 27, no. 1, pp. 4-15, 2010.

[46] M. E. J. Newman, "The spread of epidemic disease on networks," *Phys. Rev. E*, vol. 66, pp. 016128-1 – 016128-11, 2002.

[47] H. Rahmandad and J. Sterman, "Heterogeneity and network structure in the dynamics of diffusion: comparing Agent-based and Differential Equation models," *Management Science*, vol. 54, no. 5, pp. 998-1014, 2008.

[48] T. Tassier, "The economics of epidemiology," in *SpringerBriefs in Public Health*, 1, Ed.,

Springer, 2013.

[49] S. Aral and D. Walker, "Creating social contagion through viral product design: a randomized trial of peer influence in networks," *Management Science,* vol. 57, no. 9, pp. 1623 - 1639, 2011.

[50] S. Aral and D. Walker, "Identifying influential and susceptible members of social networks," *Science,* vol. 337, no. 6092, pp. 337-341, 2012.

[51] J.-P. Onnela and F. Reed-Tsochas, "Spontaneous emergence of social influence in online systems," *Proceedings of the National Academy of Sciences,* vol. 107, no. 43, pp. 18375 - 18380, 2010.

[52] R. Pastor-satorras and A. Vespignani, "Epidemic Spreading in Scale-Free Networks," *Phys. Rev. Lett.,* vol. 86, pp. 3200-3203, 2001.

[53] O. Hinz, B. Skiera, C. Barrot and J. U. Becker, "Seeding strategies for viral marketing: an empirical comparison," *Journal of Marketing,* vol. 75, no. 6, pp. 55-71, 2011.

[54] D. J. Watts and P. S. Dodds, "Influentials, networks, and public opinion formation," *Journal of Consumer Research,* vol. 34, no. 4, pp. 441-458, 2007.

[55] Y. Dover, J. Goldenberg and D. Shapira, "Network traces on penetration: uncovering degree distribution from adoption data," *Marketing Science,* vol. 31, no. 4, p. 689–712, 2012.

[56] L. Weng, A. Flammini, A. Vespignani and F. Menczer, "Competition among memes in a world with limited attention," *Scientific Reports,* vol. 2, no. 335, 2012.

[57] V. Raghavan, G. Steeg, A. Galstyan and A. Tartakovsky, "Modeling temporal activity patterns in dynamic social networks," *Computational Social Systems, IEEE Transactions on,* vol. 1, no. 1, pp. 89-107, March 2014.

[58] MarketingCharts, "Which Content Categories Are Being Shared on Which Social Networks?," 9 July 2014. [Online]. Available: http://www.marketingcharts.com/online/which-content-categories-are-being-shared-on-which-social-networks-43953/. [Accessed 10 12 2014].